Risk prediction in diffuse large B-cell lymphoma improves when combining baseline PET features with interim PET response

Jakoba J. Eertink,[1,2] Martijn W. Heymans,[3,4] Sanne E. Wiegers,[1,2] Annelies L. Bes,[1,2] Ulrich Dührsen,[5] Andreas Hüttmann,[5] Lars Kurch,[6] Sally F. Barrington,[7] George N Mikhaeel,[8] Pieternella J. Lugtenburg,[9] Luca Ceriani,[10,11] Emanuele Zucca,[11,12] Tamas Györke,[13] Sándor Czibor,[13] Gerben J.C. Zwezerijnen,[2,14] Ronald Boellaard,[2,14] Josée M. Zijlstra[1,2] and Christine Hanoun;[5] on behalf of the PETRA Consortium

[1]Amsterdam UMC location Vrije Universiteit Amsterdam, Hematology, Amsterdam, the Netherlands; [2]Cancer Center Amsterdam, Imaging and Biomarkers, Amsterdam, the Netherlands; [3]Amsterdam UMC location Vrije Universiteit Amsterdam, Epidemiology and Data Science, Amsterdam, the Netherlands; [4]Amsterdam Public Health Research Institute, Methodology, Amsterdam, the Netherlands; [5]Department of Hematology and Stem Cell Transplantation, West German Cancer Center, University Hospital Essen, University of Duisburg-Essen, Essen, Germany; [6]Klinik und Poliklinik für Nuklearmedizin, Universitätsklinikum Leipzig, Leipzig, Germany; [7]King's College London and Guy's and St Thomas' PET Centre, School of Biomedical Engineering and Imaging Sciences, King's Health Partners, Kings College London, London, UK; [8]Department of Clinical Oncology, Guy's Cancer Centre and School of Cancer and Pharmaceutical Sciences, King's College London University, London, UK; [9]Erasmus MC Cancer Institute, University Medical Center Rotterdam, Department of Hematology, Rotterdam, the Netherlands; [10]Department of Nuclear Medicine and PET/CT Centre, IIMSI - Imaging Institute of Southern Switzerland, Bellinzona, Università della Svizzera Italiana, Bellinzona, Switzerland; [11]SAKK - Swiss Group for Clinical Cancer Research, Bern, Switzerland; [12]Department of Oncology, IOSI - Oncology Institute of Southern Switzerland, Bellinzona, Università della Svizzera Italiana, Bellinzona, Switzerland; [13]Department of Nuclear Medicine, Medical Imaging Centre, Semmelweis University, Budapest, Hungary and [14]Amsterdam UMC location Vrije Universiteit Amsterdam, Radiology and Nuclear Medicine, Amsterdam, the Netherlands

Correspondence: C. Hanoun
christine.hanoun@uk-essen.de

Received: December 21, 2024.
Accepted: May 8, 2025.
Early view: May 15, 2025.

https://doi.org/10.3324/haematol.2024.287241

Abstract

Accurate detection of patients at high risk of treatment failure following first-line immunochemotherapy in diffuse large B-cell lymphoma (DLBCL) is of paramount importance as patients might benefit from early treatment escalation. Recently, we introduced the International Metabolic Prognostic Index (IMPI) based on metabolic tumor volume (MTV), age and stage that outperformed the International Prognostic Index. However, radiomic features such as the maximum distance between the largest lesion and another lesion ($Dmax_{bulk}$) or the peak standardized uptake value (SUV_{peak}) along with early treatment response at interim positron emission tomography (iPET) based on ΔSUV_{max} may have additional predictive value. We tested different models for risk prediction aiming to develop a dynamic risk tool. All patients within the PETRA database with newly diagnosed DLBCL treated with R-CHOP, who had available clinical data, baseline PET and iPET scans were included. The optimal transformation of $Dmax_{bulk}$, SUV_{peak} and ΔSUV_{max} was determined by choosing the best fitting Cox regression model with lowest Akaike Information Criterion (AIC), while the cross-validated c-index was obtained as a measure for discrimination. Risk models were developed using clinical, baseline PET and iPET data. The best risk model was compared to the IMPI and our subsequent ClinicalPET model. A total of 1,014 patients were included in the analyses. Best baseline model included age, MTV and $Dmax_{bulk}$ (AIC 3208.89, c-index 0.70). Adding iPET response further improved outcome prediction (AIC 3140.36, c-index 0.74) with wider segregation of Kaplan-Meier curves and improved rates of correct risk classification, supporting the value of a dynamic risk assessment in DLBCL.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of lymphoma. Although most patients are sensitive to standard immunochemotherapy based on rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone (R-CHOP),[1,2] up to 30% of all patients are primary refractory or suffer from relapse after first-line therapy. Thus, there is the utmost need for early detection of patients who are likely to fail front-line therapy in order to optimize treatment strategies.

The International Prognostic Index (IPI) is the most popular risk stratification tool for DLBCL and is based on five clinical factors counted on a dichotomous scale: older age defined as >60 years, advanced stage defined as stage III/IV, elevated serum lactate dehydrogenase (LDH), poor performance status defined as Eastern Cooperative Oncology Group (ECOG) >1, and involvement of more than one extranodal site.[3] The IPI is commonly used in daily routine, yet fails to detect individual patients at high risk.[4] Positron emission tomography/computed tomography (PET/CT) using ^{18}F-fluorodeoxglucose (FDG) is considered standard of care in DLBCL for diagnosis and treatment response.[5] Pretreatment PET/CT enables accurate staging and also allows for quantitative and radiomic measurements, such as determination of the total metabolic tumor volume (MTV), the quantification of FDG uptake measured as maximum and peak (hottest 1 mL volume of tumor) standardized uptake value (SUV_{max} and SUV_{peak}) as well as the maximum distance between the largest lesion and the lesion furthest away ($Dmax_{bulk}$).[6-8] Several studies have shown that MTV in DLBCL is predictive for both progression-free survival (PFS) and overall survival (OS).[9-11] Recently, the PETRA consortium introduced the International Metabolic Prognostic Index (IMPI) as a new pretreatment risk stratification tool consisting of three baseline parameters: age, MTV, and stage. Using age and MTV as continuous variables and stage as a categorical variable, the IMPI individualizes risk prediction in DLBCL and is more accurate than the IPI in identifying patients at high risk of failing first-line treatment.[12] Furthermore, MTV, SUV_{peak} and $Dmax_{bulk}$, when incorporated with clinical factors (age and performance status) into a ClinicalPET model at baseline, were also shown to independently predict outcome in DLBCL.[7,13]

Early metabolic response to chemotherapy assessed using an interim PET/CT (iPET) scan performed after two (iPET-2) or four (iPET-4) cycles of chemoimmunotherapy is also strongly associated with improved outcome[14] using two methods for iPET response assessment. The visual Deauville score (DS) is based on a 5-point scale comparing the residual FDG uptake in the hottest lymphoma lesion to the physiological uptake in the liver and mediastinum on the same scan, and is recommended in current guidelines.[5,8,15] The quantitative ΔSUV_{max} method refers to the residual FDG uptake in the hottest lesion in the iPET scan compared to the hottest lesion in the baseline scan. Several studies have found ΔSUV_{max} to be superior to the DS, especially regarding its ability to predict early treatment failures.[14,16] An 'optimal' threshold for ΔSUV_{max} of ≥66% at iPET-2 and ≥70% at iPET-4 has been proposed to define a favorable iPET response,[17] but whether ΔSUV_{max} might provide better discrimination as a continuous variable in this setting has not been explored.

The aims of this study were to analyze whether a dynamic risk tool incorporating early metabolic response at iPET compared with baseline patient characteristics and radiomic features could further improve outcome prediction in an era of evolving clinical and radiomic risk prediction.

Methods

Study population

Patients from the PETRA consortium database (https://petralymphoma.org) were originally treated within one of five clinical studies (GSTT15,[9] HOVON-84,[18] NCRI,[19] PETAL,[20] and SAKK[21]) and included in this analysis if patients had newly diagnosed DLBCL, and were treated with R-CHOP with available clinical data and baseline PET and iPET scans, the latter performed after cycle 2 or cycle 4. All clinical studies were approved by institutional review boards and/or ethics committees. The use of all data within the PETRA imaging database has been approved by the institutional review board of the VU University Medical Center (JR/20140414).

PET parameters

Baseline PET scans were assessed for quality using criteria included in EANM guidelines.[22] Scans were excluded if: 1) incomplete; or 2) essential DICOM data were missing; or 3) the liver SUV_{mean} was outside the acceptable range of 1.3-3.0 and the total image activity (MBq), was outside 50-80% of the total injected FDG activity.

Lesions were delineated at baseline using a fully automated preselection defined by SUV ≥4.0 and a volume threshold of ≥3 mL using the Accurate tool.[23] Physiological uptake (e.g., bladder, kidneys) was deleted and lymphoma lesions <3 mL were added with single mouse-clicks. All scans were reviewed by a nuclear medicine physician who was blinded to outcome. Based on these delineations, MTV (in mL), SUV_{peak} and $Dmax_{bulk}$ (in mm) were extracted for all patients using RaCAT software, an open-source radiomic calculator measuring the maximum distance between the largest lesion and the lesion furthest away.[24] ΔSUV_{max} was calculated based on the highest SUV_{max} for both baseline PET and iPET scans, where the reduction in SUV was calculated as: ΔSUV_{max} = (SUV_{max} iPET scan - SUV_{max} baseline PET)/ SUV_{max} baseline PET).

Statistical analysis

The primary endpoint of this analysis was 3-year PFS, de-

fined as time from baseline PET to progression, relapse or death from any cause. Secondary endpoints were 3-year OS defined as time from baseline PET to death. For both PFS and OS, patients were censored after three years. Further analyses were performed using 1-year PFS and 1-year OS to test for the ability of the final prediction model to identify patients with primary refractory disease.

A multi-step approach was used to examine how to incorporate $Dmax_{bulk}$, ΔSUV_{max} and SUV_{peak} to the predictive models. For all models, MTV was included using a linear spline model, and age was used as continuous variable, as described previously.[12] Detailed statistical methods are provided in the *Online Supplementary Appendix*. Briefly, best spline models of $Dmax_{bulk}$, ΔSUV_{max} and SUV_{peak} were compared to other transformations to test for the best transformation model by choosing the model with the highest R^2 and lowest Akaike information criterion (AIC). Cox regression models incorporating clinical factors and PET parameters tested for best baseline predictors before the additive value of iPET response were evaluated. The thus-derived prediction model was then compared to the IMPI and ClinicalPET model before internal validation was performed.

Results

Starting with the IMPI population (N=1,241 patients), a total of 1,014 patients met the inclusion criteria for the subsequent analyses. Patients with no, incomplete or false timing of iPET scans (N=70) and patients who were not treated with R-CHOP at first-line therapy (N=21) were excluded. Of the remaining 1,150 patients, 136 patients were further excluded. Reasons for exclusion were: no FDG uptake at baseline (N=4), incomplete/missing dicom header information (N=11), ^{18}F-FDG–PET quality control out of range (N=44), scans outside quality range (N=51), short follow-up under two years (N=22), missing relevant clinical data (N=3); for one patient, we were unable to distinguish tumor from tissue.

Patient characteristics are presented in Table 1. Median follow-up was 4.57 years with a 3-year PFS rate of 75.6% (95% confidence interval (CI): 72.9-78.3%) and a 3-year OS rate of 83.0% (95% CI: 80.7- 85.3%). Survival rates according to the three IMPI subgroups were comparable between the present cohort and the original IMPI cohort (*Online Supplementary Table S1*), suggesting our cohort to be representative for further outcome analyses.

Table 1. Patient characteristics of included patients.

Characteristic	Total (N=1,014)	PETAL (N=408)	HOVON-84 (N=278)	GSTT (N=84)	SAKK (N=111)	NCRI (N=133)
Age, years, median, (range)	62 (18-86)	61 (18-82)	64 (23-80)	61 (26-86)	58 (18-81)	61 (21-80)
Age, years, N (%)						
≤60	459 (45.2)	200 (49.0)	92 (33.1)	42 (50.0)	62 (55.9)	63 (47.4)
>60	555 (54.7)	208 (51.0)	186 (66.9)	42 (50.0)	49 (44.1)	70 (52.6)
Stage, N (%)						
1	120 (11.8)	89 (21.8)	-	9 (10.7)	14 (12.6)	8 (6.0)
2	244 (24.1)	88 (21.6)	47 (16.9)	19 (22.6)	39 (35.1)	51 (38.3)
3	217 (21.4)	83 (20.3)	64 (23.0)	10 (11.9)	25 (22.5)	35 (26.3)
4	433 (42.7)	148 (36.3)	167 (60.0)	46 (54.8)	33 (29.7)	39 (29.3)
LDH, N (%)						
< ULN	411 (40.5)	182 (44.6)	91 (32.7)	33 (39.2)	55 (49.5)	50 (37.6)
>ULN	603 (59.5)	226 (55.4)	187 (67.3)	51 (60.7))	56 (50.0)	83 (62.4)
ECOG, N (%)						
0	503 (49.6)	178 (43.6)	157 (56.5)	29 (34.5)	64 (57.7)	75 (56.4)
1	392 (38.6)	192(47.1)	86 (30.9)	31 (36.9)	39 (35.1)	44 (33.1)
2	101 (10.0)	29 (7.1)	35 (12.6)	15 (17.9)	8 (7.2)	14 (10.5)
3	18 (1.8)	9 (2.2)	-	9 (10.7)	-	-
EN, N (%)						
≤1	689 (67.9)	288 (70.6)	166 (59.7)	43 (51.2)	86 (77.5)	106 (79.7)
>1	325 (32.1)	120 (29.4)	112 (40.3)	41 (48.8)	25 (22.5)	27 (20.3)
IPI, N (%)						
Low	335 (33.0)	154 (37.7)	47 (16.9)	25 (29.8)	57 (51.3)	52 (39.1)
Low-intermediate	230 (22.7)	104 (25.5)	67 (24.1)	9 (10.7)	22 (19.8)	28 (21.1)
High-intermediate	267 (26.3)	91 (22.3)	98 (35.3)	26 (31.0)	17 (15.3)	35 (26.3)
High	182 (17.9)	59 (14.5)	66 (23.7)	24 (28.6)	15 (13.5)	18 (13.5)

ECOG: Eastern Cooperative Oncology Group; EN: extranodal involvement >1 site; IPI: International Prognostic Index; LDH: lactate dehydrogenase; N: number; ULN: upper normal level.

Step 1 - Transformation of $Dmax_{bulk}$, ΔSUV_{max} and SUV_{peak}

We compared the best LSP models for ΔSUV_{max}, $Dmax_{bulk}$ and SUV_{peak} with other regression spline models and identified only minor differences between different transformation methods for all chosen endpoints (*Online Supplementary Table S2.1-2*). Thus, the linear version was chosen for further analyses.

Step 2 - Combining clinical characteristics and PET metrics at baseline

Baseline characteristics that were significantly associated with survival were age (P=0.01), MTV (P<0.01), and $Dmax_{bulk}$ (P<0.01), while SUV_{peak}, stage and ECOG did not show any significant impact on outcome prediction (Tables 2 and 3, respectively). The optimal baseline model was the model that included MTV, age and $Dmax_{bulk}$.

Step 3 - Adding iPET response to the best baseline model

Combining age, MTV and $Dmax_{bulk}$ with iPET response markedly improved outcome prediction and was superior to the baseline model alone, which resulted in the best model fit with the highest c-index and lowest AIC. This model consisted of age, MTV, $Dmax_{bulk}$ and ΔSUV_{max} and resulted in an AIC of 3140.36 using 3-year PFS (Table 4). The baseline and iPET model that used ΔSUV_{max} as a continuous variable outperformed visual DS for response assessment (AIC of 3140.36 compared to 3177.44, respectively) and led to a sharper segregation of KM-curves compared to iPET response alone (*Online Supplementary Figure S1*).

There was a significant interaction between ΔSUV_{max} and timing of iPET, which was true for both 3-year OS and 3-year PFS (P<0.01). Patients with the same reduction in intensity, e.g., the same ΔSUV_{max} values, had an inferior outcome if the iPET assessment was performed later (i.e., after 4 cycles compared to 2 cycles) (*Online Supplementary Figure S2A, B),*), this was also true if we used dichotomous cut-offs.

Step 4 - Comparisons of the new model to the IMPI and the ClinicalPET model

Both our optimal baseline model and our optimal baseline-iPET model outperformed the IPI, IMPI and ClinicalPET models with respect to AIC, R^2 and c-index values for all chosen endpoints (Table 4). Focusing on the baseline models, the baseline model (age, MTV and $Dmax_{bulk)}$ outperformed the other baseline models. This model is very comparable to the IMPI also including age and MTV, but replacing stage by $Dmax_{bulk}$.

The baseline-iPET model (using ΔSUV_{max}) had the lowest value for AIC and highest c-index for all chosen endpoints resulting in a wider segregation of the Kaplan-Meier curves (Table 4, Figures 1 and 2).

Most importantly, our new baseline-iPET model was superior in detecting patients at high risk, with a shorter 3-year PFS of 31.58% (95% CI: 23.68-42.13%) compared to the high-risk patients according to the IMPI and ClinicalPET models (3-year PFS: 47.08% [95% CI: 38.25-57.96%] and 44.45% [95% CI: 35.71-55.31%], respectively) (Figure 1).

Patients in the intermediate-risk group in the baseline-iPET model had a 3-year PFS of 67.59% (95% CI: 62.51-73.08%),

Table 2. Multivariable Cox regression analyses for baseline clinical and PET characteristics as well as iPET response based on ΔSUV_{max} with respect to 3-year progression-free survival.

Parameter	HR	95% CI	*P*
MTV	1.00	1.00-1.01	<0.01
MTV'	1.00	-	<0.01
Age	1.01	1.00-1.02	0.01
Stage 2	0.81	0.42-1.53	0.51
Stage 3	0.64	0.33-1.26	0.20
Stage 4	0.83	0.43-1.60	0.59
ECOG >1	0.97	0.68-1.38	0.84
Extranodal involvement >1 site	0.78	0.56-1.09	0.14
ΔSUV_{max}	0.98	0.98-0.99	<0.01
$Dmax_{bulk}$	1.00	1.00-1.00	<0.01
SUV_{peak}	0.98	0.97-1.00	0.07

ΔSUV_{max}: delta maximum standardized uptake value; 95% CI: 95% confidence interval; $Dmax_{bulk}$: maximum distance between the largest lesion and the lesion furthest away; ECOG: Eastern Cooperative Oncology Group; HR: Hazard ratio; iPET: interim positron emission tomography; MTV: metabolic tumor volume; PET: positron emission tomography; PFS: progression-free survival; SUV_{peak}: peak standardized uptake value. Hazard ratios for the Cox regression model using 3-year PFS are presented for MTV per mL, age per year, ΔSUV_{max} per percentage change, and $Dmax_{bulk}$ per mm. MTV is included using a linear spline at 307.9 mL, therefore, MTV' refers to the HR for MTV values above 307.9.

Table 3. Multivariable Cox regression analyses for baseline clinical and PET characteristics as well as iPET response based on ΔSUVmax with respect to 3-year overall survival.

Parameter	HR	95% CI	*P*
MTV	1.00	1.00-1.00	<0.01
MTV'	1.00	0.99-1.00	<0.01
Age	1.04	1.02-1.05	<0.01
Stage 2	1.45	0.58-3.62	0.42
Stage 3	1.15	0.45-2.96	0.76
Stage 4	1.85	0.73-4.70	0.40
ECOG>1	1.07	0.71-1.61	0.74
Extranodal involvement >1 site	0.67	0.45-0.98	0.04
ΔSUV_{max}	0.98	0.98-0.98	<0.01
$Dmax_{bulk}$	1.00	1.00-1.00	<0.01
SUV_{peak}	0.99	0.96-1.01	0.18

ΔSUV_{max}: delta maximum standardized uptake value; 95% CI: 95% confidence interval; $Dmax_{bulk}$: maximum distance between the largest lesion and the lesion furthest away; ECOG: Eastern Cooperative Oncology Group; HR: hazard ratio; MTV: metabolic tumor volume; iPET: interim positron emission tomography; PET: positron emission tomograpy; SUV_{peak}: peak standardized uptake value. Hazard ratios for the Cox regression model using 3-year overall survival are presented for MTV per mL, age per year, ΔSUV_{max} per percentage change, and $Dmax_{bulk}$ per mm. MTV is included using a linear spline at 307.9 mL, therefore, MTV' refers to the HR for MTV values above 307.9.

which was comparable to survival rates of intermediate-risk patients identified by the IMPI and ClinicalPET models (3-year PFS: 68.65% [95% CI: 63.62-74.08%] and 66.48% [95% CI: 61.35-72.03%], respectively).

Lastly, patients assigned to the low-risk group according to the baseline-iPET model had a superior 3-year PFS of 86.87% (95% CI: 84.22-89.62%) compared to low-risk patients in the IMPI and ClinicalPET models (3-year PFS: 83.73% [95% CI: 80.83-86.73%] and 85.26% [95% CI: 82.48-88.14%], respectively).

Data were comparable for 3-year OS (Figure 2); survival rates are presented in *Online Supplementary Table S3*.

We also determined changes in progression risk classification and found an improvement using the baseline-iPET model of 22% for 3-year PFS and 16% for 3-year OS compared to the IMPI, meaning that 22% of the patients were reclassified to a correct risk category when using our baseline-iPET model compared to the IMPI. The same effects were seen for the comparison to the ClinicalPET model, with a reclassification rate of 9% for 3-year PFS

Table 4. Different prediction models based on baseline clinical and PET characteristics alone and combined with iPET response.

Baseline models	R2		C-index		AIC	
	3-year PFS	3-year OS	3-year PFS	3-year OS	3-year PFS	3-year OS
IPI (age, stage, LDH, ECOG, EN involvement)	0.06	0.07	0.64	0.66	3,280.65	2,256.93
IMPI (MTV, age, stage)	0.10	0.11	0.68	0.71	3,247.09	2,224.28
MTV, age, $Dmax_{bulk}$	0.13	0.13	0.70	0.71	3,208.89	2,200.13
ClinicalPET model (age, ECOG, $Dmax_{bulk}$, MTV, SUV_{peak})	0.13	0.12	0.70	0.71	3,216.07	2,208.75
Baseline and iPET combined	**R2**		**c-Index**		**AIC**	
	3-year PFS	3-year OS	3-year PFS	3-year OS	3-year PFS	3-year OS
MTV, age, stage, iPET response (ΔSUV_{max})	0.16	0.17	0.72	0.75	3,177.44	2,165.59
MTV, age, $Dmax_{bulk}$, iPET response (ΔSUV_{max})	0.18	0.12	0.74	0.77	3,140.36	2,146.41
MTV, age, $Dmax_{bulk}$, iPET response (DS 1-3 *vs.* 4-5)	0.15	0.15	0.71	0.75	3,182.96	2,176.31

AIC: Akaike information criterion; $Dmax_{bulk}$: maximum distance between the largest lesion and the lesion furthest away; DS: Deauville Score; ECOG: Eastern Cooperative Oncology Group; EN: extranodal involvement >1 site; IMPI: International Metabolic Prognostic Index; iPET: interim positron emission tomography; IPI: International Prognostic Index; LDH: lactate dehydrogenase; MTV: metabolic tumor volume; OS: overall survival; PET: positron emission tomograpy; PFS: progression-free survival; SUV_{peak}: peak standardized uptake value; ΔSUVmax: delta maximum standardized uptake value.

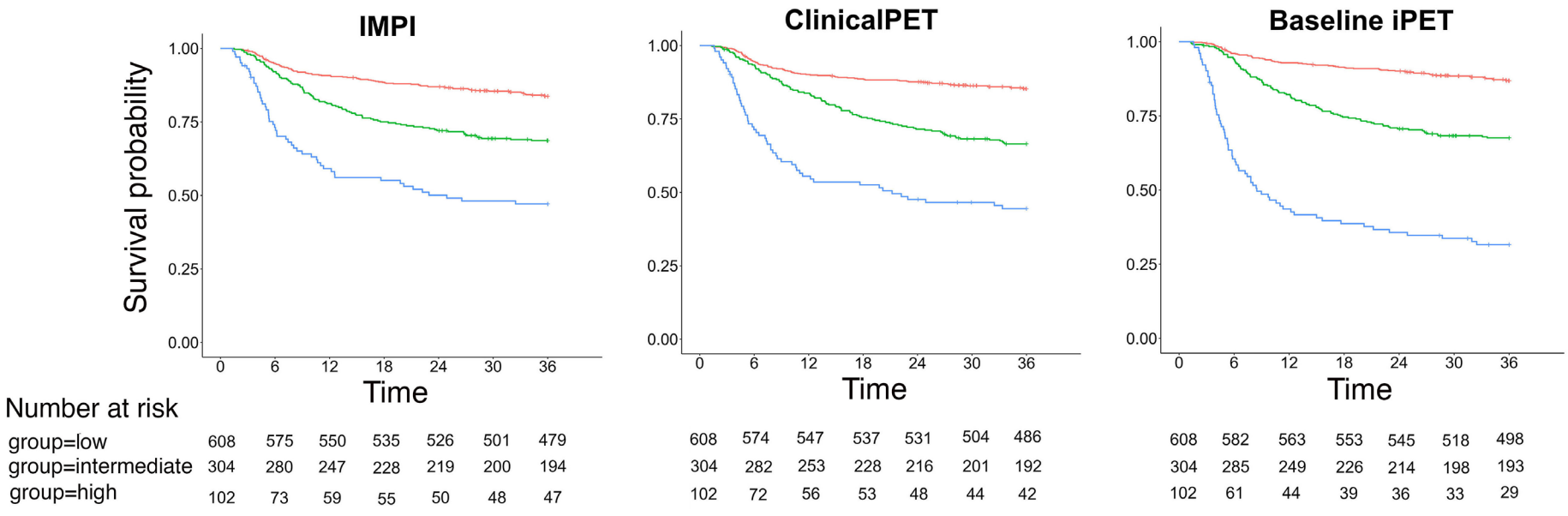

Figure 1. Kaplan-Meier curves displaying progression-free survival according to different risk models based on risk groups comprising 60% (low risk), 30% (intermediate risk), and 10% (high risk) of patients. Combining baseline metabolic tumor volume, age and the maximum distance between the largest lesion and the lesion furthest away ($Dmax_{bulk}$) with interim positron emission tomography (iPET) response to a baseline-iPET model improves risk prediction in diffuse large B-cell lymphoma resulting in a wider segregation of Kaplan-Meier curves compared to the International Metabolic Prognostic index (IMPI) and the baseline ClinicalPET model with respect to progression-free survival.

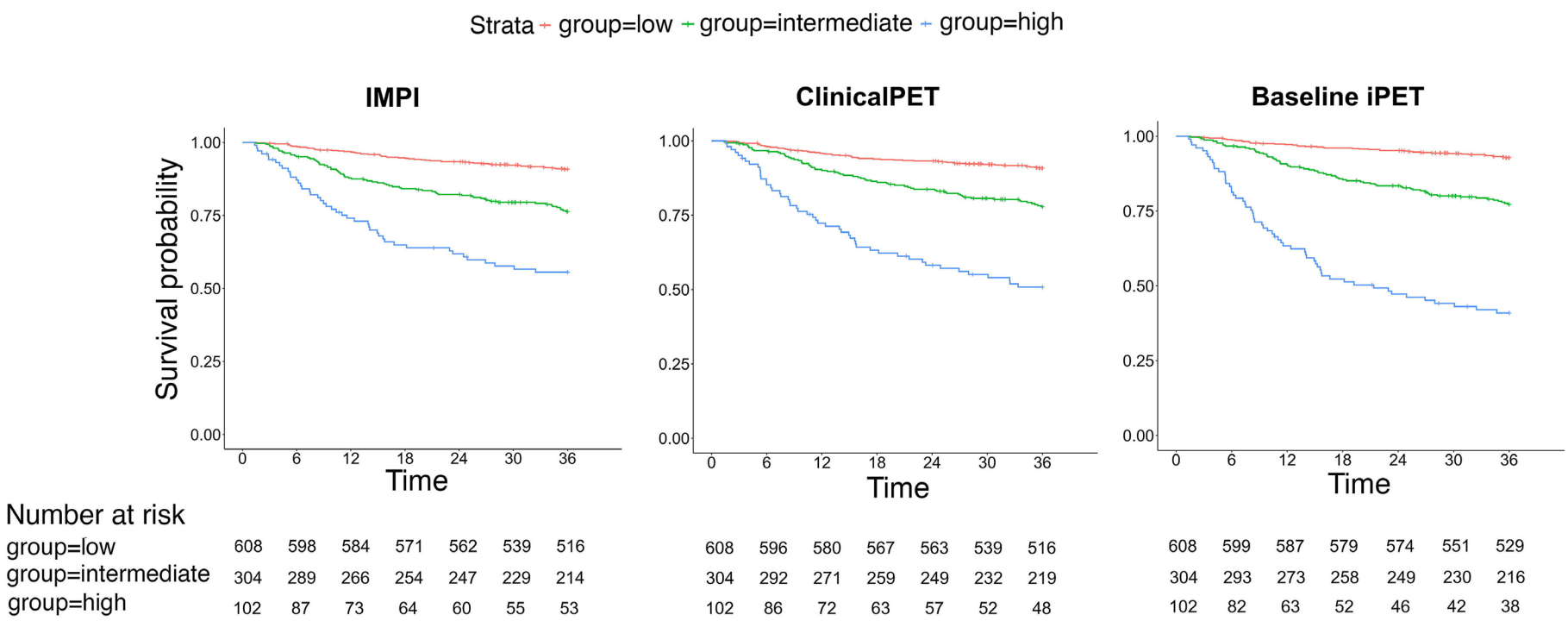

Figure 2. Kaplan-Meier curves displaying overall survival according to different risk models based on risk groups comprising 60% (low risk), 30% (intermediate risk), and 10% (high risk) of patients. Combining baseline metabolic tumor volume, age and the maximum distance between the largest lesion and the lesion furthest away ($Dmax_{bulk}$) with interim positron emission tomography (iPET) response to a baseline-iPET model improves risk prediction in diffuse large B-cell lymphoma resulting in a wider segregation of Kaplan-Meier curves compared to the International Metabolic Prognostic index (IMPI) and the baseline ClinicalPET model with respect to overall survival.

and 10% for 3-year OS, as well as 20.8% for 3-year PFS and 30% compared to the IPI, respectively (Table 5). Moreover, detection of primary refractory patients was more precise according to our new model compared to the other risk prediction models, with a survival rate of 44% for the high-risk group according to our baseline-iPET model and 59% of the high-risk group according to the IMPI (*Online Supplementary Table S4*).

For a rapid risk estimation, a PFS calculator is provided in the *Online Supplementary Appendix*, including patient examples.

Step 5 – Cross-validation of models

Internal validation based on the 'leave-one-study-out' approach confirmed our findings and further supported superiority of our new baseline-iPET model compared to the IMPI and the ClinicalPET model (*Online Supplementary Table S5*). The 'leave-one-study-out' cross-validation of the final model showed a calibration slope value of 0.97, which showed that the model is robust against overfitting across application in different cohorts. Notably, only the HOVON84 trial performed the iPET scan after cycle 4, while all others performed iPET scan after cycle 2. Thus, we were not able to validate our findings using different time points for iPET assessment.

Discussion

We present a dynamic risk assessment tool that enables individualized risk prediction for patients with DLBCL, by combining baseline clinical and PET parameters with interim PET response performed during immunochemotherapy based on R-CHOP.

In our analyses, we aimed to assess whether risk prediction could be further improved by combining iPET response with the best baseline model. ΔSUV_{max} was superior to visual DS1-3 *versus* 4-5 for response. The important addition of this compared to previous work is both the independence of baseline features and interim response and their additive prognostic value. This places our model in the unique situation where interim response is not evaluated in isolation, but rather in the context of the baseline disease-burden characteristics. Combining the best baseline model with interim PET response improved the ability to identify patients at high risk of disease progression and early death, outperforming baseline IPI, baseline IMPI and the baseline ClinicalPET model. The baseline-iPET model improved correct risk re-classification rates by 22% compared to the IMPI, 21% compared to the IPI and 9% compared to the ClinicalPET model. Superiority in detecting the primary 10% refractory patients may allow for an early treatment switch for patients who are very likely to fail front-line immunochemotherapy.

In line with previous findings,[14,16,25] and notably, the time point of iPET response assessment was important in this setting. Patients with a ΔSUV_{max} reduction of >66% after iPET-2 had a comparable outcome to patients with a ΔSUV_{max} reduction >70% after iPET-4.[17] However, as iPET-4 was only performed in one of the five studies we were not able to further deepen our analyses. Though the additive predictive

Table 5. Net reclassification displaying the rate of correct progression risk classification of our baseline-iPET model compared to the International Metabolic Prognostic Index, the ClinicalPET model and the International Prognostic Index.

	Baseline-iPET model, *versus* IMPI Estimates (95% CI)		Baseline-iPET model, *versus* ClinicalPET model, Estimates (95% CI)		Baseline-iPET model, *versus* IPI Estimates (95% CI)	
	3-year PFS	3-year OS	3-year PFS	3-year OS	3-year PFS	3-year OS
Net reclassification index (event)	0.14 (0.03-0.24)	0.13 (0.01-0.24)	0.04 (-0.11-0.15)	0.09 (-0.02-0.18)	0.06 (-0.05-0.50)	0.29 (0.05-0.45)
Net reclassification index (no event)	0.08 (0.04-0.13)	0.03 (-0.01-0.07)	0.06 (0.01-0.12)	0.02 (-0.03-0.04)	0.15 (-0.11-0.25)	0.01 (-0.06-0.23)
Net reclassification index (total)	0.22 (0.12-0.31)	0.16 (0.06-0.26)	0.09 (-0.03-0.20)	0.10 (0.01-0.192)	0.21 (0.09-0.48)	0.30 (0.19-0.43)

95% CI: 95% confidence interval; IMPI: International Metabolic Prognostic Index; iPET: interim positron emission tomography; IPI: International Prognostic Index; PFS: progression-free survival; OS: overall survival.

value of iPET response to baseline parameters has been reported before,[9,10,26] this is the first study incorporating ΔSUV_{max} as a continuous instead of dichotomous variable. The best baseline model in our analysis incorporated age, MTV and $Dmax_{bulk}$, all used as continuous variables. Beside age, no other IPI factors contributed further additive predictive value at baseline, supporting the assumption that more than one extranodal involvement, LDH and ECOG act as surrogates for tumor burden, which can be better assessed using MTV. Unlike the IMPI,[12] stage was not included in our best baseline model, but replaced by $Dmax_{bulk}$ as new radiomic feature, which is defined as maximum distance between the lymphoma lesion with the largest volume and the lesion furthest apart from it. Thus, $Dmax_{bulk}$ may better reflect the spread of lymphoma than stage and be an independent risk factor for lymphoma burden, as previously suggested.[6,13]

Strengths of our study are the large patient population from five high quality clinical trials, all with quality assured baseline PET scans and uniform assessment of MTV using the same segmentation method of SUV4. Using the 'leave-one-study-out' approach we were able to support our findings by internal validation. A PFS calculator for rapid and individualized risk estimation is provided in the *Online Supplementary Appendix*. However, whilst uniform treatment with R-CHOP is in principle a strength, making data more comparable, patients at higher risk of treatment failure (defined as IPI score 3-5) are nowadays recommended to be treated with Pola-R-CHP based on findings of the POLARIX-trial.[2] This is a limitation, as risk prediction rates may vary in the setting of new treatment protocols, requiring further analysis. Other limitations are that neither assessment of MTV nor the assessment of $Dmax_{bulk}$ has yet been standardized. There are different methods for MTV determination, including methods using relative or fixed thresholds. However, approaches for MTV standardization are advanced, with good inter-reader reproducibility using SUV4 as fixed threshold, as performed in our analyses.[27-29]

Nevertheless, data on interobserver reproducibility of $Dmax_{bulk}$ are still lacking; here, a consensus for a standardized assessment is urgently needed before this radiomic feature can be integrated into clinical routine.

We also did not include information on genetic risk profiles, that may have additional prognostic value in this setting. Lastly, before translation into clinical routine, our proposed risk model needs an external validation in a prospective clinical trial. This validation should also include a Pola-R-CHP treated cohort and may be set up in a risk-based treatment protocol incorporating an early access to chimeric antigen receptor T-cell therapies or use of bispecific antibodies for patients likely to fail front-line therapy in order to directly test for outcome improvements.

In conclusion, age, MTV and $Dmax_{bulk}$ as baseline features combined with quantitative iPET response resulted in a dynamic and individualized risk prediction identifying patients at good, intermediate and poor risk. While patients at good risk may be suitable for treatment de-escalation, patients at high risk may benefit from alternative treatments by adding or switching to bispecific antibodies or chimeric antigen receptor T-cell therapies. Prospective validation of our risk assessment model, including validation in a Pola-R-CHP treated cohort, is needed before translation into clinical routine.

Disclosures

SFB is in receipt of departmental funding from Bristol Myers Squibb international corporation, Pfizer Inc. and Takeda. JMZ received support from Roche. All the other authors have no conflicts of interest to disclose.

Contributions

JJE contributed to the concept and design of this study, performed PET/CT analyses, and performed statistical analyses. MWH contributed to the concept and design of this study and performed statistical analyses. NGM, SFB, JMZ and CH contributed to the concept and design of this study. UD, AH, SFB, NGM, EZ, TG and PJL were responsible for data acquisition. GJCZ, SEW, SFB, CH, LK, LC and SC performed PET/CT analyses. All authors contributed to the interpreta-

tion of the data and critically reviewed and approved the manuscript.

Acknowledgments
The authors thank all patients who participated in the trials and the collaborating investigators who kindly supplied their data. The authors also thank all data managers who collected the clinical data and 18F-FDG–PET/CT scans for individual studies.

Funding
This study was financially supported by the Dutch Cancer Society (VU 2018–11648). The PETAL trial was supported by grants from Deutsche Krebshilfe (107592 and 110515). SFB acknowledges the support from the National Institute for Health and Care Research (RP-2-16-07-001). King's College London and the UCL Comprehensive Cancer Imaging Centre are funded by the CRUK and EPSRC in association with the MRC and the Department of Health and Social Care (England). This work was also supported by core funding from the Wellcome/EPSRC Centre for Medical Engineering at King's College London [WT203148/Z/16/Z] and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London and/or the NIHR Clinical Research Facility.

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

References

1. Sehn LH, Donaldson J, Chhanabhai 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(22):5027-5033.
2. Tilly H, Morschhauser F, Sehn LH, et al. Polatuzumab vedotin in previously untreated diffuse large B-cell lymphoma. N Engl J Med. 2022;386(4):351-363.
3. 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.
4. Sehn LH, Berry B, Chhanabhai M, et al. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. Blood. 2007;109(5):1857-1861.
5. 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.
6. Cottereau AS, Nioche C, Dirand AS, et al. (18)F-FDG PET dissemination features in diffuse large B-cell lymphoma are predictive of outcome. J Nucl Med. 2020;61(1):40-45.
7. 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.
8. 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.
9. 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.
10. 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.
11. Toledano MN, Desbordes P, Banjar A, et al. Combination of baseline FDG PET/CT total metabolic tumour volume and gene expression profile have a robust predictive value in patients with diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging. 2018;45(5):680-688.
12. 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.
13. Eertink JJ, Zwezerijnen GJC, Heymans MW, et al. Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma. Blood. 2023;141(25):3055-3064.
14. 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.
15. Ricard F, Cheson B, Barrington S, et al. Application of the Lugano Classification for initial evaluation, staging, and response assessment of Hodgkin and Non-Hodgkin lymphoma: the PRoLoG Consensus Initiative (Part 1-Clinical). J Nucl Med. 2023;64(1):102-108.
16. Rekowski J, Hüttmann A, Schmitz C, et al. Interim PET evaluation in diffuse large B-cell lymphoma using published recommendations: comparison of the Deauville 5-point scale and the ΔSUV(max) method. J Nucl Med. 2021;62(1):37-42.
17. 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(1):37-43.
18. 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.
19. Mikhaeel NG, Cunningham D, Counsell N, et al. FDG-PET/CT after two cycles of R-CHOP in DLBCL predicts complete remission but has limited value in identifying patients with poor outcome - final result of a UK National Cancer Research Institute prospective study. Br J Haematol. 2021;192(3):504-513.
20. 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.
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. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328-354.
23. Boellaard R. Quantititave oncology molecular analysis suite: ACCURATE. J Nucl Med. 2018;59(Suppl 1):1753.
24. Pfaehler E, Zwanenburg A, de Jong JR, Boellaard R. RaCaT: an open source and easy to use radiomics calculator tool. PLoS One. 2019;14(2):e0212223.
25. Kurch L, Huttmann A, Georgi TW, et al. Interim PET in diffuse large B-cell lymphoma. J Nucl Med. 2021;62(8):1068-1074.
26. Michaud L, Bantilan K, Mauguen A, Moskowitz CH, Zelenetz AD, Schöder H. Prognostic value of (18)F-FDG PET/CT in diffuse large B-cell lymphoma treated with a risk-adapted immunochemotherapy regimen. J Nucl Med. 2023;64(4):536-541.
27. Barrington SF, Zwezerijnen B, de Vet HCW, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method is most successful? A study on behalf of the PETRA Consortium. J Nucl Med. 2021;62(3):332-337.
28. Barrington SF, Meignan M. Time to prepare for risk adaptation in lymphoma by standardizing measurement of metabolic tumor burden. J Nucl Med. 2019;60(8):1096-1102.
29. 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.