PET-Derived Metabolic Volume Metrics in the Hodgkin Lymphoma

4

Lale Kostakoglu and Stephane Chauvie

4.1 Methodological Considerations

4.1.1 PET-Derived Quantitative Metrics

4.1.1.1 Standardized Uptake Value (SUV)

SUV is the most frequently used semiquantitative PET metric for measuring tumor glucose metabolism. It is defined as the ratio of the decay corrected FDG concentration in a volume of interest (VOI) to the injected dose normalized to the patient's body weight. Besides body weight-based SUV, various other SUVs have been introduced to account for the different bio-distribution of FDG in different body compositions (Table 4.1). The most commonly used is SUL, which is SUV corrected per the lean body mass (LBM) defined, respectively, for male and female as

$$\mathrm{LBM} = 1.1 \times \mathrm{weight} - 120 \times \left(\mathrm{weight} / \mathrm{height}\right)^2$$

$$\mathrm{LBM} = 1.07 \times \mathrm{weight} - 148 \times \left(\mathrm{weight} / \mathrm{height}\right)^2$$

This index takes into account the different biodistribution of FDG in the fat tissue. Even if several recommendations exist to use SUL, e.g., in the treatment response evaluation [4], actually it is not of widespread use because of the familiarity established with SUVmax. The general advice, also furnished by the EANM-SNM guidelines on FDG-PET use [1, 3], is to collect SUL along with SUV data to further the understanding of its relevance to in both clinical practice and experimental settings. The SUV, being an index of PET tracer uptake in any tissue should be measured in a known volume of interest (VOI), because with different VOIs its measure considerably varies.

SUV_{max} The SUV_{max} is defined as the maximum value for SUV in a VOI. The rationale is choosing the single point that represents the hottest uptake or highest metabolism in the tumor. This rationale is quite strong and moreover the SUV_{max} is simple to measure. However, being a single voxel measurement, SUV_{max} is intrinsically vulnerable to image noise (Fig. 4.1). Consequently, repeated tumor SUV_{max} measurements showed an intra-patient bias of 5–30 % [5].

SUV_{mean} The SUV_{mean} is the average value of different measurements of SUVs within the VOI drawn for the tumor. It is much less vulnerable to

L. Kostakoglu, MD, MPH (✉)
Nuclear Medicine and Molecular Imaging, Department of Radiology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1141, New York, NY 10029, USA
e-mail: Lale.kostakoglu@mssm.edu

S. Chauvie
Department of Medical Physics, 'Santa Croce e Carle' Hospital, Cuneo, Italy

A. Gallamini (ed.), *PET Scan in Hodgkin Lymphoma*, DOI 10.1007/978-3-319-31797-7_4

Table 4.1 Pros and cons of different SUV measures

	SUV_{max}	SUV_{peak}	SUV_{mean}	MTV	TGV
Volume of interest (VOI)	Single point	1 cm^3	Variable	Variable	Variable
Number of voxels in VOI	1	10–30	Hundreds–thousands	Hundreds–thousands	Hundreds–thousands
Intra-patient repeatability	5–30 % [4]	1–11 % [4]	Depends on segmentation method	Depends on segmentation method	Depends on segmentation method
Affected by image noise	Strongly	Slightly	Moderately	Slightly	Slightly

image noise, but it heavily depends on the delineation method used for drawing the VOI and the selected region within the tumor volume [6]. Defining a VOI in a tumor mass may have different meanings depending on its coverage within the mass. Generally one would like to provide SUV_{mean} of the entire lesion but this requires the knowledge of the exact dimension and borders with respect to the background, but this is often not the case in routine applications. An alternative approach includes delineation of a VOI inside the tumor far from its border to minimize the effect of the background uptake on the SUV measurement. Nonetheless, VOI delineation is subjective; tumors are usually heterogeneous and/or sometimes associated with necrotic centers; finally, the rationale for selecting only a part of the tumor without including the hottest part defeats the purpose of obtaining an accurate measurement.

SUV_{peak} The SUV_{peak} represents the maximum tumor activity within a 1 cm^3 VOI in the hottest part of the tumor volume [3, 6]. The rationale is to have an index measurement associated with the hottest part of the tumor, i.e., SUV_{max}, but in a standard volume of 1 cm^3. The SUV_{peak} characteristically is less affected by the noise compared to SUV_{max} and does not require definition of tumor boundaries which is a necessary step for obtaining an SUV_{mean}. Repeat tumor SUV_{peak} measurements yields a lower within-patient bias (1–11 %) compared to those of SUV_{max} [5]. The SUV_{peak} is the proposed measurement in the definition of therapy response for PET response criteria in solid tumors (PERCIST) developed by Wahl et al. [4]. Nonetheless, despite being relatively simple, this method requires the careful use of custom software on a dedicated workstation to be accurately calculated.

4.1.1.2 Sources of Errors in SUV Measurements

Common sources of errors involved in SUV measurements from technical and host-related factors are summarized in Table 4.2. Extensive review in literature exists to discuss these factors [1, 2], and recommendations have also been released by the US and European nuclear medicine associations [1, 3]. The recommendations provided should be considered to be minimal standards to abide by and should be followed by all imaging centers. While several recommendations are easy to adopt in clinical practice, e.g., maintaining a rest state during uptake time, others are more difficult to achieve in a busy clinic, e.g., the consistent time interval for the uptake period. Importantly, the higher the level of standardization reached, the simpler it will be to compare PET metrics acquired at different time points (intra-patient) and between different patients (inter-patient) either at a single site or across multiple centers.

Technical (Site) Factors

Several factors are patient independent and/or dependent only on the equipment and the procedure used by the site to perform PET/CT imaging studies. The requirements to limit the influence of these factors on SUV measurements should be fulfilled on one occasion and verified periodically (Table 4.2). The cross-calibration of PET scanners and activity calibrators are essential to

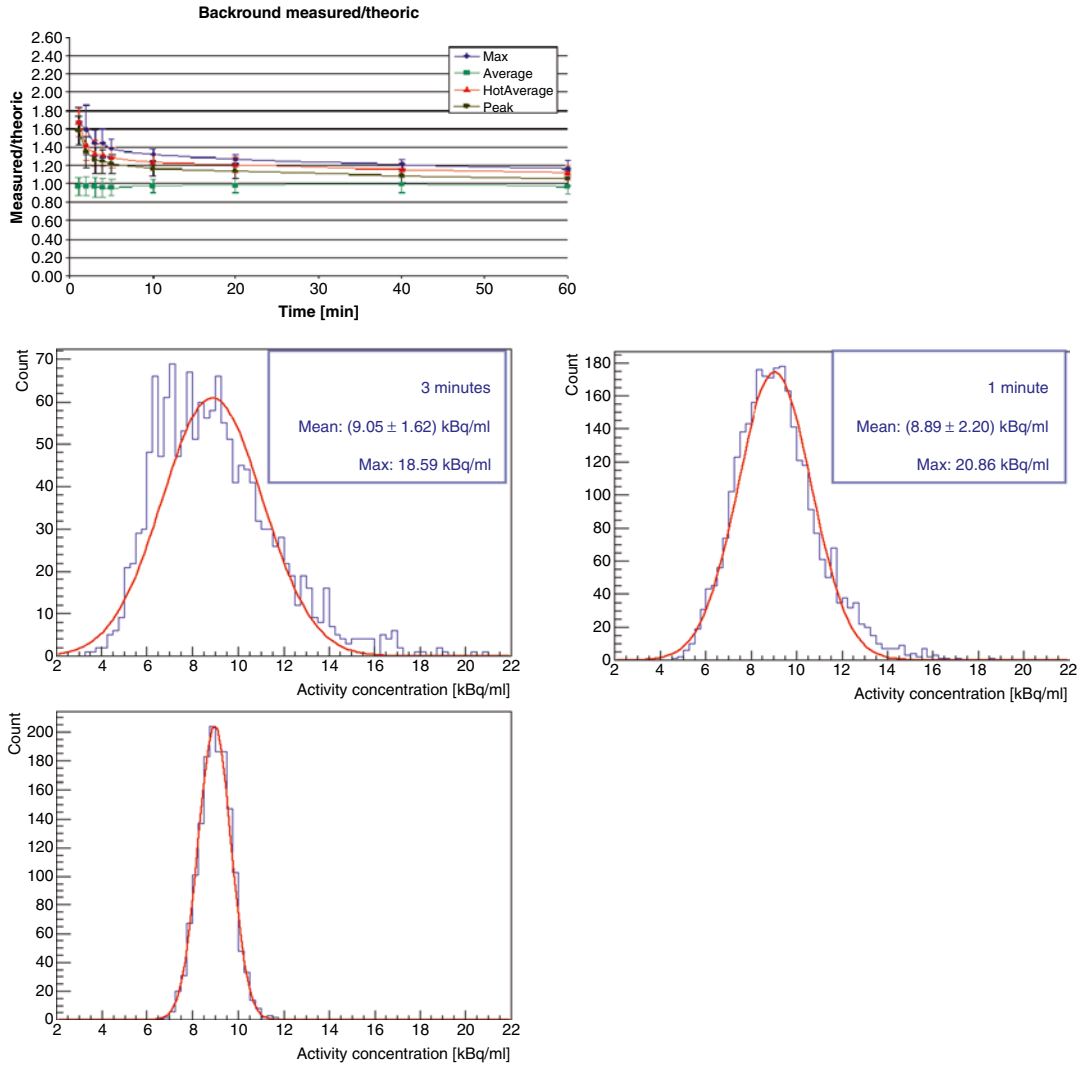

Fig. 4.1 The detailed characteristics of the noise affecting PET images are often not well known. Typically, it is assumed that overall the noise may be characterized as Gaussian. Noise levels observed in PET images complicate their interpretation; since the measurement of uniformity is strictly connected with the noise. In the figure the different metrics used for measuring activity concentration are "max," the highest pixel value; "hot," the average pixel value in a 1 cm diameter region around the "max;" "peak," the average pixel value in a 1 cm diameter ROI in the hottest region; and "average," the average of all the pixel value encompassed in the region of interest. In this figure one can see how the metrics described in the text with the acquisition time changes inside a large uniform VOI placed in the center of a uniform phantom. "Max," "hot," and "peak" have a similar trend and are the most influenced by statistics. When increasing the acquisition duration, these values decrease and the measured activity concentration become closer to the estimated values. On the other hand "average" does not change with the scan duration, and the value of the ratio between the measured and the expected activity is always about one. Errors are larger for "max," "hot," and "peak" at lower statistics, while "average" is more or less constant because its value is averaged on a large number of pixels. SUVmean is a good description of the expected value while "peak" and "max" are always overestimating the real value. And this changes dramatically when the acquisition time is small. This is well explained in the following histogram describing the SUV distribution inside the same VOI. While mean value is constant independent of the acquisition time "peak" and "max" are much larger at small time

Table 4.2 Common sources of error in SUV calculations

Source of errors	Typical errors (routine)[a]	Expected errors (controlled environment)	Action to be carried out[b]	Performed
Scanner and activity meter calibration	30 %	10 %	Scanner calibration: mean activity concentration in a homogenous area of phantom should be within ±3 % of the expected value. Activity-meter calibration: mean activity of a radioactive source should be within ±3 % of expected value	Per scanner, annually
Syringe residual	5 %	0.1 %	Rinsing syringe with saline is strongly recommended. Injected activity can be corrected with measured residual activity in the syringe. Alternatively a fixed value of the residual activity should be used if an evaluation of the mean activity in the syringe is carried out in the site on a statistically sufficient number of patients	Per patient
Clock time differences	10 %	1 %	Clock used to measure time of injected and residual activity measurements and injection and acquisition time should be synchronized (±30 s)	Per site, annually
Quality of administration	50 %	0 % (with patient exclusion)	An evaluation of extravasation should be carried out including the point of injection of the tracer in PET scan. For example, if intravenous injection is carried out in the harm, the patient could position the harms over the head, but the PET scan must include them in the longitudinal field of view	Per patient
Imaging parameters[c]	30 %	10 %	Mean recovery coefficient in the hot sphere should fulfill EANM guidelines	Per scanner
Contrast media	15 %	0 %	Avoid pre-PET administration of IV contrast media. If contrast-enhanced CT is needed, IV contrast should be administered after low-dose CT for AC[d]	Per patient
Region of interest (ROI)	50 %	0 %	Statistical bias due to different methods of defining ROI for SUVs should be reduced by using the same algorithm for PET segmentation. Systematic bias could persist since no method assures to measure the “true” SUV of a lesion	
Uptake time	15 %	8 %	For clinical trials consistently use uptake time of 55’ to 75’ after injection	Per patient

Table 4.2 (continued)

Source of errors	Typical errors (routine)[a]	Expected errors (controlled environment)	Action to be carried out[b]	Performed
Motion and respiration artifacts	30 %	–	Particular care should be devoted if quantitative measurements of lesions are carried out in the thoracic region. Use breathing control device if available	Per patient
Blood glucose level	15 %	Unknown	Blood glucose level should be measured and reported. Guidelines should be used for managing hyperglycemia. Linear SUV correction is applied retrospectively	Per patient
Patient's weight	8 %	2 %	Patient's weight should be measured with 1 kg calibrated weight and reported	Per patient
Patient's height	–	–	Patient's height should be measured and reported for SUL calculation	Per patient

[a]Typical errors, such those encountered in routine clinical practice
[b]If proper actions are carried out, as listed in the last column, the errors could be reduced to the value indicated in column 3
[c]Acquisition and reconstruction parameters
[d]Attenuation correction

minimize SUV variability. The procedure for calibration of the PET scanner is depicted in Fig. 4.2. Although cumbersome, this approach proved effective in increasing the accuracy of tracer uptake measurements by 5–10 % [6]. This is well below the range of 10–25 % observed variations even in a controlled environment of a multicenter clinical trial [4]. Particular care should be taken to use the same activity calibrator to measure the activity used for calibrating the scanner. If more than one calibrator is used, they should all be cross-calibrated with a traceable radioactive source. PET sites not equipped with dose calibrators cannot get reliable SUV measures. Indeed, the activity injected in the patient must be always measured with the calibrator that is cross-calibrated with the PET/CT scanner used for imaging. If the activity is measured elsewhere, for example, at the radiopharmaceutical production site, this process is not necessary. Cross calibration in a multicenter framework generally permits to achieve a variability less than 10 %, while 5 % should be a requirement for using PET/CT in a quantitative way [7, 8]. An optimal inter-scanner variability of 3 % has been reached when comparing two [8] PET/CT scanners requiring new cross-calibration strategies. Imaging parameters, such as scan duration per bed position, acquisition mode, ^{18}F-FDG dose, and reconstruction methods directly affect the image quality and quantitative results [3]. These parameters should be preset to fulfill the guidelines [1] for the recovery coefficient curve. The recovery coefficients are calculated as the ratio of measured and expected activity concentration in hot spheres of different radius in a phantom (Fig. 4.3). In addition to the above parameters, the actual administered activity and the accuracy of patient's weight and height influence the variability in SUV measurements. The injected activity is the difference between activity measured with the activity calibrator in the syringe and the syringe and administration lines residual. If the line is flushed with saline, the residual activity is usually lower than 1–3 MBq, and its measurement could be definitively omitted. Clock synchronization should be carried out on all the clocks of the department with respect to the scanner and the dose calibrator clocks to avoid bias in time and, consequently, SUV assessments.

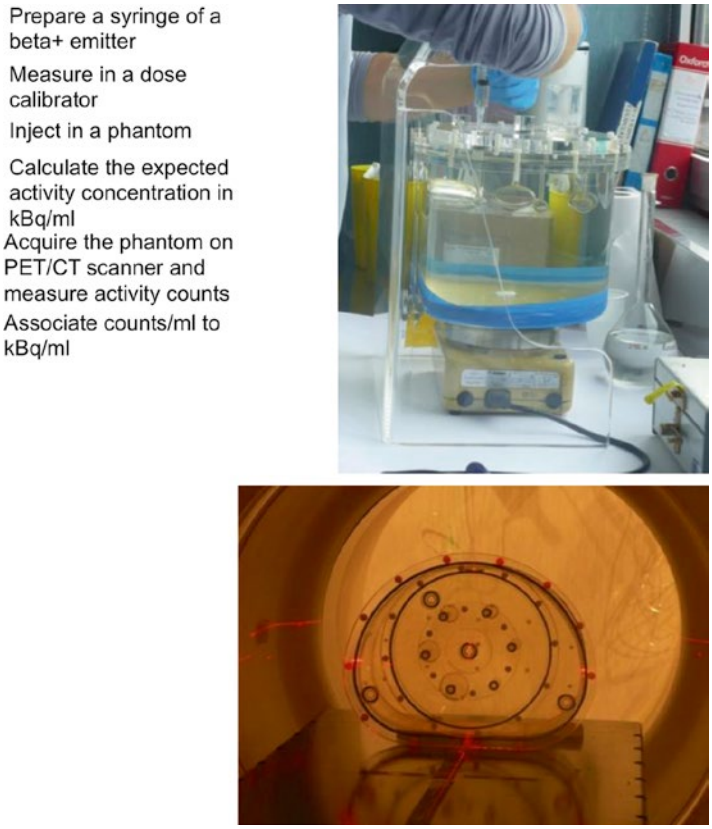

Fig. 4.2 PET scanner calibration. PET scanner electronics measure the count rate of annihilation events. PET scanner calibration is carried out to associate an activity to this count rate. This is done by injecting a known activity, measured in an activity calibrator, in a cylindrical uniform phantom and scanning it with PET

Data are given to reproduce the curves

x	Min	Max	opt	non opt
10	0.29	0.52	0.44	0.54
13	0.51	0.74	0.65	0.65
17	0.75	0.94	0.92	0.9
22	0.77	1.01	0.97	0.8
28	0.85	1.05	1	0.96
37	0.88	1.08	1	1

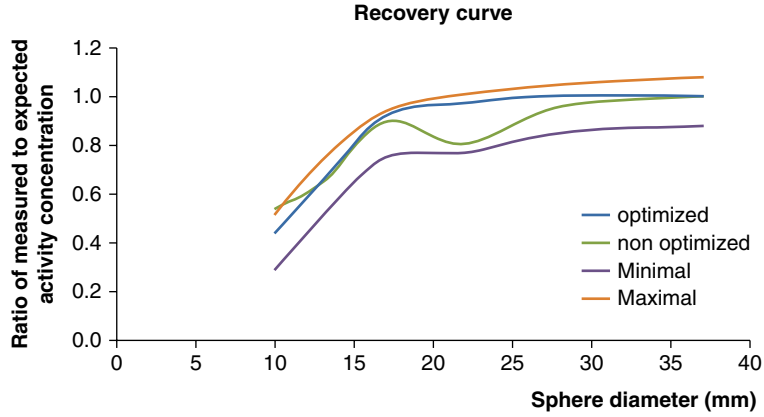

Fig. 4.3 Recovery coefficient. An example of a recovery coefficient curve is provided for non-optimized (*left*) and optimized (*right*) PET parameters

Intravenously administered contrast media could alter SUV of a lesion if a diagnostic CT is performed as part of the PET/CT [3]. While specific recommendations could be found elsewhere [3], a general recommendation is to perform low-dose scan CT for attenuation correction before the PET scan and the full dose diagnostic CT after the PET scan. For the calculation of SULs, the patient's weight should be routinely entered in the PET dicom dataset with the calibration factors to avoid errors in SUV calculations (e.g., 5 kg difference in an 80 kg patient lead to a 5 % error in SUV).

Host Factors

Several patient-dependent factors from patient preparation to scan acquisition affect SUVs and must be verified on the single patient (Table 4.2). The biological factors include uptake time, plasma glucose levels, and patient motion or breathing artifacts. For most of these factors, clear recommendations [3] have been provided, as they directly affect SUVs and also image interpretation. In addition, SUV measurements are affected by tumor perfusion and hypoxia, inflammatory cell infiltrates in the tumor microenvironment, which cannot be controlled by extrinsic manipulations. The SUVs decrease in normal tissues with the increase of uptake time [9] with a linear decrease of SUV in all three compartments. The FDG uptake from the same lesion in images acquired at different time intervals after the radiotracer injection is influenced by the recirculating FDG. It is, hence, fundamental to use the same uptake time for all time points when sequentially imaging the same patient to maintain intra-patient consistency and to reduce the uptake time changes in longitudinal scans. As a general recommendation, a patient requiring quantitative PET measurements should be scheduled as the first patient of the day to minimize delays in acquisition times, which occur frequently later in the day. Moreover, in order to get comparable data in longitudinal studies, the PET scanner technicians should annotate the actual uptake time, to ensure reproducible results in the next scans. Elevated plasma glucose levels result in decreased FDG uptake by the tumor, leading to erroneously low SUV values [1, 8]. Consequently, variable plasma glucose levels in longitudinal studies of the same patient will likely cause artificial SUV changes. A constant plasma glucose level in the range of 4–7 mmol/L in an individual patient across all longitudinal studies and tracks of the measured values are an achievable goal with a concerted team effort. There are several strategies for dealing with plasma glucose levels in SUV calculations, but further research is needed to understand whether intra-subject or inter-subject standardization is required. Patient's physical or breathing motion can also significantly influence SUV measurements [3]. To minimize this negative effect, the PET and CT fusion images should be visually analyzed to identify possible patient motion nearby a lesion. Patient breathing particularly influences the lesions in the thoracic area. Correction techniques are being introduced in PET/CT scanners using dynamic acquisition and breathing control devices; however, until then the data associated with motion should not be used for SUV measurements.

4.1.1.3 Metabolic Tumor Volume Measurements

Other proposed PET-derived functional metrics include metabolically active tumor volume (MTV) and total lesion glycolysis (TLG). The tumor volume concept has been developed in the late 1990s [10] but not evolved until recently because of the lack of necessary software developments. These volume-based PET parameters measure metabolic activity in an entire tumor mass designed to reflect tumor biology.

MTV

The MTV measure the total volume of the metabolically active tumor included within a VOI, both for a single lesion both for multiple lesions and expressed in cm^3 or ml. The rationale is the assumption of a metabolic activity higher than the surrounding healthy tissue to be able to accurately define the tumor volume. MTV is slightly affected by noise since it includes hundreds or thousands of voxels.

TLG

The TLG is the product of SUV_{mean} in the defined VOI and the MTV; the rationale is to combine tumor burden and its metabolic activity to obtain an index that is correlated to the tumor volume and the uptake within the entire volume. The routine application of these parameters is challenging because the quantification process requires complex calculations, is conducive to subjective definitions of VOIs, and is rather time consuming. There are several segmentation algorithm definitions, relying on manual (by an expert) or semiautomatic methods for tumor delineation (Table 4.3). With the recent development of software-assisted automated VOI assessments, volume-based metabolic quantitative parameters have become increasingly available. Although these metrics are potentially useful clinical parameters for assessing treatment response and survival, they are not ready for clinical applications at the moment because they are yet to be standardized and validated [8–10]. The advantages and disadvantages of these methods are provided in Table 4.3.

4.1.1.4 Variability of PET-Derived Quantitative Metrics

The first prerequisite to reliably measure a PET-derived tumor volume is to assure a robust and reproducible method to accurately determine SUV-based parameters, overcoming the above-mentioned sources of error related to physical, technical, and biological factors [7]. In particular, all quantitative PET metrics are affected by user-defined factors including image acquisition settings, i.e., duration of acquisition, thickness of

Table 4.3 Pros and cons of the various categories of PET image segmentation techniques

Category	Characteristics	Limitations
Manual techniques	Visual interpretation and manual delineation of contours	Time consuming. Susceptibility to window level settings
	Very simple to use. Tools to transfer RT objects to treatment planning systems available from most vendors	Suffer from intra- and inter-observer variability. Consensus reading by nuclear medicine physician and radiation oncologist hardly practical in busy clinical departments
Thresholding techniques	Most frequently used due to their simple implementation and high efficiency	Hard decision making. Too sensitive to PVE, tumour heterogeneity and motion artifacts. Some methods focus on volume, others focus on intensity differences. Combination of both seems to provide best results [92]
Variational approaches	Subpixel accuracy, boundary continuity and relatively efficient. They are mathematically well developed and allow for incorporation of priors such as shape	Sensitive to image noise. As a PDE, stability and convergence could be subject to numerical fluctuations, especially if the parameters are not properly selected
Learning methods	Utilize pattern recognition power. Two main types: supervised (classification) and unsupervised (clustering)	Computational complexity especially in supervised methods, which require time-consuming training. Feature selection besides commonly used intensity is a flexibility but cab also be a challenge
Stochastic models	Exploit statistical differences between tumour uptake and surrounding tissues. Most natural to deal with the noisy nature of PET	Effect of initialization and convergence to local optimal solutions are concerns, especially when compromises are made to improve efficiency

Extracted by *Eur J Nucl Med Mol Imaging* (2010) 37:2165–2187, table 2, p. 2175 [32]

the slice, acquisition mode (2D vs. 3D), reconstruction algorithm, and the correction herein applied, i.e., attenuation, scattered and random coincidences, and dead time correction. To minimize SUV variability, it is necessary to cross-calibrate the PET scanners and ancillary instruments. Though cumbersome, this approach proved effective in increasing the accuracy of tracer uptake measurements, reducing inter-scanner variability of the measured activity to 5–10 % [7, 11–16], which is a major achievement, compared to 10–25 % variation observed even in a controlled environment of a multicenter clinical trial [17].

4.1.1.5 PET Test–Retest Reproducibility

Reported variability of SUV in patient test–retest studies differed from the desired range of ≤10 % [18–23]. The largest repeatability study of 62 patients with gastrointestinal malignancies reported an intra-subject coefficient of variation decrease in SUV measurements from 16 % to 11 % after applying a centralized quality control assessment and analysis [24, 25]. These studies showed that the variance of SUV is greater in clinical practice than in clinical trials even in a single site experience: the threshold criteria for a difference of a second scan in respect to baseline at 95 % confidence level were 49 % and 44 % for SUV_{max} and SUV_{mean}, respectively. A recent meta-analysis by De Langen et al. showed that SUV_{mean} had a slightly better repeatability than SUV_{max}, with a better reproducibility in larger lesions [26]. However, a recent study comparing SUV_{max}, SUV_{mean}, SUV_{peak}, and TGV found that different SUV definitions yielded 20 % variation in tumor response values for an individual tumor and variation of up to 90 % for a single SUV measurement [27]. Another study showed that mean percentage difference in SUV_{max} measurements in 100 patients with a known chest lesion obtained on subsequent scans was 0.9 ± 7.8 with a coefficient of variation of 4.3 % [28]. This variability was much lower than that reported in previous studies with a range of 2.5–8.2 % [7, 11, 29]. Besides SUV, Leijenaar et al. [30] demonstrated a high test–retest reproducibility of various radiomics features as well as a high (91 %) interobserver variability. Based on the results of these studies, minimal protocol variation should be ensured when performing repeated scans on the same patient required to improve the reliability of SUV measurements.

4.1.2 Segmentation Methods for Volume Calculations

Different segmentation techniques for PET-derived volumes have been proposed with a varying complexity (Table 4.3). Hence, comparing the performance of different methods from published data is almost impossible given the variety of algorithms used and degree of operator manipulations [31, 32]. To date, there is no consensus on a reproducible, accurate, and practical method that should be preferred for tumor segmentation. The existing methodologies are described in the following paragraph.

Manual Technique The manual contouring by an experienced imaging expert is the first methods applied in this field and it is still widely used. However, this procedure is cumbersome, and time consuming, particularly in patients with disseminated disease. This method is technically least sophisticated but economically less demanding and expectedly leads to significant interobserver variability in the range of 5–137 % [33].

Thresholding Method The most widely used method to define a tumor volume is the thresholding method that requires identification of voxels exceeding a predefined threshold [34]. The thresholding can be performed using fixed or adaptive methodologies. In general, application of the proper threshold technique is a challenging task because of the limited resolution of PET images. Blurring due to partial volume effect [35] (Fig. 4.4) or motion artifacts and noise fluctuations due to limited photon counts can degrade segmentation accuracy.

Percentage threshold The earliest thresholding method was based on a percentage SUV, mainly using a cutoff of 40–50 % of the SUV_{max} [36].

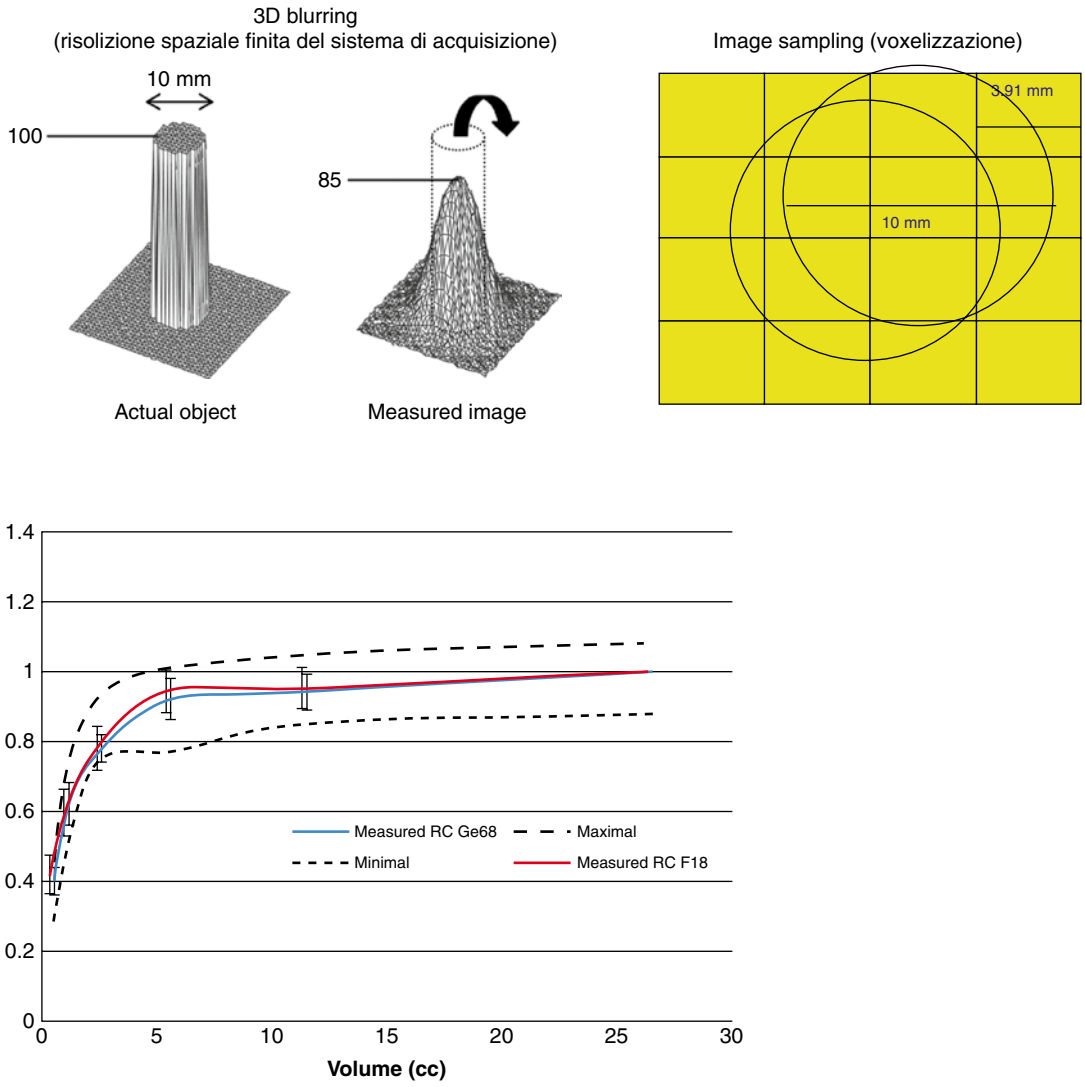

Fig. 4.4 Partial volume effect. Partial volume effect refers to both image blurring due to scanner finite spatial resolution (*left* in figure) and to voxel sampling (*right* in figure). It affects small lesions and is negatively affected by tumor heterogeneity. The SUV_{max} and SUV_{mean} measurements in a lesion volume of 2.5 ml (about 1 cm diameter) could be underestimated by up to 50 % (Fig. 4.2), and complete recovery of the actual SUV is done for lesions greater than 5 ml. Many strategies have been developed in the past to correct for partial volume effects but none of them reached a daily practice maturity [19]. Only recently, new algorithms have been applied directly to reconstruction algorithms in modern scanners. It should be emphasized that small tumor volumes do not necessarily imply small number of cells since the tumors become visible at about 10^5–10^6 malignant cells considering the resolution limits of the PET scanners

This method was simply based on phantom studies of static spheres. Subsequently, a value of 40 % was adopted by several groups for tumor delineation in radiotherapy planning of non-small cell lung cancer (NSCLC) [37], cervical cancer [38], and head and neck squamous cell carcinoma (HNSCC) [39]. The principal drawback of this method is that the optimal threshold is influenced by the size of the tumor volume; the surrounding background is not taken into account and is often "scanner specific" because of the strong dependence on the spatial resolution of the instrument. Based on the available data suggesting an insufficient

tumor coverage using fixed thresholding methods, this method was no longer recommended, particularly for RT planning purposes [40].

Fixed threshold As an alternative method, an absolute SUV threshold can be used for tumor segmentation. However, tumor inhomogeneity and motion artifacts may hinder the application of this approach by failing to provide adequate tumor delineation in nearly half of the cases, in particular for lesions showing a low tumor-to-background ratio [41]. Moreover, fixed thresholding techniques take neither the background nor the tumor size into consideration [42] thus being inappropriate to define a tumor volume.

Adaptive threshold To address the background-dependent variability, some investigators suggested adapting the threshold to tumor-to-background (TBR) ratio [43, 44]. Subsequently, a more developed system based on an iterative technique was introduced to optimize the thresholding for the TBR approach [44–47]. The rationale is to change TBR threshold iteratively till when an optimal threshold is generated by the convergence algorithm. This is a reasonable and logic approach. However, the coexistence of several operator-dependent thresholding methods, based on different morphologic aspects of radiotracer concentration in tumors, justifies the search for an automatic threshold computing software.

Gradient technique This technique measures gradient differences between the lesion and the surrounding background with a good spatial accuracy and efficiency [48, 49]. Gradient methodology includes simple edge or ridge detectors [50] or watershed method [51]. More recently deformable active contour models have been applied to PET segmentation with the assumption that contours are characterized by sharp variations in the image intensity [52, 53]. Despite being intuitive, the gradient technique suffers considerably from image noise and often requires filtering of the images with a blurring effect [54].

More sophisticated techniques To overcome all the difficulties originating from thresholding and gradient techniques, several authors have explored more sophisticated approaches used in other science domains such as active deformable models, learning methods, and stochastic approaches [55] and those using a pattern recognition algorithm [56]. Learning methods based on classification require training of the method moving from data with known labels (known ground truth). However, this is a challenging task due to variability of PET tracer uptakes and biodistribution in tissues, which in turn depends on the biomarker concentration in the blood (e.g., glucose concentration for FDG) and other technical factors. In addition, PET images need to be properly drawn to identify the ground truth for training purposes (e.g., the structures contoured by a panel of experienced radiologists). Therefore, behind the ground truth, the application of these methods requires a number of other information with a thoroughly checked source. Stochastic models offer the advantage of incorporating the variable of the voxel's intensity directly into the model. However, these models are based on a proper predefined noise model, which has not been yet defined for PET and is strongly influenced by the parameters and type of the reconstruction algorithm. In general, the Gaussian assumption is used because it simplifies the computational burden and speeds up convergence.

Comparison between methods Reproducibility is a key issue associated with segmentation methods. Different methods give rise to variations in the calculated PET volumes in the range of 40–50 % [9, 31, 58, 59], and this variability can even reach 400 % [32]. The performance of tumor delineation methods, in turn, largely depends on variations in the TBR, image resolution, and image noise level. Evaluating the accuracy of the segmentation methods is rather difficult because it is virtually impossible to rely on a ground truth as comparator. Studies have been proposed using phantoms, morphological images (CT or MRI), and pathology specimens [57], but there is no consensus among scientists on the optimal method. Despite the heterogeneity of clinical behavior and aggressiveness of the malignant processes, there is preliminary evidence to suggest

that MTV and TLG have independent prognostic value across different types of cancers, including lymphoma [60]. It is therefore important to pursue validation studies to establish the real value of these methodologies and also prove their reproducibility in large prospective data sets.

4.1.2.1 Applications in Radiation Oncology

Radiation therapy (RT) is one of the pillars of combined-modality treatment for the Hodgkin lymphoma. Successive technological progresses achieved over the past decade have revolutionized the definition of the target tumor volume and the boundaries of the radiation field. These new methods increased the effectiveness of this treatment modality which delivered much smaller doses to critical organs such as the lung, heart, and breast [61]. RT treatments can be classified as total lymphoid/nodal irradiation (TLI/TNI), extended field RT (EFRT), involved-field RT (IFRT), and involved node RT (INRT) (see Chapter 5). In the modern era of conformal radiotherapy, TNI and the EFRT are no longer in practice and supplanted by limited-field radiation therapy: IFRT, if the RT field encompasses all of the clinically involved nodal regions, and INRT, with an assumption to deliver the dose only to the initially involved nodes, rather than including the entire region of the involved nodal chain. Consequently, the current guidelines for combined chemotherapy and RT indicate that the delineation of the target volume should always be carried out on the affected regions [62, 63].

Field delineation in RT planning is one the most important applications of PET/CT imaging (see Chapter V). In recent years, a large number of studies and methodological research projects were performed to develop and validate automatic and semiautomatic algorithms for accurate and robust delineation of RT target volumes. So far only a few clinical trials have been conducted in which dose escalation was prescribed on an FDG avid area within the GTV [64–67].

Recent studies proved high observer variability in clinical target volume (CTV) delineation for HL [68–70], thus, highlighting the need for a robust and operator-independent methodology for target definition. A considerable improvement in treatment volume definition on simulation CT has been obtained by integrating the information provided by the FDG-PET/CT, acquired before chemotherapy for diagnosis and staging purposes [67, 71–74]. In order to combine the FDG-PET/CT outcome with the CT-based CTV delineation, the common practice is the visual assessment. Briefly, the physician compares the two imaging modalities displayed on two different screens and confirms the matching on anatomical landmarks. However, this approach is time consuming and operator dependent. Some authors proposed methods based on rigid image coregistration and overlay (image fusion) highlighting favorable results if the FDG-PET/CT is acquired in the treatment position [73–75].

Dedicated PET/CT planning is already available in some centers, but care must be taken when fusing diagnostic and planning scans because of the need for a deformable registration, which is yet to become a standardized procedure. Nonetheless, there are practical obstacles in routine practice such as the scanning position of the patient (position of the arms and/or neck) and the use of different scanners. In addition, weight loss and lymph node shrinkage occurring between the two imaging stages represent particularly challenging issues for PET/CT matching based on rigid registration.

Similar to other cancers, PET/CT manual contouring is the standard technique in lymphoma [76]. To increase reproducibility, the use of a flat table for PET/CT imaging is advisable. Due to the relatively simple geometry of the lymphoma lymph nodal masses in axial CT and PET/CT sections, a PET segmentation algorithm has been rarely used instead of manual contouring for RT planning.

4.1.3 Conclusions

There is a large variability in computational complexity and level of user interaction required by the various image segmentation techniques. In the near future, the development of more sophis-

ticated and robust tools for PET segmentation will probably help physicians to use these quantitative methods with higher precision and accuracy. However, it is imperative to adopt standardized acquisition, reconstruction, and analysis protocols for the clinical use of PET quantitative metrics.

4.2 Clinical Applications in Lymphoma

4.2.1 Why Should Quantitative Methodology Be Preferred Over Qualitative?

The widely utilized anatomic imaging parameters rely on tumor size change as a measure for treatment response. Nevertheless, functional imaging lends itself as a better surrogate metric for demonstrating a biological tumor response. Although visual assessment of FDG-PET/CT has been successfully integrated into clinical practice for therapy monitoring, high rate of false-positive results even in the hands of expert readers have raised concerns [77–81] for its usefulness, particularly, for interim PET-adapted therapeutic strategies. With the emphasis on the liver as a reference background adopted by D 5PS criteria [82], the inter-patient variability and intra-patient fluctuations of hepatic FDG uptake during therapy [83–85] have become a focus of concern. More importantly, the depth of tumor response categorization by visual criteria may lead to suboptimal differentiation between response categories by oversimplification. Furthermore, visual assessment is proven to be a reproducible and efficacious method for treatment response assessment in HL [159, 160] and FL [179] but its role is less substantiated with the currently available data in other lymphoma subsets [153]. Quantitative analysis allows for an objective assessment of treatment response, thereby minimizing interobserver variations and more suitable for a continuous measure of response which is also one of the most effective ways to reduce sample size [86]. In order to minimize potential treatment-associated morbidity, and unnecessary interventions, the tumor metabolic response can be used as a practical early clinical end point to substitute survival end points, which may counteract the high cost and lengthy process attendant with the regulatory approval of the novel drugs. Functional imaging provides an earlier and faster readout for treatment response compared to morphologic imaging; consequently, it is preferable for early and accurate evaluation of the efficacy of novel treatments. With the recent insurgence of sophisticated software programs, tumor volumes can be determined with much less effort than otherwise. Thus, MTV as a measure of the viable tumor fraction or TLG, as a product of MTV and mean SUV within the volume, may better predict ultimate patient outcome than anatomical imaging either at baseline or early during therapy. MTV is and may better estimate tumor burden. Hence, there is a strong interest in the development of various quantitative metabolic PET metrics in an effort to decrease the rate of false-positive results, increase reproducibility, and maximize statistical power.

4.2.2 PET-Derived Quantitative Metrics in Clinical Practice

4.2.2.1 Standardized Uptake Values (SUV)

As alluded in the previous section, SUV_{max} has been investigated as a quantitative PET parameter to provide an objective measure for assessing tumor metabolic activity in tissues.

Baseline Tumor Characterization

The advent of genomic and proteomic technologies have been shifting traditional cancer management toward an individualized treatment strategy. However, these methods are impractical in a routine setting and do not allow for a complete characterization of the tumor because tumor tissues are spatially and temporally heterogeneous. Noninvasive assessment of tumor behavior with the use of imaging may provide a more comprehensive guidance for improving therapy decisions in cancer patients. Among all indications, differentiation between a malignant and

benign etiology or a low-grade phenotype from that of a high grade using an objective imaging tool would be clinically desirable. In this regard, although limited and not validated, the existing published data showing correlation between the SUV_{max} and tumor histologic characteristics, surgical stage, and prognosis are summarized in the following section.

Diagnosis of different tumor phenotypes Considering the need for a more aggressive treatment for transformed low-grade lymphomas (LGL) compared to LGLs [87], early identification of transformation to an aggressive phenotype would be clinically consequential. There is sufficient evidence that FDG-PET/CT can detect transformation of chronic lymphocytic leukemia (CLL) to diffuse large B-cell lymphoma (DLBCL), the so-called Richter's transformation [88–91]. In a retrospective study by Bruzzi et al. ($n=37$), SUV_{max} of >5.0 was considered highly suggestive of Richter's transformation with an overall sensitivity and negative predictive value (NPV) of 91 % and 97 %, respectively [88]. Recently, Falchi et al. evaluated and reported that SUVmax ≥10, international prognostic score (IPS) ≥2, bulky disease, and age ≥65 were independently associated with shorter OS in CLL patients ($n=332$) [89]. SUV_{max} ≥10 strongly correlated with overall survival (OS) (OS: 57 vs. 7 months). Corroborating these results, Michallet et al. identified a threshold of tumor SUV_{max} >10 as the most effective discriminating cutoff value which yielded a sensitivity and specificity of 91 % and 95 %, respectively, for identifying transformation by PET in CLL patients ($n=250$) [90].

The transformation to large B-cell aggressive lymphoma is also a critical event for patients with follicular lymphoma (FL), which warrants a more aggressive therapy approach than de nov*o* FL. The value of FDG-PET/CT diagnosing transformation has been well established for guiding lymph node biopsy when transformation is suspected. Although there is lack of consistency for defining an exact SUV_{max} cutoff, a transformation is suggested at a SUV_{max} of 10–15 [91–96]. But it should be emphasized that thresholds indicating transformation should be investigated in homogeneous patient cohorts because the cutoff value will be different for different subtypes of indolent lymphomas [94]. Because proliferation is a hallmark of transformation, 3′-deoxy-3′-[^{18}F]fluorothymidine (FLT), as a specific surrogate for proliferation [97], is hypothesized to be superior to FDG for early detection of progression to a more aggressive histology (see Chapter 1: the newer tracers). Nonetheless, there are conflicting reports and this premise has not yet been proven [95, 98]. In a comparative study ($n=26$) by Wondergem et al., the ability of FDG to discriminate between FL and transformed FL was superior to that of FLT with a SUV_{max} of 14.5 aiming at 100 % sensitivity with a maximum specificity (82 %) [95]. At the optimal sensitivity, the specificity of FLT was only 36 % that would imply an unacceptably high proportion of patients requiring a biopsy to exclude transformation. The poor performance of FLT begs the question of its specificity for cell proliferation or Ki-67 expression. Therefore, the clinical impact of FLT remains to be determined in ongoing research studies.

The nodular lymphocyte predominant HL (NLPHL) is an uncommon subtype that invariably expresses CD20 with excellent OS, but unlike classical HL (cHL), late relapses may occur. In addition to staging and response assessment, determination of a disparate phenotype may be clinically relevant to because NLPHL has a propensity to be associated with concurrent or transformation to an aggressive B-cell non-Hodgkin lymphoma that would require long-term follow-up and image-guided rebiopsy. Hence, recognizing the imaging features of this entity is important. NLPHL is FDG avid, although SUVs are generally lower than those observed in cHL [99, 100]. A study by Hutchings et al. (n = 60) found that the mean SUV_{max} was 8.0 vs. 11–15 for cHL, $p=0.002$ [99]. In a retrospective design ($n=12$), NLPHL patients were also found to have lower FDG SUV_{max} compared to those with T-cell/histiocyte-rich large B-cell lymphoma (THR-LBCL) (mean SUV_{max}, 6.9 vs. 16.6, $p=0.055$) [101].

Tumor heterogeneity The spatial and temporal tumor heterogeneity limits the accuracy of tissue-based molecular assays. However, algorithms of image characterizations may capture intratumor heterogeneity as a signature of gene expression patterns, particularly, with the use of quantitative methods [102, 103]. The heterogeneity of tumor morphology largely accounts for an idiosyncratic treatment response within a single or across different neoplastic disorders. Genetic and epigenetic differences between cancer cells within a tumor might explain why some tumor cells are resistant to therapy, while others are sensitive and can be eradicated after an effective treatment.

Radiomics is an emerging field and refers to the comprehensive evaluation of the entire tumor volume using quantitative image evaluation of tumor phenotypes [102, 104, 105]. Recently, the data published by Aerts et al. suggested that radiomics decoded a general prognostic phenotype existing in multiple cancer types by revealing associations with the underlying gene expression patterns [106]. In one series of mixed cancers including DLBCL, integrating image textural features with SUV measurements significantly improved the prediction accuracy of morphological changes (Spearman correlation coefficient=0.87, $p<0.0002$) [107]. Some of the textural image features (such as entropy and maximum probability) were superior in predicting morphological changes of radiotracer uptake regions longitudinally, compared to SUVmax. In another pilot study, voxel distribution of FDG uptake demonstrated no significant differences in the heterogeneity indices between responders and nonresponders, while the heterogeneity of the intratumoral distribution of ^{111}In-ibritumomab tiuxetan was correlated with the tumor response in this cohort of 16 NHL patients [108]. In this study, pre-therapeutic FDG SUVmax was predictive of the tumor response to ^{90}Y-ibritumomab tiuxetan therapy on a lesion-by-lesion basis. This result is consistent with a previous report [109], while in another prior report, pre-therapeutic FDG SUVmax was not predictive of the tumor response to ^{90}Y-ibritumomab tiuxetan therapy [110]. This may be because of the small number of patients and different analysis methods. Nonetheless, in radionuclide therapy, the nonuniformity of the absorbed dose by the tumor may be a key issue for treatment success or failure. Pre-therapeutic FDG SUVmax in combination with heterogeneity of ^{111}In-ibritumomab tiuxetan might enhance the predictive values for tumor response and long-term outcome, which will be clarified in further studies. Radiomics may have a large clinical impact providing a wealth of extractable additional information that can be quantified for monitoring phenotypic changes during treatment. However, it is still in an early phase of development, and there are multiple technical issues that still need to be streamlined and validated to prove its clinical relevance.

Assessment of Bone Marrow Involvement (BMI)

Although it is widely recognized that a unilateral iliac crest BMB could underestimate lymphoma infiltration, bone marrow biopsy (BMB) has been the standard conventional method to evaluate bone marrow (BM) involvement in lymphomas (see Chapter 1: the need for bone marrow biopsy). However, BMB is associated with complications such as bleeding, anxiety, and pain [111, 112]. To overcome these disadvantages, the high sensitivity provided by whole body PET/CT imaging is exploited for effectively diagnosing BMI. According to the new Lugano guidelines, if a PET/CT is performed, a BMB is no longer required for the routine evaluation of patients with HL because of the low incidence of BMI [113, 114]. In DLBCL, if the scan is negative, a BMB is indicated to identify involvement by discordant histology if relevant for a clinical trial or patient management [113, 115, 116]. Several studies investigated whether visual and quantitative PET-based BM assessment can replace blind BMBs in various lymphoma subtypes.

Non-Hodgkin Lymphoma Adams et al. reported the inability of FDG-PET/CT to replace BMB in newly diagnosed DLBCL because PET-based BM assessments, including SUVs, were prognostically inferior to BMB ($n=78$). Multivariate analysis

showed that only BMB status was an independent predictive factor of PFS ($P=0.016$ and OS $P=0.004$) [117]. The design of this study, however, was not optimal because of retrospective analysis and the use of BMB as the only reference standard for the diagnosis of BMI, which only allowed for the calculation of patient-based sensitivity of FDG-PET/CT. The same group of investigators subsequently reported that head-to-head comparison with BMB, the diagnostic value of both visual and quantitative PET/CT for the detection of BMI, is low in a cohort of 40 DLBCL patients [118]. The SUV_{mean}, SUV_{max}, and SUV_{peak} of BMB-negative patients (1.4±0.49, 2.2±0.69, and 1.7±0.59, respectively) considerably overlapped with those of BMB-positive patients (1.8±0.53, 2.7±0.71, and 2.2±0.61, respectively).

Contrary to these results, in patients with FL, quantitative PET analysis was more beneficial in diagnosing BMI than visual assessment in a preliminary study of 22 patients. Optimal SUV_{max} cutoff of 2.1 yielded sensitivity and specificity combinations of approximately 87% [119]. In another study, of 41 patients with grade 1-3a FL and diffuse BM uptake, using a SUV_{mean} cutoff of ≥2 resulted in approximately 30% improved sensitivity at no cost to specificity. Moreover, using the ratio SUV_{mean}/MBP ≥1, the sensitivity of PET/CT to detect BM involvement improved to 83% [120]. As a limitation, this study was retrospective and included both staging and restaging patient groups which added heterogeneity to the data.

Hodgkin lymphoma Although the value of qualitative analysis and the rareness of BMI in HL have been addressed previously, several studies investigated the added value of a quantitative PET approach in the detection of BMI by HL. SUV_{max} evaluation did not have an incremental value to the visual evaluation in a retrospective study included 26 HL patients [121]. In another retrospective study of 106 HL patients, Salaun et al. reported that multivariate analysis revealed an independent correlation between sacral SUV_{max} and Ann Arbor stage ($p=0.005$). No BMI was found in patients who presented with SUV_{max} below 3.4 [122].

In summary, because the qualitative interpretation of PET may be marred by the physiologic accumulation of FDG within the BM, there is a need for an objective whole body technique to yield quantifiable results that may simulate BMB. At first glance, the distinction between these potentially overlapping conditions may be easy, considering that only focal FDG uptake is considered to represent BMI in HL [113, 114]. However, this distinction is challenging in NHL where BMI can present with both focal and diffuse patterns of FDG uptake [116, 117]. In this regard, development of a quantitative PET approach may be particularly relevant in patients with newly diagnosed NHL. However, a number of unsettled issues still exists, i.e., what extent of increase in BM uptake should be considered suggestive of BMI, if this increase could be quantifiable how should it be corrected by the actual BM volume that individually varies from one patient to another, how to factor in the differences in the BM volume in different parts of the body, and, finally, what would be the methods to minimize an overlap between reactive BM hyperplasia and diffuse BMI. With the wealth of available software programs, further work is underway to address these viable concerns to determine the actual role of a quantitative PET approach.

4.2.2.2 Quantitative PET-Derived Metrics Beyond SUVs

As discussed at length in the previous section, SUV can be biased by the count variability and tumor heterogeneity in a volume of interest because of the reliance on a single voxel measurement. Furthermore, besides the anatomic finding of high tumor burden in a disseminated disease, which is frequently recorded at baseline in lymphoma, a methodology able to assess and quantify the metabolic activity of a given tumor burden would be more clinically relevant. In an effort to reduce bias, increase reproducibility, and improve the predictive value of PET results, functional volume parameters, i.e., metabolic tumor volume (MTV) and tumor lesion glycolysis (TLG) have been under investigation [1, 8–10].

Prognostic Value of PET-Derived Quantitative Metrics at Baseline

If the baseline whole body disease volume is proven to be an independent prognostic factor, high-risk patients may be objectively identified for treatment intensification. However, there is paucity of clinical data for the establishment of a prognostic system that is based on pre-therapy quantitative PET metrics affecting clinical outcomes of lymphoma patients. The available literature in both HL and NHL is discussed in the following section and summarized in Table 4.4.

Hodgkin lymphoma Tumor bulk is a significant negative prognostic factor in early-stage HL [113, 123–125]. However, not only the exact definition of tumor bulk remains a controversial topic but also an objective method to measure whole body tumor burden is yet to be established for a patient-tailored management. Thus far, the practice has relied on the indirect measures of tumor burden, i.e., the extent of involved sites used by the Ann Arbor staging system, and integrated factors including number of disease sites, stage, and LDH used by the prognostic systems including the international prognostic score (IPS) to stratify risk categories [126–129]. In a prior study of HL patients treated on standard protocols, the mean tumor burden normalized to body surface area based on CT measurements was found to be largely superior to all prognostic models as a predictor of complete remission and survival [124, 125]. Given the coverage of the entire body, metabolic volume determination may be a better surrogate for response and survival by representing overall tumor functionality.

Several retrospective studies using various methodologies calculating the tumor volume showed that there may be a benefit to use PET quantitative metrics to predict survival [130–132]. In a study by Song et al. in 127 early-stage HL patients (20 % bulky) treated with six cycles of ABVD, with or without involved-field radiotherapy (IFRT), the multivariate analysis showed that only older age, B symptoms, and high MTV status were independently associated with PFS and OS (PFS, $p=0.008$; OS, $p=0.007$) [130]. In this study, a fixed threshold method of $\geq SUV_{max}$ 2.5 was used to determine the disease volume. In another single-center study, Kanoun et al. showed that pre-therapy MTV was predictive of patient outcomes in a cohort of 59 HL patients (92 % stage II–IV, 60 % IPS > 2), who were treated with an anthracycline-based therapy with or without IFRT [131]. The patients with a low MTV had a significantly better 4-year PFS than those with a high MTV (85 % vs. 42 %, p = 0.001, 88 % vs. 45 %, $p=0.0015$, respectively). MTVs were measured with a semiautomatic method using a 41 % SUVmax threshold. In multivariate analysis only baseline MTV ($p<0.006$, RR 4.4) and ΔSUV_{max} at PET2 (71 %, $p=0.0005$, RR 6.3) remained independent predictors of PFS when tumor bulk (≥10 cm) did not reach statistical significance. In contrast to these findings, Tseng D. et al. reported that at a median follow-up of 50 months, baseline absolute PET metrics including SUV_{max}, SUV_{mean}, and MTV did not predict survival in 30 HL patients (stage IIB-IV 63 %, 30 % IPS > 2) treated with varying chemotherapy regimens with or without IFRT when IPS was associated with PFS ($p<0.05$) and OS ($p<0.01$) [132]. On the contrary, the ΔMTV ($p<0.01$), ΔSUV_{max} ($p=0.01$), and ΔSUV_{mean} ($p<0.05$) at interim PET were associated with PFS and OS. This divergent result compared to others may be on the basis of a small patient cohort and the differences in methodologies, patient population, (stage, risk factors) and therapy protocol. However, all of the above reviewed studies had suboptimal designs marred by the retrospective design, which was inherently prone to biases because of non-standardized protocols and patient preparation (see previous section). Also the use of various segmentation methods and resultant MTV cut-offs that varied between 200 and 500 ml led to non-comparable and non-generalizable results. Moreover, a fixed threshold that was used by all of these studies is not considered optimal for volumetric assessment as discussed in the previous technical section.

In a retrospective analysis of prospectively acquired data in 89 cHL patients whose findings were reported previously by Hutchings et al.

Table 4.4 Published studies in lymphoma using metabolic tumor volume as a measure of outcome

	Patients	No. pts		Multicenter	Mono, multi or equalized scanners	Therapy	PET time	Segmentation method	Segmentation performed by	Cutoff
Kanoun et al. [131]	HL, excluding nodular lymphocyte predominant lymphoma	59	RE	No	MU (2 scanners)	4–6 cycles of an anthracycline-based chemo plus 20–36 Gy of IF-RT	Baseline, interim[2]	41 % SUV_{max} threshold, manual adjusted	2 blinded experts, consensus if discrepant	MTV 225 cm^3
Song et al. [130]	Early HL	127	RE	Yes	MU	6× ABVD plus 30 Gy of IF-RT plus 10 Gy on initial bulk	Baseline	Visual and fixed threshold of 2.5 SUV	Locally and reviewed centrally by an expert	MTV 198 cm^3
Tseng et al. [132]	Early and advanced HL,	30	RE	No	MO	Stanford V, ABVD, VAMP, or BEACOPP plus RT (20 Gy, 25.5 Gy, or 36 Gy)	Baseline, interim[2]	Region-growing algorithm[69]	–	NS
Hussien et al. [162]	Pediatric HL	54	PRO	Yes	EQ	GPOH-HD2002P, GPOH-HD2003, EuroNet-PHL-C1 plus IFRT	Baseline, interim2	D 5PS and fixed threshold of 2.5 SUV (body weight, body surface) and at a threshold of mean liver plus two standard deviations SUV (lean body mass)	Two blinded experts	
Esfahani et al. [137]	DLBCL	20	RE	No	MO	6× or 8× R-CHOP	Baseline, interim[2]	Threshold 1.5 liver SUV_{mean} plus 2.5 standard deviation of liver SUV. Contours were manually adjusted in case of tumor exceeding contrast-enhanced CT volumes	Two blinded experts	MTV = 379. and 5.95 TLG = 705 and 96.5 at baseline and at interim PET2, respectively

Kim et al. [141]	DLBCL	140	RE	No	MO	6–8 cycles of R-CHOP plus 36 Gy of radiotherapy to bulky disease	Baseline	Visual and a percentage threshold at 25 %, 50 %, and 75 % of SUVmax	Three experts	TLG (505) = 415.5
Sasanelli et al. [140]	DLBCL	114	RE	Yes	MU	R-CHOP21, RCHOP14, and ASCT	Baseline	41 % SUVmax threshold	One expert, subset of 50 by another	MTV 550 cm^3 TLG -
Gallicchio et al. [145]	DLBCL	52	RE	No	MO	R-CHOP		Visual and percentage threshold of 42 %	Three blinded experts in consensus for visual and subsets of 18 patients in double for segmentation	MTV 16.1 cm^3 TLG 589.5
Adams et al. [146]	DLBCL	73	RE	No	MO, EQ with NEMA/IEC IQ	R-CHOP	Baseline	40 % of the SUVmax	Single blinded expert	MTV 272.3 cm^3 TLG 2955.4
Tateischi et al. [170]	Relapsed or refractory DLBCL	55	PRO	Yes	EQ [27] with NEMA/IEC IQ phantom	Bendamustine–rituximab	Baseline, interim2, and EoT	Visual with D 5PS and fixed threshold of 2.5 SUV	Two blinded experts, third if discrepant	Δ66–68 % at interim 2 and Δ61–66 % at EoT for MTV and TLG
Malek et al. [167]	DLBCL	140	PRO			R-CHOP or R-DA-EPOCH	Interim (2–4 cycles)	D 5PS, a 37 % threshold of SUVmax and a gradient technique method	One expert	ΔSUVmax > 72 % and ΔMTV 52 %
Ceriani et al. [151]	PMBCL	103	PRO	Yes	MU	R-CHOP and R-VACOB-P and IFRT	Baseline	25 % threshold of SUVmax, manually corrected	One expert centrally	MTV =703 cm^3 , TLG 5814

No. pts number of patients, *MTV* metabolic tumor volume, *TLG* total lesion glycolysis, *D 5PS* Deauville 5-point score

[133], during a median follow-up was 52 months, no baseline clinical parameters correlated with PFS but both baseline and interim quantitative PET parameters correlated with PFS [134]. The MTV was the strongest predictor of PFS at baseline ($p=0.002$) and D-5PS at PET1 ($p<0.0001$) (unpublished data). However, these data were obtained in a mixture of early- and advanced-stage patients, with as much as 54 % of the original series of 126 patients having a limited-stage disease (IA-IIB (Fig. 4.5)). Further investigations should include a more homogeneous data for definitive conclusions on the role of quantitative PET in the determination of HL outcomes. In view of the existing promising data, there is a need for more prospective large datasets to definitively determine the complementary or independent role of quantitative FDG-PET metrics at baseline for predicting prognosis and guiding treatment decisions in cHL.

Diffuse large B-cell lymphoma For NHL, there are no universally accepted or validated criteria for defining "bulky" disease, although 6 cm was suggested as the best cutoff for FL [135] and 6–10 cm for DLBCL [136]. A more streamlined and objective tumor burden measure would be preferred to better guide management. The pre-treatment FDG-PET metrics have been investigated as a potential predictor of survival in patients with DLBCL treated with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone (R-CHOP) [137–146]. In a retrospective study of 169 patients with stage II–III (74 % IPI ≤2) de novo DLBCL, prior to R-CHOP therapy (6–8 cycles), Song et al. found in a multivariate analysis that the whole body tumor burden was a more important prognostic parameter for PFS than Ann Arbor staging (HR = 5.3; OS, HR = 7.0, both $p<0.001$) [138]. MTV was defined with a thresholding intensity based on SUV_{max} ≥2.5. With a median follow-up of 36 months, the 3-year estimates of PFS and OS were significantly higher in the low MTV than in the high group (PFS, 90 % vs. 56 %; OS, 93 % vs. 58.0 %, both $p<0.001$). The same group of investigators found similar results in 165 early-stage (71 % IPI ≤2) primary gastrointestinal DLBCL patients [147]. During a median follow-up of 37 months, MTV was a better predictor of survival than SUV_{max} as determined by the receiver operator curve (ROC) analysis (0.92 vs. 0.70). Multivariate analysis revealed that a high IPI score ($p=0.001$) and high MTV ($p<0.001$) were independent prognostic factors for both PFS and OS, while other known prognostic factors were not significant. In another study of 140 DLBCL patients who received R-CHOP therapy followed by RT to bulky disease, after a median follow-up of 28.5 months, the TLG at the threshold of 50 % ΔSUV_{max} was significantly associated with PFS and OS (HR = 4.4; $p=0.008$ for PFS and HR = 3.1; 95 % CI = 1.0–9.6; $p=0.049$ for OS) [141]. High IPI score and Ann Arbor stage of III/V did not significantly shorten PFS. Similarly, in a retrospective study of 114 DLBCL patients [140] enrolled in previously reported International Validation Study [148], Sasanelli et al., using a 41 % SUV_{max} threshold, found that MTV was the only independent predictor of OS ($p=0.002$) and PFS ($p=0,03$) compared with other pre-therapy indices including tumor bulk (≥10 cm), LDH, stage, and age-adjusted IPI. The 3-year estimates of PFS were 77 % in the low metabolic burden group and 60 % in the high metabolic tumor burden group ($p=0.04$), and prediction of OS was even better (87 % vs. 60 %, $p=0.0003$). TLG failed to predict PFS and was less predictive of OS than MTV, in contrast to prior results. This multicenter study, however, was flawed by the absence of a protocol harmonization and cross-calibration of scanners across participating canters, variability of therapy protocols, and also the lack of comparative analysis between volumetric results and SUVs. More lately, Kim et al. reported that the higher MTV inferred a significantly inferior EFS compared with the lower MTV group during a median follow-up of 28 months in 96 DLBCL patients who were treated with R-CHOP [142]. In this study, MTV was defined with a fixed threshold of 2.5. There was no difference in EFS between patients with stage II and III patients ($n=53$), but the higher MTV group showed significantly inferior EFS in this group of patients compared with the lower MTV group. Likewise, Xie et al. demonstrated that according

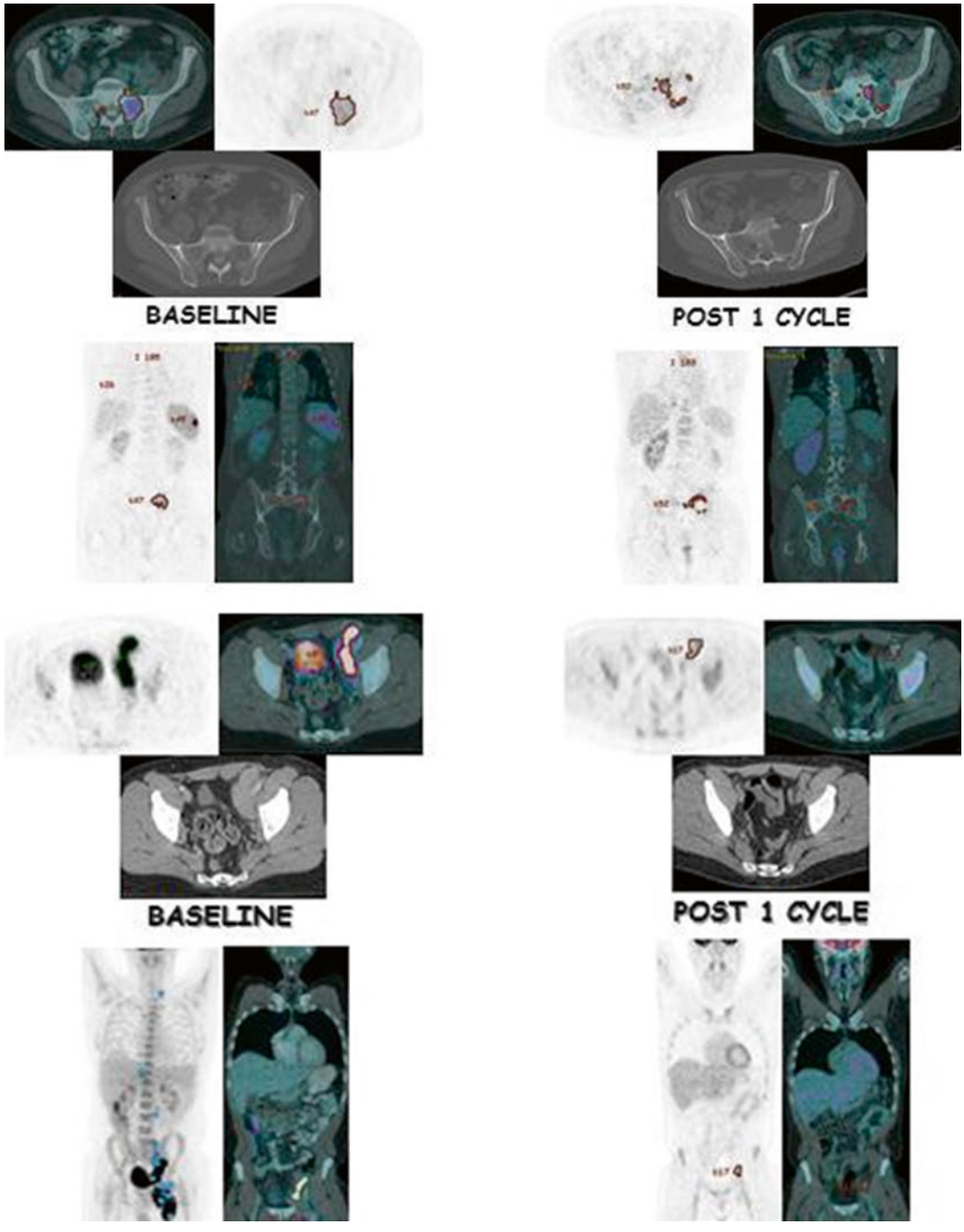

Fig. 4.5 Patient examples

to the cutoff determined from ROC analysis, lower MTV and TLG values prior to therapy were highly predictive of favorable PFS in DLBCL ($n=60$) [144]. The multivariate analysis determined that the MTV and TLG values and number of enlarged lymph nodes predicted PFS independent of the National Comprehensive Cancer Network International Prognostic Index (NCCN-IPI) score and lactate dehydrogenase (LDH) level.

There are several studies whose results contradict with the previously reported studies [145, 146]. Gallicchio et al. suggested that the baseline SUV_{max} was a better predictor of EFS ($P=0.0002$, HR 0.13) during a median 18-month follow-up than MTV and TLG in a study of 52 DLBCL patients with intermediate IPI scores, who were treated with R-CHOP [145]. Only the IPI score 3 was slightly but significantly associated with poor outcome. The metabolic volume was determined with a 42 % threshold. It is conceivable that patients with intermediate IPI score presenting high SUV_{max} would respond better since the magnitude of glycolytic activity rather than the amount of metabolically active burden appears to be the key determinant. Adams et al. retrospectively investigated the pretreatment PET/CT in 73 patients with newly diagnosed DLBCL who had undergone R-CHOP immunochemotherapy [146]. On univariate Cox regression analysis, only the NCCN-IPI was a significant predictor of PFS ($P=0.024$), and only the NCCN-IPI and MTV were significant predictors of OS ($P=0.039$ and $P=0.043$, respectively). Therefore, the NCCN-IPI was suggested to remain the most important prognostic tool in this disease.

Combined results of a systematic review of seven retrospective studies involving 703 DLBCL patients [149] which included some of the above referenced studies [137, 138, 140, 141, 143, 146] suggested that SUV_{max} and MTV are significant prognostic factors for PFS (HR 1.61; $p=0.038$ and 2.18; $p=0.000$, respectively). Similarly, high MTV and TLG values unfavorably influenced the 3-year OS (OR, 5.40 and 2.19, respectively). For OS, only high MTV was a strong predictor of poor prognosis in DLBCL with HR 2.99 ($p=0.000$). Overall this meta-analysis found that the outcomes of the included studies were inconsistent. Although the principle treatment protocol in six trials was R-CHOP [137, 138, 140, 143, 146, 150], there were inhomogeneous treatments in one trial conducted by Sasanelli et al. [140] with 55 % of patients had received R-CHOP only, 45 % of patients had received R-ACVBP, and an additional 18 % of patients had undergone autologous stem cell transplantation. Additionally, the use of different risk scoring systems also impacted the homogeneity of the analysis. Five trials used the old IPI scoring system for risk stratification [137, 138, 141, 143, 150], one trial used the age-adjusted IPI scoring system [144], and the other used the recently proposed NCCN-IPI scoring system [146]. Except one study by Oh et al. [150], most patients of six trials had low-intermediate or high-intermediate risk according to IPI system. Thirdly, the varying inclusion and exclusion criteria might have led to the heterogeneity of the results. Moreover, each study varied widely in the optimal cutoff values for survival prediction, with the cutoff values ranging from 11 to 30 for ΔSUV_{max}, from 220 to 550 ml for MTV and from 415 to 2955 for TLG. The trials also differed in the Cox proportional hazard regression methods. Moreover, the small number of patients might have influenced the reliability of results. These are collectively the probable reasons leading to the high heterogeneity of the combined results. When the outcomes from other ongoing trials are published, a further meta-analysis will be needed.

In a prospective cohort of 103 primary mediastinal large B-cell lymphoma (PMBCL) patients enrolled in the International Extranodal Lymphoma Study Group (IELSG), Ceriani et al., reported that only TLG retained statistical significance for both PFS ($P<.001$) and OS ($P=.001$) in a multivariate analysis, who received combination chemo-immunotherapy [151]. The MTV was estimated using a threshold method based on 25 % of the SUV_{max}, which was lower than other proposed thresholds [132, 152]. The 5-year OS was 100 % for patients with low TLG vs. 80 % for those with high TLG ($p=.0001$), whereas PFS was 99 % vs. 64 %,

respectively ($P<.0001$). Nonetheless, this was a retrospective evaluation in a group of 21 centers using various scanners. Additionally, despite a $p < 0.0001$, the HR for TLG was only 1.36 for increments of 10^3. Although considered preliminary, these results indicate that TLG at staging PET/CT could be a useful index in predicting outcomes in high-grade NHL including PMBCL treated with standard first-line chemotherapy regimens. Although it is premature to define the role of volumetric measurements in predicting outcomes, as a preliminary conclusion metabolic tumor volumes tend to be superior to ΔSUV_{max} in predictive values of survival, and a high MTV is significantly associated with reduced survival in DLBCL patients treated with R-CHOP. Because of the heterogeneity of the presently published data, these results should be interpreted with caution. This area of research will benefit from future large-scale prospective studies and further development in segmentation methodologies.

Predictive Value of PET-Derived Quantitative Metrics During or After Therapy

Taking a step forward from the traditional risk stratification systems, efforts have been concentrated on the interim PET results as a tool for guidance in early therapy modifications. However, the prognostic value of interim PET remains controversial in DLBCL patients with qualitative assessment variably correlated with outcome. The high false-positive rate associated with visual scoring systems, including the Deauville 5-point scale (D 5PS), has laid the grounds for quantitative PET initiatives when there is no existent optimal evaluation method for early assessment of response.

ΔSUV-based evaluation. Based on the results of multiple studies published by the Groupe d'Etude des Lymphomes de l'Adulte (GELA), it was suggested that the percentage reduction in SUV_{max} between baseline and interim PET (ΔSUV_{max}) improves both the interpretation accuracy and the interobserver reproducibility and better predicts patient outcome than visual analysis [153–155]. This group of investigators demonstrated that a 66 % reduction in SUV_{max} between baseline (PET0) and two cycles of chemotherapy (PET2) better predicted event-free survival (EFS) by reducing false-positive results of visual analysis. Other subsequent studies published corroborative results in DLBCL patients, treated with an anthracycline-based regimens plus rituximab [156, 157]. However, opposing results have also been reported by Pregno et al. in DLBCL patients treated with R-CHOP when the ΔSUV_{max} (both 66 % cutoff and median) at PET2 to PET4 was rather weakly correlated with outcome ($p=0.113$) [80]. Although it was in a homogeneous cohort, the limitation of this study included a small sample size, different time point analysis, and later than optimal time point preference (PET3 to PET4 vs. PET2). A similar quantitative approach was applied by Rossi C et al. to HL patients and showed that ΔSUV_{max} at PET2 was more accurate than the D 5PS in the prediction of outcome [158]. In this retrospective cohort of 59 consecutive HL patients who were treated with 4–8 cycles of anthracycline-based chemotherapy, PET2 ΔSUV_{max} >71 % was considered a favorable response. Although visual PET2 positivity was related to a lower 4-year PFS (45 % vs. 81 %, $p<0.002$), ΔSUV_{max} was more accurate for identifying patients with different 4-year PFS (82 % vs. 30 %; $p<0.0001$). In a multivariate analysis using the IPI and ΔSUVmax as covariates, ΔSUV_{max} remained the unique independent predictor for PFS (RR, 8.1 and $p=0.0001$). Quantitative interpretation of PET may lend itself as a more pragmatic tool to guide clinicians in lymphoma management but, the results of available data only pointed to the need for larger prospective trials and optimization and standardization of criteria for interim PET evaluation to assess the real prognostic value of interim PET results.

Tumor metabolic volume evaluation Although ΔSUV_{max} measurements partially improve on visual criteria and decrease the rate of false-positive results, a uniformly applicable ΔSUV_{max} cutoff has not been established to accurately predict clinical outcome. One can hypothesize that volumetric quantitative PET metrics have a better

predictive value early during therapy beyond that of ΔSUV_{max} as well as traditional risk factors in lymphoma. The results are summarized under two topics, "HL" and "DLBCL," respectively, in the following section. In general, a judicious approach should be adopted when reporting these studies because of the fact that the majority of these studies were retrospective, and no detailed information was provided on the quality assurance of the investigated data as well as on scanner calibration, image reconstruction algorithms, and patient scanning protocols (see previous section). Another flaw in design of prior studies included the presence of mixed population of early- and advanced-stage disease. It has been long established that Ann Arbor staging is one of the most important pre-therapy prognostication system and an essential component of prognostic models such as IPI and IPS. Therefore, evaluation of the additional value of PET quantitative metrics in distinct categories of early- and advanced-stage patients is necessary to derive a clinically meaningful prognostic information.

Hodgkin lymphoma PET-derived quantitative metrics can improve the robustness of response assessment for therapy adaptation in HL patients. There are several studies designed to address this objective [131, 132, 134]. The results of the study by Kanoun et al. revealed that both baseline MTV and ΔSUV_{max} at PET2 were independent predictors of PFS in a mixed early- and advanced-stage HL population [131]. The combination of MTV and ΔSUV_{max} made it possible to identify three subsets of HL patients with different PFS outcomes ($p<0.0001$). These included ΔSUV_{max} >71 % and MTV ≤225 ml, ΔSUV_{max} ≤71 % or MTV >225 ml, and ΔSUV_{max} ≤71 % and MTV >225 ml. In these three groups, the 4-year PFS rates were 92 %, 49 %, and 20 % ($p<0.0001$), respectively. In another retrospective study by Tseng et al., 30 HL patients (53 % stages III–IV and 67 % had IPS ≥ 2) were treated with varying chemotherapy regimens [Stanford V (67 %), ABVD (17 %), VAMP (10 %), or BEACOPP (7 %)] with or without radiation therapy [132]. Interim-treatment scans were performed at a median of 55 days from the staging PET. At a median follow-up of 50 months, baseline absolute PET parameters did not predict survival while the ΔMTV ($p=0.01$), ΔTLG ($p < 0.01$), and ΔSUV_{max} ($p=0.02$) were associated with PFS. In this study, all calculated PET parameters were further associated with OS. IPS was also associated with PFS and OS ($p<0.05$ and $p<0.01$, respectively). These results suggest that the chemosensitivity of the tumor as measured by PET early during treatment is more predictive of clinical outcome than the initial tumor bulk which gives further credence to prior validation studies [159, 160]. However, on the basis of inclusion of relapsed patients and various chemotherapy regimens inclusive of intensive treatments, these data are not conducive to reproducible results with firm conclusions. The quantitative PET results were also investigated in pediatric HL patients [161–164]. Similar to adult population, response assessment after two cycles improved the specificity of response assessment by 30 % using ΔSUV_{max} with a cutoff of 58 % [163, 164]. Contrary to these results, however, multiple other studies did not confirm the high predictive power of PET status early during therapy [79–81]. In a recent study by Hussien et al. in 54 pediatric HL patients treated on treatment optimization protocols, all quantitative PET measures (SUV_{max}, SUV_{mean}, MTV, and TGV) fared significantly better than the qualitative response assessment using D 5PS at PET2 [162]. ΔSUV_{max} was the most powerful predictor of treatment outcome (area under the curve, 0.92; $p<0.001$). The tumor volumes were determined with a fixed threshold of 2.5 SUV and at a threshold of mean liver plus two standard deviations SUV. In this study, technical parameters were better controlled than other studies, all PET scanners were cross-calibrated, and scan protocols followed EANM guidelines. However, sophisticated volumetric PET measures did not perform significantly better than the previously proposed ΔSUV_{max} in early response assessment [1, 3]. In summary, in the pediatric HL population, similar to the adult population, these results are preliminary and larger cohorts are needed to investigate this observation for a better definition of the role of PET/CT imaging. Recently, Hasenclever et al.

used a continuous scale by assigning D 5PS categories to certain quantitative PET cutoff values using the quotient of SUV_{peak} of the area with the most FDG avid residual uptake and the ΔSUV_{mean} of the liver in 898 pediatric HL patients after two chemotherapy cycles [165]. The borderlines for D 5PS 3, 4, and 5 at quantitative PET values corresponded to 0.95, 1.3, and 2.0, respectively, and quantitative PET of <1.3 excluded an unfavorable response with a high sensitivity. This method warrants a prospective validation study to be potentially used in clinical settings.

Diffuse large B-cell lymphoma Several retrospective studies investigated the value of quantitative PET-derived metrics in DLBCL, with the majority of data showing encouraging results [166, 167]. Park et al. investigated ΔSUV_{max}, TLG and Δ after 2 or 3 cycles in R-CHOP-treated DLBCL patients ($n=100$) including 57 patients with an IPI score of 1–3: the absolute values of baseline and interim SUVs calculated as the sum of values from 5 lesions (SUV_{sum}) and interim ΔSUV_{max} were significantly correlated with PFS [166]. While the ΔSUV_{max} and ΔTLG after 2 or 3 cycles were not associated with prognosis, the segmentation algorithm was based on mediastinal blood pool (MBP) threshold, which might have yielded larger MTVs than other thresholding methods would yield. The result of this study, although retrospective in design, highlights the potential of a quantitative approach to better delineate patient risk groups, particularly, in those with IPI scores of 1–3 which consists of the overlapping risk categories in which true low-risk patients should be better separated from the high-risk group to individualize therapies. These results could serve as a basis for future studies for the use of PET/CT in clinical practice, as an adjunct to IPI. Gradient-based methods appear to be more accurate compared with source-to-background ratio methods for segmenting FDG-PET images [43]. Malek et al. performed a retrospective study to correlate the ΔMTV and ΔSUV_{max} on interim PET with PFS after 2–4 cycles in 140 DLBCL patients using a gradient-based method rendered assessment of a greater tumor volume compared with the threshold-based method [167]. During a median follow-up of 37 months and with the use of R-CHOP and R-DA-EPOCH (rituximab-dose-adjusted etoposide, prednisone, Oncovin, cyclophosphamide, hydroxydaunorubicin) as the first-line therapy, D 5PS did not correlate with PFS ($P=0.37$). Compared with the threshold-based method, the gradient-based method resulted in a statistically significant greater MTV in pretreatment, as well as interim PET images. However, no significant difference was noted between the two methods. ΔMTV predicted PFS better than ΔSUV_{max} as the AUC for ΔMTV was significantly larger compared with that for ΔSUV_{max} ($AUC^{\Delta MTV}$: 0.713 and $AUC_{\Delta SUVmax}$: 0.873; P: 0.0324). Briefly, ΔMTV by either method after initial treatment was a better predictor of PFS compared with ΔSUV_{max}. Further analysis also revealed the underlying importance of ΔMTV on interim PET to predict PFS for patients who had also achieved a significant ΔSUV_{max}. MTV assessment (by either gradient- or threshold-based methods) may provide a more optimal methodology to accurately predict PFS as it incorporates the metabolic and volumetric information as a measure of tumor burden. Contrary to the aforementioned results, in a cohort of newly diagnosed 73 DLBCL patients, Adams et al. showed that the NCCN-IPI [168] was the most important prognostic tool for PFS ($p=0.024$) and OS ($p=0.039$) compared to PET-derived metrics including SUV_{max}, MTV, and TLG [146]. In this retrospective study, the authors used a threshold setting of 40 % of the SUV_{max} for volume delineation by a single expert. Median values of SUV_{max}, MTV, and TLG were used as cutoff values for group discrimination. Compared to prior studies, these significantly different results might have stemmed from methodological differences, different patient populations, shortcoming of the use of non-cross-calibrated scanners, and the overestimation of MTV and TLG through the use of a retrospective cutoff value in ROC analysis. In a pilot study of pediatric NHL patients ($n=16$), Furth et al. showed a limited predictive value for PET2 due to considerably high false-positive findings, especially in patients suffering from bulky disease [169]. With a mean follow-up of

60.2 months, the Kaplan–Meier survival analysis revealed no significant differences in 5-year PFS neither for conventional imaging modality (CIM) (76.9 % vs. 66.7 %; $p=0.67$) nor for visual PET (85.7 % vs. 66.7 %; $p=0.34$) nor for ΔSUV_{max}-based analysis (88.9 % vs. 57.1 %; $p=0.12$). In relapsed or refractory DLBCL, in a multicenter clinical trial of 55 patients treated with bendamustine–rituximab, Tateischi et al. demonstrated that the ΔTLG can be used to quantify the response to treatment and can predict PFS after the last treatment cycle [170]. In this study, scanners were cross-calibrated using a NEMA/IEC image quality phantom. MTV was calculated with a fixed threshold SUV_{max} >2.5. The percentage change in all PET parameters except for the area under the curve of the cumulative SUV-volume histogram was significantly greater in complete responders than in non-complete responders after two cycles and after the last cycle. The percentage change of the sum of total lesion glycolysis after the last cycle (relative risk, 5.24; $P=0.003$) was an independent predictor of PFS. An early PET scan after two cycles of treatment can effectively predict the outcome in patients with DLBCL treated with rituximab and anthracycline-based chemotherapy by using either a visual or quantitative approach. If its validity is proven in prospective studies, the interim ΔSUV_{max} approach may better serve clinicians to design a risk-adapted therapeutic strategy in DLBCL patients.

Radiation Therapy (RT) Planning

A limitation of FDG-PET in RT for HL is the variability in delineation of tumor volumes. Automatic or semiautomatic segmentation methods including thresholding based on a percent tumor ΔSUV_{max} may decrease variability in tumor delineations, but there is limited data in lymphoma using tumor volume segmentation methodologies. In a preliminary study using 15–40 % ΔSUV_{max} threshold segmentation method, on average, there was a 7.6-fold increase in PET volume between 15 % and 40 % ΔSUV_{max} x. There was a clinically significant decrease in dose to normal structures when the involved site radiation therapy (ISRT) plans were generated using the 15 % ΔSUV_{max} × volumes compared with the 40 % ΔSUV_{max} [171]. If these results can be reproduced, a streamlined approach may be developed using segmentation methods for conformal therapies. Moreover, the increased functional volume could be an artifact when contrast-enhanced CT is used for attenuation correction. In this case, it is recommended that the delineation volume using the relative or adaptive method should be preferred when contrast media are used for PET/CT [172].

The use of FDG-based PET data for target volume delineation in ISRT and IFRT planning requires a mindful utilization of automatic segmentation methods in conformal field designs such as ISRT, in which variations in pre-chemotherapy GTVs may lead to clinically significant changes as a result of different SUV_{max} thresholds. Clinical judgment is still required for the delineation of target volumes, and no segmentation method can reliably discern between FDG uptake caused by neoplastic processes and by physiological or inflammatory processes. The most accurate method for target volume definition in HL remains the manual generation of the volumes by a skilled radiation oncologist with input from a nuclear medicine physician when needed. This field is in evolution and further robust data are required to determine a reliable segmentation methodology to optimize treatment volumes and dose to normal structures.

4.2.3 New Technology

Magnetic resonance imaging (MRI) using diffusion-weighted technique (DWI) has been suggested as a useful method in the assessment of lymphoma lesions, particularly those with multiple conglomerate lymph nodes. There is preliminary evidence that the glycolytic rate as measured by FDG-PET and changes in water compartmentalization and water diffusion as measured by the apparent diffusion coefficients on DWI (ADC) are independent biological phenomena in newly diagnosed DLBCL [173, 174]. In one series, however, there was no significant correlation between ΔSUV_{max} and ΔADC after initiation of

the first cycle of chemotherapy in patients with HL or DLBCL. Thus, these data did not support the replacement of FDG-PET with DW-MRI for response evaluation in lymphoma patients [175]. ADC values were also found to discriminate between indolent and aggressive NHL, and this finding can be useful in assessing possible transformation from indolent to aggressive NHL [176]. There is also pilot data showing that the accuracy of DWI was significantly higher than that with PET/CT for mediastinal and hilar lymphadenopathy in differentiating between malignant and benign conditions [177]. In other cohorts, DW-MRI provided results comparable with or complementary to those of PET/CT for staging and early response assessment in DLBCL [178–180].

In summary, the literature is not mature to definitively prove or refute a diagnostic role for this modality compared to PET imaging in lymphoma patients. Further studies are warranted to assess the complementary roles of these different imaging biomarkers in the evaluation and follow-up of lymphoma.

Conclusions

The quantitative assessment with PET-derived volumes is still evolving and these preliminary findings suggest that it can be potentially useful in the prediction of clinical outcome and may improve on the predictive value of conventional risk-stratifying systems. However, currently, there is significant heterogeneity in the published data on the prognostic value of quantitative PET; thus, these results should be interpreted with caution because of their limited retrospective design, insufficient representation of risk and stage groups, differences in treatment strategies, as well as the varying methodologies used to measure MTVs. Currently, there is no consensus regarding the most optimal quantitative index to assess the metabolical activity disease burden using PET/CT imaging. Hence, the prognostic and predictive value of functional tumor volume remains to be further investigated with standardized, prospective, multicenter studies to validate as to what extent these parameters could improve individualized treatment approach in lymphoma.

References

1. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, 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.
2. Makris NE, Huisman MC, Kinahan PE, Lammertsma AA, Boellaard R. Evaluation of strategies towards harmonization of FDG PET/CT studies in multicentre trials: comparison of scanner validation phantoms and data analysis procedures. Eur J Nucl Med Mol Imaging. 2013;40:1507–15.
3. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54.
4. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50 Suppl 1(5):122S–50.
5. Lodge MA, Chaudhry MA, Wahl RL. Noise considerations for PET quantification using maximum and peak standardized uptake value. J Nucl Med. 2012;53:1041–7.
6. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. J Nucl Med. 2004;45:1519–27.
7. Boellaard R. Methodological aspects of multicenter studies with quantitative PET. Methods Mol Biol. 2011;727:335–49.
8. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of 18 F-FDG and 3′-deoxy-3′-18 F-fluorothymidine PET tumor volume measurements. J Nucl Med. 2010;51:1368–76.
9. Cheebsumon P, Yaqub M, van Velden FH, et al. Impact of [(18)F]FDG PET imaging parameters on automatic tumour delineation: need for improved tumour delineation methodology. Eur J Nucl Med Mol Imaging. 2011;38:2136–44.
10. Larson SM, Erdi Y, Akhurst T, et al. Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG Imaging. The visual response score and the change in total lesion glycolysis. Clin Positron Imaging. 1999;2:159–71.
11. Geworski L, Knoop BO, de Wit M, et al. Multicenter comparison of calibration and cross calibration of PET scanners. J Nucl Med. 2002;43:635–9.
12. Boellaard R, Hristova I, Ettinger S, et al. EARL FDG-PET/CT accreditation program: feasibility, overview and results of first 55 successfully accredited sites. J Nucl Med. 2013;54 Suppl 2:2052.

13. Zijlstra JM, Boellaard R, Hoekstra OS. Interim positron emission tomography scan in multi-center studies: optimization of visual and quantitative assessments. Leuk Lymphoma. 2009;50:1748–9.
14. Scheuermann JS, Saffer JR, Karp JS, Levering AM, Siegel BA. Qualification of PET scanners for use in multicenter cancer clinical trials: the American College of Radiology Imaging Network experience. J Nucl Med. 2009;50:1187–93.
15. Christian P. Use of a precision fillable clinical simulator phantom for PET/CT scanner validation in multi-center clinical trials: the SNM Clinical Trials Network (CTN) Program. J Nucl Med. 2012;53(Suppl):437.
16. Sunderland JJ, Christian PE. Quantitative PET/CT Scanner performance characterization based upon the SNMMI Clinical Trial Network oncology clinical simulator phantom. J Nucl Med. 2015;56:145–52.
17. Erlandsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. Phys Med Biol. 2012;57:R119–59.
18. Weber WA, Ziegler SI, Thödtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med. 1999;40:1771–7.
19. Krak NC, Boellaard R, Hoekstra OS, Twisk JWR, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. Eur J Nucl Med Mol Imaging. 2005;32:294–301.
20. Minn H, Zasadny K, Quint L, Wahl R. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. Radiology. 1995;196:167–73.
21. Nakamoto Y, Zasadny KR, Minn H, Wahl RL. Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[18 F]fluoro-D-glucose. Mol Imaging Biol. 2002;4:171–8.
22. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18 F-FDG PET in malignant tumors. J Nucl Med. 2008;49:1804–8.
23. Takahashi Y, Oriuchi N, Otake H, Endo K, Murase K. Variability of lesion detectability and standardized uptake value according to the acquisition procedure and reconstruction among five PET scanners. Ann Nucl Med. 2008;22:543–8.
24. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. J Nucl Med. 2009;50(10):1646–54.
25. Kumar V, Nath K, Berman CG, et al. Variance of SUVs for FDG-PET/CT is greater in clinical practice than under ideal study settings. Clin Nucl Med. 2013;38(3):175–82.
26. De Langen AJ, Vincent A, Velasquez LM, et al. Repeatability of 18 F-FDG uptake measurements in tumors: a metaanalysis. J Nucl Med. 2012;53(5):701–8.
27. Vanderhoek M, Perlman SB, Jeraj R. Impact of different standardized uptake value measures on PET-based quantification of treatment response. J Nucl Med. 2013;54:1188–94.
28. Lindholm H, Brolin F, Jonsson C, Jacobsson H. The relation between the blood glucose level and the FDG uptake of tissues at normal PET examinations. EJNMMI Res. 2013;3(1):50. doi:10.1186/2191-219X-3-50.
29. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18 F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. Eur J Cancer. 1999;35(13):1773–82.
30. Leijenaar RTH, Carvalho S, Velazquez ER, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52:1391–7.
31. Tylski P, Stute S, Grotus N, et al. Comparative assessment of methods for estimating tumor volume and standardized uptake value in (18)F-FDG PET. J Nucl Med. 2010;51:268–76.
32. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging. 2010;37:2165–87.
33. Riegel AC, Berson AM, Destian S, Ng T, Tena LB, Mitnick RJ, Wong PS. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. Int J Radiat Oncol Biol Phys. 2006;65(3):726–32.
34. Otsu N. A thresholding selection method from gray-level histograms. IEEE Trans Syst Man Cybern. 1979;9:62–6.
35. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. J Nucl Med. 2007;48:932–45.
36. Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer. 1997;80:2505–9.
37. Bradley J, Thorstad WL, Mutic S, Miller TR, Dehdashti F, Siegel BA, et al. Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer. Int J Radiat Oncol Biol Phys. 2004;59:78–86.
38. Miller TR, Grigsby PW. Measurement of tumor volume by PET to evaluate prognosis in patients with advanced cervical cancer treated by radiation therapy. Int J Radiat Oncol Biol Phys. 2002;53:353–9.
39. Scarfone C, Lavely WC, Cmelak AJ, Delbeke D, Martin WH, Billheimer D, et al. Prospective feasibility trial of radiotherapy target definition for head and neck cancer using 3-dimensional PET and CT imaging. J Nucl Med. 2004;45:543–52.
40. Nestle U, Weber W, Hentschel M, Grosu AL. Biological imaging in radiation therapy: role of

positron emission tomography. Phys Med Biol. 2009;54(1):R1–25.
41. Brambilla M, Matheoud R, Secco C, Loi G, Krengli M, Inglese E. Threshold segmentation for PET target volume delineation in radiation treatment planning: the role of target-to-background ratio and target size. Med Phys. 2008;35:1207–13.
42. Black QC, Grills IS, Kestin LL, Wong CY, Wong JW, Martinez AA, et al. Defining a radiotherapy target with positron emission tomography. Int J Radiat Oncol Biol Phys. 2004;60:1272–82.
43. Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Grégoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiother Oncol. 2003;69:247–50.
44. Jentzen W, Freudenberg L, Eising EG, Heinze M, Brandau W, Bockisch A. Segmentation of PET volumes by iterative image thresholding. J Nucl Med. 2007;48:108–14.
45. van Dalen JA. A novel iterative method for lesion delineation and volumetric quantification with fdg pet. Nucl Med Commun. 2007;28:485–93.
46. Nehmeh SA, El-Zeftawy H, Greco C, Schwartz J, Erdi YE, Kirov A, et al. An iterative technique to segment PET lesions using a Monte Carlo based mathematical model. Med Phys. 2009;36:4803–9.
47. Marr D, Hildreth E. Theory of edge detection. Proc R Soc Lond B Biol Sci. 1980;207:187–217.
48. Huertas A, Medioni G. Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks. IEEE Trans Pattern Anal Mach Intell. 1986;8:651–64.
49. Drever LA, Roa W, McEwan A, Robinson D. Comparison of three image segmentation techniques for target volume delineation in positron emission tomography. J Appl Clin Med Phys. 2007;8:93–109.
50. Geets X, Lee J, Bol A, Lonneux M, Grégoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imaging. 2007;34:1427–38.
51. Hsu C-Y, Liu C-Y, Chen C-M. Automatic segmentation of liver PET images. Comput Med Imaging Graph. 2008;32:601–10.
52. Li H, Thorstad WL, Biehl KJ, Laforest R, Su Y, Shoghi KI, et al. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. Med Phys. 2008;35:3711–21. Erratum pp 5958.
53. Long DT, King MA, Sheehan J. Comparative evaluation of image segmentation methods for volume quantitation in SPECT. Med Phys. 1992;19:483–9.
54. Aristophanous M, Penney BC, Martel MK, Pelizzari CA. A 53. Gaussian mixture model for definition of lung tumor volumes in positron emission tomography. Med Phys. 2007;34:4223–35.
55. Belhassen S, Zaidi H. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. Med Phys. 2010;37:1309–24.
56. Chiti A, Kirienko M, Grégoire V. Clinical use of PET-CT data for radiotherapy planning: what are we looking for? Radiother Oncol. 2010;96:277–9.
57. Kirov AS, Fanchon LM. Pathology-validated PET image data sets and their role in PET segmentation. Clin Transl Imaging. 2014;2(3):253–67. doi:10.1007/s40336-014-0068-9.
58. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rübe C, et al. Comparison of different methods for delineation of 18 F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. J Nucl Med. 2005;46:1342–8.
59. Shepherd T, Teras M, Beichel RR, et al. Comparative study with new accuracy metrics for target volume contouring in PET image guided radiation therapy. IEEE Trans Med Imaging. 2012;31:2006–24.
60. Gallamini A, Zwarthoed C, Borra A. Positron Emission Tomography (PET) in Oncology. Cancers (Basel). 2014;6(4):1821–89.
61. Zaffino P, Ciardo D, Piperno G, et al. Radiotherapy of Hodgkin and non-Hodgkin lymphoma. A non-rigid image-based registration method for automatic localization of prechemotherapy gross tumor volume. Technol Cancer Res Treat. 2015. pii: 1533034615582290.
62. Specht L, Yahalom J, Illidge T, et al. Modern radiation therapy for Hodgkin lymphoma: field and dose guidelines from the international lymphoma radiation oncology group. Int J Radiat Oncol Biol Phys. 2014;89:854–62.
63. Illidge T, Specht L, Yahalom J, et al. Modern Radiation Therapy for Nodal Non-Hodgkin Lymphoma Target Definition and Dose Guidelines from the International Lymphoma Radiation Oncology Group, from the International Lymphoma Radiation Oncology Group. Int J Radiat Oncol Biol Phys. 2014;89:49–58.
64. van Elmpt W, De Ruysscher D, van der Salm A, Lakeman A, van der Stoep J, Emans D, et al. The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer. Radiother Oncol. 2012;104:67–71.
65. Heukelom J, Hamming O, Bartelink H, Hoebers F, Giralt J, Herlestam T, et al. Adaptive and innovative radiation treatment for improving cancer treatment out- come (ARTFORCE); a randomized controlled phase II trial for individualized treatment of head and neck cancer. BMC Cancer. 2013;13:84. doi:10.1186/1471-2407-13-84.
66. Madani I, Duprez F, Boterberg T, Van de Wiele C, Bonte K, Deron P, et al. Maximum tolerated dose in a phase I trial on adaptive dose painting by numbers for head and neck cancer. Radiother Oncol. 2011;101:351–5.
67. Girinsky T, van der Maazen R, Specht L, et al. Involved-node radiotherapy (INRT) in patients with early Hodgkin lymphoma: concepts and guidelines. Radiother Oncol. 2006;79:270–7.

68. Genovesi D, Cèfaro GA, Vinciguerra A, et al. Interobserver variability of clinical target volume delineation in supra-diaphragmatic Hodgkin's disease. A multi-institutional experience. Strahlenther Onkol. 2011;187:357–66.
69. Lütgendorf-Caucig C, Fotina I, Gallop-Evans E, et al. Multicenter evaluation of different target volume delineation concepts in pediatric Hodgkin's lymphoma. A case study. Strahlenther Onkol. 2012;188:1025–30.
70. Shikama N, Oguchi M, Isobe K, et al. Quality assurance of radiotherapy in a clinical trial for lymphoma: individual case review. Anticancer Res. 2007;27:2621–5.
71. Yahalom J. Transformation in the use of radiation therapy of Hodgkin lymphoma: new concepts and indications lead to modern field design and are assisted by PET imaging and intensity modulated radiation therapy (IMRT). Eur J Haematol Suppl. 2005;75(s66):90–7.
72. Hutchings M, Loft A, Hansen M, et al. Clinical impact of FDGPET/CT in the planning of radiotherapy for early-stage Hodgkin lymphoma. Eur J Haematol. 2007;78:206–12.
73. Terezakis SA, Hunt MA, Kowalski A, et al. [^{18}F] FDG-positron emission tomography coregistration with computed tomography scans for radiation treatment planning of lymphoma and hematologic malignancies. Int J Radiat Oncol Biol Phys. 2011;81:615–22.
74. Eich HT, Müller RP, Engenhart-Cabillic R, et al. Involved-node radiotherapy in early-stage Hodgkin's lymphoma. Definition and guidelines of the German Hodgkin Study Group (GHSG). Strahlenther Onkol. 2008;184:406–10.
75. Robertson VL, Anderson CS, Keller FG, et al. Role of FDG-PET in the definition of involved-field radiation therapy and management for pediatric Hodgkin's lymphoma. Int J Radiat Oncol Biol Phys. 2011;80:324–32.
76. Konert T, Vogel W, MacManus MP, et al. PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014. Radiother Oncol. 2015;116:27–34.
77. Gallivanone F, Canevari C, Gianolli L, et al. A partial volume effect correction tailored for 18 F-FDG-PET oncological studies. Biomed Res Int. 2013;780458.
78. Hatt M, Le Pogam A, Visvikis D, et al. Impact of partial-volume effect correction on the predictive and prognostic value of baseline 18 F-FDG PET images in esophageal cancer. J Nucl Med. 2012;53(1):12–20.
79. Moskowitz CH, Schöder H, Teruya-Feldstein J. Risk-adapted dose-dense immunochemotherapy determined by interim FDG-PET in Advanced-stage diffuse large B-Cell lymphoma. J Clin Oncol. 2010;28:1896–903.
80. Pregno P, Chiappella A, Bellò M. 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.
81. 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.
82. 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;32:3048–58.
83. Ceriani L, Suriano S, Ruberto T, et al. 18 F-FDG uptake changes in liver and mediastinum during chemotherapy in patients with diffuse large B-cell lymphoma. Clin Nucl Med. 2012;37:949–52.
84. Groheux D, Delord M, Rubello D, et al. Variation of liver SUV on (18)FDG-PET/CT studies in women with breast cancer. Clin Nucl Med. 2013;38:422–5.
85. Rubello D, Gordien P, Morliere C, Guyot M, Bordenave L, Colletti PM, Hindié E. Variability of hepatic 18 F-FDG uptake at interim PET in patients with Hodgkin lymphoma. Clin Nucl Med. 2015;40:e405–10.
86. Gagne J. Innovative research methods for studying treatments for rare diseases: methodological review. BMJ. 2014;349:g6802.
87. Tsimberidou AM, Keating MJ. Richter syndrome: biology, incidence, and therapeutic strategies. Cancer. 2005;103:216–28.
88. Bruzzi JF, Macapinlac H, Tsimberidou AM, et al. Detection of Richter's transformation of chronic lymphocytic leukemia by PET/CT. J Nucl Med. 2006;47:1267–73.
89. Falchi L, Keating MJ, Marom EM, et al. Correlation between FDG/PET, histology, characteristics, and survival in 332 patients with chronic lymphoid leukemia. Blood. 2014;123(18):2783–90.
90. Michallet AS, Sesques P, Rabe KG, et al. An 18 F-FDG-PET maximum standardized uptake value >10 represents a novel valid marker for discerning Richter's Syndrome. Leuk Lymphoma. 2015;24:1–10.
91. Conte MJ, Bowen DA, Wiseman GA, et al. Use of positron emission tomographycomputed tomography in the management of patients with chronic lymphocytic leukemia/small lymphocytic lymphoma. Leuk Lymphoma. 2014;55(9):2079–84.
92. Schöder H, Noy A, Gönen M, et al. Intensity of 18fluorodeoxyglucose uptake in positron emission tomography distinguishes between indolent and aggressive non-Hodgkin's lymphoma. J Clin Oncol. 2005;23:4643–51.
93. Bodet-Milin C, Kraeber-Bodéré F, Moreau P, Campion L, Dupas B, Le Gouill S. Investigation of FDG-PET/CT imaging to guide biopsies in the detection of histological transformation of indolent lymphoma. Haematologica. 2008;93:471–2.

94. Noy A, Schöder H, Gönen M, et al. The majority of transformed lymphomas have high standardized uptake values on positron emission tomography scanning similar to diffuse large B-cell lymphoma. Ann Oncol. 2009;20:508–12.
95. Wondergem MJ, Rizvi SN, Jauw Y, et al. 18F-FDG or 3′-deoxy-3′-18F-fluorothymidine to detect transformation of follicular lymphoma. J Nucl Med. 2015;56(2):216–21.
96. Novelli S, Briones J, Flotats A, Sierra J. PET/CT assessment of follicular lymphoma and high grade B cell lymphoma – good correlation with clinical and histological features at diagnosis. Adv Clin Exp Med. 2015;24:325–30.
97. Shields AF, Grierson JR, Dohmen BM, et al. Imaging proliferation in vivo with [F-18]FLT and positron emission tomography. Nat Med. 1998;4:1334–6.
98. Buck AK, Bommer M, Stilgenbauer S, et al. Molecular imaging of proliferation in malignant lymphoma. Cancer Res. 2006;66:11055–61.
99. Hutchings M, Loft A, Hansen M, Ralfkiaer E, Specht L. Different histopathological subtypes of Hodgkin lymphoma show significantly different levels of FDG uptake. Hematol Oncol. 2006;24(3):146–50.
100. Ansquer C, Hervouët T, Devillers A, et al. 18-F FDG-PET in the staging of lymphocyte-predominant Hodgkin's disease. Haematologica. 2008;93:128–31.
101. Barber NA, Loberiza Jr FR, Perry AM, et al. Does functional imaging distinguish nodular lymphocyte-predominant Hodgkin Lymphoma from T-cell/histiocyte-rich large B-cell lymphoma? Clin Lymphoma Myeloma Leuk. 2013;13:392–7.
102. Lambin P, Rios-Velazquez E, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6.
103. Chicklore S, Goh V, Siddique M, et al. Quantifying tumour heterogeneity in 18 F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging. 2013;40:133–40.
104. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30:1234–48.
105. Tixier F, et al. Intratumor heterogeneity characterized by textural features on baseline 18 F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med. 2011;52:369–78.
106. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006. doi:10.1038/ncomms5006.
107. Bagci U, Yao J, Miller-Jaster K, Chen X, Mollura DJ. Predicting future morphological changes of lesions from radiotracer uptake in 18 F-FDG-PET images. PLoS One. 2013;8, e57105.
108. Hanaoka K, Hosono M, Tatsumi Y, et al. Heterogeneity of intratumoral (111)In-ibritumomab tiuxetan and (18)F-FDG distribution in association with therapeutic response in radioimmunotherapy for B-cell non-Hodgkin's lymphoma. EJNMMI Res. 2015;5:10.
109. Lopci E, Santi I, Tani M, Maffione AM, Montini G, Castellucci P, et al. FDG PET and 90Y ibritumomab tiuxetan in patients with follicular lymphoma. Q J Nucl Med Mol Imaging. 2010;54:436–41.
110. Jacene HA, Filice R, Kasecamp W, Wahl RL. 18 F-FDG PET/CT for monitoring the response of lymphoma to radioimmunotherapy. J Nucl Med. 2009;50:8–17.
111. Brunetti GA, Tendas A, Meloni E, et al. Pain and anxiety associated with bone marrow aspiration and biopsy: a prospective study on 152 Italian patients with hematological malignancies. Ann Hematol. 2011;90:1233–5.
112. Bain BJ. Morbidity associated with bone marrow aspiration and trephine biopsy: a review of UK data for 2004. Haematologica. 2006;91:1293–4.
113. 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;32:3059–68.
114. 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-naïve patients with Hodgkin lymphoma. J Clin Oncol. 2012;30:4508–14.
115. Khan AB, Barrington SF, Mikhaeel NG, et al. PET-CT staging of DLBCL accurately identifies and provides new insight into the clinical significance of bone marrow involvement. Blood. 2013;122:61–7.
116. Adams HJ, Kwee TC, de Keizer B, et al. FDG PET/CT for the detection of bone marrow involvement in diffuse large B-cell lymphoma: Systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2014;41:565–74.
117. Adams HJ, Kwee TC, Fijnheer R, et al. Bone marrow 18 F-fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography cannot replace bone marrow biopsy in diffuse large B-cell lymphoma. Am J Hematol. 2014;89:726–31.
118. Adams HJ, Kwee TC, Fijnheer R, et al. Direct comparison of visual and quantitative bone marrow FDG-PET/CT findings with bone marrow biopsy results in diffuse large B-cell lymphoma: does bone marrow FDG-PET/CT live up to its promise? Acta Radiol. 2015;56(10):1230–5.
119. Adams HJ, Kwee TC, Fijnheer R, et al. Utility of quantitative FDG-PET/CT for the detection of bone marrow involvement in follicular lymphoma: a histopathological correlation study. Skeletal Radiol. 2014;43:1231–6.
120. El-Najjar I, Montoto S, McDowell A, et al. The value of semiquantitative analysis in identifying diffuse bone marrow involvement in follicular lymphoma. Nucl Med Commun. 2014;35:311–5.

121. Adams HJ, Kwee TC, Fijnheer R, et al. Bone marrow FDG-PET/CT in Hodgkin lymphoma revisited: do imaging and pathology match? Ann Nucl Med. 2015;29:132–7.
122. Salaun PY, Gastinne T, Bodet-Milin C, et al. Analysis of 18F-FDG PET diffuse bone marrow uptake and splenic uptake in staging of Hodgkin's lymphoma: a reflection of disease infiltration or just inflammation? Eur J Nucl Med Mol Imaging. 2009;36:1813–21.
123. Bradley AJ, Carrington BM, Lawrance JA, et al. Assessment and significance of mediastinal bulk in Hodgkin's disease: comparison between computed tomography and chest radiography. J Clin Oncol. 1999;17:2493–8.
124. Gobbi PG, Ghirardelli ML, Solcia M, Di Giulio G, et al. Image-aided estimate of tumor burden in Hodgkin's disease: evidence of its primary prognostic importance. J Clin Oncol. 2001;19:1388–94.
125. Gobbi PG, Broglia C, Di Giulio G, et al. The clinical value of tumor burden at diagnosis in Hodgkin lymphoma. Cancer. 2004;101:1824–34.
126. Lister TA, Crowther D, Sutcliffe SB, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds Meeting. J Clin Oncol. 1989;7:1630–6.
127. Hasenclever D, Diehl V. A prognostic score for advanced Hodgkin's disease: International Prognostic Factors Project on Advanced Hodgkin's Disease. N Engl J Med. 1998;339:1506–14.
128. Diehl V, Thomas RK, Re D. Part II: Hodgkin's lymphoma: diagnosis and treatment. Lancet Oncol. 2004;5:19–26.
129. Hoppe RT, Advani RH, Bierman PJ, et al. NCCN Hodgkin disease clinical practice guidelines in oncology. 2006 v.1. Available at: http://www.nccn.org. Last accessed 6 Jan 2006.
130. 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;104:1656–61.
131. 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;41:1735–43.
132. Tseng D, Rachakonda LP, Su Z, et al. Interim-treatment quantitative PET parameters predict progression and death among patients with Hodgkin's disease. Radiat Oncol. 2012;7:5.
133. Hutchings M, Kostakoglu L, Zaucha JM, et al. In vivo treatment sensitivity testing with positron emission tomography/computed tomography after one cycle of chemotherapy for Hodgkin lymphoma. J Clin Oncol. 2014;32:2705–11.
134. Knight-Greenfield A, Cotter R, Marshall R, et al. Interim FDG PET/CT predicts response and progression free survival (PFS) better than baseline clinical and metabolic parameters in Hodgkin's lymphoma (HL): Correlation with various methodologies. J Nucl Med. 2013;54:69. Available at: http://jnm.snmjournals.org/content/54/supplement_2/69.abstract.
135. Federico M, Bellei M, Marcheselli L, et al. Follicular Lymphoma International Prognostic Index 2: a new prognostic index for follicular lymphoma developed by the International Follicular Lymphoma Prognostic Factor Project. J Clin Oncol. 2009;27:4555–62.
136. Pfreundschuh M, Ho AD, Cavallin-Stahl E, et al. Prognostic significance of maximum tumour (bulk) diameter in young adults 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.
137. Esfahani SA, Heidari P, Halpern EF, Hochberg EP, Palmer EL, Mahmood U. Baseline total lesion glycolysis measured with (18)F-FDG PET/CT as a predictor of progression-free survival in diffuse large B-cell lymphoma: a pilot study. Am J Nucl Med Mol Imaging. 2013;3:272–81.
138. Song MK, Chung JS, Shin HJ, Lee SM, Lee SE, Lee HS, Lee GW, Kim SJ, Lee SM, Chung DS. 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.
139. Manohar K, Mittal BR, Bhattacharya A, Malhotra P, Varma S. Prognostic value of quantitative parameters derived on initial staging 18F-fluorodeoxyglucose positron emission tomography/ computed tomography in patients with high-grade non-Hodgkin's lymphoma. Nucl Med Commun. 2012;33(9):974–81.
140. Sasanelli M, Meignan M, Haioun C, Berriolo-Riedinger A, Casasnovas RO, Biggi A, Gallamini A, Siegel BA, Cashen AF, Vera P, Tilly H, Versari A, Itti E. Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging. 2014;41:2017–22.
141. Kim TM, Paeng JC, Chun IK, Keam B, Jeon YK, Lee SH, Kim DW, Lee DS, Kim CW, Chung JK, Kim IH, Heo DS. 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.
142. 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;48:187–95.
143. Chihara D, Oki Y, Onoda H, et al. High maximum standard uptake value (SUVmax) on PET scan is associated with shorter survival in patients with diffuse large B cell lymphoma. Int J Hematol. 2011;93:502–8.
144. Xie M, Zhai W, Cheng S, Zhang H, Xie Y, He W. Predictive value of F-18 FDG PET/CT quantization parameters for progression-free survival in patients

with diffuse large B-cell lymphoma. Hematology. 2016;21(2):99–105.
145. Gallicchio R, Mansueto G, Simeon V, et al. F-18 FDG PET/CT quantization parameters as predictors of outcome in patients with diffuse large B-cell lymphoma. Eur J Haematol. 2014;92:382–9.
146. Adams HJ, de Klerk JM, 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;94:532–9.
147. Song MK, Chung JS, Shin HJ, et al. Prognostic value of metabolic tumor volume on PET / CT in primary gastrointestinal diffuse large B cell lymphoma. Cancer Sci. 2012;103:477–82.
148. 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 ΔSUVmax. Eur J Nucl Med Mol Imaging. 2013;40:1312–20.
149. Xie M, Wu K, Liu Y, Jiang Q, Xie Y. Predictive value of F-18 FDG PET/CT quantization parameters in diffuse large B cell lymphoma: a meta-analysis with 702 participants. Med Oncol. 2015;32:446.
150. Oh MY, Oh SB, Seoung HG, et al. Clinical significance of standardized uptake value and maximum tumor diameter in patients with primary extranodal diffuse large B cell lymphoma. Korean J Hematol. 2012;47:207–12.
151. Ceriani L, Martelli M, Zinzani PL, et al. Utility of baseline 18FDG-PET/CT functional parameters in defining prognosis of primary mediastinal (thymic) large B-cell lymphoma. Blood. 2015;126:950–6.
152. Lee PWD, Lavori P, Quon A, Hara W, Maxim P, Le QT, Wakelee H, Donington J, Graves E, Loo BW. Metabolic tumor burden predicts for disease progression and death in lung cancer. Int J Radiation Oncology Biol Phys. 2007;69:328–33.
153. Lin C, Itti E, Haioun C, et al. Early 18 F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. J Nucl Med. 2007;48:1626–32.
154. Itti E, Lin C, Dupuis J, et al. Prognostic value of interim 18 F-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.
155. 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.
156. Safar V, Dupuis J, Itti E, et al. Interim [18 F]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.
157. Nols N, Mounier N, Bouazza S, et al. Quantitative and qualitative analysis of metabolic response at interim positron emission tomography scan combined with International Prognostic Index is highly predictive of outcome in diffuse large B-cell lymphoma. Leuk Lymphoma. 2014;55:773–80.
158. Rossi C, Kanoun S, Berriolo-Riedinger A, et al. Interim 18 F-FDG PET SUVmax reduction is superior to visual analysis in predicting outcome early in Hodgkin lymphoma patients. J Nucl Med. 2014;55:569–73.
159. Gallamini A, Barrington SF, Biggi A, et al. The predictive role of interim Positron Emission Tomography on Hodgkin lymphoma treatment outcome is confirmed using the 5-point scale interpretation criteria. Haematologica. 2014;99:1107–13.
160. Biggi A, Gallamini A, Chauvie S, et al. International validation study for interim PET in ABVD-treated, advanced-stage Hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. J Nucl Med. 2013;54:683–90.
161. Sharma P, Gupta A, Patel C, et al. Pediatric lymphoma: metabolic tumor burden as a quantitative index for treatment response evaluation. Ann Nucl Med. 2012;26:58–66.
162. Hussien AE, Furth C, Schönberger S, et al. FDG-PET response prediction in pediatric Hodgkin's lymphoma: impact of metabolically defined tumor volumes and individualized SUV measurements on the positive predictive value. Cancers (Basel). 2015;7:287–304.
163. Furth C, Steffen IG, Amthauer H, et al. Early and late therapy response assessment with [18 F]fluorodeoxyglucose positron emission tomography in pediatric Hodgkin's lymphoma: analysis of a prospective multicenter trial. J Clin Oncol. 2009;27:4385–91.
164. Furth C, Meseck RM, Steffen IG, et al. SUV-measurements and patient-specific corrections in pediatric Hodgkin-lymphoma: is there a benefit for PPV in early response assessment by FDG-PET? Pediatr Blood Cancer. 2012;59:475–80.
165. Hasenclever D, Kurch L, Mauz-Körholz C, et al. qPET – a quantitative extension of the Deauville scale to assess response in interim FDG-PET scans in lymphoma. Eur J Nucl Med Mol Imaging. 2014;41:1301–8.
166. Park S, Moon SH, Park LC, et al. The impact of baseline and interim PET/CT parameters on clinical outcome in patients with diffuse large B cell lymphoma. Am J Hematol. 2012;87:937–40.
167. Malek E, Sendilnathan A, Yellu M, et al. Metabolic tumor volume on interim PET is a better predictor of outcome in diffuse large B-cell lymphoma than semi-quantitative methods. Blood Cancer J. 2015;5, e326.
168. 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;123:837–42.

169. Furth C, Steffen IG, Erdrich AS, et al. Explorative analyses on the value of interim PET for prediction of response in pediatric and adolescent non-Hodgkin lymphoma patients. EJNMMI Res. 2013;3(1):71.
170. Tateishi U, Tatsumi M, Terauchi T, et al. Prognostic significance of metabolic tumor burden by positron emission tomography/computed tomography in patients with relapsed/refractory diffuse large B-cell lymphoma. Cancer Sci. 2015;106:186–93.
171. Walker AJ, Chirindel A, Hobbs RF, et al. Use of standardized uptake value thresholding for target volume delineation in pediatric Hodgkin lymphoma. Pract Radiat Oncol. 2015;5:219–27.
172. Vera P, Modzelewski R, Hapdey S, et al. Does enhanced CT influence the biological GTV measurement on FDG-PET images? Radiother Oncol. 2013;108:86–90.
173. Punwani S, Prakash V, Bainbridge A, et al. Quantitative diffusion weighted MRI: a functional biomarker of nodal disease in Hodgkin lymphoma? Cancer Biomark. 2010;7(4):249–59.
174. de Jong A, Kwee TC, de Klerk JM, et al. Relationship between pretreatment FDG-PET and diffusion-weighted MRI biomarkers in diffuse large B-cell lymphoma. Am J Nucl Med Mol Imaging. 2014;4:231–8.
175. Hagtvedt T, Seierstad T, Lund KV, et al. Diffusion-weighted MRI compared to FDG PET/CT for assessment of early treatment response in lymphoma. Acta Radiol. 2015;56:152–8.
176. Mosavi F, Wassberg C, Selling J, Molin D, Ahlström H. Whole-body diffusion-weighted MRI and (18)F-FDG PET/CT can discriminate between different lymphoma subtypes. Clin Radiol. 2015;70:1229–36.
177. Usuda K, Maeda S, Motono N, et al. Diagnostic performance of diffusion – weighted imaging for multiple hilar and mediastinal lymph nodes with FDG accumulation. Asian Pac J Cancer Prev. 2015;16:6401–6.
178. Siegel MJ, Jokerst CE, Rajderkar D, et al. Diffusion-weighted MRI for staging and evaluation of response in diffuse large B-cell lymphoma: a pilot study. NMR Biomed. 2014;27:681–91.
179. Punwani S, Taylor SA, Saad ZZ, et al. Diffusion-weighted MRI of lymphoma: prognostic utility and implications for PET/MRI? Eur J Nucl Med Mol Imaging. 2013;40:373–85.
180. Trotman J, Luminari S, Boussetta S, et al. Prognostic value of PET-CT after first-line therapy in patients with follicular lymphoma: a pooled analysis of central scan review in three multicentre studies. Lancet Haematol. 2014;1:e17–27.