haematologica
Journal of the Ferrata Storti Foundation

Improving the predictive value of end-of-treatment PET/CT in diffuse large B-cell lymphoma

by Anne L. Bes, Gerben J.C. Zwezerijnen, Martijn W. Heymans, Ulrich Dührsen, Jakoba J. Eertink, Sanne E. Wiegers, Pieternella J. Lugtenburg, Andreas Hüttmann, Lars Kurch, Christine Hanoun, N. George Mikhaeel, Luca Ceriani, Emanuele Zucca, Sándor Czibor, Tamás Györke, Martine E.D. Chamuleau, Stefano Fanti, Sze Ting Lee, Otto S. Hoekstra, Josée M. Zijlstra, Sally F. Barrington and Ronald Boellaard

Received: July 31, 2025.
Accepted: December 24, 2025.

Citation: Anne L. Bes, Gerben J.C. Zwezerijnen, Martijn W. Heymans, Ulrich Dührsen, Jakoba J. Eertink, Sanne E. Wiegers, Pieternella J. Lugtenburg, Andreas Hüttmann, Lars Kurch, Christine Hanoun, N. George Mikhaeel, Luca Ceriani, Emanuele Zucca, Sándor Czibor, Tamás Györke, Martine E.D. Chamuleau, Stefano Fanti, Sze Ting Lee, Otto S. Hoekstra, Josée M. Zijlstra, Sally F. Barrington and Ronald Boellaard. Improving the predictive value of end-of-treatment PET/CT in diffuse large B-cell lymphoma. Haematologica. 2026 Jan 8. doi: 10.3324/haematol.2025.288821 [Epub ahead of print]

Improving the predictive value of end-of-treatment PET/CT in diffuse large B-cell lymphoma

*Anne L. Bes,[1,2] *Gerben J.C. Zwezerijnen,[2,3] Martijn W. Heymans,[4] Ulrich Dührsen,[5] Jakoba J. Eertink,[1,2] Sanne E. Wiegers,[2,3] Pieternella J. Lugtenburg,[6] Andreas Hüttmann,[5] Lars Kurch,[7] Christine Hanoun,[5] N. George Mikhaeel,[8] Luca Ceriani,[9,10] Emanuele Zucca,[10-12] Sándor Czibor,[13] Tamás Györke,[13] Martine E.D. Chamuleau,[1,2] Stefano Fanti,[14] Sze Ting Lee,[15] Otto S. Hoekstra,[2,3] Josée M. Zijlstra,[1,2] Sally F. Barrington,[16] and Ronald Boellaard,[2,3] on behalf of the PETRA consortium

[1]Department of Hematology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[2]Imaging and Biomarkers, Cancer Center Amsterdam, Amsterdam, The Netherlands
[3]Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[4]Department of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[5]Department of Hematology, West German Cancer Center, University Hospital Essen, University of Duisburg– Essen, Essen, Germany
[6]Department of Hematology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands
[7]Clinic and Polyclinic for Nuclear Medicine, Department of Nuclear Medicine, University of Leipzig, Leipzig, Germany
[8]Department of Clinical Oncology, Guy's Cancer Centre and School of Cancer and Pharmaceutical Sciences, King's College London, London, United Kingdom
[9]Department of Nuclear Medicine and PET/CT Centre, Imaging Institute of Southern Switzerland–EOC, Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland
[10]SAKK Swiss Group for Clinical Cancer Research, Bern, Switzerland
[11]Department of Oncology, Oncology Institute of Southern Switzerland–EOC, Faculty of Biomedical Sciences, Università della Svizzera Italiana, Bellinzona, Switzerland
[12]Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland
[13]Department of Nuclear Medicine, Medical Imaging Centre, Semmelweis University, Budapest, Hungary
[14]Nuclear Medicine department, Sant'Orsola-Malpighi Hospital, Bologna, Italy
[15]Department of Molecular Imaging and Therapy, Austin Health, Melbourne, Australia
[16]School of Biomedical Engineering and Imaging Sciences, King's College London and Guy's and St Thomas' PET Centre, King's Health Partners, King's College London, London, United Kingdom

*****ALB and GJCZ contributed equally as co-first authors.

Running heads: EOT-PET enhances outcome prediction in DLBCL.

Corresponding author: Anne L. Bes; a.l.bes@amsterdamumc.nl

Data-sharing statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements: The authors gratefully thank all the patients who participated in the trials and the collaborating investigators who kindly supplied their data. A complete list of the members of the PETRA consortium can be found on: https://petralymphoma.org

Funding: This work was financially supported by the Dr. Werner Jackstädt Stiftung. They had no role in gathering, analyzing, or interpreting the data.

Disclosures: The authors declare no relevant conflicts of interest.

Author Contributions: A.L.B., G.J.C.Z., M.W.H., U.D., J.J.E., N.G.M., O.S.H., J.M.Z., S.F.B., and R.B contributed to the concept and design of this study; U.D., P.J.L., A.H., N.G.M., L.C., E.Z., T.G., S.C., M.E.D.C., N.G.M., S.F., S.L., J.M.Z., and S.F.B. were responsible for data acquisition; A.L.B., G.J.C.Z., J.J.E., S.E.W., L.K., C.H., S.C., and S.F.B. performed PET/CT analyses; A.L.B., and M.W.H. performed statistical analyses; and all authors contributed to the interpretation of the data. All authors critically reviewed and approved the manuscript.

Abstract

The 5-point Deauville score (DS) assesses end-of-treatment (EOT) response on PET/CT in diffuse large B-cell lymphoma patients, categorizing scans as ‘positive’ or ‘negative’ for complete metabolic response. However, the positive predictive value (PPV) is suboptimal at 60%. We evaluated whether quantitative PET parameters combined with clinical data could improve prediction of treatment failure in EOT PET-positive patients. Baseline and EOT PET/CT scans of 138 DS4–5 patients were analyzed. Lesions were segmented using a semi-automated adaptive method (SUV4.0 or MV3). PET parameters, including total metabolic tumor volume (TMTV), number of lesions (NOL), tumorSUV/liverSUV-ratio (TLR), the maximum distance between the largest and any other lesion (DmaxBulk), and changes over time, were obtained. Two Cox regression models predicted 2-year progression-free survival. Clinical data were combined with EOT PET in model 1, and baseline, EOT and delta values in model 2. After internal bootstrapping, models were evaluated for classification using different risk-of-progression cutoffs. Sensitivity, specificity, PPV and negative predictive values (NPV) were determined. Using forward selection, model 1 comprised two variables: the NOL and the tumorSUVpeak/liverSUVmean (TLRpeakmean) at EOT (AIC=690.072, c-index=0.747). Model 2 incorporated NOL, TLRpeakmean (EOT) and baseline SUVmean (AIC=687.064, c-index=0.762). The PPV improved to over 85% without compromising the NPV. False positives dropped from 54 (39%, by DS) to 9 (7%) and 6 (4%) for models 1 and 2, respectively. Adding baseline features did not notably impact the models’ performance. Our models could support more accurate response-adapted treatment decisions, reducing unnecessary subsequent false positive-directed treatments to just 7%.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most prevalent aggressive non-Hodgkin lymphoma.[1] First-line immunochemotherapy has a curative efficacy of 60-70%, but one-third of patients experience refractory disease or relapse.[2] Fluorine^{18}-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography-computed tomography (PET/CT) is recommended for initial staging and end-of-treatment (EOT) response assessment.[3, 4]

Currently, the post-therapy response is assessed by the 5-point visual Deauville score (DS),[5] which classifies metabolic outcome as complete (DS1-3) or incomplete (DS4-5). The simplicity of the DS, which uses the ratio between the FDG uptake in the hottest residual lymphoma lesion and liver, is desirable for interpretation but may also limit its predictive power.[5] While the negative predictive value (NPV) stands at 85%, the positive predictive value (PPV) remains suboptimal at 60% due to a high number of false positives, suggesting that nearly half of the patients with a DS4-5 are cured despite their positive final scan.[5, 6] An incorrect prognosis can be impactful as patients may be selected for subsequent therapies, such as consolidative radiotherapy. Patients may unnecessarily be subjected to potential risks and anxiety that come with receiving further treatment, biopsies or serial imaging.[7-11] Better criteria at EOT are thus essential to improve patient selection for further treatment.

Several research groups have proposed more precise response criteria at EOT by defining quantitative cutoff values based on changes in the maximum standardized uptake value (ΔSUVmax) or tumor-to-liver ratios higher than one.[12-14] At staging, there is increasing evidence supporting the prognostic potential of other quantitative parameters such as the total metabolic tumor volume (TMTV) and total lesion

glycolysis (TLG).[15-17] Recently, factors that reflect the dissemination of disease, such as the maximum distance between lesions, have also been reported as strong prognosticators,[17] which combined with TMTV can identify high-risk groups before treatment.

The PETRA consortium previously demonstrated that a combination of baseline tumor (TMTV, SUVpeak and DmaxBulk) and clinical (performance status and age) predictors can greatly enhance the PPV and accurately stratify high-risk patients at baseline.[18] However, few studies have focused on utilizing these quantitative features at EOT to predict the risk of relapse and need for second-line treatment.

Our aim was to improve the prediction of 2-year progression-free survival (2-yr-PFS) compared to the DS by focusing on increasing the PPV without compromising the NPV by (1) identifying quantitative EOT PET parameters that predict PFS, (2) developing a model combining EOT PET and clinical parameters and (3) exploring whether integrating baseline PET quantitative features could improve prediction.

Methods

Study population

Patients with DLBCL from 5 prospective studies (HOVON-84,[19] HOVON-130,[20] SAKK,[21] PETAL,[22] IAEA[23]), 2 retrospective studies (BOLOGNA,[24] GSTT15[25]) and real-world data (Austin Health, Melbourne) in the PETRA database[26], that had a baseline and positive EOT scan (DS4-5), were included. PMBL patients were excluded upfront. Patients with a complete metabolic response (CMR; DS1-3) were included in the sensitivity analysis. All trials had institutional review board approval.

Quantitative and clinical features

Quality control and lesion delineation methods are detailed in *Methods S1.* PET features were extracted from total metabolic tumor volume (TMTV) segmentations including TMTV (mL), SUVpeak, SUVmean and SUVmax. Lesional SUV was compared with liver uptake in a 3cm diameter sphere, including tumorSUVpeak/liverSUVmean ratio (TLRpeakmean), tumorSUVpeak/liverSUVpeak ratio (TLRpeakpeak) and tumorSUVmax/liverSUVmax ratio (TLRmaxmax). TLG was calculated as TMTV*SUVmean. Dissemination was assessed using the number of lesions (NOL) and DmaxBulk, defined as the distance between the largest lesion and the most distant lesion. The NOL could include multiple lesions in the same nodal area. Absolute (Δ) and percent (Δ%) changes from baseline to EOT were calculated, with Δ% defined as 100*((Baseline-EOT)/Baseline). Clinical feature collection is described in *Methods S2.*

Model development

Two models were developed: an EOT model (model 1), using clinical data and quantitative features extracted from EOT PET scans, and an EOT + baseline model (model 2) that additionally incorporated baseline (BL) and change in PET values. The primary outcome was 2-yr-PFS, defined as the time from the baseline PET scan to progression, relapse or death.

Univariate Cox regression explored the association between variables and PFS. Cubic spline transformations were applied for non-linearity if necessary. Highly correlating factors (Spearman r>0.9) were removed to avoid multicollinearity. Final models were constructed using forward selection to evaluate independent prognostic predictors.

Model fitting was assessed using the Akaike information criterion (AIC) and c-index. Subsequently, models were internally validated through bootstrapping with 500 generated datasets, adjusting regression coefficients for optimism by multiplying them by a shrinkage factor.[27]

Patient classification

The risk-of-progression at 24 months was estimated for each patient using both models. Risk-of-event cutoffs from 20 to 90% were explored to classify patients into low and high-risk groups. For each cutoff, predicted classifications were compared against observed outcomes to determine sensitivity, specificity, PPV, NPV and the number of false positives and false negatives. To estimate NPV for the entire population we included 2-yr-PFS retrospective outcome data from DS1-3 EOT patients. Since tumor segmentations were not available for DS1-3 patients, we assumed these patients would be classified as low-risk by the models. The cutoff yielding the best balance between sensitivity and specificity was explored further using Kaplan-Meier survival analysis.

The added relevance of our EOT model was assessed by comparing it to a published baseline clinical PET model,[18] which used MTV, SUVpeak, DmaxBulk, age and ECOG to identify high-risk patients at baseline.

Assessment of confounding therapy

Information about second-line therapy was available for 122 patients, though the indications were mostly undocumented. The impact of radiotherapy on model performance was evaluated in this subset.

Statistical analyses were performed using R (version 4.2.3), with a p-value <0.05 considered statistically significant.

Results

Study population

Within the PETRA database, 847 DLBCL patients were identified who underwent an EOT scan with an assigned DS, of whom 225 were classified as 'PET-positive' with an incomplete response (DS4-5), and 622 as having CMR (DS1-3, *Figure 1*).

The predictive models were built solely on PET-positive patients (n=225). Patients were deemed non-eligible due to missing or unusable scans (n=54), invalid clinical data (n=14), non-compliance with quality control standards (n=10) or having tumor uptake ≤DS3 at revision (n=9, *Figure 1*). A total of 138 EOT PET-positive patients were included, with 62 patients classified as DS4 and 76 patients as DS5. Patient characteristics are summarized in *Table 1,* and described in greater detail for each study in *Table S1.*

The median age was 61 years, ranging from 18 to 88 years. Most patients (n=135, 97.8%) received R-CHOP or a combination thereof, while only 3 patients received R-CEOP treatment. The median follow-up time was 53 months. Eighty-four patients (60.9%) had an event, of which 80 either progressed or relapsed, and 4 died. Among the DS4 patients, 63% remained event-free at 2 years, compared to 20% of patients with DS5 (*Figure 2*).

The majority of patients with CMR were scored as DS1 (n=303, 48.7%), then DS2 (n=167, 26.9%) and DS3 (n=152, 24.4%, *Table S2*). Notably, 13.5% of these patients developed progression.

Model development

The descriptive statistics for quantitative PET variables in PET-positive patients are listed in *Table S3*. All variables showed a reduction from baseline to EOT, with the largest changes in TMTV and TLG.

Highly correlated variables were eliminated, leaving TMTV-BL (at baseline), TMTV-EOT (at end-of-treatment), Δ%TMTV (relative difference), SUVmean-BL, TLRpeakmean-BL, TLRpeakmean-EOT, Δ%TLRpeakmean, NOL-BL, NOL-EOT, Δ%NOL, and DmaxBulk-BL, Δ%DmaxBulk and ΔDmaxBulk (absolute difference), and the clinical features age, sex, stage, IPI-score, IPI-stage, IPI-EN, IPI-ECOG and IPI-LDH, which were taken forward into multivariate models. Details on the univariate analysis and spline transformations are given in *Results S1*. TMTV was selected over TLG due to its wide usage and ease of interpretation. TLRpeakmean was favored over SUVpeak because of its independence from the administered activity and patient body weight. It is also less sensitive to noise and different PET systems when compared to TLRmaxmax.

Finally, after applying forward selection, model 1 comprised two (splined) variables expressing the tumor-to-liver ratio and the number of lesions: TLRpeakmean-EOT and NOL-EOT (AIC=690.072, c-index=0.753, R^2=0.436; after bootstrapping: c-index=0.747, R^2=0.410). Model 2 incorporated the same two features with the addition of the mean SUV at baseline (SUVmean-BL; AIC=687.064, c-index=0.771, R^2=0.456; after bootstrapping: c-index=0.762, R^2=0.452). No clinical features were retained by the models. Shrinkage factors of 0.935 (model 1) and 0.922 (model 2) were obtained after internal bootstrapping validation to adjust the regression coefficients (*Table 2*).

Patient classification

A risk-to-progression estimate was calculated for every patient. The individual risk estimates were fairly similar for the two models (r=0.97, p=7.61e-87, *Figure S2*). Examples of patients with different risk predictions are shown in *Figure 3.*

The performance of models 1 and 2 are summarized in *Tables 3* and *4,* using various risk-to-progression cutoff values. Increasing the risk threshold led to higher specificity but lowered sensitivity. No patient had a risk score <10%. A 50% threshold for model 1 and 60% threshold for model 2 classified the highest number of patients correctly and achieved a PPV above 85%. The corresponding NPVs were 67% and 68%, respectively. However, after including the PFS from 622 patients with CMR (DS1-3), comprising 538 true negatives and 84 false negatives, the NPV increased to 85% (*last column of Tables 3 and 4*). Model 2 fit the data better than the simpler model 1 (χ^2(df=1) = 5.009, p=0.025), which resulted in 2 more correctly classified patients.

The survival curves (*Figure 4*) showed a clear separation in 2-yr-PFS between DS1-3 and DS4-5 groups (81.4% versus 37.0%). Upon applying model 1 with a 50% risk threshold, the DS4-5 group further separated into two distinct subgroups: a low-risk group (<50%) with a 2-yr-PFS of 64.2%, and a high-risk group (>50%) with a PFS of 11.2%. Stratification using model 1 also separated the curves better compared to DS4 and DS5 separately (*Figure S3*). The low-risk (<50%) group had a 2-yr-PFS of 64.2%, compared to 58.1% in the DS4 group. Similarly, the high-risk (>50%) group had a 2-yr-PFS of 11.2%, whereas the DS5 group was 19.7%.

Furthermore, model 1 (EOT) has improved prognostic power when compared to the previously published clinical PET model.[18] Following the same methodology as for

model 1, a 20% risk-of-progression cutoff was chosen for the clinical PET model. The sensitivity, specificity, PPV and NPV of the clinical PET model all decreased in contrast to model 1 (*Table S5*). Notably, the number of false negatives was substantially lower for model 1 (22 versus 53). The baseline model identified 43 patients at high risk of having an event within 2 years, compared to 74 patients using model 1. The two models overlapped in selecting 27 patients.

Lastly, an overview of patients receiving second-line therapy can be viewed in *Table S6-8*. The inclusion of radiotherapy status did not significantly enhance the performance of either model, suggesting limited additional predictive value (*Results S2*).

Discussion

The reliability of the DS at the end of first-line treatment for DLBCL has been questioned due to its low PPV caused by a high number of false positives. Our study aimed to address this by identifying quantitative PET features that could improve the PPV for 2-yr-PFS. Two relatively simple prediction models were developed: an EOT model (1) including the NOL and TLRpeakmean at EOT, and an EOT + baseline model (2) that incorporated SUVmean at baseline as an additional feature. Both models outperformed DS for correctly classifying the risk-of-progression and enhanced the PPV to over 85% without compromising the NPV.

The importance of TLRpeakmean in our model is not surprising, as it represents the tumor-to-liver ratio similar to the DS, but in a semi-quantitative manner to minimize the risk of visual misinterpretation. Although TLRmaxmax is more commonly used to quantify the DS, we favored TLRpeakmean because it is more robust to noise and

image reconstruction differences. This definition of TLR was also recommended in the recent EANM guidelines on FDG oncology imaging.[28] SUVpeak is less sensitive to noise and differences in image resolution between PET systems, making it more generalizable across scanner generations, especially with the increasing use of high-resolution scanners,[29-31] whereas the SUVmean reflects the most reproducible measure for uniform liver uptake.[32]

Others have also demonstrated that EOT TLR cutoffs can identify patients with inferior PFS and OS.[12, 14, 33] Nevertheless, the reported cutoff values varied widely according to the patient cohort. In contrast, the TLR in our model is expressed as a continuous variable, meaning it can be applied to different populations. When combined with a simple measure, the number of lesions at EOT, which may partially serve as a surrogate for disease dissemination, the improvement in prognostic value is substantial. The number of lesions at EOT has not commonly been used in predictive models for DLBCL, but the number of extra nodal lesions in DLBCL and nodal lesions in follicular lymphoma at baseline are well-known risk parameters.[34, 35]

Despite the significant association of TMTV at baseline and EOT to PFS (univariate), and its recognition as an important prognostic factor,[25, 36-38] it was not selected in our final models. In our cohort of only PET-positive patients, the selected variables thus appeared to be stronger prognosticators. Furthermore, none of the clinical features contributed to the predictive power. This aligns with a recent study where radiomics-only models outperformed those integrating clinical parameters.[39]

The use of baseline quantitative PET features in predictive models has been reported for estimating individual patient outcomes,[18, 40] but models that incorporate EOT data are scarce. Cui et al[41] developed a model combining clinical, baseline,

EOT and delta PET features, that outperformed three models: an IPI-model, a clinical features only model (with and without DS), and a PET radiomics model (BL, EOT and delta) for time to progression. Their best model (c-index of 0.853) surpassed our model performances (c-index=0.747 and 0.762 respectively). However, this comparison is challenging due to differences in feature selection. Additionally, their dataset included both PET-negative and -positive patients with relatively few cases of progression.

We examined the potential of our models to correctly classify PET-positive patients using a series of risk-of-progression cutoffs. A clear trade-off was seen between sensitivity and specificity. Cutoff risks of 50% and 60% for model 1 and 2, respectively, yielded objectively the best balance between specificity with sensitivity. An important advantage of our models is their ability to provide a continuous probability, rather than a dichotomous assessment such as the DS. A more suitable threshold can thus be chosen depending on the clinical context. For example, instead of the proposed 50% threshold for model 1, a higher threshold could be chosen for a less sensitive but highly specific patient selection.

In total, 107 (77.5%) patients were classified correctly by model 1 and 109 (79.0%) by model 2 compared to 84 (61%) using the DS. The NPV first appeared low (~67%) but this is to be expected in a PET-positive only group. We tried to simulate the NPV for the entire DLBCL patient population by incorporating PFS data from 622 patients with CMR to our model outcome. The NPV then reached an expected value of around 85%.[5, 6] This approach may be optimistic, as it does not account for SUVmean at baseline, the number of lesions or TLRpeakmean for assessing risk in these patients.

Model 2 performed slightly better than model 1, resulting in the correct classification of two additional patients. Given the limited performance improvement and the practical advantage of relying on quantitative data from a single timepoint, model 1 has our preference. We also demonstrated that model 1 (EOT) outperformed our previously published clinical PET model, in terms of PPV and NPV,[18] which shows that a baseline features-only model does not select the same patients and is less useful in this context.

Model 1 separated DS4-5 patients into two distinct risk groups, a low-risk group (<50% risk, 2-yrs-PFS of 64.2%) and a high-risk group (>50% risk, 2-yr-PFS of 11.2%). Compared to the classification based solely on DS, which resulted in 54 false positive patients (39%), model 1 significantly reduced the number of false positives to just 9 patients (7%). When the low-risk group, as defined by the model, was compared directly to DS4, the survival increased from 58.1% to 64.2%. Even though this is a significant improvement, 24 patients (35.8%) were still not classified correctly and would be 'undertreated' if only high-risk patients would receive secondary treatment. Future research targeting the DS4 subgroup is warranted, although this may be challenging due to the need for a large dataset and integration of advanced radiomics with clinical or diagnostic features.

This preliminary analysis suggests that the quantitative model improves risk discrimination and may better inform post-first-line treatment decisions. A two-step risk assessment approach might be worth further exploration: patients would first undergo visual assessment to distinguish CMR from PET-positive cases, after which DS4-5 patients could be further stratified using the EOT model to identify those at higher risk. Such an approach might help refine selection for secondary treatment, such as radiotherapy, while allowing lower-risk patients to be monitored

conservatively. However, the latter approach remains hypothetical as it was solely based on internal data. We used bootstrapping for internal validation to mitigate overfitting and assess coefficient shrinkage, but external validation remains necessary to ensure prognostic reliability of the model.

Our study was limited by the sample size, which may have affected the robustness of the model development. However, the number of enrolled PET-positive EOT patients was substantial, considering over 70% of patients respond to first-line treatment. The PETRA dataset combines several studies, which gives the advantage of a larger dataset, but may contain larger variations in terms of therapy choices and genetic variability. For example, the HO130 study enrolled patients with MYC oncogene rearrangements who are known to have a poor prognosis.[42] Although the protocol strongly recommended to confirm each relapse by biopsy, it cannot be ensured this was the case for all events. Despite this limitation, the overall results are expected to remain largely unaffected as false positives in non-biopsy proven cases would also have influenced the classification based on visual assessment using the DS. Nevertheless, the model demonstrates improved patient classification compared to the DS. Furthermore, among a subset of 122 patients within the PETRA cohort, 27 patients (22%) received radiotherapy after first-line treatment, although the criteria for their selection were not always specified. PET-positive DLBCL patients who have received radiotherapy at EOT usually have a favorable outcome.[43] However, radiotherapy had no impact on our model performance. The decision for radiotherapy thus might have been made by the treating physician instead of the EOT PET result, or the sample size may have been too small to detect an effect on the model. Finally, patients with DS5 were slightly overrepresented in our dataset, which could affect the generalizability of our results.

In conclusion, we developed a quantitative PET model comprising tumorSUVpeak/TumorSUVmean and number of lesions at EOT, with the optional inclusion of SUVmean at baseline that improves the PPV for 2-year progression-free survival to over 85%, while maintaining a strong NPV. Our model could help guide response-adapted therapy after initial treatment, reducing the number of patients who might receive unnecessary secondary treatment to 7%.

References

1. Al-Hamadani M, Habermann TM, Cerhan JR, Macon WR, Maurer MJ, Go RS. Non-Hodgkin lymphoma subtype distribution, geodemographic patterns, and survival in the US: A longitudinal analysis of the National Cancer Data Base from 1998 to 2011. Am J Hematol. 2015;90(9):790-795.
2. Raut LS, Chakrabarti PP. Management of relapsed-refractory diffuse large B cell lymphoma. South Asian J Cancer. 2014;3(1):66-70.
3. Coughlan M, Elstrom R. The use of FDG-PET in diffuse large B cell lymphoma (DLBCL): predicting outcome following first line therapy. Cancer Imaging. 2014;14(1):34.
4. 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):3059-3068.
5. 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):3048-3058.
6. Juweid ME, Mueller M, Alhouri A, MZ AR, Mottaghy FM. Positron emission tomography/computed tomography in the management of Hodgkin and B-cell non-Hodgkin lymphoma: An update. Cancer. 2021;127(20):3727-3741.
7. Pirani M, Marcheselli R, Marcheselli L, Bari A, Federico M, Sacchi S. Risk for second malignancies in non-Hodgkin's lymphoma survivors: a meta-analysis. Ann Oncol. 2011;22(8):1845-1858.
8. Smith-Bindman R, Lipson J, Marcus R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med. 2009;169(22):2078-2086.
9. Rock CB, Chipman JJ, Parsons MW, et al. Second Primary Malignancies in Diffuse Large B-cell Lymphoma Survivors with 40 Years of Follow Up: Influence of Chemotherapy and Radiation Therapy. Adv Radiat Oncol. 2022;7(6):101035.
10. Derry-Vick HM, Heathcote LC, Glesby N, et al. Scanxiety among Adults with Cancer: A Scoping Review to Guide Research and Interventions. Cancers (Basel). 2023;15(5):1381.
11. Thompson CA, Charlson ME, Schenkein E, et al. Surveillance CT scans are a source of anxiety and fear of recurrence in long-term lymphoma survivors. Ann Oncol. 2010;21(11):2262-2266.
12. Ferrari C, Pisani AR, Masi T, et al. Lesion-to-Liver SUVmax Ratio to Improve the Prognostic Value of the End of Treatment PET/CT in Diffuse Large B-Cell Lymphoma. J Clin Med. 2022;11(19):5541.
13. Allioux F, Gandhi D, Vilque JP, et al. End-of-treatment (18)F-FDG PET/CT in diffuse large B cell lymphoma patients: DeltaSUV outperforms Deauville score. Leuk Lymphoma. 2021;62(12):2890-2898.
14. Li YH, Zhao YM, Jiang YL, et al. The prognostic value of end-of-treatment FDG-PET/CT in diffuse large B cell lymphoma: comparison of visual Deauville criteria and a lesion-to-liver SUV(max) ratio-based evaluation system. Eur J Nucl Med Mol Imaging. 2022;49(4):1311-1321.
15. Shagera QA, Cheon GJ, Koh Y, et al. Prognostic value of metabolic tumour volume on baseline (18)F-FDG PET/CT in addition to NCCN-IPI in patients with diffuse large B-cell lymphoma: further stratification of the group with a high-risk NCCN-IPI. Eur J Nucl Med Mol Imaging. 2019;46(7):1417-1427.
16. Zhang YY, Song L, Zhao MX, Hu K. A better prediction of progression-free survival in diffuse large B-cell lymphoma by a prognostic model consisting of baseline TLG and %DeltaSUV(max). Cancer Med. 2019;8(11):5137-5147.
17. Cottereau AS, Meignan M, Nioche C, et al. Risk stratification in diffuse large B-cell lymphoma using lesion dissemination and metabolic tumor burden calculated from baseline PET/CT(dagger). Ann Oncol. 2021;32(3):404-411.

18. Eertink JJ, van de Brug T, Wiegers SE, et al. (18)F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging. 2022;49(3):932-942.
19. Lugtenburg PJ, de Nully Brown P, van der Holt B, et al. Rituximab-CHOP With Early Rituximab Intensification for Diffuse Large B-Cell Lymphoma: A Randomized Phase III Trial of the HOVON and the Nordic Lymphoma Group (HOVON-84). J Clin Oncol. 2020;38(29):3377-3387.
20. Chamuleau MED, Burggraaff CN, Nijland M, et al. Treatment of patients with MYC rearrangement positive large B-cell lymphoma with R-CHOP plus lenalidomide: results of a multicenter HOVON phase II trial. Haematologica. 2020;105(12):2805-2812.
21. 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(23):2523-2529.
22. Duhrsen U, Muller 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(20):2024-2034.
23. Carr R, Fanti S, Paez D, et al. Prospective international cohort study demonstrates inability of interim PET to predict treatment failure in diffuse large B-cell lymphoma. J Nucl Med. 2014;55(12):1936-1944.
24. Zinzani PL, Gandolfi L, Broccoli A, et al. Midtreatment 18F-fluorodeoxyglucose positron-emission tomography in aggressive non-Hodgkin lymphoma. Cancer. 2011;117(5):1010-1018.
25. Mikhaeel NG, Smith D, Dunn JT, et al. Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL. Eur J Nucl Med Mol Imaging. 2016;43(7):1209-1219.
26. Eertink JJ, Burggraaff CN, Heymans MW, et al. Optimal timing and criteria of interim PET in DLBCL: a comparative study of 1692 patients. Blood Adv. 2021;5(9):2375-2384.
27. Frank E. Harrell J. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. 2nd ed. Springer Cham; 2015.
28. Boellaard R, Herrmann K, Barrington SF, et al. [18F]FDG PET/CT: EANM procedure guidelines for tumour imaging: version 3.0. EANM J. 2025;1:100006.
29. Sher A, Lacoeuille F, Fosse P, et al. For avid glucose tumors, the SUV peak is the most reliable parameter for [(18)F]FDG-PET/CT quantification, regardless of acquisition time. EJNMMI Res. 2016;6(1):21.
30. Kaalep A, Sera T, Rijnsdorp S, et al. Feasibility of state of the art PET/CT systems performance harmonisation. Eur J Nucl Med Mol Imaging. 2018;45(8):1344-1361.
31. van Sluis J, Boellaard R, Dierckx R, Stormezand GN, Glaudemans A, Noordzij W. Image Quality and Activity Optimization in Oncologic (18)F-FDG PET Using the Digital Biograph Vision PET/CT System. J Nucl Med. 2020;61(5):764-771.
32. Zwezerijnen GJC, Eertink JJ, Ferrandez MC, et al. Reproducibility of [18F]FDG PET/CT liver SUV as reference or normalisation factor. Eur J Nucl Med Mol Imaging. 2023;50(2):486-493.
33. 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(2):e0211649.
34. Solal-Celigny P, Roy P, Colombat P, et al. Follicular lymphoma international prognostic index. Blood. 2004;104(5):1258-1265.
35. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. N Engl J Med. 1993;329(14):987-994.
36. Vercellino L, Cottereau AS, Casasnovas O, et al. High total metabolic tumor volume at baseline predicts survival independent of response to therapy. Blood. 2020;135(16):1396-1405.
37. Ceriani L, Gritti G, Cascione L, et al. SAKK38/07 study: integration of baseline metabolic heterogeneity and metabolic tumor volume in DLBCL prognostic model. Blood Adv. 2020;4(6):1082-1092.

38. Tout M, Casasnovas O, Meignan M, et al. Rituximab exposure is influenced by baseline metabolic tumor volume and predicts outcome of DLBCL patients: a Lymphoma Study Association report. Blood. 2017;129(19):2616-2623.
39. Ferrandez MC, Golla SSV, Eertink JJ, et al. Validation of an Artificial Intelligence-Based Prediction Model Using 5 External PET/CT Datasets of Diffuse Large B-Cell Lymphoma. J Nucl Med. 2024;65(11):1802-1807.
40. Mikhaeel NG, Heymans MW, Eertink JJ, et al. Proposed New Dynamic Prognostic Index for Diffuse Large B-Cell Lymphoma: International Metabolic Prognostic Index. J Clin Oncol. 2022;40(21):2352-2360.
41. Cui Y, Jiang Y, Deng X, et al. (18)F-FDG PET-Based Combined Baseline and End-Of-Treatment Radiomics Model Improves the Prognosis Prediction in Diffuse Large B Cell Lymphoma After First-Line Therapy. Acad Radiol. 2023;30(7):1408-1418.
42. Rosenwald A, Bens S, Advani R, et al. Prognostic Significance of MYC Rearrangement and Translocation Partner in Diffuse Large B-Cell Lymphoma: A Study by the Lunenburg Lymphoma Biomarker Consortium. J Clin Oncol. 2019;37(35):3359-3368.
43. Freeman CL, Savage KJ, Villa DR, et al. Long-term results of PET-guided radiation in patients with advanced-stage diffuse large B-cell lymphoma treated with R-CHOP. Blood. 2021;137(7):929-938.

Tables

Table 1. Patient characteristics of PET-positive patients (n=138)

Characteristic		Value
Sex	Male	80 (58.0%)
	Female	58 (42.0%)
Age at diagnosis (years)	Median (range)	61 (18-88)
Follow-up (months)	Median	53
DS	4	62 (44.9%)
	5	76 (55.1%)
PFS (years)	≥ 2	54 (39.1%)
	< 2	84 (60.9%)
R-CHOP (cycles)	5	2 (1.4%)
	6	69 (50%)
	7	3 (2.2%)
	8	61 (44.2%)
R-CEOP (cycles)	6	3 (2.2%)
Ann Arbor Stage	I	8 (5.8%)
	II	14 (10.1%)
	III	32 (23.2%)
	IV	84 (60.9%)
IPI	Low-risk (IPI 0-1)	22 (15.9%)
	Low-intermediate (IPI 2)	29 (21.0%)
	High-intermediate (IPI 3)	43 (31.2%)
	High-risk (IPI 4-5)	44 (31.9%)
IPI-Age (years)	≤ 60	66 (47.8%)
	> 60	72 (52.2%)
IPI-Stage	I/II	22 (15.9%)
	III/VI	116 (84.1%)
IPI-EN (sites involved)	0-1	67 (48.6%)
	> 1	71 (51.4%)
IPI-ECOG	< 2	112 (81.2%)
	≥ 2	26 (18.8%)
IPI-LDH	LDH ≤ ULN	32 (23.2%)
	LDH > ULN	106 (76.8%)

DS=Deauville Score; PFS=progression-free survival; IPI=International Prognostic Index; EN=involvement of extra-nodal sites; ECOG=performance status according to the Eastern Cooperative Oncology Group; LDH=lactase dehydrogenase; R-CHOP=standard immunochemotherapy based on rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone; R-CEOP=rituximab, cyclophosphamide, etoposide, vincristine, and prednisolone; ULN=upper limit of normal

Table 2. Hazards of final models

	Model 1			Model 2		
Variable	HR (95% CI)	Coefficient after bootstrap	p-value	HR (95% CI)	Coefficient after bootstrap	p-value
TLRpeakmean-EOT	2.0181 (1.491-2.732)	0.657	5.45e-06	1.942 (1.430-2.638)	0.612	2.17e-05*
TLRpeakmean-EOT'	0.2786 (0.135-0.577)	-1.195	0.00058	0.327 (0.154-0.695)	-1.030	0.004*
NOL-EOT	1.1763 (0.975-1.420)	0.152	0.09083	1.160 (0.962-1.399)	0.137	0.120
NOL-EOT'	0.7110 (0.340-1.488)	-0.320	0.36548	0.733 (0.351-1.533)	-0.286	0.409
SUVmean-BL	-	-	-	0.912 (0.839-0.991)	-0.085	0.030*

'=splined variable; HR=hazard ratio; TLRpeakmean=tumorSUVpeak/liverSUVmean ratio; -EOT=feature at end-of-treatment; NOL=number of lesions; -BL=feature at baseline

Table 3. Sensitivity scores using model 1 (end-of-treatment), varying the risk-to-progression cutoff value

Risk to progression (%)	Correctly classified (n)	False positives (n)	False negatives (n)	Sensitivity	Specificity	Accuracy	PPV	NPV	NPV on whole group*
20	85	52	1	0.988	0.037	0.616	0.615	0.667	0.864
30	102	28	8	0.905	0.482	0.739	0.731	0.765	0.860
40	107	14	17	0.798	0.741	0.775	0.827	0.702	0.851
50	**107**	**9**	**22**	**0.738**	**0.833**	**0.775**	**0.873**	**0.672**	**0.846**
60	105	7	26	0.691	0.870	0.761	0.892	0.644	0.842
70	102	5	31	0.631	0.907	0.739	0.914	0.613	0.836
80	94	3	41	0.512	0.944	0.681	0.935	0.554	0.825
90	81	0	57	0.321	1.000	0.587	1.000	0.487	0.808

*The last column was calculated using data from n=138 patients with Deauville score 4-5 and n=622 with Deauville score 1-3; PPV=positive predictive value; NPV=negative predictive value

Table 4. Sensitivity scores using model 2 (end-of-treatment + baseline), varying the risk to progression cutoff value

Risk to progression (%)	Correctly classified (n)	False positives (n)	False negatives (n)	Sensitivity	Specificity	Accuracy	PPV	NPV	NPV on whole group*
20	88	50	0	1.000	0.074	0.638	0.627	1.000	0.866
30	101	30	7	0.917	0.444	0.732	0.720	0.774	0.861
40	107	16	15	0.821	0.704	0.775	0.812	0.717	0.853
50	107	10	21	0.750	0.815	0.775	0.863	0.677	0.847
60	**109**	**6**	**23**	**0.726**	**0.889**	**0.790**	**0.910**	**0.676**	**0.846**
70	101	4	33	0.607	0.926	0.732	0.927	0.602	0.834
80	90	4	44	0.476	0.926	0.652	0.909	0.532	0.821
90	85	0	54	0.369	1.000	0.616	1.000	0.505	0.812

*The last column was calculated using data from n=138 patients with Deauville score 4-5 and n=622 with Deauville score 1-3

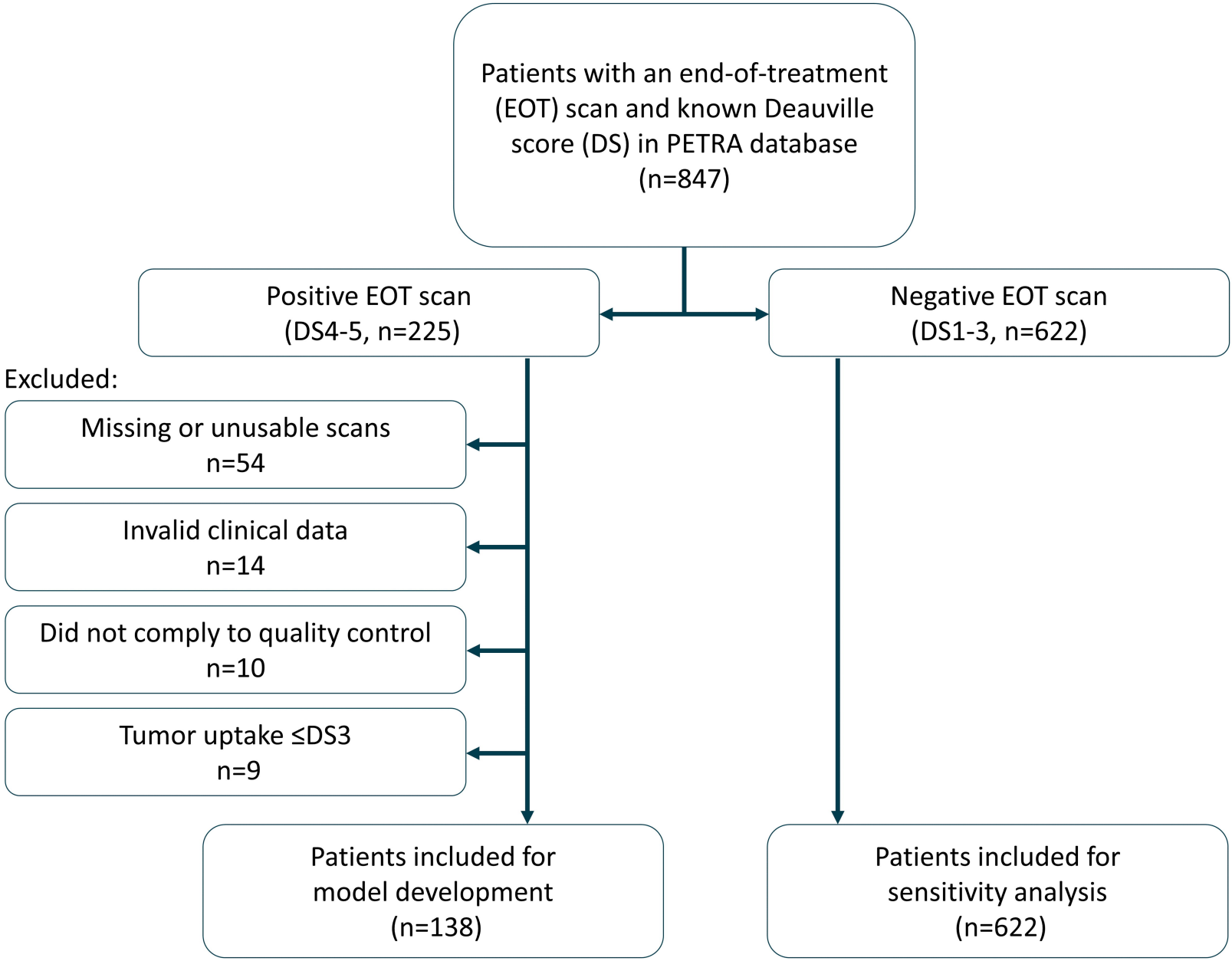
Figure legends

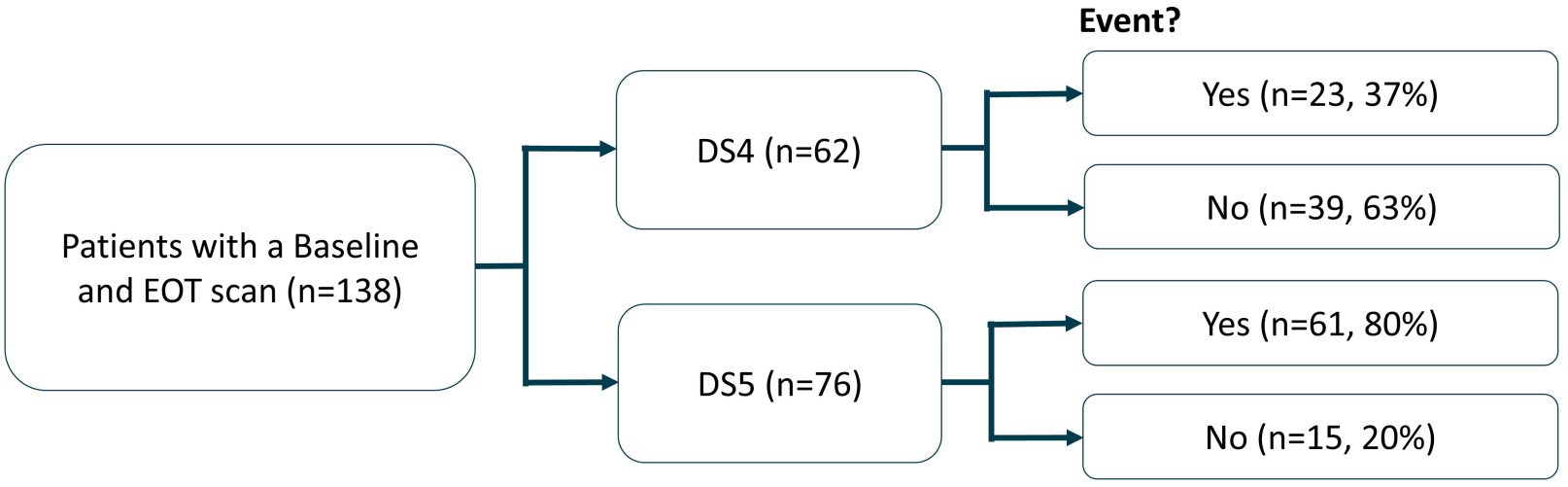
Figure 1. Consort diagram for study population

Figure 2. Overview of events in PET-positive patients

Figure 3. Examples of baseline and end-of-treatment scans with different risk predictions for models 1 and 2

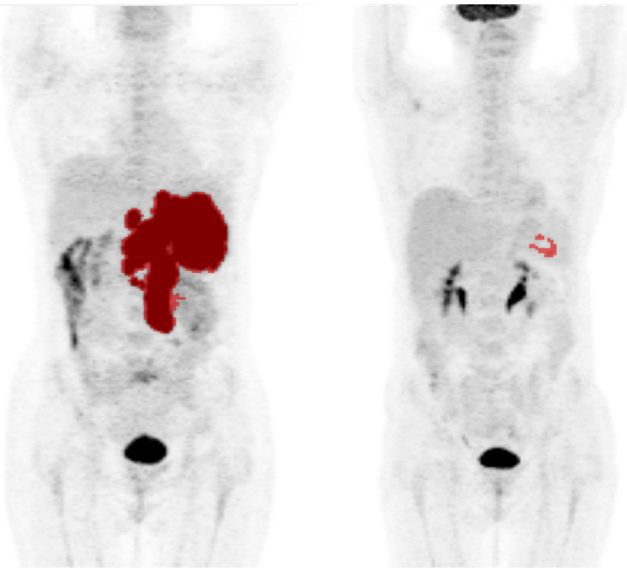
Figure 4. Kaplan-Meier survival curves comparing patients with Deauville score 1-3 and 4-5 to the model 1 <50% and >50% risk groups for 2-year progression-free survival. The survival curves show a clear separation in 2-yr-PFS between Deauville score (DS) 1-3 and 4-5 groups. After applying model 1 with a 50% risk threshold, the DS4-5 group further separated into two distinct subgroups: a low-risk group (<50%) and a high-risk group (>50%).


Patients with an end-of-treatment (EOT) scan and known Deauville score (DS) in PETRA database (n=847)
Positive EOT scan (DS4-5, n=225)
Negative EOT scan (DS1-3, n=622)
Excluded:
Missing or unusable scans n=54
Invalid clinical data n=14
Did not comply to quality control n=10
Tumor uptake ≤DS3 n=9
Patients included for model development (n=138)
Patients included for sensitivity analysis (n=622)


Event?
Patients with a Baseline and EOT scan (n=138)
DS4 (n=62)
DS5 (n=76)
Yes (n=23, 37%)
No (n=39, 63%)
Yes (n=61, 80%)
No (n=15, 20%)

Baseline
SUVmean: 12.23

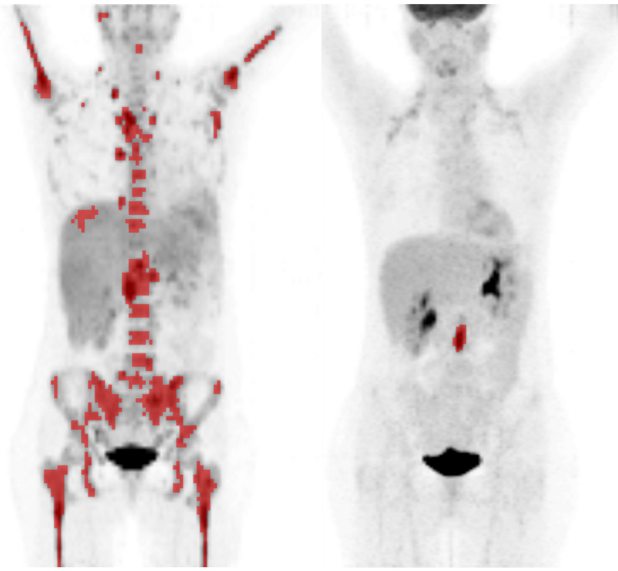
EOT
TLR: 1.63
NOL: 2



Risk model 1: 0.303
Risk model 2: 0.257

Baseline
SUVmean: 5.22

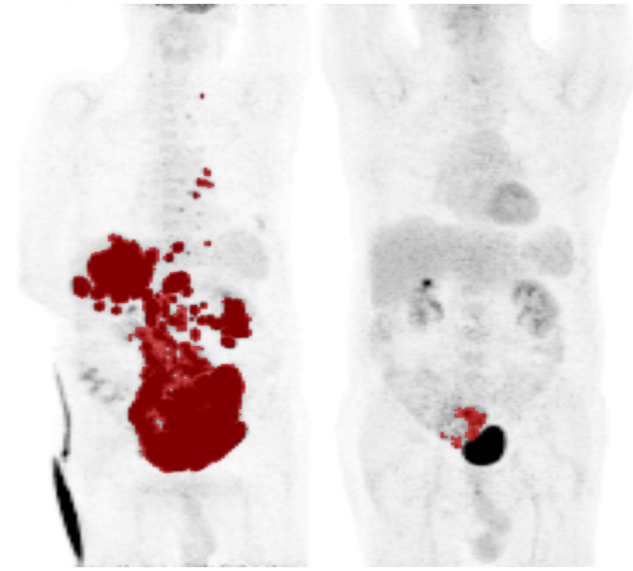
EOT
TLR: 3.53
NOL: 1



Risk model 1: 0.584
Risk model 2: 0.715

Baseline
SUVmean 9.77

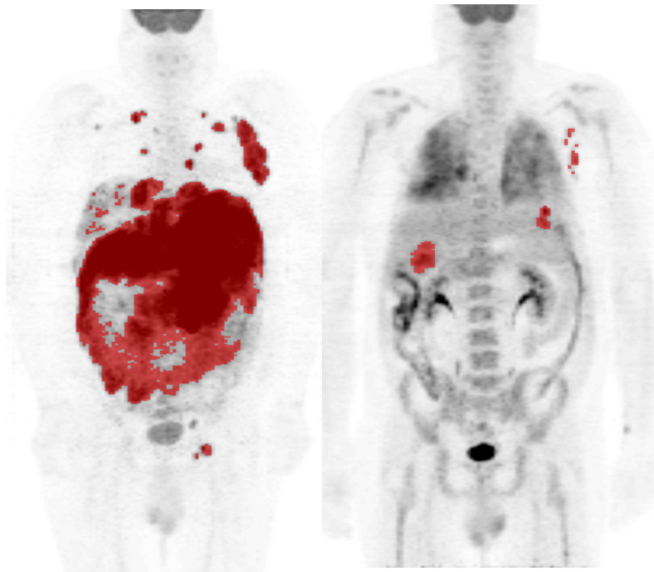
EOT
TLR: 4.05
NOL: 4



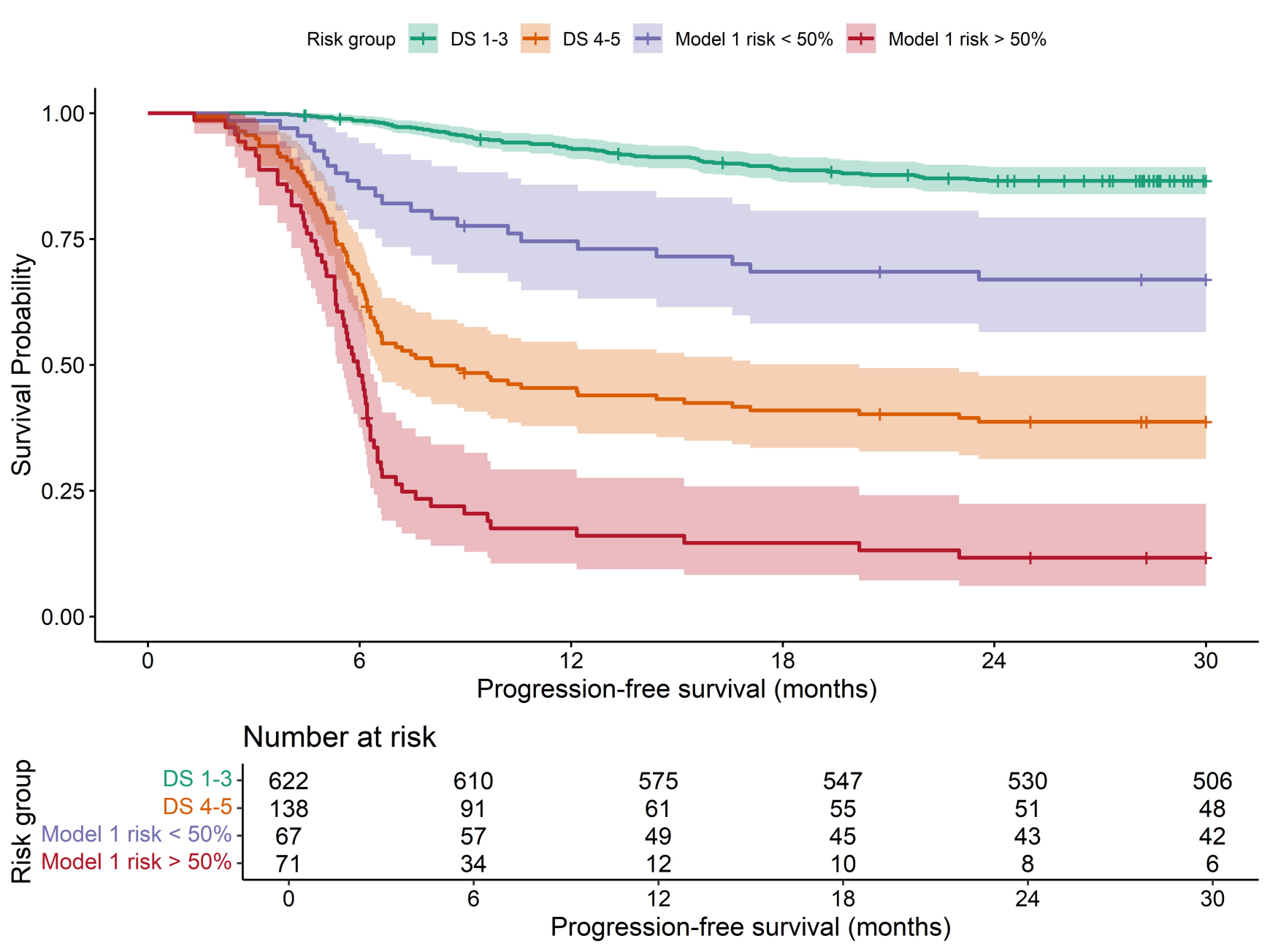
Risk model 1: 0.794
Risk model 2: 0.771

Baseline
SUVmean 7.98

EOT
TLR: 2.56
NOL: 8



Risk model 1: 0.689
Risk model 2: 0.715


Risk group
DS 1-3
DS 4-5
Model 1 risk < 50%
Model 1 risk > 50%
Survival Probability
1.00
0.75
0.50
0.25
0.00
0
6
12
18
24
30
Progression-free survival (months)
Number at risk
Risk group
DS 1-3 622 610 575 547 530 506
DS 4-5 138 91 61 55 51 48
Model 1 risk < 50% 67 57 49 45 43 42
Model 1 risk > 50% 71 34 12 10 8 6
0
6
12
18
24
30
Progression-free survival (months)

Methods S1: Quality control and image analysis

The quality control followed EANM guidelines,[1] requiring a liver SUVmean between 1.3 and 3.0 and a plasma glucose lower than 11 mmol/L. If the liver SUVmean was outside the acceptable range, but the total image activity was 50-80% of the total injected FDG activity (in MBq), scans were still included. Scans were required to be complete and contain all essential DICOM data. All scans were centrally reviewed using standard visual Deauville score criteria applied by each site.

Baseline scans were segmented semi-automatically using the ACCURATE tool[2] following the current benchmark method for DLBCL: a fixed SUV4.0 threshold[3] (https://petralymphoma.org/accurate-tool/). However, lesions at end-of-treatment (EOT) often have a low SUV uptake, which cannot be delineated using SUV4.0. Therefore, we used the lesional based L2A method, which was previously tested in interim-PET to successfully delineate low-uptake tumors.[4] This lesion-based approach segments each lesion depending on their uptake values, applying either the SUV4.0 (if SUVmax ≥ 10) or MV3 method (if SUVmax < 10). MV3 is a majority vote method that includes voxels detected by at least 3 of the following delineation thresholds: SUV4.0, SUV2.5, 41% of SUVmax or 50% of SUVpeak.[5] This way it resembles the baseline method for DS5 lesions and, at the same time, is more optimal for lesions with lower uptakes (DS4). A fixed threshold SUV2.5 was used in the few cases the L2A method failed. If necessary, non-lymphoma FDG uptake was manually removed. Segmentations were performed by a trained researcher (ALB) and reviewed by a nuclear medicine physician (GJCZ), including confirmation of lesion uptake classification (DS4-5).

Methods S2: Clinical features

Clinical predictors including age (continuous), sex, Ann Arbor stage (I-IV), IPI score (International Prognostic Index, 4 risk groups) and the five IPI components as binary variables: age (≤60 versus >60), stage (I-II versus III-IV), EN (involvement of extra-nodal sites, 0-1 versus >1), ECOG (performance status according to the Eastern Cooperative Oncology Group <2 versus ≥2) and LDH (lactase dehydrogenase ≤ upper limit of normal (ULN) versus LDH > ULN) were collected. Follow-up and outcome data were extracted from the clinical databases.

Table S1. Patients characteristics of PET-positive patient stratified by study (n=138)

		Australian	GSTT15	HOVON130	HOVON84	PETAL	IAEA	SAKK
Inclusion	n	15	11	21	43	27	3	18
Sex	Male	10 (66.7%)	5 (45.5%)	3 (14.3%)	17 (39.5%)	15 (55.6%)	0 (0.0%)	8 (44.4%)
	Female	5 (33.3%)	6 (54.5%)	18 (85.7%)	26 (60.5)	12 (44.4%)	3 (100.0%)	10 (55.6%)
Age at diagnosis (years)	Median (range)	70 (55-88)	53 (31-72)	58 (28-76)	64 (23-79)	53 (24-78)	21 (18-83)	57 (26-77)
Follow-up (months)	Median	53	75	32	58	43	-	65
DS	4	2 (13.3%)	8 (72.7%)	5 (23.8%)	26 (60.5%)	11 (40.7%)	0 (0.0%)	10 (55.6%)
	5	13 (86.7%)	3 (27.3%)	16 (76.2%)	17 (39.5%)	16 (59.3%)	3 (100.0%)	8 (44.4%)
PFS (years)	≥ 2	7 (46.7%)	4 (36.4%)	3 (14.3%)	21 (48.8%)	12 (44.4%)	0 (0.0%)	7 (38.9%)
	< 2	8 (53.3%)	7 (63.6%)	18 (85.7%)	22 (51.2%)	15 (55.6%)	3 (100.0%)	11 (61.1%)
Chemo therapy	R-CHOP	15	8	21	43	20	10	18
	5 cycles	0 (0.0%)	0 (0.0%)	1 (4.8)	1 (2.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	6 cycles	6 (40.0%)	8 (100.0%)	1 (4.8)	19 (44.2%)	17 (85.0%)	0 (0.0%)	18
	7 cycles	0 (0.0%)	0 (0.0%)	3 (14.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	8 cycles	9 (60.0%)	0 (0.0%)	16 (76.2%)	23 (53.5%)	3 (15%)	10 (100.0%)	0 (0.0%)
	R-CEOP *(6 cycles)*	0	3	0	0	0	0	0
Ann Arbor Stage	I	0 (0.0%)	1 (9.1%)	0 (0.0%)	0 (0.0%)	4 (14.8%)	0 (0.0%)	3 (16.7%)
	II	0 (0.0%)	0 (0.0%)	3 (14.3%)	2 (4.7%)	4 (14.8%)	1 (33.3%)	4 (22.2%)
	III	6 (40.0%)	1 (9.1%)	6 (28.6%)	8 (18.6%)	7 (25.9%)	0 (0.0%)	4 (22.2%)
	IV	9 (60.0%)	9 (81.8%)	12 (57.1%)	33 (76.7%)	12 (44.4%)	2 (66.7%)	7 (38.9%)
IPI	Low-risk (IPI 0-1)	0 (0.0%)	1 (9.1%)	3 (14.3%)	3 (7.0%)	8 (29.6%)	1 (33.3%)	6 (33.3%)
	Low-intermediate (IPI 2)	2 (13.3%)	2 (18.2%)	4 (19.0%)	9 (20.9%)	7 (25.9%)	0 (0.0%)	5 (27.8%)
	High-intermediate (IPI 3)	7 (46.7%)	4 (36.4%)	12 (57.1%)	12 (27.9%)	4 (14.8%)	0 (0.0%)	4 (22.2%)
	High-risk (IPI 4-5)	6 (40.0%)	4 (36.4%)	2 (9.5%)	19 (44.2%)	8 (29.6%)	2 (66.7%)	3 (16.7%)
IPI-Age (years)	≤ 60	1 (6.7%)	7 (63.6%)	13 (61.9%)	15 (34.9%)	17 (63.0%)	2 (66.7%)	11 (61.1%)
	> 60	14 (93.3%)	4 (36.4%)	8 (38.1%)	28 (65.1%)	10 (37.0%)	1 (33.3%)	7 (38.9%)
IPI-Stage	I/II	0 (0.0%)	1 (9.1%)	3 (14.3%)	2 (4.7%)	8 (29.6%)	1 (33.3%)	7 (38.9%)
	III/VI	15 (100.0%)	10 (91.0%)	18 (85.7%)	41 (95.3%)	19 (70.4%)	2 (66.7%)	11 (61.1%)
IPI-EN (sites involved)	0-1	8 (53.3%)	3 (27.3%)	11 (52.4%)	16 (37.2%)	16 (59.3%)	1 (33.3%)	12 (66.7%)
	> 1	7 (46.7%)	8 (72.7%)	10 (47.6%)	27 (62.8%)	11 (40.7%)	2 (66.7%)	6 (33.3%)
IPI-ECOG	< 2	12 (80.0%)	7 (63.6%)	19 (90.5%)	35 (81.4%)	21 (77.8%)	2 (66.7%)	16 (88.9%)
	≥ 2	3 (20.0%)	4 (36.4%)	2 (9.5%)	8 (18.6%)	6 (22.2%)	1 (33.3%)	2 (11.1%)
IPI-LDH	LDH ≤ ULN	5 (33.3%)	3 (27.3%)	3 (14.3%)	8 (18.6%)	7 (25.9%)	1 (33.3%)	5 (27.8%)
	LDH > ULN	10 (66.7%)	8 (72.7%)	18 (85.7%)	35 (81.4%)	20 (74.1%)	2 (66.7%)	13 (72.2%)

DS=Deauville Score; PFS=Progression free survival; IPI=International Prognostic Index; EN=involvement of extra-nodal sites; ECOG=performance status according to the Eastern Cooperative Oncology Group; LDH=lactase dehydrogenase

Table S2. Patient characteristics of patients with complete metabolic response

		n=622 (100%)
Deauville score	1	303 (48.7%)
	2	167 (26.9%)
	3	152 (24.4%)
Chemotherapy	R-CHOP	622 (100%)
PFS	≥ 2 years	538 (86.5%)
	< 2 years	84 (13.5%)

R-CHOP=standard immunochemotherapy based on rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone

Table S3. Distribution of quantitative PET descriptives

Variable	Median	Q1	Q3
TMTV-BL (ml)	635.42	293.90	1351.92
TMTV-EOT (ml)	11.73	3.61	43.16
ΔTMTV (ml)	559.18	246.87	1317.19
Δ%TMTV	97.58	90.93	99.30
TLG-BL	6120.68	2472.96	12033.09
TLG-EOT	59.36	13.80	378.83
ΔTLG	5214	1820	11521
Δ%TLG	93.32	98.44	99.67
SUVpeak-BL	18.13	12.99	24.91
SUVpeak-EOT	5.76	3.65	12.44
ΔSUVpeak	9.94	3.69	15.75
Δ%SUVpeak	62.85	27.47	79.36
SUVmean-BL	8.90	6.93	10.64
SUVmean-EOT	5.13	3.30	7.41
ΔSUVmean	3.17	0.83	5.71
Δ%SUVmean	39.84	12.03	61.52
SUVmax-BL	21.57	15.92	30.51
SUVmax-EOT	8.82	4.91	17.62
ΔSUVmax	10.71	2.71	18.54
Δ%SUVmax	55.11	13.65	74.20
TLRpeakpeak-BL	8.26	5.83	12.02
TLRpeakpeak-EOT	2.30	1.47	4.94
ΔTLRpeakpeak	5.11	2.15	8.09
Δ%TLRpeakpeak	64.44	38.93	82.65
TLRpeakmean-BL	9.70	6.72	13.87
TLRpeakmean-EOT	2.64	0.98	5.69
ΔTLRpeakmean	6.38	2.70	9.99
Δ%TLRpeakmean	67.20	37.69	82.49
TLRmaxmax-BL	8.26	5.84	11.46
TLRmaxmax-EOT	2.88	1.71	6.09
ΔTLRmaxmax	4.53	1.91	7.49
Δ%TLRmaxmax	58.71	24.29	77.66
NOL-BL (n)	9.50	3.00	19.00
NOL-EOT (n)	2.00	1.00	4.00
ΔNOL (n)	4.50	1.00	13.75
Δ%NOL	66.67	34.62	87.50
DmaxBulk-BL (mm)	308.90	188.10	432.60
DmaxBulk-EOT (mm)	19.60	0.00	156.00
ΔDmaxBulk (mm)	233.20	66.30	352.60
Δ%DmaxBulk	86.01	86.01	100.00

Δ%=percent reduction; -BL=feature at baseline; -EOT=feature at end-of-treatment; TMTV=total metabolic tumor volume; TLG=total lesion glycolysis; TLRpeakpeak=tumorSUVpeak/liverSUVpeak ratio TLRpeakmean=tumorSUVpeak/liverSUVmean ratio; TLRmaxmax=tumorSUVmax/liverSUVmax ratio, NOL=number of lesions; DmaxBulk= distance between the largest lesion and the most distant lesion

Results S1: Univariate Cox and spline transformations

Spline transformations were relevant for TMTV-EOT, Δ%TMTV, TLG-EOT, Δ%TLG, SUV-EOT, Δ%SUV, TLR-EOT, Δ%TLR, NOL-BL and NOL-EOT, ΔNOL and ΔDmaxBulk**.** The transformed variables were further used in the model, see *Table S4*.

The univariate analysis showed statistical significance for the following linear or transformed variables: TMTV-BL, TMTV-EOT and Δ%TMTV, TLG-EOT and Δ%TLG, TLR-EOT, SUV-EOT and Δ%SUV, TLR-EOT and Δ%TLR, NOL-BL, NOL-EOT and Δ%NOL, DmaxBulk-BL, DmaxBulk-EOT and Δ%DmaxBulk. None of the clinical features, except for the DS, were significant.

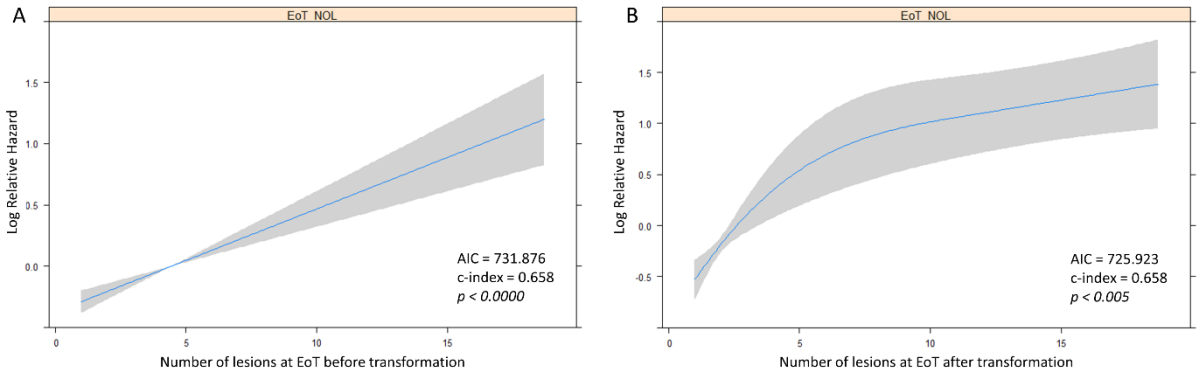


Figure S1. Relative log hazard for the number of lesions at end-of-treatment before (A) and after (B) cubic spline transformation.

Table S4. Univariate cox regression results before and after spline transformation in DS4 and 5 patients (n=138)

Variable	Before spline transformation				After spline transformation			
	HR (95% CI)	c-index	AIC	p-value	HR (95% CI)	c-index	AIC	p-value
DS-EOT	3.433 (2.113-5.576)	0.640	734.372	6.24e-07*	-	-	-	-
Age	0.994 (0.979-1.010)	0.514	762.061	0.448	-	-	-	-
Stage	1.249 (0.957-1.630)	0.552	759.691	0.102	-	-	-	-
Sex	0.707 (0.452-1.106)	0.531	760.259	0.129	-	-	-	-
IPI	1.146 (0.934-1.406)	0.543	760.895	0.193	-	-	-	-
IPI-Age	0.979 (0.638-1.503)	0.506	762.617	0.922	-	-	-	-
IPI-Stage	1.405 (0.745-2.650)	0.520	761.425	0.293	-	-	-	-
IPI-EN	1.160 (0.755-1.781)	0.524	762.166	0.498	-	-	-	-
IPI-ECOG	0.964 (0.551-1.685)	0.495	762.609	0.896	-	-	-	-
IPI-LDH	1.700 (0.971-2.975)	0.546	758.784	0.063	-	-	-	-
TMTV-BL (ml)	1.000 (1.000-1.000)	0.565	757.361	0.014*	1.000 (0.998-1.001)	0.565	759.040	0.574
TMTV-EOT (ml)	1.003 (1.002-1.004)	0.695	743.728	4.46e-08*	0.849 (0.796-0.906)	0.695	722.794	7.69e-07*
ΔTMTV (ml)	1.000 (1.000-1.000)	0.535	759.597	0.065	1.000 (1.000-1.000)	0.535	761.504	0.760
Δ%TMTV (%)	0.991 (0.984-0.997)	0.639	756.017	0.003*	0.956 (0.932-0.981)	0.638	746.840	0.001*
TLG-BL	1.000 (1.000-1.000)	0.542	760.171	0.103	1.000 (1.000-1.000)	0.542	762.037	0.716
TLG-EOT	1.000 (1.000-1.000)	0.705	748.546	2.01e-06*	0.966 (0.956-0.978)	0.705	719.025	3.72e-09*
ΔTLG	1.000 (1.000-1.000)	0.513	761.644	0.309	1.000 (1.000-1.000)	0.533	763.545	0.752
Δ%TLG	0.992 (0.986-0.998)	0.671	757.141	0.006*	0.941 (0.917-0.965)	0.670	738.481	2.94e-06*
SUVpeak-BL	0.994 (0.970-1.020)	0.532	762.433	0.661	1.013 (0.932-1.100)	0.527	764.348	0.769
SUVpeak-EOT	1.109 (1.077-1.142)	0.710	721.111	3.57e-12*	0.554 (0.384-0.800)	0.710	712.791	0.002*
ΔSUVpeak	0.936 (0.911-0.962)	0.665	737.660	2.20e-06*	1.019 (0.936-1.109)	0.665	739.483	0.670
Δ%SUVpeak	0.983 (0.978-0.988)	0.714	727.082	1.71e-10*	0.975 (0.961-0.990)	0.714	716.866	0.001*
SUVmean-BL	0.948 (0.881-1.019)	0.554	760.445	0.148	1.008 (0.803-1.265)	0.554	762.440	0.945
SUVmean-EOT	1.255 (1.168-1.349)	0.692	728.729	5.79e-10*	0.593 (0.393-0.896)	0.692	724.355	0.013*
ΔSUVmean	0.827 (0.773-0.884)	0.681	729.729	3.27e-08*	0.913 (0.729-1.144)	0.681	731.066	0.428
Δ%SUVmean	0.983 (0.978-0.988)	0.703	729.848	2.95e-10*	0.973 (0.957-0.990)	0.703	720.293	0.002*
SUVmax-BL	0.992 (0.972-1.012)	0.537	762.017	0.439	1.001 (0.926-1.081)	0.537	764.017	0.985
SUVmax-EOT	1.073 (1.050-1.095)	0.702	727.022	5.63e-11*	0.738 (0.612-0.889)	0.702	718.026	0.001*
ΔSUVmax	0.949 (0.929-0.969)	0.670	735.753	9.39e-07*	0.996 (0.926-1.070)	0.670	737.739	0.906
Δ%SUVmax	0.984 (0.979-0.989)	0.707	727.139	3.45e-10*	0.978 (0.964-0.992)	0.707	719.379	0.003*
TLRpeakpeak-BL	0.998 (0.954-1.045)	0.506	762.621	0.940	0.895 (0.738-1.086)	0.545	763.315	0.262
TLRpeakpeak-EOT	1.249 (1.177-1.326)	0.722	722.002	2.54e-13*	0.229 (0.107-0.490)	0.722	707.862	1.46e-04*
ΔTLRpeakpeak	0.906 (0.860-0.955)	0.625	747.591	2.41e-04*	1.035 (0.872-1.228)	0.625	749.441	0.696
Δ%TLRpeakpeak	0.987 (0.982-0.991)	0.702	736.406	4.24e-09*	0.971 (0.958-0.986)	0.702	720.026	7.65e-05*
TLRpeakmean-BL	1.000 (0.963-1.037)	0.504	762.626	0.984	0.933 (0.803-1.084)	0.536	763.767	0.362
TLRpeakmean-EOT	1.215 (1.153-1.280)	0.722	722.094	3.12e-13*	0.271 (0.139-0.528)	0.722	707.624	1.26e-04*
ΔTLRpeakmean	0.926 (0.886-0.968)	0.614	749.464	6.11e-04*	1.041 (0.908-1.195)	0.614	751.136	0.562
Δ%TLRpeakmean	0.986 (0.981-0.990)	0.701	736.209	1.04e-08*	0.971 (0.957-0.985)	0.701	720.071	6.29e-05*
TLRmaxmax-BL	0.991 (0.945-1.040)	0.514	762.488	0.711	0.902 (0.733-1.111)	0.534	763.517	0.333
TLRmaxmax-EOT	1.214 (1.153-1.280)	0.721	723.322	3.29e-13*	0.409 (0.249-0.673)	0.721	711.100	4.30e-04*
ΔTLRmaxmax	0.885 (0.838-0.935)	0.650	740.985	1.32e-05*	1.059 (0.879-1.275)	0.650	742.636	0.548
Δ%TLRmaxmax	0.986 (0.982-0.991)	0.704	730.927	1.21e-10*	0.977 (0.964-0.990)	0.704	718.900	0.001*
NOL-BL (n)	1.008 (0.998-1.019)	0.586	760.514	0.124	0.920 (0.847-1.000)	0.587	758.568	0.049*
NOL-EOT (n)	1.092 (1.064-1.119)	0.658	731.876	8.56e-12*	0.339 (0.162-0.710)	0.658	725.923	0.004*
ΔNOL	0.998 (0.985-1.011)	0.502	762.549	0.783	1.112 (1.017-1.217)	0.503	759.781	0.020*
Δ%NOL (%)	0.996 (0.993-0.999)	0.571	755.871	0.004*	1.000 (0.992-1.009)	0.571	757.871	0.995
DmaxBulk-BL	1.001 (1.000-1.002)	0.565	757.197	0.018*	0.999 (0.997-1.002)	0.565	759.062	0.715
DmaxBulk-EOT	1.004 (1.002-1.005)	0.651	734.642	1.46e-09*	0.994 (0.988-1.001)	0.651	733.912	0.100
ΔDmaxBulk	0.999 (0.998-1.000)	0.545	760.226	0.126	1.003 (1.000-1.005)	0.531	758.962	0.058*
Δ%DmaxBulk	0.992 (0.988-0.996)	0.611	748.764	5.82e-05*	0.999 (0.990-1.008)	0.611	750.690	0.787

AIC=Akaike information criterion

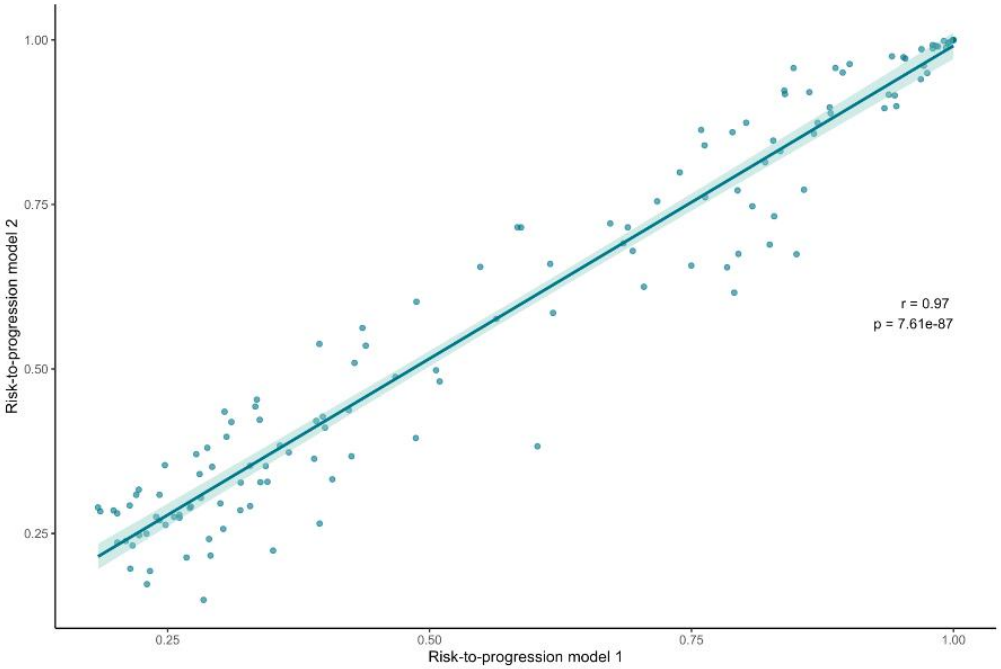


Figure S2. Correlation plot comparing risk-of-progression between model 1 and 2

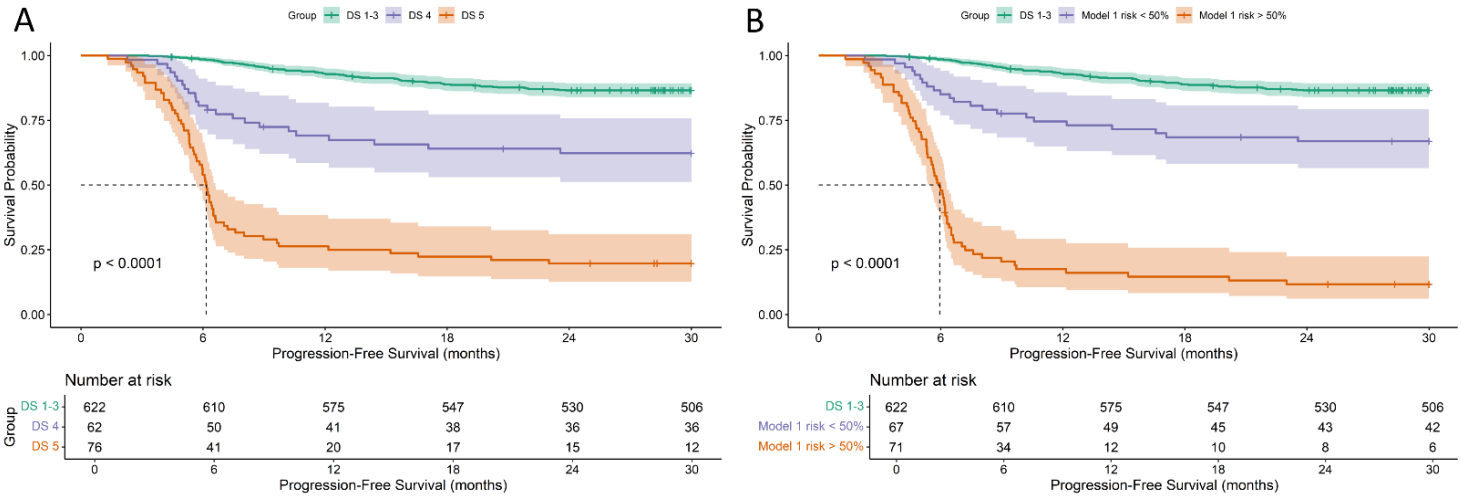


Figure S3. Kaplan-Meier survival curves comparing Deauville score (A) to the model 1 <50% and >50% risk groups for 2 year progression-free survival (B)

Table S5. Sensitivity scores using the clinical PET model

Risk to progression (%)	Correctly classified (n)	False positives (n)	False negatives (n)	Sensitivity	Specificity	Accuracy	PPV	NPV
5	84	37	17	0.798	0.315	0.609	0.644	0.500
10	79	22	37	0.560	0.593	0.573	0.681	0.464
20	73	12	53	0.369	0.778	0.529	0.721	0.442
30	62	6	70	0.167	0.889	0.449	0.700	0.407
40	59	6	78	0.071	0.982	0.428	0.857	0.405

Table S6. Overview of second-line therapy

Type	n = 138	Time in months between EOT PET and start therapy (median, IQR)
No second-line therapy	50	-
Radiotherapy (RT)	18	1.15 (0.46-1.25)
Chemo(immuno)therapy	29	1.08 (0.44-4.90)
Chemo(immuno)therapy + ASCT (+ RT*)	17	1.12 (0.62-6.47)
Chemo(immuno)therapy + RT	4	1.95 (1.28-3.24)
Unknown	20	-

*5 patients in this group had also received radiotherapy (RT); ASCT=Autologous stem cell transplantation

Table S7. Distribution of PFS for patients receiving radiotherapy

	Event	No Event	Total
Radiotherapy	12	15	27 (19.6%)
No Radiotherapy	63	32	95 (68.8%)
Unknown	9	7	16 (11.6%)
Total	84	54	138 (100.0%)

Table S8. Distribution of PFS for patients receiving second-line therapy (general)

	Event	No Event	Total
Second-line therapy	53	18	71 (51.4%)
No Second-line therapy	21	29	50 (36.2%)
Unknown	10	7	17 (12.3%)
Total	84	54	138 (100.0%)

Results S2: Consolidating radiotherapy

A subgroup of 122 patients in our database had data available on receiving consolidating radiotherapy. In this group 27 (22%) patients received radiotherapy after first-line treatment, of which 14 patients had a single lesion at EOT and 12 patients showed progression within 2 years. The radiotherapy status (binary) did not improve the performance of model 1 (p=0.340) and 2 (p=0.603), suggesting limited additional predictive value (*Table S9)*.

Table S9. Hazards of models after addition of radiotherapy in a subset of n=122 patients

	Model 1 (AIC=600.163, c-index=0.752)		Model 2 (AIC=597.707, c-index=0.773)	
Variable	HR (95% CI)	p-value	HR (95% CI)	p-value
TLRpeakmean-EOT	2.007 (1.454-2.773)	2.33e-05*	1.906 (1.368-2.656)	1.37e-04*
TLRpeakmean-EOT'	0.254 (0.111-0.582)	0.001*	0.316 (0.132- 0.753)	0.009*
NOL-EOT	1.185 (0.962-1.458)	0.110	1.164 (0.946-1.432)	0.152
NOL-EOT'	0.703 (0.317-1.556)	0.384	0.735 (0.332-1.627)	0.447
SUVmean-BL	-	-	0.912 (0.835-0.996)	0.041
Radiotherapy	0.733 (0.387-1.388)	0.340	0.840 (0.435-1.621)	0.603

'= splined variable; HR=hazard ratio

References:

1. Boellaard R, Herrmann K, Barrington SF, et al. [18F]FDG PET/CT: EANM procedure guidelines for tumour imaging: version 3.0. The EANM Journal. 2025;1(
2. Boellaard R. Quantitative oncology molecular analysis suite: ACCURATE [abstract]. J Nucl Med. 2018;59(supplement 1):1753.
3. Boellaard R, Buvat I, Nioche C, et al. International Benchmark for Total Metabolic Tumor Volume Measurement in Baseline (18)F-FDG PET/CT of Lymphoma Patients: A Milestone Toward Clinical Implementation. J Nucl Med. 2024;65(9):1343-1348.
4. Zwezerijnen GJC, Eertink JJ, Burggraaff CN, et al. Interobserver Agreement on Automated Metabolic Tumor Volume Measurements of Deauville Score 4 and 5 Lesions at Interim (18)F-FDG PET in Diffuse Large B-Cell Lymphoma. J Nucl Med. 2021;62(11):1531-1536.
5. Burggraaff CN, Rahman F, Kassner I, et al. Optimizing Workflows for Fast and Reliable Metabolic Tumor Volume Measurements in Diffuse Large B Cell Lymphoma. Mol Imaging Biol. 2020;22(4):1102-1110.