OPEN

Interobserver Reliability of the 5-Point Deauville Score and SUV-Based Quantification of FDG PET/CT Scans in High-Risk Pediatric Hodgkin Lymphoma

A Report From the Children's Oncology Group

Inki Lee, MD,† Sarah A. Milgrom, MD,‡ Hyung-Jun Im, MD,*§ Qinglin Pei, PhD,‖ Lindsay Renfro, PhD,¶** Bradford S. Hoppe, MD,†† Kathleen M. McCarten, MD,‡‡ Cindy L. Schwartz, MD,§§ Peter D. Cole, MD,‖‖ Kara M. Kelly, MD,¶¶ and Steve Y. Cho, MD*****

Purpose: In lymphomas, the Deauville score (DS) standardizes ^{18}F-FDG PET treatment response assessments. We assessed interobserver agreement of the DS and developed a quantitative PET-based parameter to minimize subjectivity of response assessments.

Methods: PET scans performed within the Children's Oncology Group AHOD0831 study were analyzed. One hundred one scans obtained after the first cycle of chemotherapy (PET1) and 83 obtained after the second cycle (PET2) were eligible for inclusion. Each scan was assigned a DS by 2 nuclear medicine radiologists and a Hodgkin lymphoma index (HLI), defined as the ratio of the tumor's SUV_{max} or SUV_{peak} to the liver's SUV_{mean}. Cohen κ coefficient measured the rate of agreement of the DS assigned by the 2 radiologists. Receiver operating characteristic curves and the Youden index were used with the consensus DS to identify an optimal HLI cutoff value to distinguish PET-negative from PET-positive images. For each HLI cutoff value, sensitivity, specificity, and accuracy were calculated.

Results: Interobserver agreement in DS was moderate (PET1: $\kappa = 0.480$, $P < 0.001$; PET2: $\kappa = 0.428$, $P < 0.001$). On PET1, the optimal cutoff values for HLI_{max} were 1.19 for DS2/DS3 and 1.78 for DS3/DS4, and for HLI_{peak} were 1.19 for DS2/DS3 and 1.44 for DS3/DS4. On PET2 images, the corresponding values were 1.18, 1.58, 0.94, and 1.26. The most accurate predictor was HLI_{max} on PET2 when scans were dichotomized as DS1–DS3 versus DS4–DS5 (accuracy, 92.8%; sensitivity, 100%; specificity, 91.8%).

Conclusion: The HLI provides an objective, quantitative measure of disease response and thus may improve interreader agreement for lymphoma response assessments.

Key Words: PET, Hodgkin lymphoma, Deauville score, quantitative, SUV, Hodgkin Lymphoma Index, HLI

(*Clin Nucl Med Open* 2024;1: e00011)

^{18}F-FDG PET/CT plays a critical role in monitoring lymphomas' response to therapy.[1,2] FDG PET findings are associated with prognosis, prompting the current generation of protocols in pediatric Hodgkin lymphoma (HL) that incorporate response assessments to guide therapy. Treatment may be intensified in children with an unfavorable response and/or de-escalated in those with a favorable response.[3–8] Therefore, accurate response assessments are critical to ensure that patients receive appropriate therapy.

However, the interpretation of PET/CT scans is subjective, depending on the nuclear medicine physician's visual assessment.[9] In order to standardize the interpretation of PET scans for lymphoma response assessments, the 5-point scale Deauville score (DS) was proposed at the 2009 First International Workshop on Interim PET Scans in Lymphoma.[10] This score compares the FDG uptake within sites of lymphoma to reference tissues including the liver and mediastinal blood pool. Currently, the DS is regarded as the international standard for response assessment of lymphomas using FDG PET/CT. It has been incorporated in the Lugano Criteria guidelines for both HL and non-Hodgkin lymphoma.

Although the DS method aims to minimize interreader variability, it still depends on subjective visual assessments. Interobserver disagreement in measuring DS has been reported as a significant issue in many studies, especially for differentiation between DS2 and DS3.[9,11] Therefore, a more objective and reproducible response assessment method is needed. With this goal, quantitative methods have been reported that incorporate a ratio of tumor and liver PET standardized uptake value (SUV) parameters.[11,12]

In this study, we evaluated the interobserver reliability of DS in pediatric patients with high-risk HL, and we developed a quantitative measure based on SUV ratios for objective disease response assessments.

From the *Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, WI; †Department of Nuclear Medicine, Korea Cancer Center Hospital, Korea Institute of Radiological and Medical Sciences, Seoul, Republic of Korea; ‡Department of Radiation Oncology, University of Colorado, Aurora, CO; §Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea; ‖Children's Oncology Group, Statistics and Data Center, Department of Biostatistics, University of Florida, Gainesville, FL; ¶Division of Biostatistics, University of Southern California, Los Angeles; and **Children's Oncology Group, Monrovia, CA; ††Department of Radiation Oncology, Mayo Clinic, Jacksonville, FL; ‡‡Imaging and Radiation Oncology Core, Lincoln, RI; §§Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI; ‖‖Department of Pediatrics, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ; ¶¶Department of Pediatric Oncology, Roswell Park Comprehensive Cancer Center and University at Buffalo Jacobs School of Medicine and Biomedical Sciences, Buffalo, NY; and ***Department of Radiology, University of Wisconsin Carbone Cancer Center, Madison, WI.

I.L. and S.A.M. contributed equally to this study.

Conflicts of interest and sources of funding: none declared. This study was supported by NCTN Operations Center grant U10CA180886, NCTN Statistics & Data Center grant U10CA180899, and St Baldricks Foundation.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Trial: Children's Oncology Group (COG) AHOD0831, NCI-2011-01994.

Correspondence to: Steve Y. Cho, MD, Department of Radiology, Nuclear Medicine and Molecular Imaging Section, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Suite MC3252 Madison, WI 53792. E-mail: scho@uwhealth.org.

ISSN: 2994-8908/24/0102–0000

DOI: 10.1097/nm9.0000000000000011

MATERIALS AND METHODS

Patients

From December 2009 to January 2011, a prospective, multicenter Children's Oncology Group study was initiated for children and adolescents younger than 21 years, with high-risk HL (AHOD0831, NCT01026220). On this trial, all patients received an initial 2 cycles of AVBE-PC (doxorubicin, bleomycin, vincristine, etoposide, prednisone, and cyclophosphamide) chemotherapy, followed by FDG PET/CT response assessment. Patients with a rapid response to therapy, as determined by the FDG PET/CT, received 2 more courses of ABVE-PC, whereas patients with a slow response to therapy received an additional 2 cycles of ifosfamide and vinorelbine chemotherapy followed by 2 more cycles of ABVE-PC. Patients underwent radiation therapy to initially bulky sites and to slowly responding nodal sites as defined by early FDG PET/CT. A total of 164 patients enrolled on this trial were reviewed retrospectively. Of these, patients with FDG PET/CT scans that were of high quality and were amenable to quantitative analyses were included in the current study.

PET/CT Imaging Acquisition

In the multicenter AHO0831 trial, FDG PET/CT scans were obtained at participating sites using standard-of-care clinical PET/CT imaging.[13] FDG PET/CT recommendations were provided in the protocol that included the following: (1) fast for at least 4 hours before the PET/CT scan, (2) deliver FDG intravenously with at a dose of 0.140 to 0.200 mCi/kg, (3) obtain the PET scan 60 ± 10 minutes after the injection of FDG using a PET/CT scanner, (4) acquire a low-dose CT scan for attenuation correction, and (5) collect PET emission data with at least 5 minutes per bed position for bismuth germanate, lutetium oxyorthosilicate, and gadolinium oxyorthosilicate systems operated in the 2D mode; at least 3 minutes per bed position for bismuth germanate, lutetium oxyorthosilicate, and gadolinium oxyorthosilicate systems operated in the 3D mode; and at least 6 minutes per bed position for sodium iodide systems.[14]

Visual Assessment of FDG PET Images

Each PET1 and PET2 was assigned a DS by 2 experienced nuclear medicine physicians independently.[1,2] If the physicians assigned different scores to the same PET image, then a consensus DS was decided after a convened read between the 2 physicians.

Quantification of SUV on FDG PET Images

The SUV was obtained for the residual tumor (Mirada RTx; Mirada Medical, Denver, CO). The SUV_{max} was the maximal single pixel uptake value. The SUV_{peak} was the average SUV_{max} in a 1-cm^3 spherical tumor volume with the highest FDG uptake. For normalization, the liver was selected as a reference tissue. The SUV_{mean} of liver was calculated with a 3-cm-diameter region of interest placed in the normal liver. Using SUV_{max} of the tumor, SUV_{peak} of the tumor, and SUV_{mean} of the liver, a quantitative value for DS was calculated that was defined as the Hodgkin Lymphoma Index (HLI). HLI_{max} was defined as the ratio of SUV_{max} of the tumor to SUV_{mean} of liver. HLI_{peak} was defined as the ratio of SUV_{peak} of the tumor to SUV_{mean} of liver. HLI_{max} and HLI_{peak} were measured on both PET1 and PET2 images.

PET-Negative and PET-Positive Dichotomization Based on the DS

Scans were dichotomized into 2 groups based on the 5-point DS. When the cutoff was between DS2 and DS3, the 2 subgroups were DS1–DS2 (PET-negative) and DS3–DS5 (PET-positive). When the cutoff was between DS3 and DS4, the 2 subgroups were DS1–DS3 (PET-negative) and DS4–DS5 (PET-positive).

Statistical Analysis

The rate of agreement and discrepancy of visual DS was analyzed using Cohen κ coefficient (SPSS version 23.0; IBM Corp, Chicago, IL). The degree of agreement was decided based on the reference.[15] For the cutoff value of HLI between PET-negative and PET-positive images, receiver operating characteristic curves were used with the HLI and consensus DS by 2 physicians (Medcalc version 10.1.7.0; Medcalc Software, Mariakerke, Belgium). The optimal cutoff value of HLI was calculated based on the Youden index of the receiver operating characteristic curve. For each cutoff value of HLI, sensitivity, specificity, and accuracy were calculated. For the density distribution, R package *ggplot2* was used (https://cran.r-project.org/).

RESULTS

A total of 101 patients had high-quality scans that were performed after the first cycle of AVBE-PC chemotherapy (PET1). Of these, 83 patients also had high-quality scans that were performed after the second cycle (PET2). These PET/CT scans were included for analyses. The characteristics of the patient cohort are summarized in Table 1.

TABLE 1. Patient and Disease Characteristics

Characteristics	N = 101
Age, y	
Mean (SD)	15.0 (3.1)
Median (IQR)	15.7 (14.0, 16.9)
Range	5.2–21.4
Sex	
Female	42 (42%)
Male	59 (58%)
Race	
American Indian or Alaska Native	2 (2.0%)
Asian	4 (4.0%)
Black or African American	17 (17%)
Unknown	8 (7.9%)
White	70 (69%)
Ethnicity	
Hispanic or Latino	18 (18%)
Not Hispanic or Latino	82 (81%)
Unknown	1 (1.0%)
Pathology	
Hodgkin lymphoma, not otherwise specified	17 (17%)
Lymphocyte-rich	8 (7.9%)
Lymphocyte depletion	3 (3.0%)
Mixed cellularity	12 (12%)
Nodular lymphocyte predominance	1 (1.0%)
Nodular sclerosis	60 (59%)
Stage	
III	46 (46%)
IV	55 (54%)
Bulk	
No	16 (16%)
Yes	85 (84%)
PET2 reviewed	83 (82%)

Values are presented as n (%), unless specified otherwise.

TABLE 2. Agreement for Deauville Score (DS) Between 2 Physicians

	κ	
	PET1	**PET2**
Five point scales: DS1, DS2, DS3, DS4, DS5	0.480	0.428
Two subgroups: DS1–DS2/DS3–DS5	0.629	0.553
Two subgroups: DS1–DS3/DS4–DS5	0.759	0.765

Interobserver Agreement in DS

Table 2 shows the overall agreement in the DS obtained independently by 2 readers. DS on PET1 and PET2 images showed moderate interobserver agreement (κ = 0.480, P < 0.001; and κ = 0.428, P < 0.001, respectively). The 5-point DS was dichotomized to define PET scans as negative or positive. When subgroups were divided as DS1–DS2 (PET-negative) and DS3–DS5 (PET-positive), there was good agreement (PET1: κ = 0.629, P < 0.001; PET2: κ = 0.553, P < 0.001). Likewise, when subgroups were divided as DS1–DS3 (PET-negative) and DS4–DS5 (PET-positive), there was good agreement (PET1: κ = 0.759, P < 0.001; PET2: κ = 0.765, P < 0.001).

Distribution of HLI

The distribution of DS based on the 2 independent physicians' consensus for PET1 (101 scans) was 2.0% DS1, 15.8% DS2, 22.8% DS3, 36.6% DS4, and 22.8% DS5. For PET2 (83 scans), the distribution of DS was 22.8% DS1, 36.6% DS2, 12.9% DS3, 6.9% DS4, and 3.0% DS5. Figure 1 shows the distribution of HLI_{max} and HLI_{peak} for PET1 and PET2. These figures depict the HLI values and the consensus DS. The optimal cutoff values of HLI_{max} on PET1 images were 1.19 for DS2/DS3 and 1.78 for DS3/DS4 (Table 3). The cutoff values of HLI_{peak} on PET1 images were 1.19 for DS2/DS3 and 1.44 for DS3/DS4 (Table 3). On PET2 images, the HLI_{max} cutoff values were 1.18 for DS2/DS3 and 1.58 for DS3/DS4; the HLI_{peak} cutoff values were 0.94 for DS2/DS3 and 1.26 for DS3/DS4 (Table 4).

Accuracy of HLI

Tables 3 and 4 show the accuracy of predicting DS using HLI, dichotomized by the optimal cutoff values. On PET1, the accuracy of HLI_{max} was 88.1% using 1.19 as a cutoff (DS1–DS2/DS3–DS5) and 85.2% using 1.78 as a cutoff (DS1–DS3/DS4–DS5) (Table 3). On PET1, the accuracy of HLI_{peak} was 79.2% using 1.19 as a cutoff (DS1–DS2/DS3–DS5) and was 83.2% using 1.44 as a cutoff (DS1–DS3/DS4–DS5) (Table 3). On PET2, HLI_{max} had an accuracy of 81.8% and 92.8% for cutoff values of 1.18 (DS1–DS2/DS3–DS5) and 1.58 (DS1–DS3/DS4–DS5), respectively (Table 4). On PET2, HLI_{peak} had an accuracy of 71.1% and 90.4% for the cutoff values of 0.94 (DS1–DS2/DS3–DS5) and 1.26 (DS1–DS3/DS4–DS5), respectively (Table 4). In all cases, accuracy was higher when scans were dichotomized as DS1–DS3 and DS4–DS5, rather than DS1–DS2 and DS3–DS5, with the sole exception of HLI_{max} on PET1 (Tables 3 and 4).

DISCUSSION

This study aimed to evaluate the interobserver reliability of FDG PET DS with visual assessment and to compare an objective PET SUV-based quantitative index with DS. Although DS is useful for monitoring the treatment response of lymphomas and enables standardization in multicenter trials, its weakness is that interpretations are subjective and have variable interobserver reliability. In this study, the visual assessment by 2 experienced nuclear medicine radiologists for measuring DS was variably discordant, which was consistent with prior reports that the interobserver reliability of

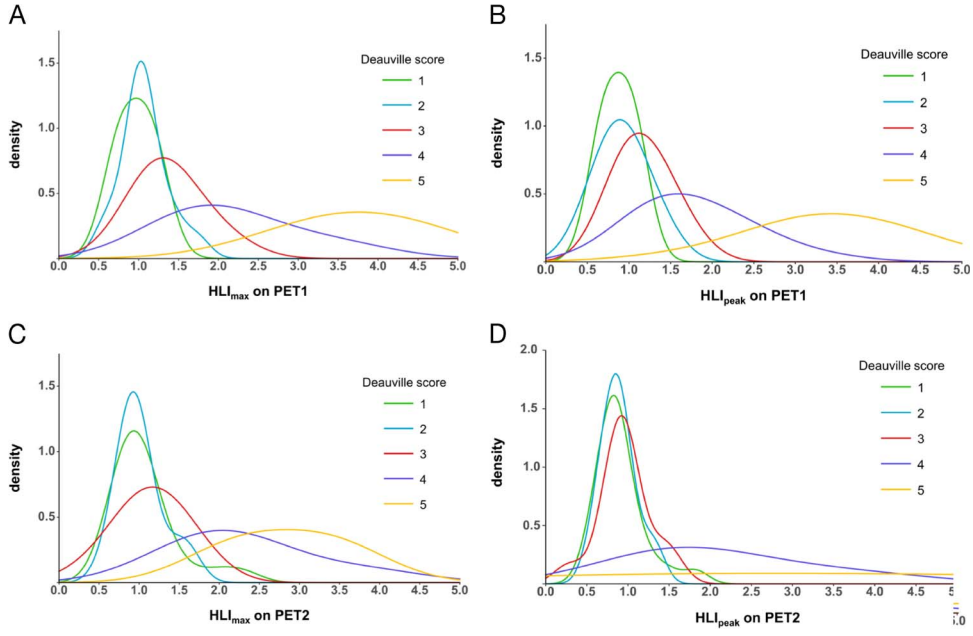

FIGURE 1. The distribution curves of the HLI_{max} and HLI_{peak} on PET1 and PET2. HLI_{max} on PET1 separates the Deauville score (DS) by thresholds of 1.19 for DS2/DS3 and 1.78 for DS3/DS4 (A). HLI_{peak} on PET1 shows thresholds of 1.19 and 1.44 for DS2/DS3 and DS3/DS4, respectively (B). HLI_{max} on PET2 displays thresholds of 1.18 and 1.58 for DS2/DS3 and DS3/DS4, respectively (C). HLI_{peak} on PET2 has thresholds of 0.94 and 1.26 for DS2/DS3 and DS3/DS4, respectively (D).

TABLE 3. Diagnostic Accuracy of the Hodgkin Lymphoma Index for Deauville Score After 1 Cycle of Chemotherapy (PET1)

	Sensitivity, %	Specificity, %	Accuracy, %	95% Confidence Interval
HLI_{max}				
Cutoff value of 1.19 (DS1–DS2/DS3–DS5)	90.4	83.3	88.1	0.847–0.963
Cutoff value of 1.78 (DS1–DS3/DS4–DS5)	80.0	92.7	85.2	0.866–0.973
HLI_{peak}				
Cutoff value of 1.19 (DS1–DS2/DS3–DS5)	78.3	88.9	79.2	0.826–0.952
Cutoff value of 1.44 (DS1–DS3/DS4–DS5)	80.0	87.8	83.2	0.849–0.965

DS can be poor.[9] The visual assessment by each physician may be different, despite using the 5-point scale DS that has been developed to assist with objective visual assessments. In this study, rates of agreement between the 2 physicians were lower using the 5-point DS compared with the 2 dichotomized subgroups (PET-positive vs PET-negative). Thus, treatment decisions based on the 5-point DS may be affected by the reader's subjectivity.

Interobserver discrepancy in DS reporting can have significant implications for patients. Specifically, patients treated according to PET response–adapted algorithms may receive inappropriate therapy, resulting in increased rates of disease relapse or of treatment-related toxicity. For example, if patients with an incomplete response are misclassified as having achieved a complete response and receive less intensive therapy, then they would have a greater risk of a disease relapse. Alternatively, if patients with a complete metabolic response are misclassified and receive more intensive treatment, they would have a greater risk of adverse treatment-related effects. Therefore, great caution should be used when implementing PET-adapted treatment approaches, and minimizing subjectivity in response assessments is crucial.

It is typically not difficult to assign a DS4 or DS5 using the 5-point scale, because these categories are defined by definite FDG uptake; however, distinguishing DS2 from DS3 may be difficult because differences between FDG uptake of the mediastinal blood pool and liver relative to the tumor may be subtle. Therefore, an objective system to quantify residual lymphoma FDG uptake, as a surrogate for the visual DS, is desirable. Hasenclever et al defined the quantitative PET value as the ratio of the SUV_{peak} of the residual tumor to the SUV_{mean} of the liver.[12] They reported that quantitative PET makes the DS a continuous scale and provides cutoff values between positive and negative PET images in lymphoma.[12] The cutoff values between DS2/DS3, DS3/DS4, and DS4/DS5 were reported as 0.95, 1.3, and 2.0, respectively.[12] Annunziata et al developed the rPET, defined as the ratio between the SUV_{max} of the residual tumor and liver.[11] They reported that rPET with a cutoff value of 1.14 is a more accurate prognostic factor in HL than the 5-point DS.[11]

In this study, the SUV_{max} and SUV_{peak} of the residual tumor were compared. SUV_{max} is defined as the SUV of the single hottest voxel, and SUV_{peak} is defined as an average SUV of multiple voxels around the single hottest voxel.[16] Thus, SUV_{peak} may be a more robust value than SUV_{max}.[17] However, tumor SUV_{max} is more readily available in standard clinical practice than SUV_{peak}. Furthermore, in this study, we found that the accuracy of the HLI_{max}, based on the SUV_{max}, was higher than that of the HLI_{peak}, based on the SUV_{peak}. Therefore, our findings suggest that SUV_{max} can be used for the evaluation of HLI. Furthermore, a limitation of SUV_{peak} is that it varies, depending on the size of the region of interest.[17]

The difference between rPET and HLI is the reference value. In the rPET study, the SUV_{max} of the liver was used for normalization,[11] whereas, in this study, the SUV_{mean} of the liver was used. FDG uptake in the liver is often heterogeneous, so SUV_{mean} may be less influenced by noise than SUV_{max}. It has been reported that the value of SUV_{mean} is more stable than SUV_{max}.[18] Therefore, liver SUV_{mean} may be a more appropriate value to use for normalization than SUV_{max}.

A dichotomized DS was used to distinguish PET-negative and PET-positive images, with a cutoff of DS2/3 or DS3/4.[9,11,12,19] DS1 and DS2 are accepted as a complete metabolic response. DS3 is also considered to likely represent a complete

TABLE 4. Diagnostic Accuracy of the Hodgkin Lymphoma Index for Deauville Score After 2 Cycles of Chemotherapy (PET2)

	Sensitivity, %	Specificity, %	Accuracy, %	95% Confidence Interval
HLI_{max}				
Cutoff value of 1.18 (DS1–DS2/DS3–DS5)	73.8	78.3	81.8	0.673–0.861
Cutoff value of 1.58 (DS1–DS3/DS4–DS5)	100.0	91.8	92.8	0.923–0.998
HLI_{peak}				
Cutoff value of 0.94 (DS1–DS2/DS3–DS5)	73.9	71.7	71.1	0.666–0.856
Cutoff value of 1.26 (DS1–DS3/DS4–DS5)	100.0	90.4	90.4	0.923–0.998

metabolic response; however, as stated in the imaging companion article to the Lugano criteria by Barrington et al,[2] in clinical trials that de-escalate treatment based on FDG PET response, it may be preferable to consider a DS3 as an inadequate response to avoid undertreating patients.[2,20] In this study, the overall predictive accuracy of HLI for PET-negative or PET-positive images was higher when scans were dichotomized as DS1–DS3/DS4–DS5, rather than DS1–DS2/DS3–DS5. These findings are concordant with previously reported results.[9] As the cutoff value was changed from DS2/DS3 to DS3/DS4, the rates of agreement between the 2 physicians increased. Therefore, interobserver agreement may be improved if the DS1–DS3 subgroup is defined as PET-negative and the DS4–DS5 subgroup as PET-positive.

DSs on PET1 and PET2 were evaluated independently by the 2 readers in this study. The overall agreement between the 2 physicians was higher on PET1 than on PET2. We hypothesize that as the treatment progresses, the intensity of FDG uptake in the tumor decreases, complicating the visual assessment. Therefore, an objective, quantitative response assessment based on SUV may be particularly valuable later during therapy.

This study is not without limitations. First, it is a retrospective study of scans from a multicenter trial, so the PET parameters may be affected by the different PET/CT scanners.[21] However, this effect is minimized by normalizing to the SUV_{mean} of the liver on each individual scan. Second, it is difficult to set an optimal cutoff value between PET-negative and PET-positive images with this limited number of patients. The cutoff values that we identified are different than values obtained in other previous studies.[11,12] A study with a large number of cases from multiple centers is necessary to define the optimal cutoff values. In addition, future research should assess for an association between HLI and event-free survival. Lastly, additional research is warranted to explore disease response assessments in pediatric HL using PET/MRI,[22] rather than PET/CT, to minimize radiation exposure from diagnostic imaging.

CONCLUSION

In this cohort of pediatric patients with high-risk HL, the 5-point DS had interobserver variability due to the physicians' subjective visual assessments. HLI was developed to provide an objective, quantitative measure of disease response. HLI values may improve interreader agreement of FDG PET/CT scans for lymphoma response assessments. Thus, HLI may contribute to optimizing response-adapted therapy in HL.

REFERENCES

1. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol* 2014;32(27):3059–68.
2. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the international conference on malignant lymphomas imaging working group. *J Clin Oncol* 2014;32(27):3048–58.
3. Mauz-Korholz C, Landman-Parker J, Balwierz W, et al. Response-adapted omission of radiotherapy and comparison of consolidation chemotherapy in children and adolescents with intermediate-stage and advanced-stage classical Hodgkin lymphoma (EuroNet-PHL-C1): a titration study with an open-label, embedded, multinational, non-inferiority, randomised controlled trial. *Lancet Oncol* 2022;23(1):125–137.
4. Mauz-Korholz C, Landman-Parker J, Fernandez-Teijeiro A, et al. Response-adapted omission of radiotherapy in children and adolescents with early-stage classical Hodgkin lymphoma and an adequate response to vincristine, etoposide, prednisone, and doxorubicin (EuroNet-PHL-C1): a titration study. *Lancet Oncol* 2023;24(3):252–261.
5. Castellino SM, Pei Q, Parsons SK, et al. Brentuximab vedotin with chemotherapy in pediatric high-risk Hodgkin's lymphoma. *N Engl J Med* 2022;387(18):1649–1660.
6. Metzger ML, Link MP, Billett AL, et al. Excellent outcome for pediatric patients with high-risk Hodgkin lymphoma treated with brentuximab vedotin and risk-adapted residual node radiation. *J Clin Oncol* 2021;39(20):2276–2283.
7. Kelly KM, Cole PD, Pei Q, et al. Response-adapted therapy for the treatment of children with newly diagnosed high risk Hodgkin lymphoma (AHOD0831): a report from the Children's Oncology Group. *Br J Haematol* 2019;187(1):39–48.
8. Friedman DL, Chen L, Wolden S, et al. Dose-intensive response-based chemotherapy and radiation therapy for children and adolescents with newly diagnosed intermediate-risk Hodgkin lymphoma: a report from the Children's Oncology Group study AHOD0031. *J Clin Oncol* 2014;32(32):3651–8.
9. Kluge R, Chavdarova L, Hoffmann M, et al. Inter-reader reliability of early FDG-PET/CT response assessment using the Deauville scale after 2 cycles of intensive chemotherapy (OEPA) in Hodgkin's lymphoma. *PLoS One* 2016;11(3):e0149072.
10. Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET-scan in lymphoma. *Leuk Lymphoma* 2009;50(8):1257–60.
11. Annunziata S, Cuccaro A, Calcagni ML, Hohaus S, Giordano A, Rufini V. Interim FDG-PET/CT in Hodgkin lymphoma: the prognostic role of the ratio between target lesion and liver SUV_{max} (rPET). *Ann Nucl Med* 2016;30(8):588–592.
12. Hasenclever D, Kurch L, Mauz-Korholz 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(7):1301–8.
13. 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(2):328–54.
14. Fahey FH, Palmer MR, Strauss KJ, Zimmerman RE, Badawi RD, Treves ST. Dosimetry and adequacy of CT-based attenuation correction for pediatric PET: phantom study. *Radiology* 2007;243(1):96–104.
15. Altman DG. *Practical Statistics for Medical Research*. New York: Taylor & Francis; 1990.
16. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol* 2010;195(2):310–320.
17. Vanderhoek M, Perlman SB, Jeraj R. Impact of the definition of peak standardized uptake value on quantification of treatment response. *J Nucl Med* 2012;53(1):4–11.
18. Kuhnert G, Boellaard R, Sterzer S, et al. Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis. *Eur J Nucl Med Mol Imaging* 2016;43(2):249–258.
19. Itti E, Meignan M, Berriolo-Riedinger A, et al. An international confirmatory study of the prognostic value of early PET/CT in diffuse large B-cell lymphoma: comparison between Deauville criteria and ΔSUV_{max}. *Eur J Nucl Med Mol Imaging* 2013;40(9):1312–1320.
20. Meignan M, Barrington S, Itti E, Gallamini A, Haioun C, Polliack A. Report on the 4th International Workshop on Positron Emission Tomography in Lymphoma held in Menton, France, 3–5 October 2012. *Leuk Lymphoma* 2014;55(1):31–7.
21. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med* 2009;50 suppl 1:11s–20s.
22. Veit-Haibach P, Ahlstrom H, Boellaard R, et al. International EANM-SNMMI-ISMRM consensus recommendation for PET/MRI in oncology. *Eur J Nucl Med Mol Imaging* 2023;50(12):3513–3537.