Journal of Nuclear Medicine, published on May 8, 2020 as doi:10.2967/jnumed.120.244145

Interim PET evaluation in diffuse large B-cell lymphoma employing published recommendations: Comparison of the Deauville 5-point scale and the ΔSUV_{max} method

Jan Rekowski[1], Andreas Hüttmann[2], Christine Schmitz[2], Stefan P. Müller[3], Lars Kurch[4], Jörg Kotzerke[5], Christiane Franzius[6], Matthias Weckesser[7], Frank M. Bengel[8], Martin Freesmeyer[9], Andreas Hertel[10], Thomas Krohn[11], Jens Holzinger[12], Ingo Brink[13], Uwe Haberkorn[14], Fonyuy Nyuyki[15], Daniëlle M. E. van Assema[16], Lilli Geworski[17], Dirk Hasenclever[18], Karl-Heinz Jöckel[1], Ulrich Dührsen[2]

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

Short running title: Interim PET evaluation in DLBCL

Keywords: Diffuse large B-cell lymphoma; Early metabolic response to therapy; Interim positron emission tomography; Deauville score; deltaSUVmax approach

First/corresponding author: Dr. Jan Rekowski; Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Universität Duisburg-Essen; Hufelandstraße 55, 45147 Essen, Germany; jan.rekowski@uk-essen.de; +49-201-723-77215

ORCID: Jan Rekowski (0000-0002-5207-3864); Andreas Hüttmann (0000-0003-2230-3873); Christine Schmitz (0000-0003-2766-1696); Matthias Weckesser (0000-0003-4253-2390); Frank M. Bengel (0000-0003-4529-7963)

ABSTRACT

The value of interim 18F-fluorodeoxyglucose positron emission tomography (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 deltaSUVmax (ΔSUV_{max}) approach – two methods to assess early metabolic response to standard chemotherapy in DLBCL. Methods: Of 609 DLBCL patients participating in the Positron Emission Tomography-guided Therapy of Aggressive non-Hodgkin Lymphomas (PETAL) trial, iPET scans of 596 patients originally evaluated using the ΔSUV_{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 ΔSUV_{max} approach is characterized as a relative reduction of the maximum standardized uptake value between baseline and iPET staging of less than or equal to 66%. We investigated the two methods' correlation and concordance by Spearman's 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 cut-offs. Time-dependent receiver operating curve analysis provided information on the methods' respective discrimination performance. Results: Deauville score and ΔSUV_{max} approach differed in their iPET-based prognosis. The ΔSUV_{max} approach outperformed the Deauville score in terms of discrimination performance – most likely due to a high number of false-positive decisions by the Deauville score. Cut-off-independent discrimination performance remained low for both methods, but cut-off-related analyses showed promising results. Both favored the ΔSUV_{max} approach, e.g. for the segregation by iPET response, where the event-free survival hazard ratio was 3.14 (95% confidence interval (CI): 2.22 – 4.46) for ΔSUV_{max} and 1.70 (95% CI: 1.29 – 2.24) for the Deauville score. Conclusion: When considering treatment intensification, the currently used Deauville score cut-off of an uptake above that of the liver seems to be inappropriate and associated with potential harm for DLBCL patients. The ΔSUV_{max} criterion of a relative

reduction of the maximum standardized uptake value of less than or equal to 66% should be considered as an alternative.

INTRODUCTION

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 six cycles of cyclophosphamide, doxorubicin, vincristine, prednisone, and, for patients positive for the cluster of differentiation molecule 20, rituximab (R-CHOP) (*1*). While approximately one third of all patients progress after six cycles of R-CHOP, a substantial proportion of patients might be overtreated (*2,3*). Thus, risk adapted treatment approaches are urgently needed but demand for precise and reliable tools to guide therapy.

18F-Fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET) has been shown to predict outcome in aggressive lymphomas (*4*). After one to four cycles of treatment, an interim PET/CT (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 mainly based on a visual comparison between the glucose uptake of lymphoma tissue and uptake of liver or mediastinum, respectively (*6*). A cut-off 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 deltaSUVmax (ΔSUV_{max}) approach utilizing the maximum standardized uptake value (SUV_{max}) of the hottest tumor lesion (*7*). This approach compares SUV_{max} at baseline and interim PET staging. An unfavorable iPET result is defined as relative SUV_{max} reduction of less than or equal to 66% – a cut-off that has been confirmed in several studies (*3*,*8–10*). Advantages of the Deauville score are that it is easy to apply and only requires the iPET scan. It is, however, associated with an increased false-positive rate and susceptibility for inter-reader variability (*11–13*). A disadvantage of the ΔSUV_{max} approach is that it requires a baseline scan as a reference. Moreover, it has been 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

semi-quantitative assessment that is independent of any background noise and less prone to inter-reader variability.

Despite their competing nature, little work on a direct comparison of the two methods is available in the literature (*9,10,12*). The Positron Emission Tomography-guided Therapy of Aggressive non-Hodgkin Lymphomas (PETAL) trial has recently shown that iPET response predicts outcome when assessed using the ΔSUV_{max} approach. In a post-hoc analysis of this study, iPET scans were re-assessed 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 ΔSUV_{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 three weeks between cycles two and three to prevent false-positive results in iPET staging that uniformly took place after the second cycle (*16*). Patients with favorable iPET response (ΔSUV_{max} >66%) either received four more cycles of R-CHOP or the same treatment plus two extra doses of rituximab. In patients with unfavorable iPET response (ΔSUV_{max} ≤66%),

treatment options included continuation of R-CHOP for six additional cycles and receipt of six blocks of a more intensive protocol originally intended for Burkitt's lymphoma (*17*). Outcome, however, remained unaffected by treatment changes which provided an opportunity to use the entire study population to assess the prognostic value of iPET (*4*).

18F-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*). Performance of interim scanning was required to be 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 minutes after tracer injection, patients had to be fasted for at least four hours, and blood glucose levels were not allowed to exceed 200mg/dl. The median chemotherapy-free interval prior to iPET scanning was 20 days and no patient's individual chemotherapy-free interval was shorter than ten days.

Interim 18F-FDG PET/CT Evaluation

During the trial, iPET scans were evaluated de-centrally by local nuclear medicine physicians using the ΔSUV_{max} method. An iPET response was regarded as unfavorable when the relative SUV_{max} reduction compared to baseline was ≤66% (*4,7*). Unfavorable iPET scans without unphysiological ^{18}F-FDG uptake according to visual criteria were also regarded as negative. This modification of the ΔSUV_{max} approach considered that a return to physiological activity may require less than a 66% SUV_{max} reduction in patients with an iPET lacking unphysiological ^{18}F-FDG uptake. After conclusion of the trial, for 502 of 609 DLBCL patients, iPET scans were re-evaluated by either one of three experienced nuclear medicine physicians (SPM, LK, DMEvA) employing the Deauville scale (DS) and defining an unfavorable iPET result as DS >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 ΔSUV_{max} evaluation was uniformly performed de-centrally, 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 pre-specified 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 non-protocol 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. Spearman's rank correlation coefficient assessed the association between the two iPET methods in general, whereas agreement in classification indicated concordance between the subgroups defined by the cut-offs of Deauville score and ΔSUV_{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 ΔSUV_{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 two methods' predictive values (*18*). We here made use of the nearest neighbor estimator with the time point of interest being two years after iPET staging. A simple bootstrap with 10,000 iterations allowed for the construction of empirical 95% confidence intervals (CI) for all measures of discrimination performance. In terms of these discrimination measures, we defined an unfavorable iPET response with any of the two methods as a positive test result.

Note that for the analyses relating to the 66% cut-off and dividing the population into two parts (concordance with the Deauville score cut-off, Kaplan-Meier estimation, hazard ratio, sensitivity, specificity, and predicted values) the above-mentioned modification of the ΔSUV_{max} approach for patients with an iPET lacking unphysiological ^{18}F-FDG uptake was used. For the correlation with the ordinal Deauville score variable as well as for the ROC curve and the AUC, however, this was not feasible as these analyses are based on the continuous ΔSUV_{max} variable not making any binary distinction into good and poor prognosis. We used R, version 3.5.1 (R Core Team), to carry out 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 out of 609 (97.9%) DLBCL patients participating in the PETAL trial (Figure 1). Median follow-up time in the restricted population was 51.4 months (95% CI: 49.7–53.7 months), 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 (cf. 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 two years after iPET staging. Kaplan-Meier curves for the entire cohort can be found for all endpoints in Supplemental Figure 1.

Ninety-two of 596 patients had an SUV_{max} reduction ≤66%. In 29 of these, iPET scans were devoid of unphysiological ^{18}F-FDG uptake resulting in their re-assignment to the favorable prognosis group according to the modification of the ΔSUV_{max} method described before. Patients thus re-classified tended

to have very low baseline SUV_{max} (median: 7.2; 1st quartile: 5.4; 3rd quartile: 9.8). Their outcome resembled that of patients with an SUV_{max} reduction >66% (Supplemental Figure 2).

Correlation and Concordance

Spearman's rho between Deauville score and ΔSUV_{max} approach was 0.31 (95% CI: 0.23–0.38). The number of patients with unfavorable iPET response given the respective cut-offs was more than four times higher with the Deauville score (45.3%; 270/596) than with the ΔSUV_{max} approach (10.4%; 62/596). Cut-off-based concordance was 63.1% (376/596) – with more than a third of the patients having a ΔSUV_{max} favorable but Deauville score unfavorable iPET response (Figure 2A). Looking at the event-free survival curves by concordance, patients with double-favorable iPET response had the best outcome and double-unfavorable patients had the worst. The event-free survival curve of patients with ΔSUV_{max} favorable but Deauville score unfavorable iPET response, however, was rather close to the survival curve of double-favorable patients (Figure 2B).

Discrimination Performance

The event-free survival Kaplan-Meier estimator at the ROC analysis time point of interest (two years after iPET) was 71.6% (95% CI: 67.0–76.6%). Global cut-off-independent discrimination performance as indicated by the AUC was poor for both approaches in all four endpoints but for the ΔSUV_{max} approach consistently higher than for the Deauville score (Figure 3A). Accordingly, both ROC curves tended to be flat for all endpoints (Supplemental Figure 3). Regarding the given cut-offs, Kaplan-Meier event-free survival curves graphically showed more pronounced segregation of patients with favorable and unfavorable iPET response with the ΔSUV_{max} approach than with the Deauville score (Figure 4). The same was true for the three secondary endpoints (Supplemental Figure 4). Associated hazard ratios were in line with these findings, e.g., for event-free survival with a hazard ratio between unfavorable and favorable patients of more than three with the ΔSUV_{max} approach and less than two with the Deauville score (Figure 3B). Sensitivity was higher for the Deauville score (52.5% (95% CI: 45.5–59.3%)) than for the ΔSUV_{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 ΔSUV_{max} cut-off over all possible realizations of the unknown event prevalence. In contrast, the negative predictive value slightly favored the Deauville cut-off over ΔSUV_{max} (Supplemental Figure 5). For all endpoints, numerical 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 ΔSUV_{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 area under the ROC curve as well as to the hazard ratios between subgroups defined by the pre-specified cut-offs of ΔSUV_{max} approach and Deauville score. Concordance of iPET response with regard to the two methods' most commonly used definitions was relatively low – most likely due to a high false-positive rate associated with the Deauville score definition of an unfavorable iPET response of an uptake greater than liver SUV_{max}.

Our observations complement smaller studies comparing ΔSUV_{max} and Deauville score for evaluation of iPET scans after two cycles of R-CHOP or similar regimens. In one study, iPET was prognostic only when the scans were evaluated by the ΔSUV_{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 ΔSUV_{max} approach, but an effect of iPET on overall survival was only seen 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 ΔSUV_{max} cut-off 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 cut-off for the Deauville score, that is, DS >2 defined as an uptake greater than that of the mediastinum. Given its definition, the false-positive rate observed with DS >3 even increases with DS >2 (Supplemental Figure 6) so that we do not recommend its use in the identification of R-CHOP non-responders. Two smaller studies suggest cut-offs above liver activity to be more appropriate in segregating the DLBCL population after two treatment cycles than DS >3, but neither of the two cut-offs 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 ΔSUV_{max} approach to have more potential cut-offs to choose from (*22,23*).

Although the cut-off-based Kaplan-Meier curves and associated hazard ratios between unfavorable and favorable patients indicate good segregation, the area under the ROC curve is relatively poor for both Deauville score and ΔSUV_{max} approach. In our opinion, this is a negligible concern as the aim is not to achieve high global discrimination performance across all possible cut-offs as assessed by the area under the ROC curve. 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 two methods' cut-offs. Another limitation of our comparison of ΔSUV_{max} approach and Deauville score is the modification of the ΔSUV_{max} approach in patients with an iPET lacking unphysiological ^{18}F-FDG uptake. Although it does not affect concordance, hazard ratios, sensitivity, specificity, or predictive values, it may have an impact on correlation coefficient, ROC curve, and area under the ROC curve. 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 re-classified to the favorable prognosis group due to lack of unphysiological ^{18}F-FDG uptake according to visual criteria. Overall, this re-classification occurred in 29 patients with 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 cut-off-based analyses as the question of a therapy switch requires a binary yes or no decision. And, although numbers are small, the outcome of re-classified patients appeared to be similar to that from patients with an SUV_{max} reduction >66%. Thus, the modification of the ΔSUV_{max} method introduced in the PETAL trial may also be of value in future investigations.

Despite the good local discrimination performance of the ΔSUV_{max} approach, much is still unknown about its properties. Although several authors (*3*,*8*–*10*) confirmed the 66% SUV_{max} reduction cut-off originally proposed by Lin et al. (*7*), available data on inter-rater reliability and reproducibility of the ΔSUV_{max} approach is scarce. In our investigation, the Deauville score assessment is a pre-dominantly centralized post-hoc analysis and, consequently, different nuclear medicine specialists were involved in Deauville score and ΔSUV_{max} assessments, respectively. Given the higher number of ΔSUV_{max} than Deauville score raters, this gives room for possibly increased inter-rater variation with the ΔSUV_{max} approach. Itti et al., however, found the 66% SUV_{max} reduction to be associated with a higher inter-rater reproducibility than DS >3; overall they rated inter-observer agreement as "almost perfect" with the ΔSUV_{max} approach but only "substantial" when the Deauville score was applied (*12*). In the PETAL trial, 10% of all iPET scans ΔSUV_{max} results were re-evaluated by nuclear medicine physician from other trial sites, where the

concordance between the first and second reader was 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 ΔSUV_{max} definition of an unfavorable iPET response of a relative SUV_{max} reduction ≤66% 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 of an uptake above that of the liver (DS >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 ΔSUV_{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 in combination with other patient, tumor, or treatment characteristics requires further study.

ACKNOWLEDGEMENTS

We thank the patients and investigators for their participation, and Deutsche Krebshilfe (nos. 107592 and 110515) for financial support.

DISCLOSURE

All authors declare they are not aware of any conflicts of interest that could have influenced the results of this investigation in either way.

KEY POINTS

Question: Should the Deauville score (DS >3) or the ΔSUV_{max} approach ($\Delta SUV_{max} \leq 66\%$) be the preferred method to evaluate early response to standard R-CHOP therapy in diffuse large B-cell lymphoma?

Pertinent Findings: In a post-hoc analysis of the PETAL trial, the ΔSUV_{max} approach presented with higher discrimination performance than the Deauville score. This was especially true for local discrimination measures associated with the two methods' most commonly used cut-offs – due to an increased false-positive rate of the Deauville score.

Implications for Patient Care: To prevent diffuse large B-cell lymphoma patients with favorable prognosis from harm resulting from unjustified iPET-based treatment intensification, the ΔSUV_{max} cut-off ($\Delta SUV_{max} \leq 66\%$) 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*. 2008;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 ΔSUVmax. *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, Walewski 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;51:1857–62.
21. Fan Y, Zhang Y, Yang Z, et al. Evaluating early interim fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography with the $SUV_{max\text{-}liver}$-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 SUVmax (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.

FIGURES & TABLES

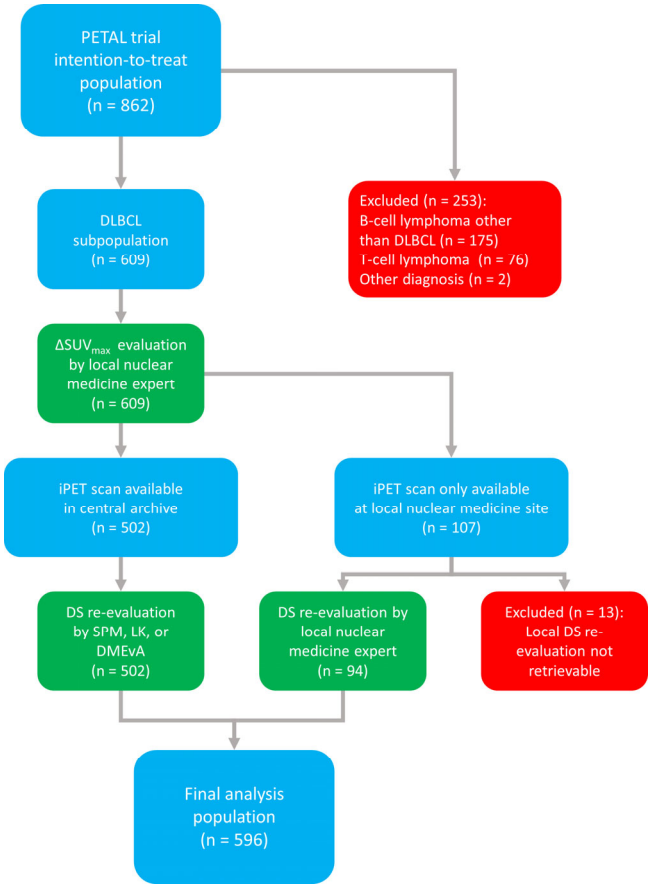


Figure 1 Flow of patients in terms of interim positron emission tomography assessments and resulting subpopulations. DLBCL: diffuse large B-cell lymphoma; DS: Deauville score; ΔSUV_{max}: deltaSUVmax; iPET: interim positron emission tomography.

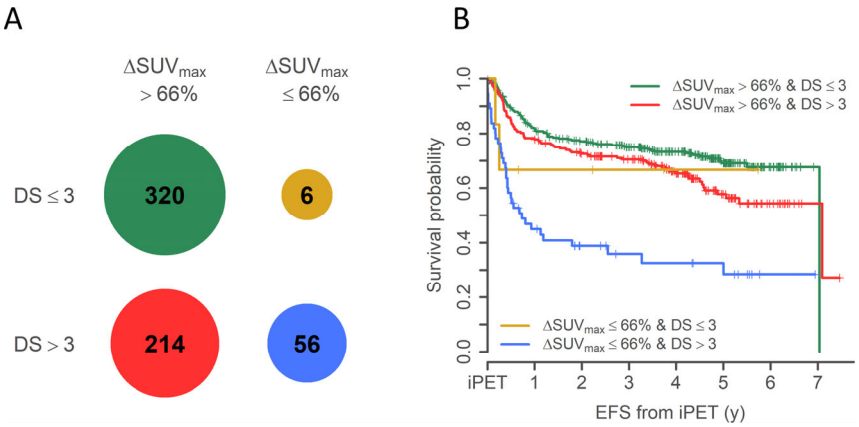


Figure 2 Concordance between ΔSUV_{max} and Deauville score cut-off (Panel A). Kaplan-Meier event-free survival curves by concordance category (Panel B). DS: Deauville score; ΔSUV_{max}: deltaSUVmax; EFS: event-free survival; iPET: interim positron emission tomography.

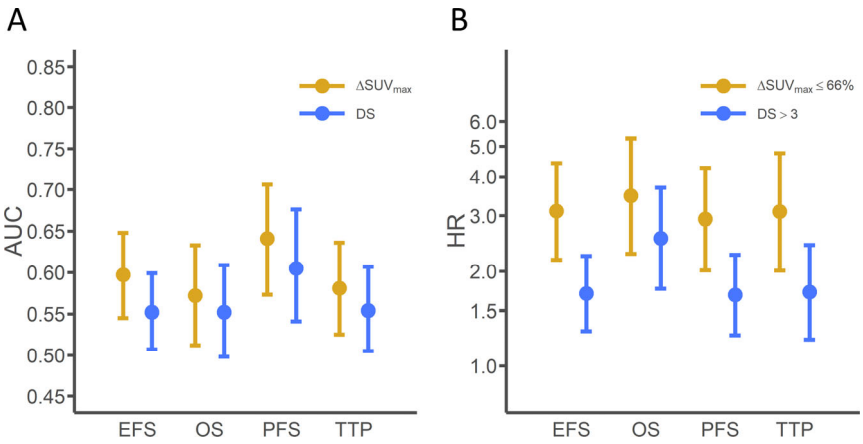


Figure 3 Area under the receiver operating characteristic curve (Panel A) and Cox regression hazard ratio (Panel B; logarithmic scale) with 95% confidence interval by method and time-to-event endpoint. AUC: area under the receiver operating characteristic curve; DS: Deauville score; ΔSUV_{max}: deltaSUVmax; EFS: event-free survival; HR: hazard ratio; OS: overall survival; PFS: progression-free survival; TTP: time to progression.

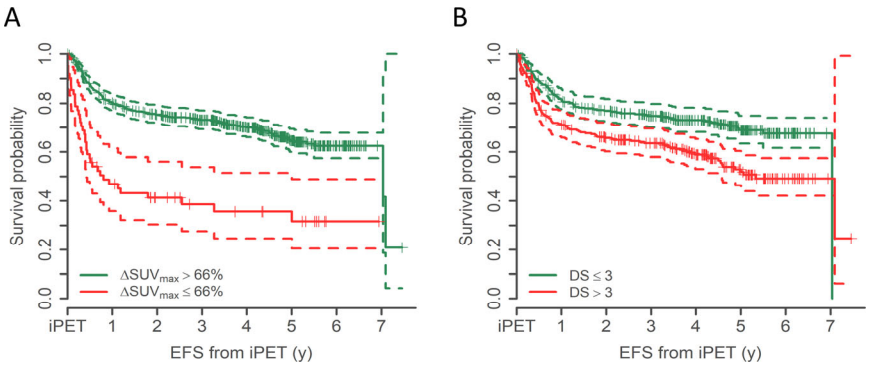


Figure 4 Kaplan-Meier event-free survival curves by iPET response according to ΔSUV_{max} (Panel A) and Deauville score (Panel B). DS: Deauville score; ΔSUV_{max}: deltaSUVmax; EFS: event-free survival; iPET: interim positron emission tomography.

Table 1 Characteristics for all evaluable patients by subpopulations defined by the two methods' cut-offs. Data on the International Prognostic Index was not available for one patient with ΔSUV_{max} >66% and DS >3.

Characteristics	All evaluable	ΔSUV_{max} >66%	ΔSUV_{max} ≤66%	DS ≤3	DS >3
No. of patients	596	534	62	326	270
Age, median (IQR) [years]	62 (51–70)	62 (51–70)	62 (50–69)	62 (52–70)	61 (50–70)
Age >60 years	308 (51.8%)	274 (51.4%)	34 (54.8%)	174 (53.5%)	134 (49.6%)
Male sex	331 (55.5%)	294 (55.1%)	37 (59.7%)	185 (56.7%)	146 (54.1%)
ECOG performance status ≥2	59 (9.9%)	47 (8.8%)	12 (19.4%)	23 (7.1%)	36 (13.3%)
Ann Arbor stage III or IV	349 (58.7%)	304 (57.0%)	45 (72.6%)	173 (53.2%)	176 (65.2%)
Extranodal sites >1	192 (32.3%)	166 (31.1%)	26 (41.9%)	94 (28.9%)	98 (36.3%)
Lactate dehydrogenase >ULN	329 (55.3%)	289 (54.2%)	40 (64.5%)	151 (46.5%)	178 (65.9%)
International Prognostic Index					
Low risk	219 (36.8%)	205 (38.5%)	14 (22.6%)	140 (43.1%)	79 (29.3%)
Low-intermediate risk	155 (26.1%)	137 (25.7%)	18 (29.0%)	83 (25.5%)	72 (26.7%)
High-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%)

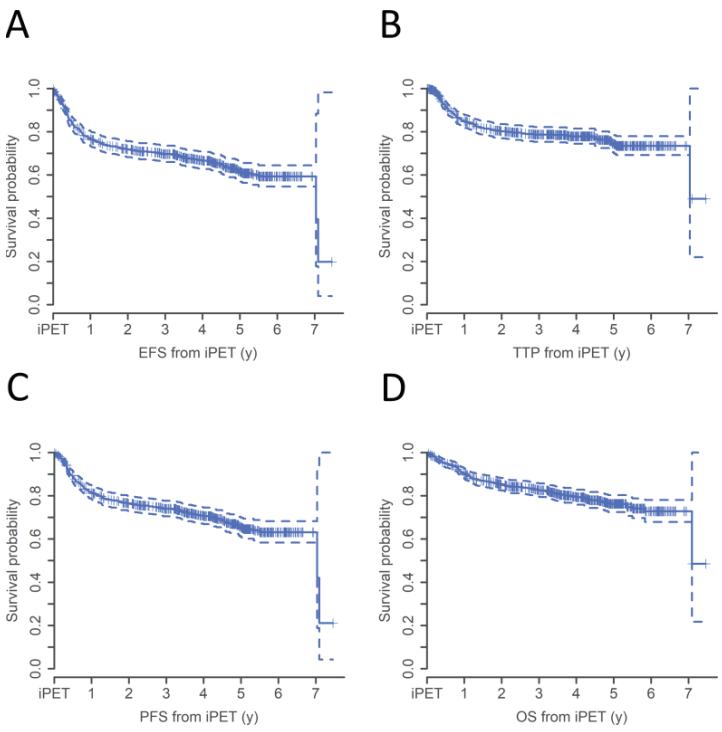
DS: Deauville score; ΔSUV_{max}: deltaSUVmax; ECOG: Eastern Cooperative Oncology Group; IQR: interquartile range; ULN: upper limit normal.

Interim PET evaluation in diffuse large B-cell lymphoma employing published recommendations:

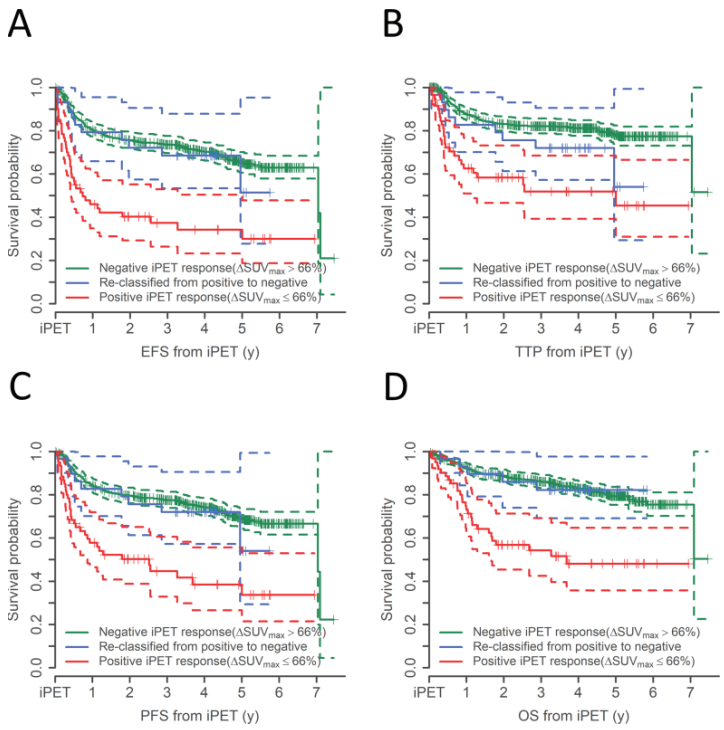
Comparison of the Deauville 5-point scale and the ΔSUV_{max} method

Jan Rekowski, Andreas Hüttmann, Christine Schmitz, Stefan Müller, Lars Kurch, Jörg Kotzerke, Christiane Franzius, Matthias Weckesser, Frank M. Bengel, Martin Freesmeyer, Andreas Hertel, Thomas Krohn, Jens Holzinger, Ingo Brink, Uwe Haberkorn, Fonyuy Nyuyki, Daniëlle M. E. van Assema, Lilli Geworski, Dirk Hasenclever, Karl-Heinz Jöckel, Ulrich Dührsen

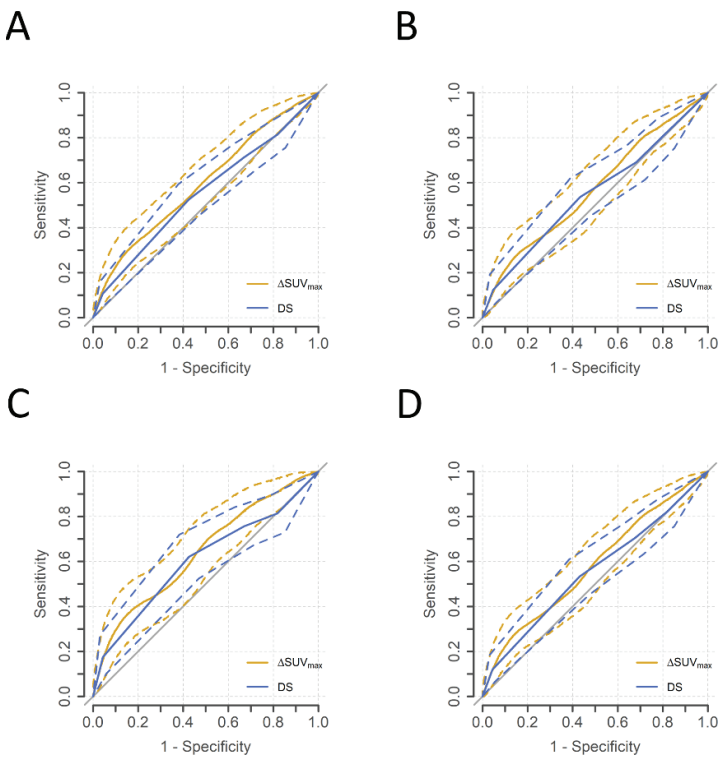
Supplemental Data



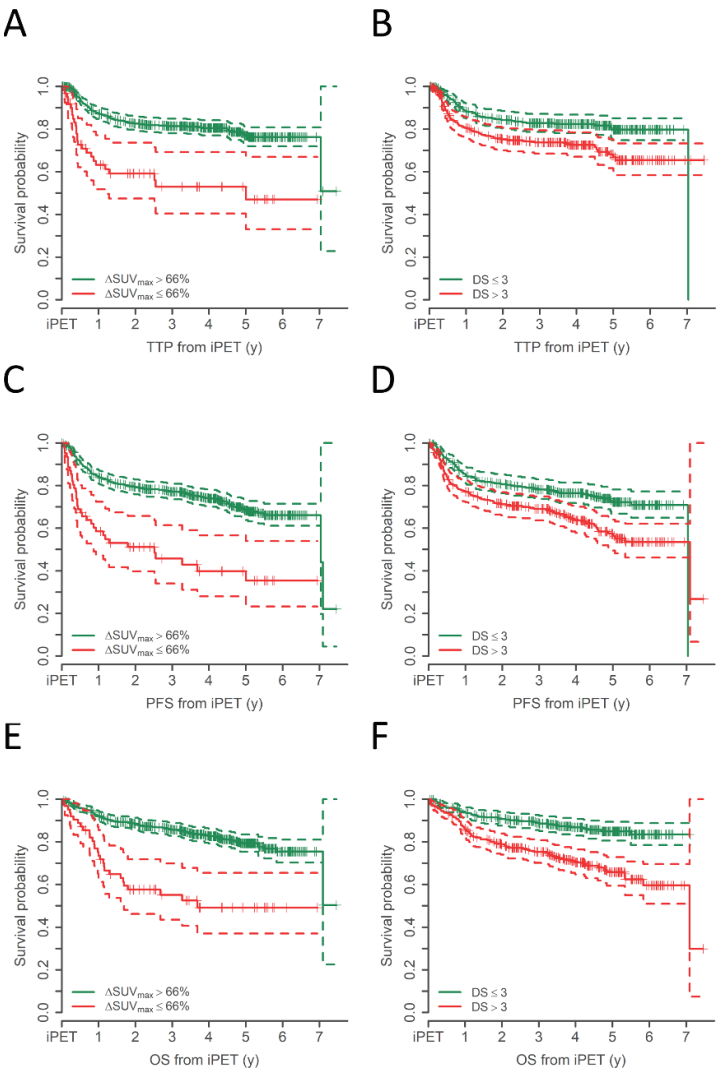
Supplemental Figure 1 Kaplan-Meier curves for the final analysis population (n = 596) for event-free survival (Panel A), time to progression (Panel B), progression-free survival (Panel C), and overall survival (Panel D). Dashed lines indicate 95% confidence intervals. EFS: event-free survival; iPET: interim positron emission tomography; OS: overall survival; PFS: progression-free survival; TTP: time to progression.



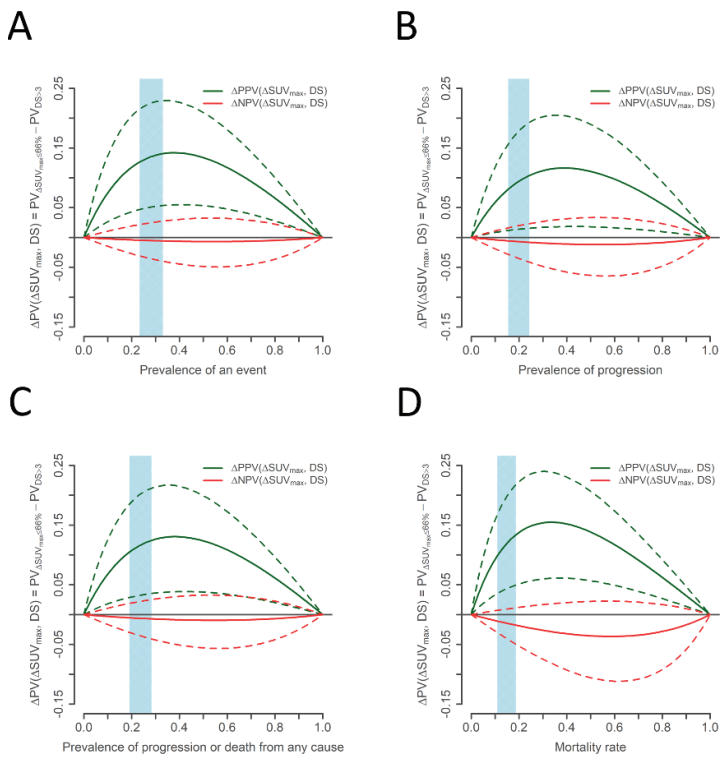
Supplemental Figure 2 Kaplan-Meier curves by ΔSUV_{max} interim positron emission tomography classification for event-free survival (Panel A), time to progression (Panel B), progression-free survival (Panel C), and overall survival (Panel D). The blue curves represent the group of patients lacking unphysiological ^{18}F-FDG uptake in interim positron emission tomography (n = 29) and being re-classified from positive response to negative response according to the modification of the ΔSUV_{max} method as described in the manuscript. Dashed lines indicate 95% confidence intervals. DS: Deauville score; ΔSUV_{max}: deltaSUVmax; EFS: event-free survival; iPET: interim positron emission tomography; OS: overall survival; PFS: progression-free survival; TTP: time to progression.



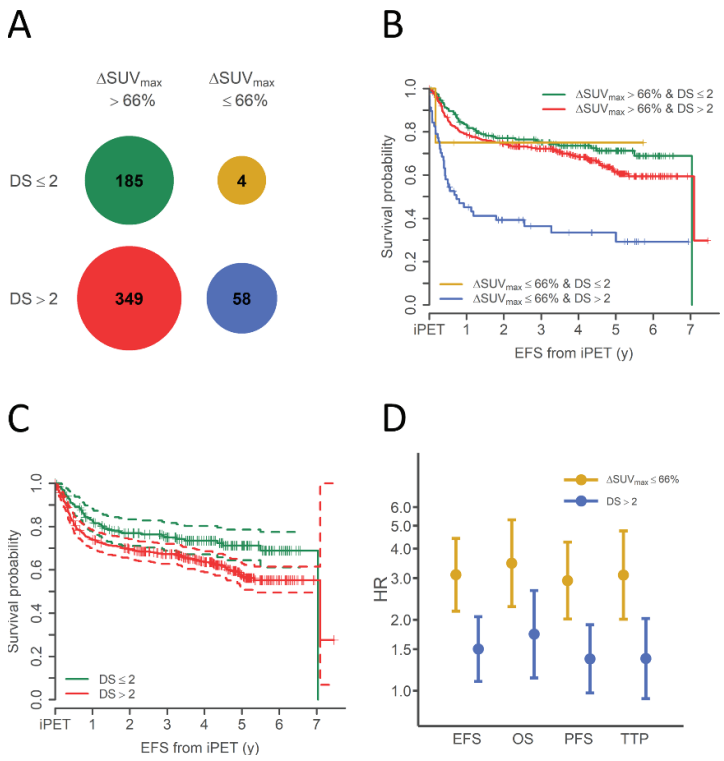
Supplemental Figure 3 Receiver operating characteristic curves for the ΔSUV_{max} approach and the Deauville score for event-free survival (Panel A), time to progression (Panel B), progression-free survival (Panel C), and overall survival (Panel D). Dashed lines indicate bootstrapped 95% confidence intervals. DS: Deauville score; ΔSUV_{max}: deltaSUVmax.



Supplemental Figure 4 Kaplan-Meier curves by early metabolic tumor response as assessed by the ΔSUV_{max} cut-off and the Deauville score cut-off for time to progression (Panels A & B), progression-free survival (Panels C & D), and overall survival (Panels E & F). Dashed lines indicate 95% confidence intervals. DS: Deauville score; ΔSUV_{max}: deltaSUVmax; EFS: event-free survival; iPET: interim positron emission tomography; OS: overall survival; PFS: progression-free survival; TTP: time to progression.



Supplemental Figure 5 Plot of predictive value differences for event-free survival (Panel A), time to progression (Panel B), progression-free survival (Panel C), and overall survival (Panel D). The x-axis spans the entire range of possible values for the unknown event prevalence of the respective endpoint (e.g., for the probability of an event in terms of overall survival, that is, mortality). The shaded blue rectangle represents a data-driven best guess for the true prevalence. Its interval on the x-axis refers to one minus the 99% confidence interval of the final analysis population's Kaplan-Meier estimator $\widehat{KM}$ at the time point of interest two years after interim positron emission tomography, e.g., for event-free survival it corresponds to the 99% confidence interval of $1-\widehat{KM}_{EFS}(t=2)$, where the left boundary x_1 of the blue rectangle refers to the lower confidence limit and its right boundary x_2 to the respective upper confidence limit. The y-axis indicates the difference between either the positive or negative predictive value of the ΔSUV_{max} ≤66% cut-off and the corresponding predictive value (positive or negative) of the Deauville score greater than three cut-off; in general notation for any of the two predictive values $\Delta PV(\Delta SUV_{max}, DS) = PV_{\Delta SUV_{max}\leq 66\%} - PV_{DS>3}$ as in the y-axis label. Plotted as a function of the unknown prevalence, the difference between the positive predictive value of the ΔSUV_{max} cut-off and the positive predictive value of the Deauville score cut-off ($\Delta PPV(\Delta SUV_{max}, DS)$ is shown in green. The difference between the negative predictive value of the ΔSUV_{max} cut-off and the negative predictive value of the Deauville score cut-off ($\Delta NPV(\Delta SUV_{max}, DS)$ is shown in red. Dashed lines indicate empirical 95% confidence intervals for the respective differences obtained by the bootstrap. DS: Deauville score; ΔSUV_{max}: deltaSUVmax; NPV: negative predictive value; PPV: positive predictive value; PV: predictive value.



Supplemental Figure 6 Results for the Deauville score cut-off defining an unfavorable early metabolic tumor response as an uptake above that of the mediastinum: Concordance between the ΔSUV_{max} and the Deauville score cut-off (Panel A). Kaplan-Meier event-free survival curves by concordance category (Panel B). Kaplan-Meier event-free survival curves by early metabolic tumor response (Panel C). Cox regression model hazard ratio with 95% confidence interval by method and time-to-event endpoint (Panel D). DS: Deauville score; ΔSUV_{max}: deltaSUVmax; EFS: event-free survival; HR: hazard ratio; iPET: interim positron emission tomography; OS: overall survival; PFS: progression-free survival; TTP: time to progression.

Supplemental Table 1 Numerical results of the time-dependent receiver operating characteristic curve analyses and the Cox regression model for event-free, progression-free, and overall survival as well as for time to progression. With regard to the area under the receiver operating characteristic curve, sensitivity, and specificity, confidence intervals relate to a simple bootstrap with 10,000 repetitions.

		EFS	TTP	PFS	OS
AUC (95% CI)	ΔSUV_{max}	0.597 (0.545–0.648)	0.581 (0.525–0.636)	0.641 (0.573–0.706)	0.572 (0.512–0.633)
	DS	0.552 (0.507–0.599)	0.554 (0.505–0.607)	0.605 (0.541–0.676)	0.552 (0.498–0.609)
HR (95% CI)	ΔSUV_{max}	3.10 (2.18–4.42)	3.09 (2.01–4.76)	2.93 (2.01–4.26)	3.47 (2.27–5.30)
	DS	1.70 (1.29–2.24)	1.72 (1.21–2.43)	1.68 (1.25–2.25)	2.54 (1.76–3.68)
Sensitivity (95% CI)	ΔSUV_{max}	0.246 (0.186–0.312)	0.247 (0.173–0.328)	0.249 (0.182–0.321)	0.325 (0.232–0.426)
	DS	0.525 (0.455–0.593)	0.537 (0.451–0.619)	0.533 (0.454–0.610)	0.621 (0.524–0.719)
Specificity (95% CI)	ΔSUV_{max}	0.888 (0.859–0.917)	0.874 (0.845–0.902)	0.881 (0.851–0.909)	0.880 (0.852–0.906)
	DS	0.575 (0.528–0.622)	0.567 (0.522–0.612)	0.570 (0.525–0.616)	0.574 (0.530–0.617)

AUC: area under the receiver operating characteristic curve; CI: confidence interval; DS: Deauville score; ΔSUV_{max}: deltaSUVmax; EFS: event-free survival; HR: hazard ratio; OS: overall survival; PFS: progression-free survival; TTP: time to progression.