
Interim PET Evaluation in Diffuse Large B-Cell Lymphoma Using Published Recommendations: Comparison of the Deauville 5-Point Scale and the $\Delta\text{SUV}_{\text{max}}$ Method

Jan Rekowski¹, Andreas Hüttmann², Christine Schmitz², Stefan P. Müller³, Lars Kurch⁴, Jörg Kotzerke⁵, Christiane Franzius⁶, Matthias Weckesser⁷, Frank M. Bengel⁸, Martin Freesmeyer⁹, Andreas Hertel¹⁰, Thomas Krohn¹¹, Jens Holzinger¹², Ingo Brink¹³, Uwe Haberkorn¹⁴, Fonyuy Nyuyki¹⁵, Daniëlle M.E. van Assema¹⁶, Lilli Geworski¹⁷, Dirk Hasenclever¹⁸, Karl-Heinz Jöckel¹, and Ulrich Dührsen²

¹Institut für Medizinische Informatik, Biometrie, und Epidemiologie, Universitätsklinikum, Essen, Germany; ²Klinik für Hämatologie, Universitätsklinikum, Essen, Germany; ³Klinik für Nuklearmedizin, Universitätsklinikum, Essen, Germany; ⁴Klinik und Poliklinik für Nuklearmedizin, Universitätsklinikum, Leipzig, Germany; ⁵Klinik für Nuklearmedizin, Universitätsklinikum Carl Gustav Carus, Dresden, Germany; ⁶Zentrum für moderne Diagnostik (Zemodi), Zentrum für Nuklearmedizin und PET/CT, Bremen, Germany; ⁷Klinik für Nuklearmedizin, Universitätsklinikum, Münster, Germany; ⁸Klinik für Nuklearmedizin, Medizinische Hochschule, Hannover, Germany; ⁹Klinik für Nuklearmedizin, Universitätsklinikum, Jena, Germany; ¹⁰Klinik für Diagnostische und Therapeutische Nuklearmedizin, Klinikum, Fulda, Germany; ¹¹Klinik für Nuklearmedizin, Universitätsklinikum, Aachen, Germany; ¹²Institut für Diagnostische Radiologie, Neuroradiologie, und Nuklearmedizin, Johannes Wesling Klinikum, Minden, Germany; ¹³Klinik für nuklearmedizinische Diagnostik und Therapie, Ernst von Bergmann Klinikum, Potsdam, Germany; ¹⁴Radiologische Klinik und Poliklinik, Universitätsklinikum, Heidelberg, Germany; ¹⁵Klinik für Nuklearmedizin, Brüderkrankenhaus St. Josef, Paderborn, Germany; ¹⁶Department of Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands; ¹⁷Stabsstelle Strahlenschutz und Abteilung Medizinische Physik, Medizinische Hochschule, Hannover, Germany; and ¹⁸Institut für Medizinische Informatik, Statistik, und Epidemiologie, Universität Leipzig, Leipzig, Germany

The value of interim ¹⁸F-FDG PET/CT (iPET)-guided treatment decisions in patients with diffuse large B-cell lymphoma (DLBCL) has been the subject of much debate. This investigation focuses on a comparison of the Deauville score and the change-in-SUV_{max} ($\Delta\text{SUV}_{\text{max}}$) approach—2 methods to assess early metabolic response to standard chemotherapy in DLBCL. **Methods:** Of 609 DLBCL patients participating in the PET-Guided Therapy of Aggressive Non-Hodgkin Lymphomas trial, iPET scans of 596 patients originally evaluated using the $\Delta\text{SUV}_{\text{max}}$ method were available for post hoc assessment of the Deauville score. A commonly used definition of an unfavorable iPET result according to the Deauville score is an uptake greater than that of the liver, whereas an unfavorable iPET scan with regard to the $\Delta\text{SUV}_{\text{max}}$ approach is characterized as a relative reduction of the SUV_{max} between baseline and iPET staging of less than or equal to 66%. We investigated the 2 methods' correlation and concordance by Spearman rank correlation coefficient and the agreement in classification, respectively. We further used Kaplan–Meier curves and Cox regression to assess differences in survival between patient subgroups defined by the pre-specified cutoffs. Time-dependent receiver-operating-characteristic curve analysis provided information on the methods' respective discrimination performance. **Results:** Deauville score and $\Delta\text{SUV}_{\text{max}}$ approach differed in their iPET-based prognosis. The $\Delta\text{SUV}_{\text{max}}$ approach outperformed the Deauville score in terms of discrimination performance—most likely because of a high number of false-positive decisions by the Deauville score. Cutoff-independent discrimination

performance remained low for both methods, but cutoff-related analyses showed promising results. Both favored the $\Delta\text{SUV}_{\text{max}}$ approach, for example, for the segregation by iPET response, where the event-free survival hazard ratio was 3.14 (95% confidence interval, 2.22–4.46) for $\Delta\text{SUV}_{\text{max}}$ and 1.70 (95% confidence interval, 1.29–2.24) for the Deauville score. **Conclusion:** When considering treatment intensification, the currently used Deauville score cutoff of an uptake above that of the liver seems to be inappropriate and associated with potential harm for DLBCL patients. The $\Delta\text{SUV}_{\text{max}}$ criterion of a relative reduction in SUV_{max} of less than or equal to 66% should be considered as an alternative.

Key Words: diffuse large B-cell lymphoma; early metabolic response to therapy; interim PET; Deauville score; deltaSUVmax approach

J Nucl Med 2021; 62:37–42

DOI: 10.2967/jnumed.120.244145

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of non-Hodgkin lymphoma, showing a widely varying response to standard chemoimmunotherapy usually encompassing 6 cycles of cyclophosphamide, doxorubicin, vincristine, prednisone, and, for patients positive for the cluster of differentiation molecule 20, rituximab (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone [R-CHOP]) (1). Although approximately one third of all patients progress after 6 cycles of R-CHOP, a substantial proportion of patients might be overtreated (2,3). Thus, risk-adapted treatment approaches are urgently needed but demand precise and reliable tools to guide therapy.

¹⁸F-FDG PET has been shown to predict outcome in aggressive lymphomas (4). After 1–4 cycles of treatment, an interim PET/CT

Received Mar. 2, 2020; revision accepted Apr. 9, 2020.

For correspondence or reprints contact: Jan Rekowski, Institut für Medizinische Informatik, Biometrie, und Epidemiologie, Universitätsklinikum Essen, Universität Duisburg-Essen; Hufelandstraße 55, 45147 Essen, Germany. E-mail: jan.rekowski@uk-essen.de
Published online May 8, 2020.

COPYRIGHT © 2021 by the Society of Nuclear Medicine and Molecular Imaging.

(iPET) scan can determine the degree of remaining glucose metabolism (5). Different methods for ^{18}F -FDG PET response assessment at interim staging exist: staging guidelines recommend the Deauville score, a 5-point ordinal scale based mainly on a visual comparison between the glucose uptake of lymphoma tissue and the uptake of liver or mediastinum, respectively (6). A cutoff for the definition of an unfavorable prognosis or a positive iPET response is commonly defined as an uptake greater than that of the liver. An alternative method is the change-in-SUV_{max} ($\Delta\text{SUV}_{\text{max}}$) approach using the SUV_{max} of the hottest tumor lesion (7). This approach compares SUV_{max} at baseline and iPET staging. An unfavorable iPET result is defined as a relative SUV_{max} reduction of less than or equal to 66%—a cutoff that has been confirmed in several studies (3,8–10). Advantages of the Deauville score are that it is easy to apply and requires only the iPET scan. It is, however, associated with an increased false-positive rate and susceptibility for interreader variability (11–13). A disadvantage of the $\Delta\text{SUV}_{\text{max}}$ approach is that it requires a baseline scan as a reference. Moreover, some have argued that it classifies too few patients to an unfavorable prognosis to be useful to guide therapy (14). In contrast to the Deauville score, it provides semiquantitative assessment that is independent of any background noise and less prone to interreader variability.

Despite their competing nature, little work on a direct comparison of the 2 methods is available in the literature (9,10,12). The PET-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL) trial has recently shown that iPET response predicts outcome when assessed using the $\Delta\text{SUV}_{\text{max}}$ approach. In a post hoc analysis of this study, iPET scans were reassessed for the Deauville score; results for the entire trial population consisting of a variety of aggressive B-cell and T-cell lymphoma subtypes have been described before (4). Here, we focus on DLBCL, providing data on the concordance between Deauville score and $\Delta\text{SUV}_{\text{max}}$ method and their respective discrimination performance.

MATERIALS AND METHODS

Study Population

The PETAL trial (registered under ClinicalTrials.gov NCT00554164 and EudraCT 2006-001641-33) was a multicenter randomized controlled study for patients with newly diagnosed aggressive non-Hodgkin lymphomas investigating treatment options in patients stratified by iPET response (15). The Federal Institute for Drugs and Medical Devices (reference no. 61-3910-4032976) and the ethics committees of all participating sites (reference no. 07-3366) approved the study, and all patients provided written informed consent.

Study Design

Patients were treated with R-CHOP-14 but with 3 wk between cycles 2 and 3 to prevent false-positive results in iPET staging, which uniformly took place after the second cycle (16). Patients with a favorable iPET response ($\Delta\text{SUV}_{\text{max}} > 66\%$) received either 4 more cycles of R-CHOP or the same treatment plus 2 extra doses of rituximab. In patients with an unfavorable iPET response ($\Delta\text{SUV}_{\text{max}} \leq 66\%$), treatment options included continuation of R-CHOP for 6 additional cycles and receipt of 6 blocks of a more intensive protocol originally intended for Burkitt lymphoma (17). Outcome, however, remained unaffected by treatment changes, thus providing an opportunity to use the entire study population to assess the prognostic value of iPET (4).

^{18}F -FDG PET/CT Imaging

In the PETAL trial, 23 nuclear medicine institutions participated. Their local nuclear medicine specialists performed and evaluated ^{18}F -FDG PET images according to the PETAL study protocol as described previously (4). iPET was required to be performed under the same

conditions as at baseline staging, and the same PET scanner and reconstruction method had to be used. All scans had to cover a body area at least from the skull base to the mid thigh, PET scans had to be acquired 60 ± 10 min after tracer injection, patients had to have fasted for at least 4 h, and blood glucose levels could not exceed 200 mg/dL. The median chemotherapy-free interval before iPET scanning was 20 d, and no patient's individual chemotherapy-free interval was shorter than 10 d.

iPET Evaluation

During the trial, iPET scans were evaluated decentrally by local nuclear medicine physicians using the $\Delta\text{SUV}_{\text{max}}$ method. An iPET response was regarded as unfavorable when the relative SUV_{max} reduction, compared with baseline, was 66% or less (4,7). Unfavorable iPET scans without unphysiologic ^{18}F -FDG uptake according to visual criteria were also regarded as negative. This modification of the $\Delta\text{SUV}_{\text{max}}$ approach considered that a return to physiologic activity may require less than a 66% SUV_{max} reduction in patients with an iPET lacking unphysiologic ^{18}F -FDG uptake. After conclusion of the trial, for 502 of 609 DLBCL patients, iPET scans were reevaluated by any 1 of 3 experienced nuclear medicine physicians using the Deauville scale and defining an unfavorable iPET result as a Deauville score of more than 3—an uptake greater than liver SUV_{max} (6). If retrievable, iPET scans not available for centralized evaluation were analyzed in the same way by local nuclear medicine experts, yielding 94 additional Deauville scale assessments. Thus, the $\Delta\text{SUV}_{\text{max}}$ evaluation was uniformly performed decentrally, whereas the Deauville scale evaluation was done in a predominantly centralized manner. A diagram providing an overview of the patient flow in terms of iPET assessments is shown in Figure 1.

Outcome Variables

The prespecified primary endpoint of the PETAL trial (event-free survival defined as the time from iPET staging to disease progression, treatment discontinuation due to excessive toxicity, switch to a nonprotocol treatment, relapse, or death from any cause) was also the main focus of this investigation. We assessed the robustness of our results across more regularly used outcomes and also included the secondary endpoints time to progression, overall survival, and progression-free survival—respectively defined as the time from iPET staging to disease progression, to death from any cause, and to disease progression or death from any cause.

Statistical Analysis

We used the reverse Kaplan–Meier method to calculate the patients' median follow-up time. The Spearman rank correlation coefficient assessed the association between the 2 iPET methods in general, whereas agreement in classification indicated concordance between the subgroups defined by the cutoffs for Deauville score and $\Delta\text{SUV}_{\text{max}}$ approach. Kaplan–Meier curves provided the possibility to investigate differences in outcome between these subgroups, and hazard ratios obtained by Cox regression quantified these differences. To characterize the discrimination performance of the $\Delta\text{SUV}_{\text{max}}$ approach and Deauville score, we used time-dependent receiver-operating-characteristic (ROC) analysis to estimate the area under the ROC curve (AUC), sensitivity, and specificity, as well as the 2 methods' predictive values (18). We here made use of the nearest-neighbor estimator with the time point of interest being 2 y after iPET staging. A simple bootstrap with 10,000 iterations allowed for the construction of empiric 95% confidence intervals (CIs) for all measures of discrimination performance. In terms of these discrimination measures, we defined an unfavorable iPET response with any of the 2 methods as a positive test result. Note that for the analyses relating to the 66% cutoff and dividing the population into 2 parts (concordance with the Deauville score cutoff, Kaplan–Meier estimation, hazard ratio, sensitivity, specificity, and predicted values),

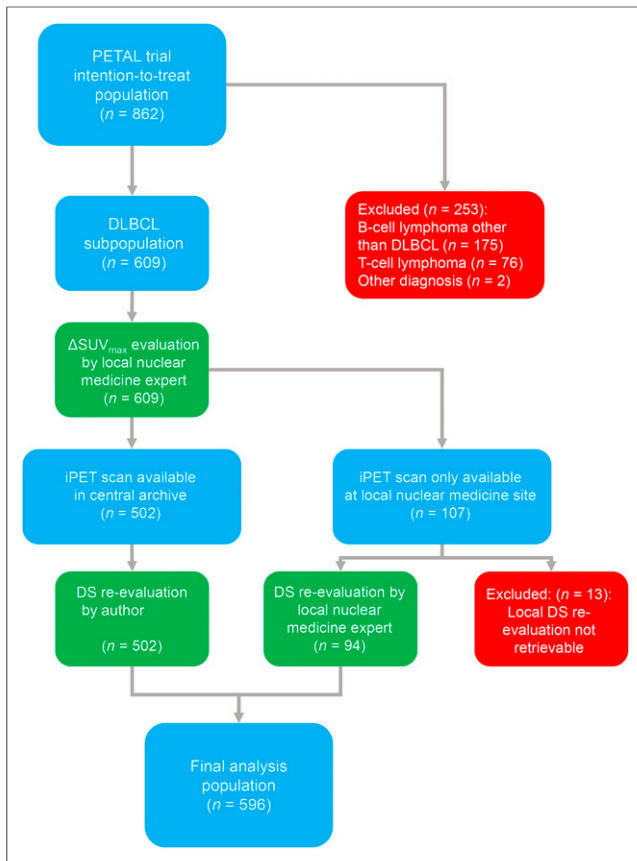


FIGURE 1. Flow of patients in terms of iPET assessments and resulting subpopulations. DS = Deauville score.

the above-mentioned modification of the $\Delta\text{SUV}_{\text{max}}$ approach for patients with an iPET lacking unphysiologic ^{18}F -FDG uptake was used. For the correlation with the ordinal Deauville score as well as for the ROC curve and the AUC, however, this was not feasible because these analyses are based on the continuous $\Delta\text{SUV}_{\text{max}}$ variable, not making any binary distinction into good and poor prognosis. We used R, version 3.5.1 (R Core Team), to perform all statistical analyses.

RESULTS

Clinical Characteristics and Follow-up

Our investigation was restricted to DLBCL patients from the PETAL intention-to-treat population with available data from post hoc Deauville score analysis, that is, 596 (97.9%) DLBCL patients participating in the PETAL trial (Fig. 1). The median follow-up time in the restricted population was 51.4 mo (95% CI, 49.7–53.7 mo), which was comparable to the whole DLBCL subgroup of the PETAL trial. Overall, differences in the characteristics of the subgroup studied here and the entire DLBCL population of the PETAL trial were negligible (Table 1 and Hüttmann et al. (16)). With regard to event-free survival, for 207 patients an event terminated their follow-up time—for 164 of them before the ROC analysis time point of interest at 2 y after iPET staging. Kaplan–Meier curves for the entire cohort can be found for all endpoints in Supplemental Figure 1 (supplemental materials are available at <http://jnm.snmjournals.org>).

Ninety-two of 596 patients had an SUV_{max} reduction of 66% or less. In 29 of these, iPET scans were devoid of unphysiologic

^{18}F -FDG uptake, resulting in their reassignment to the favorable prognosis group according to the modification of the $\Delta\text{SUV}_{\text{max}}$ method described before. Patients thus reclassified tended to have a very low baseline SUV_{max} (median, 7.2; first quartile, 5.4; third quartile, 9.8). Their outcome resembled that of patients with an SUV_{max} reduction of more than 66% (Supplemental Fig. 2).

Correlation and Concordance

The Spearman ρ between Deauville score and $\Delta\text{SUV}_{\text{max}}$ approach was 0.31 (95% CI, 0.23–0.38). The number of patients with an unfavorable iPET response given the respective cutoffs was more than 4 times higher with the Deauville score (45.3%; 270/596) than with the $\Delta\text{SUV}_{\text{max}}$ approach (10.4%; 62/596). Cutoff-based concordance was 63.1% (376/596), with more than a third of the patients having a $\Delta\text{SUV}_{\text{max}}$ -favorable but a Deauville score–unfavorable iPET response (Fig. 2A). Looking at the event-free survival curves by concordance, we found that patients with a doubly favorable iPET response had the best outcome and that doubly unfavorable patients had the worst. The event-free survival curve for patients with a $\Delta\text{SUV}_{\text{max}}$ -favorable but Deauville score–unfavorable iPET response, however, was rather close to the survival curve of doubly favorable patients (Fig. 2B).

Discrimination Performance

The event-free survival Kaplan–Meier estimator at the ROC analysis time point of interest (2 y after iPET) was 71.6% (95% CI, 67.0%–76.6%). Global cutoff-independent discrimination performance as indicated by the AUC was poor for both approaches in all 4 endpoints but for the $\Delta\text{SUV}_{\text{max}}$ approach consistently higher than for the Deauville score (Fig. 3A). Accordingly, both ROC curves tended to be flat for all endpoints (Supplemental Fig. 3). Regarding the given cutoffs, Kaplan–Meier event-free survival curves graphically showed more pronounced segregation of patients with favorable and unfavorable iPET response with the $\Delta\text{SUV}_{\text{max}}$ approach than with the Deauville score (Fig. 4). The same was true for the 3 secondary endpoints (Supplemental Fig. 4). Associated hazard ratios were in line with these findings—for example, for event-free survival with a hazard ratio between unfavorable and favorable patients of more than 3 with the $\Delta\text{SUV}_{\text{max}}$ approach and less than 2 with the Deauville score (Fig. 3B). Sensitivity was higher for the Deauville score (52.5%; 95% CI, 45.5%–59.3%) than for the $\Delta\text{SUV}_{\text{max}}$ approach (24.6%; 95% CI, 18.6%–31.2%), whereas specificity was lower for the Deauville score (57.5%; 95% CI, 52.8%–62.2%, vs. 88.8%; 95% CI, 85.9%–91.7%), indicating a higher false-positive rate. The positive predictive value favored the $\Delta\text{SUV}_{\text{max}}$ cutoff over all possible realizations of the unknown event prevalence. In contrast, the negative predictive value slightly favored the Deauville cutoff over $\Delta\text{SUV}_{\text{max}}$ (Supplemental Fig. 5). For all endpoints, the numeric results of time-dependent ROC analyses and Cox regression are available in Supplemental Table 1.

DISCUSSION

In this comparison of methods assessing early metabolic response to standard R-CHOP treatment in DLBCL patients, we showed that the $\Delta\text{SUV}_{\text{max}}$ approach outperformed the Deauville score in terms of discrimination performance for event-free survival, progression-free survival, overall survival, and time to progression. This applied to the global discrimination measure AUC as well as to the hazard ratios between subgroups defined by the prespecified cutoffs of the $\Delta\text{SUV}_{\text{max}}$ approach and Deauville score. The concordance of iPET

TABLE 1
Characteristics for All Evaluable Patients by Subpopulations Defined by Cutoffs for the 2 Methods

Characteristics	All evaluable	$\Delta\text{SUV}_{\text{max}} > 66\%$	$\Delta\text{SUV}_{\text{max}} \leq 66\%$	$\text{DS} \leq 3$	$\text{DS} > 3$
No. of patients	596	534	62	326	270
Median age (y)	62 (IQR, 51–70)	62 (IQR, 51–70)	62 (IQR, 50–69)	62 (IQR, 52–70)	61 (IQR, 50–70)
Age > 60 y (n)	308 (51.8%)	274 (51.4%)	34 (54.8%)	174 (53.5%)	134 (49.6%)
Male sex (n)	331 (55.5%)	294 (55.1%)	37 (59.7%)	185 (56.7%)	146 (54.1%)
ECOG performance status ≥ 2 (n)	59 (9.9%)	47 (8.8%)	12 (19.4%)	23 (7.1%)	36 (13.3%)
Ann Arbor stage III or IV (n)	349 (58.7%)	304 (57.0%)	45 (72.6%)	173 (53.2%)	176 (65.2%)
Extranodal sites > 1 (n)	192 (32.3%)	166 (31.1%)	26 (41.9%)	94 (28.9%)	98 (36.3%)
Lactate dehydrogenase > ULN (n)	329 (55.3%)	289 (54.2%)	40 (64.5%)	151 (46.5%)	178 (65.9%)
International Prognostic Index (n)					
Low risk	219 (36.8%)	205 (38.5%)	14 (22.6%)	140 (43.1%)	79 (29.3%)
Low to intermediate risk	155 (26.1%)	137 (25.7%)	18 (29.0%)	83 (25.5%)	72 (26.7%)
High to intermediate risk	124 (20.8%)	110 (20.6%)	14 (22.6%)	61 (18.8%)	63 (23.3%)
High risk	97 (16.3%)	81 (15.2%)	16 (25.8%)	41 (12.6%)	56 (20.7%)

DS = Deauville score; IQR = interquartile range; ECOG = Eastern Cooperative Oncology Group; ULN = upper limit of normal. Data on International Prognostic Index was not available for 1 patient with $\Delta\text{SUV}_{\text{max}} > 66\%$ and $\text{DS} > 3$.

response with regard to the 2 methods' most commonly used definitions was relatively low, most likely because of a high false-positive rate associated with the Deauville score definition that an unfavorable iPET response consists of uptake greater than liver SUV_{max} .

Our observations complement smaller studies comparing $\Delta\text{SUV}_{\text{max}}$ and Deauville score for evaluation of iPET scans after 2 cycles of R-CHOP or similar regimens. In one study, iPET was prognostic only when the scans were evaluated by the $\Delta\text{SUV}_{\text{max}}$ method, whereas application of the Deauville criteria failed to yield statistically significant outcome differences (8). In another study, the Deauville scale appeared to predict event-free survival better than the $\Delta\text{SUV}_{\text{max}}$ approach, but an effect of iPET on overall survival was seen only with the latter method (19). Any comparison of these studies with ours, however, must be exercised with caution, because they differed with regard to treatment performance, use of granulocyte colony-stimulating factor, and iPET timing (4).

The high false-positive rate of the Deauville score has been reported before and is of utter importance in the given setting (11).

The general aim of early response to treatment assessments is identifying DLBCL patients who do not respond sufficiently to standard R-CHOP therapy and guiding them to different treatment approaches. Such alternative therapies, however, are usually more aggressive or rather expensive. With the high false-positive rate of the Deauville score, more aggressive approaches would imply the ethical issue of increased toxicity in patients who would also have responded satisfactorily to standard therapy, whereas expensive treatments such as modern cellular therapies would result in a waste of scarce resources. Given its higher specificity, the $\Delta\text{SUV}_{\text{max}}$ cutoff spares these patients from this potential harm—at the price of a smaller fraction of patients being selected for alternative treatment approaches.

There is another commonly used cutoff for the Deauville score, that is, a Deauville score of more than 2, defined as an uptake greater than that of the mediastinum. Given its definition, the false-positive rate observed with a Deauville score of more than 3 even increases with a Deauville score of more than 2 (Supplemental

Fig. 6); we therefore do not recommend its use in the identification of R-CHOP nonresponders. Two smaller studies suggest cutoffs above liver activity to be more appropriate in segregating the DLBCL population after 2 treatment cycles than a Deauville score of more than 3, but neither of the 2 cutoffs proposed (1.4- and 1.6-fold liver SUV_{max} , respectively) has been validated so far (20,21). Nevertheless, it appears appealing to translate the Deauville score from its ordinal scale to a quantitative scale similar to the $\Delta\text{SUV}_{\text{max}}$ approach to have more potential cutoffs to choose from (22,23).

Although the cutoff-based Kaplan–Meier curves and associated hazard ratios between unfavorable and favorable patients indicate good segregation, the AUC is relatively poor

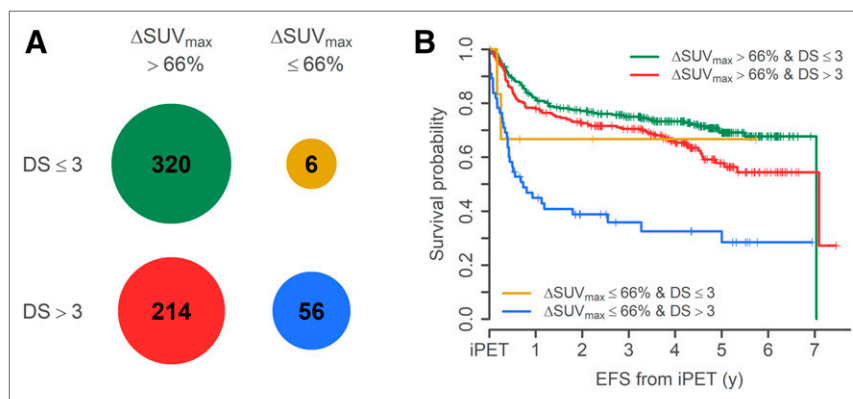


FIGURE 2. (A) Concordance between $\Delta\text{SUV}_{\text{max}}$ and Deauville score cutoff. (B) Kaplan–Meier event-free survival curves by concordance category. DS = Deauville score; EFS = event-free survival.

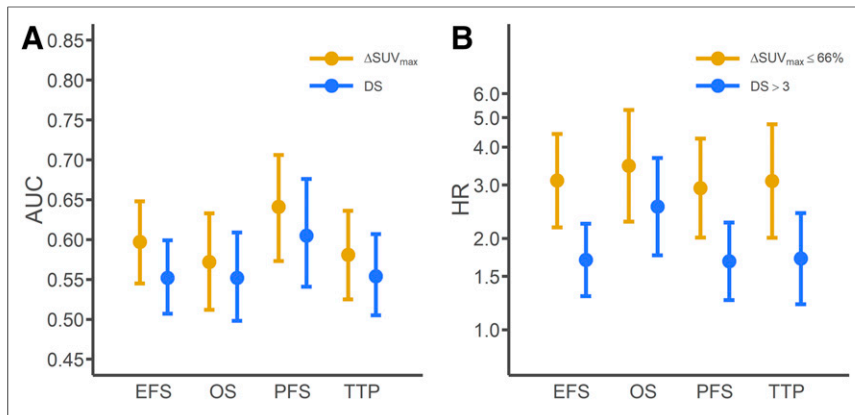


FIGURE 3. AUC (A) and Cox regression hazard ratio (B; logarithmic scale) with 95% CI by method and time-to-event endpoint. DS = Deauville score; EFS = event-free survival; HR = hazard ratio; OS = overall survival; PFS = progression-free survival; TTP = time to progression.

for both the Deauville score approach and the $\Delta\text{SUV}_{\text{max}}$ approach. In our opinion, this is a negligible concern because the aim is not to achieve high global discrimination performance across all possible cutoffs as assessed by the AUC. The aim rather is to identify patients at high risk of failing R-CHOP treatment, that is, to realize high local discrimination performance associated with a given criterion—focusing on the analyses associated with the 2 methods’ cutoffs. Another limitation of our comparison of $\Delta\text{SUV}_{\text{max}}$ approach and Deauville score is the modification of the $\Delta\text{SUV}_{\text{max}}$ approach in patients with an iPET scan lacking unphysiologic ^{18}F -FDG uptake. Although this modification does not affect concordance, hazard ratios, sensitivity, specificity, or predictive values, it may have an impact on correlation coefficient, ROC curve, and AUC. In the latter analyses, we used the relative reduction of SUV_{max} on the continuous scale for all patients, that is, also including those patients subsequently being reclassified to the favorable prognosis group because of lack of unphysiologic ^{18}F -FDG uptake according to visual criteria. Overall, this reclassification occurred in 29 patients with an unfavorable iPET response according to their actually measured relative SUV_{max} reduction. We again would like to highlight that the focus of this investigation was on the cutoff-based analyses, as the question of a therapy switch requires a binary yes or no decision. And, although numbers are small, the outcome of reclassified patients appeared to be similar to that of patients with an SUV_{max} reduction of more than

66%. Thus, the modification of the $\Delta\text{SUV}_{\text{max}}$ method introduced in the PETAL trial may also be of value in future investigations.

Despite the good local discrimination performance of the $\Delta\text{SUV}_{\text{max}}$ approach, much is still unknown about its properties. Although several authors (3,8–10) confirmed the 66% SUV_{max} reduction cutoff originally proposed by Lin et al. (7), available data on the interrater reliability and reproducibility of the $\Delta\text{SUV}_{\text{max}}$ approach are scarce. In our investigation, the Deauville score assessment is a predominantly centralized post hoc analysis, and consequently, different nuclear medicine specialists were involved in the Deauville score and $\Delta\text{SUV}_{\text{max}}$ assessments. Given the higher number of $\Delta\text{SUV}_{\text{max}}$ than Deauville score raters, this factor gives room for possibly increased interrater variation with the $\Delta\text{SUV}_{\text{max}}$ approach. Itti et al., however, found the 66% SUV_{max} reduction to be associated with a higher interrater reproducibility than a Deauville score of more than 3; overall, they rated interobserver agreement as “almost perfect” with the $\Delta\text{SUV}_{\text{max}}$ approach but only “substantial” when the Deauville score was applied (12). In the PETAL trial, 10% of all iPET scans’ $\Delta\text{SUV}_{\text{max}}$ results were reevaluated by nuclear medicine physicians from other trial sites, with the concordance between the first and second readers being 97.7% (4). By contrast, the agreement within pairs of experienced nuclear medicine physicians using the Deauville score has been reported to be 77%–90% (24). Nonetheless, this issue calls for additional investigations.

increased interrater variation with the $\Delta\text{SUV}_{\text{max}}$ approach. Itti et al., however, found the 66% SUV_{max} reduction to be associated with a higher interrater reproducibility than a Deauville score of more than 3; overall, they rated interobserver agreement as “almost perfect” with the $\Delta\text{SUV}_{\text{max}}$ approach but only “substantial” when the Deauville score was applied (12). In the PETAL trial, 10% of all iPET scans’ $\Delta\text{SUV}_{\text{max}}$ results were reevaluated by nuclear medicine physicians from other trial sites, with the concordance between the first and second readers being 97.7% (4). By contrast, the agreement within pairs of experienced nuclear medicine physicians using the Deauville score has been reported to be 77%–90% (24). Nonetheless, this issue calls for additional investigations.

CONCLUSION

The $\Delta\text{SUV}_{\text{max}}$ definition stating that an unfavorable iPET response consists of a relative SUV_{max} reduction of 66% or less appears to be a more suitable tool to assess early metabolic response to standard R-CHOP therapy in DLBCL patients than the Deauville score, as the Deauville score definition stating that an unfavorable iPET response consists of uptake above that of the liver (Deauville score > 3) seems to be associated with a high false-positive rate. When therapy intensification or a switch to an experimental treatment is considered, we recommend the $\Delta\text{SUV}_{\text{max}}$ approach instead of the Deauville score as a prognostic instrument in first-line DLBCL treatment guidance. Whether this is as a standalone tool or must be combined with other patient, tumor, or treatment characteristics requires further study.

DISCLOSURE

Deutsche Krebshilfe provided financial support (grants 107592 and 110515). No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

We thank the patients and investigators for their participation.

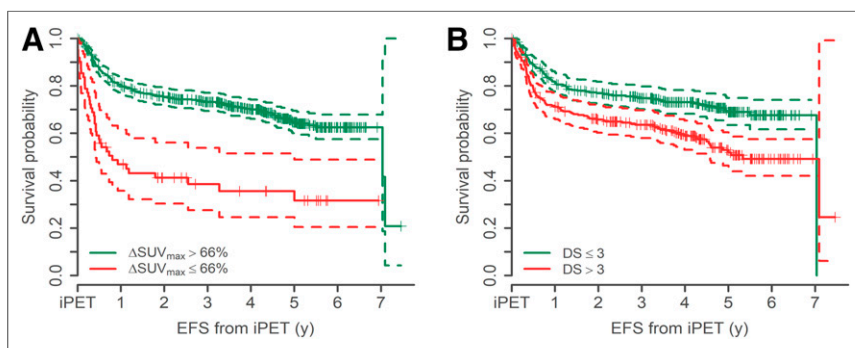


FIGURE 4. Kaplan–Meier event-free survival curves by iPET response according to $\Delta\text{SUV}_{\text{max}}$ (A) and Deauville score (B). DS = Deauville score; EFS = event-free survival.

KEY POINTS

QUESTION: Should the Deauville score (>3) or the $\Delta\text{SUV}_{\text{max}}$ approach ($\Delta\text{SUV}_{\text{max}} \leq 66\%$) be the preferred method to evaluate early response to standard R-CHOP therapy in DLBCL?

PERTINENT FINDINGS: In a post hoc analysis of the PETAL trial, the $\Delta\text{SUV}_{\text{max}}$ approach had higher discrimination performance than the Deauville score. This was especially true for local discrimination measures associated with the 2 methods' most commonly used cutoffs, because of an increased false-positive rate for the Deauville score.

IMPLICATIONS FOR PATIENT CARE: To prevent DLBCL patients with a favorable prognosis from harm resulting from unjustified iPET-based treatment intensification, the $\Delta\text{SUV}_{\text{max}}$ cutoff ($\Delta\text{SUV}_{\text{max}} \leq 66\%$) should be considered a standard tool for the assessment of early metabolic treatment response.

REFERENCES

1. Sehn LH, Donaldson J, Chanabhai M, et al. Introduction of combined CHOP plus rituximab therapy dramatically improved outcome of diffuse large B-cell lymphoma in British Columbia. *J Clin Oncol*. 2005;23:5027–5033.
2. Poeschel V, Held G, Ziepert M, et al. Four versus six cycles of CHOP chemotherapy in combination with six applications of rituximab in patients with aggressive B-cell lymphoma with favourable prognosis (FLYER): a randomised, phase 3, non-inferiority trial. *Lancet*. 2020;394:2271–2281.
3. Schmitz C, Hüttmann A, Müller SP, et al. Dynamic risk assessment based on positron emission tomography scanning in diffuse large B-cell lymphoma: post-hoc analysis from the PETAL trial. *Eur J Cancer*. 2020;124:25–36.
4. Dührsen U, Müller S, Hertenstein B, et al. Positron emission tomography-guided therapy of aggressive non-Hodgkin lymphomas (PETAL): a multicenter, randomized phase III trial. *J Clin Oncol*. 2018;36:2024–2034.
5. Burggraaff CN, de Jong A, Hoekstra OS, et al. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging*. 2019;46:65–79.
6. Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET scan in lymphoma. *Leuk Lymphoma*. 2009;50:1257–1260.
7. 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;48:1626–1632.
8. 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.
9. Toledano MN, Vera P, Tilly H, Jardin F, Becker S. Comparison of therapeutic evaluation criteria in FDG-PET/CT in patients with diffuse large-cell B-cell lymphoma: prognostic impact of tumor/liver ratio. *PLoS One*. 2019;14:e0211649.
10. Li X, Sun X, Li J, et al. Interim PET/CT based on visual and semiquantitative analysis predicts survival in patients with diffuse large B-cell lymphoma. *Cancer Med*. 2019;8:5012–5022.
11. Adams HJA, Kwee TC. Proportion of false-positive lesions at interim and end-of-treatment FDG-PET in lymphoma as determined by histology: systematic review and meta-analysis. *Eur J Radiol*. 2016;85:1963–1970.
12. 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 $\Delta\text{SUV}_{\text{max}}$. *Eur J Nucl Med Mol Imaging*. 2013;40:1312–1320.
13. 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:e0149072.
14. Merryman RW, LaCasce AS. Response-adapted therapy in aggressive lymphoma: not yet ready for clinical care. *J Clin Oncol*. 2018;36:2009–2011.
15. Dührsen U, Hüttmann A, Jöckel KH, Müller S. Positron emission tomography guided therapy of aggressive non-Hodgkin lymphomas: the PETAL trial. *Leuk Lymphoma*. 2009;50:1757–1760.
16. Hüttmann A, Rekowski J, Müller S, et al. Six versus eight doses of rituximab in patients with aggressive B cell lymphoma receiving six cycles of CHOP: results from the “Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas” (PETAL) trial. *Ann Hematol*. 2019;98:897–907.
17. Hoelzer D, Walowski J, Dohner H, et al. Improved outcome of adult Burkitt lymphoma/leukemia with rituximab and chemotherapy: report of a large prospective multicenter trial. *Blood*. 2014;124:3870–3879.
18. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56:337–344.
19. Mamot C, Klingbiel D, Hitz F, et al. Final results of a prospective evaluation of the predictive value of interim positron emission tomography in patients with diffuse large B-cell lymphoma treated with R-CHOP-14 (SAKK 38/07). *J Clin Oncol*. 2015;33:2523–2529.
20. Itti E, Juweid ME, Haioun C, et al. Improvement of early ^{18}F -FDG PET interpretation in diffuse large B-cell lymphoma: importance of the reference background. *J Nucl Med*. 2010;51:1857–1862.
21. Fan Y, Zhang Y, Yang Z, et al. Evaluating early interim fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography with the $\text{SUV}_{\text{max-liver}}$ based interpretation for predicting the outcome in diffuse large B-cell lymphoma. *Leuk Lymphoma*. 2017;58:1–9.
22. 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;41:1301–1308.
23. 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;30:588–592.
24. Burggraaff CN, Cornelisse AC, Hoekstra OS, et al. Interobserver agreement of interim and end-of-treatment ^{18}F -FDG PET/CT in diffuse large B-cell lymphoma: impact on clinical practice and trials. *J Nucl Med*. 2018;59:1831–1836.