REGULAR ARTICLE

blood advances

Baseline radiomics features and *MYC* rearrangement status predict progression in aggressive B-cell lymphoma

Jakoba J. Eertink,[1,2] Gerben J. C. Zwezerijnen,[2,3] Sanne E. Wiegers,[1,2] Simone Pieplenbosch,[2,3] Martine E. D. Chamuleau,[1,2] Pieternella J. Lugtenburg,[4] Daphne de Jong,[2,5] Bauke Ylstra,[2,5] Matias Mendeville,[2,5] Ulrich Dührsen,[6] Christine Hanoun,[6] Andreas Hüttmann,[6] Julia Richter,[7] Wolfram Klapper,[7] Yvonne W. S. Jauw,[1-3] Otto S. Hoekstra,[2,3] Henrica C. W. de Vet,[8,9] Ronald Boellaard,[2,3] and Josée M. Zijlstra,[1,2] on behalf of the PETRA consortium

[1]Department of Hematology, Amsterdam University Medical Center (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 Hematology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands; [5]Department of Pathology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [6]Department of Hematology, West German Cancer Center, University Hospital Essen, University of Duisburg-Essen, Essen, Germany; [7]Hematopathology Section, Department of Pathology, University Hospital Schleswig-Holstein, Christian-Albrecht-University of Kiel, Kiel, Germany; [8]Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; and [9]Department of Methodology, Amsterdam Public Health, Amsterdam, The Netherlands

Key Points

- Baseline radiomics features accurately predict progression in aggressive B-cell lymphoma.
- Radiomics features combined with *MYC* rearrangement status resulted in the most accurate selection of high-risk patients.

We investigated whether the outcome prediction of patients with aggressive B-cell lymphoma can be improved by combining clinical, molecular genotype, and radiomics features. *MYC*, *BCL2*, and *BCL6* rearrangements were assessed using fluorescence in situ hybridization. Seventeen radiomics features were extracted from the baseline positron emission tomography–computed tomography of 323 patients, which included maximum standardized uptake value (SUV_{max}), SUV_{peak}, SUV_{mean}, metabolic tumor volume (MTV), total lesion glycolysis, and 12 dissemination features pertaining to distance, differences in uptake and volume between lesions, respectively. Logistic regression with backward feature selection was used to predict progression after 2 years. The predictive value of (1) International Prognostic Index (IPI); (2) IPI plus *MYC*; (3) IPI, *MYC*, and MTV; (4) radiomics; and (5) *MYC* plus radiomics models were tested using the cross-validated area under the curve (CV-AUC) and positive predictive values (PPVs). IPI yielded a CV-AUC of 0.65 ± 0.07 with a PPV of 29.6%. The IPI plus *MYC* model yielded a CV-AUC of 0.68 ± 0.08. IPI, *MYC*, and MTV yielded a CV-AUC of 0.74 ± 0.08. The highest model performance of the radiomics model was observed for MTV combined with the maximum distance between the largest lesion and another lesion, the maximum difference in SUV_{peak} between 2 lesions, and the sum of distances between all lesions, yielding an improved CV-AUC of 0.77 ± 0.07. The same radiomics features were retained when adding *MYC* (CV-AUC, 0.77 ± 0.07). PPV was highest for the *MYC* plus radiomics model (50.0%) and increased by 20% compared with the IPI (29.6%). Adding radiomics features improved model performance and PPV and can, therefore, aid in identifying poor prognosis patients.

Submitted 22 July 2022; accepted 26 September 2022; prepublished online on *Blood Advances* First Edition 28 October 2022; final version published online 12 January 2023. https://doi.org/10.1182/bloodadvances.2022008629.

Presented in poster and oral form at the 8th international workshop on PET in lymphoma and myeloma, 11 September 2021 (abstract number 176). Presented in oral session 627 at the 63rd annual meeting of the American Society of Hematology, Atlanta, GA, 12 December 2021 (abstract number 451).

Deidentified individual participant data can be requested through the PETRA consortium (https://petralymphoma.org). Contact and more information can be obtained either via the contact form or the email address of the consortium (petra@amsterdamumc.nl).

The full-text version of this article contains a data supplement.

Introduction

Patients with aggressive B-cell lymphoma have a large variation in outcome, which is partly explained by genetic abnormalities, such as *MYC* oncogene rearrangements (*MYC*-R).[1] *MYC*-R occur in ~10% to 15% of patients.[1-3] Thirty percent of the patients only have *MYC*-R and are often referred to as single-hit patients (*MYC*-SH). In 70% of these cases, *MYC*-R is accompanied by a translocation of the *BCL2* and/or *BCL6* genes, which is classified as high-grade B-cell lymphoma with *MYC* and *BCL2* and/or *BCL6* rearrangement, also called double/triple hit (DH/TH).[4] For these patients, standard first-line therapy results in poor outcomes with a 2-year progression-free survival (PFS) of ~60%.[2,5,6] Therefore, patients with DH/TH are often treated with dose-intensification regimens, although no standard of care regimen has formally been established for them.[7,8]

18F-fluorodeoxyglucose (18F-FDG) positron emission tomography–computed tomography (PET/CT) is the current clinical standard for staging at baseline and response evaluation during or after treatment.[9,10] 18F-FDG PET/CT is also used to quantify the metabolic tumor volume (MTV) in patients[11,12] as an estimate of the total tumor burden. Baseline MTV is an important predictor of outcome and is inversely related to overall survival (OS) and PFS.[13,14] Moreover, we recently showed that MTV combined with Ann Arbor staging and age allows individual relapse prediction for de novo patients with diffuse large B-cell lymphoma (DLBCL) by applying the International Metabolic Prognostic Index.[15] Besides MTV, additional quantitative parameters, often referred to as radiomics features, can be extracted from 18F-FDG PET/CT scans. Radiomics features provide detailed information on the distribution of 18F-FDG–tracer uptake, morphology, and spread and texture of lesions. Radiomics features extracted from baseline 18F-FDG PET/CT scans have shown to be predictive of relapse in patients with DLBCL beyond just MTV.[16-20]

Whether baseline radiomics features differ between patients with aggressive B-cell lymphoma with molecular high-risk features, such as *MYC*-R, and patients without these high-risk features is still unknown. Moreover, the added value of radiomics features on the predictive value of *MYC*-R status has not been studied yet. Therefore, this study aimed to analyze the relation between *MYC*-R status and baseline PET parameters and to investigate the added value of radiomics features to the predictive value of *MYC*-R status in aggressive B-cell lymphomas.

Material and methods

Study population

In this posthoc analysis, we included all patients with de novo aggressive B-cell lymphoma whose tumor data on *MYC*, *BCL2*, and *BCL6* rearrangements by fluorescence in situ hybridization (FISH)[21] and baseline 18F-FDG PET scans were available from the PETRA database. 18F-FDG PET/CT scans and patient-level clinical and genetic data were collated and harmonized by the PETRA consortium.[22] All patients were originally included in the multicenter phase 2 HOVON-130 trial (https://eudract.ema.europa.eu/, #2014-002654-39),[23] the multicenter randomized phase 3 HOVON-84 trial, (https://eudract.ema.europa.eu/, #2006-005174-42)[24] and the multicenter randomized phase 3 PETAL trial (https://eudract.ema.europa.eu/, #2006-001641-33).[25] Individual trials were approved by institutional review boards and all the patients included provided informed consent. The use of all data within the PETRA imaging database has been approved by the institutional review board of the Vrije Universiteit University Medical Center (JR/20140414). Patients with wild-type *MYC* (hereafter referred to as patients with *MYC*-WT DLBCL) were treated with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP); R-CHOP intensified with rituximab (RR-CHOP); or 2 cycles of R-CHOP after a Burkitt protocol consisting of high-dose methotrexate, cytarabine, hyperfractionated cyclophosphamide and ifosfamide, split-dose doxorubicin and etoposide, vincristine, vindesine, and dexamethasone. Patients with *MYC*-SH and DH/TH were treated with R-CHOP combined with lenalidomide (R2-CHOP), R-CHOP, RR-CHOP, or the Burkitt protocol.

Pathology review

For all patients, *MYC*, *BCL2*, and *BCL6* rearrangement statuses were assessed using FISH.[21] The FISH analysis was performed according to routine procedures with the following standard commercial probes as part of the diagnostic workup: *MYC* break-apart, *BCL2* break-apart, and *BCL6* break-apart probes (Vysis/Abbott, DAKO, and Kreatech). In selected cases, FISH data were completed as part of the central pathology review process using Vysis/Abbott break-apart probes for only *BCL2* and *BCL6*.[23] Patients were classified as DH/TH according to the World Health Organization 2016 classification.[4]

PET/CT analysis

For PET/CT quality control (QC), we used the ranges as suggested by the guidelines of the European Association of Nuclear Medicine for the hepatic mean standardized uptake value (SUV_{mean}) and plasma glucose.[26] When the hepatic SUV_{mean} fell outside the suggested ranges, but the total image activity was 50% to 80% of the total injected activity, scans were still included. Moreover, QC rejected scans if (1) scans were incomplete, (2) essential Digital Imaging and Communications in Medicine information was missing, and (3) scans were from a PET-only system. Quantitative PET/CT analysis of all tumor lesions was performed using the ACCURATE tool. MTV was calculated at baseline using the fixed SUV $\geq$4.0 segmentation method.[27] Nontumor 18F-FDG avid regions (eg, brain, kidney, and bladder) adjacent to lesions were manually removed. All scans were reviewed by a nuclear medicine physician, and delineations were performed under the supervision of a nuclear medicine physician who was blinded to the outcome.

Radiomics feature extraction

MTV, SUV_{max}, SUV_{peak}, SUV_{mean}, and total lesion glycolysis (TLG) were extracted at patient level for all patients included. Furthermore, the following 12 dissemination features were extracted: the number of lesions, 4 features quantifying distance between lesions,[28] 5 features quantifying the differences in SUV_{peak} between lesions, and 3 features quantifying the differences in MTV between lesions. All image processing and feature calculations were performed using RaCaT software,[29] which complies with the imaging biomarker standardization initiative criteria.[30]

Statistical analysis

Differences in radiomics features between MYC subgroups. Differences in radiomics features between patients with *MYC*-WT DLBCL, *MYC*-SH DLBCL, and DH/TH were

assessed using the Kruskal-Wallis test by ranks. In the case of significant differences in radiomics features, Dunn's test of multiple comparisons with Benjamini-Hochberg correction for multiple testing was used as a post hoc test. Correlations between radiomics features stratified for *MYC*-R status were calculated using Spearman correlation coefficients.

Prediction models. We tested the predictive value of the following models:

1. IPI: the International Prognostic Index (IPI) using low, low-intermediate, high-intermediate, and high-risk groups.[31]
2. *MYC*: *MYC*-R status (categorical: *MYC*-WT, *MYC*-SH, and DH/TH).
3. IPI + *MYC*: a combination of IPI and *MYC*-R status (categorical).
4. IPI + *MYC* + MTV: a combination of IPI, *MYC*-R status (categorical), and MTV.
5. Radiomics: MTV, SUV_{max}, SUV_{peak}, SUV_{mean}, TLG, and 12 dissemination features.
6. Radiomics + *MYC*: MTV, SUV_{max}, SUV_{peak}, SUV_{mean}, TLG, 12 dissemination features, and *MYC*-R status.
7. Combined: MTV, SUV_{max}, SUV_{peak}, SUV_{mean}, TLG, 12 dissemination features, IPI, and *MYC*-R status.

Multivariate logistic regression with backward feature selection was used to predict the risk of progression or relapse after 2 years. Follow-up started at the time of baseline 18F-FDG PET/CT scan. We started with all potential predictors in the model and at every turn, the predictor with the highest *P* value was excluded from the model until all remaining predictors were significant. Patients who died without progression or were lost to follow-up within 2 years were excluded. Before feature selection, continuous variables that had a skewness of >0.5 were log-transformed using the natural logarithm. Model performance was assessed using repeated cross-validation (fivefold, 2000 repeats) yielding the cross-validated area under the curve of the receiver operating characteristics curve (CV-AUC). To match the prevalence of patients with *MYC*-SH and DH/TH with real-world prevalence,[1] for each repeat all 245 patients with *MYC*-WT DLBCL were included, and 10 patients with *MYC*-SH DLBCL and 20 with DH/TH were selected using random stratified sampling. Within the same cross-validation loop, we determined overfitting in the regression coefficients of the best model by applying the train linear predictor (calibration slope) in the test data sets and determined its Akaike information criterion (AIC).

The cell of origin (COO) was available for 298 patients,[32] for whom we also tested the predictive value of a prediction model with all features from the model that included IPI and *MYC*-R status (model 3) and the combined model (model 7) and COO (categorical: germinal center B cell, nongerminal center B cell, unclassified).

Relative feature importance. z scores of individual predictors were calculated to compare the relative effects of predictors that were measured on different scales for all multivariate logistic models. z scores were calculated by subtracting the mean and dividing by the standard deviation. These standardized features were used as predictors in logistic regression. The absolute values of the regression coefficients quantify the relative importance of the predictors.

Diagnostic performance. For all multivariate models, the sum of individual predictors, weighted by the regression coefficients, together with the intercept of the model resulted in the predicted probability (expressed as log odds) of progression for each patient. To calculate the diagnostic performance of the models, high- and low-risk groups were defined based on prior probability (ie, prevalence) of events.[33] For the IPI prediction model, patients with 4 or 5 adverse factors were considered high risk. For the *MYC* prediction model, patients with DH/TH were considered high risk. The diagnostic performance of the prediction models was assessed using sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Survival curves for time to progression (TTP), PFS, and OS were obtained with Kaplan-Meier analyses and compared with log-rank tests.

Statistical analysis was performed using R (version 4.0.3). A *P* value of $< .05$ was considered statistically significant.

Results

Patient characteristics

There were 458 patients with DLBCL with *MYC*, *BCL2*, and *BCL6* rearrangement status available, of whom 323 were included in this analysis. A total of 135 patients were excluded based on the following criteria/reasons: (1) no whole-body PET/CT scan available (n = 59), (2) PET/CT scan outside QC (n = 13), (3) essential Digital Imaging and Communications in Medicine information missing (n = 12), (4) no 18F-FDG avid disease (n = 6), (5) *BCL2* and *BCL6* mutation status could not be assessed (n = 9), (6) lost to follow-up within 2 years (n = 15), or (7) one of the individual IPI components was missing (n = 4). Patients who died within 2 years without signs of progression (n = 12) were excluded from the development of the prediction model and TTP survival analysis but included for PFS and OS.

In total, 245 patients with *MYC*-WT DLBCL, 24 with *MYC*-SH DLBCL, and 54 with DH/TH were included in this study (Table 1). For 3 patients with *MYC* and *BCL2* rearrangements, *BCL6* rearrangement status could not be assessed. The 2-year TTP of patients with *MYC*-WT DLBCL was 85.7% (95% confidence interval [CI], 81.4-90.2), compared with 66.7% (95% CI, 50.2-88.5) for patients with *MYC*-SH DLBCL and 57.4% (95% CI, 45.6-72.2) for patients with DH/TH. Both *MYC*-SH and DH/TH subgroups had a more pronounced male predominance compared with patients with *MYC*-WT DLBCL. Patients with *MYC*-SH DLBCL had similar baseline characteristics as those with *MYC*-WT DLBCL, whereas patients with DH/TH more often had higher frequencies of advanced-stage disease, elevated lactate dehydrogenase levels, and extranodal involvement leading to higher IPI scores compared with patients with *MYC*-SH DLBCL and *MYC*-WT DLBCL. In the *MYC*-WT DLBCL cohort, 1 patient was treated with the Burkitt protocol and the rest were treated with R-CHOP regimens. Sixty-seven percent of patients with *MYC*-SH DLBCL vs 76% of patients with DH/TH received other induction therapies than R-CHOP. Patient characteristics for individual trials are presented in supplemental Table 1.

Table 1. Patient characteristics

	Total (n = 323)	*MYC*-WT DLBCL (n = 245)	*MYC*-SH DLBCL (n = 24)	DH/TH (n = 54)
Gender				
Male	185 (57)	131 (53)	15 (63)	39 (72)
Female	138 (43)	114 (47)	9 (38)	15 (28)
Age (interquartile range), y	63 (53-71)	64 (54-71)	57 (45-66)	63 (55-71)
Ann Arbor stage				
I	23 (7)	20 (8)	2 (8)	1 (2)
II	54 (17)	44 (18)	5 (21)	5 (9)
III	73 (23)	62 (25)	3 (13)	8 (15)
IV	173 (54)	119 (49)	14 (58)	40 (74)
Lactate dehydrogenase				
Normal	124 (38)	103 (42)	9 (38)	12 (22)
>Normal	199 (62)	142 (58)	15 (63)	42 (78)
Extranodal involvement				
0-1	204 (63)	166 (68)	16 (67)	22 (41)
>1	119 (37)	79 (32)	8 (33)	32 (59)
World Health Organization performance status				
0	186 (58)	142 (58)	15 (63)	29 (54)
1	100 (31)	74 (30)	8 (33)	18 (33)
2	33 (10)	26 (11)	1 (4)	6 (11)
3	4 (1)	3 (1)		1 (2)
IPI				
Low	83 (26)	68 (28)	8 (33)	7 (13)
Low-intermediate	65 (20)	51 (21)	4 (17)	10 (19)
High-intermediate	104 (32)	75 (31)	8 (33)	21 (39)
High	71 (22)	51 (21)	4 (17)	16 (30)
Treatment				
6 × R-CHOP	99 (31)	94 (38)	1 (4)	4 (7)
8 × R-CHOP	29 (9)	24 (10)	2 (8)	3 (6)
6 × R-CHOP + 2 × R	71 (22)	67 (27)	2 (8)	2 (4)
6 × RR-CHOP	44 (14)	41 (17)	2 (8)	1 (2)
8 × RR-CHOP	22 (7)	18 (7)	1 (4)	3 (6)
Burkitt protocol	4 (1)	1 (1)	2 (8)	1 (2)
6 × R2-CHOP + 2 × R	54 (17)		14 (59)	40 (74)

All data are presented as number of patients (%), unless indicated otherwise.

Differences in radiomics features between *MYC* subgroups

Patients with *MYC*-SH DLBCL showed significantly lower intensity values (SUV_{mean} and SUV_{max}; $P < .04$) (Table 2; supplemental Tables 2 and 3) than those with *MYC*-WT DLBCL and DH/TH and more homogeneous intensity between lesions, as shown by the lower maximum difference in SUV_{peak} between 2 lesion ($DSUVpeak_{patient}$; $P = .10$) values. MTV and differences in MTV between lesions of patients with *MYC*-SH DLBCL were comparable to those with *MYC*-WT DLBCL. Patients with DH/TH had comparable uptake intensity and differences in intensity between lesions compared with patients with *MYC*-WT DLBCL. However, MTV was significantly higher ($P < .001$) (supplemental Tables 2 and 3), the spread of the disease was significantly larger (all: $P < .04$), differences in volume between lesions were significantly larger (all: $P < .001$), and the sum of differences between all lesions was significantly higher in patients with DH/TH compared with patients with *MYC*-WT DLBCL. MTV was highly correlated with TLG ($r = 0.98$-0.99) in all *MYC* subgroups, but not highly correlated with other radiomic features ($r < 0.7$ for all features). SUV_{peak} correlated highly with SUV_{mean}, SUV_{max}, $DSUVpeak_{bulk}$, and $DSUVpeak_{patient}$ for all *MYC* subgroups ($r > 0.7$ for all features).

Prediction model

The logistic regression model with IPI using 2-year TTP as outcome yielded a CV-AUC of 0.65 ± 0.07 (95% CI, 0.52-0.83) (Table 3; Figure 1). The model with IPI applied to patients with DH/TH

Table 2. Descriptive statistics of radiomics features stratified for *MYC*-R status

	MYC-WT DLBCL (n = 245)	*MYC*-SH DLBCL (n = 24)	DH/TH (n = 54)
SUV_{peak}	17.0 (11.8-22.4)	12.0 (8.7-16.7)	17.4 (12.9-23.4)
MTV	256.6 (54.9-777.0)	292.5 (15.9-1 098.7)	709.5 (304.6-1 280.1)
No. of lesions	7 (3-16)	5 (1-15)	10 (3-24)
$Spread_{patient}$	3 122.0 (156.7-25 383.9)	2 099.7 (0-18 197.4)	11 828.1 (1 034.3-77 878.9)
$Dmax_{bulk}$	28.2 (8.8-43.6)	28.6 (0-50.2)	35.3 (23.0-51.5)
$Volumemax_{patient}$	114.5 (11.0-497.0)	118.2 (0-700.7)	444.7 (88.3-752.8)
$DSUVpeak_{patient}$	10.5 (2.2-16.7)	5.0 (0-10.3)	10.8 (5.6-19.2)

All values are denoted as median (interquartile range). Corresponding *P* values between subgroups are presented in supplemental Table 3.

yielded a CV-AUC of 0.56 ± 0.15 (95% CI, 0.28-0.85). The logistic regression model with *MYC*-R status resulted in a CV-AUC of 0.58 ± 0.08 (95% CI, 0.43-0.72). The model that combined IPI with *MYC*-R status yielded a CV-AUC of 0.68 ± 0.08 (95% CI, 0.55-0.87). Adding the natural logarithm of MTV to IPI and *MYC*-R improved the CV-AUC to 0.74 ± 0.08 (95% CI, 0.59-0.87).

The highest model performance for the radiomics model after backward feature selection was observed for the natural logarithm of MTV combined with the maximum distance between the largest lesion and any other lesion ($Dmax_{bulk}$), the maximum difference in SUV_{peak} between 2 lesions ($DSUVpeak_{patient}$), and the sum of distances between all lesions ($Spread_{patient}$), yielding an improved CV-AUC of 0.77 ± 0.07 (95% CI, 0.62-0.89) (Figure 2). The same radiomics features were retained after backward feature selection when adding *MYC*-R status to the model (natural logarithm of MTV, $Dmax_{bulk}$, $DSUVpeak_{patient}$, and $Spread_{patient}$), which together with *MYC*-R status yielded comparable model performance (CV-AUC of 0.77 ± 0.07; 95% CI, 0.63-0.90; and lowest AIC). IPI was not retained in the combined model after backward feature selection, therefore, the combined model included the same features as the radiomics + MYC model thereby yielding the same CV-AUC and AIC. After the backward feature selection, the COO was not retained in the IPI + *MYC* and combined model. MTV was the most important radiomics feature in the radiomics model and the radiomics + MYC model, followed by $Dmax_{bulk}$ (supplemental Table 4).

Diagnostic performance

Sensitivity (31.8%) and PPV (29.6%) were the lowest for the IPI model. The NPV was comparable for all models and always >82% (Table 4). The PPV increased by 10% when combining radiomics features and *MYC*-R status compared with the IPI + *MYC* model (40.4% vs 50.0%) and increased by 20% in comparison with the IPI model (29.6% vs 50.0%). PPV and NPV were highest in the radiomics + *MYC* model. However, the model that only included radiomics features had comparable diagnostic performance to the radiomics + *MYC* model. Nineteen patients with DH/TH were classified as low risk by our radiomics + *MYC* model (supplemental Table 5), of which 4 patients showed progression within 2 years. Thirty-eight patients with DH/TH were classified as low risk by the IPI model, of which 15 patients showed progression within 2 years.

High-risk IPI patients had a 2-year TTP of 70.4% (95% CI, 60.6-81.9) (Table 5; Figure 3), a 2-year PFS of 64.9% (95% CI, 55.1-76.5) (Figure 3), and a 2-year OS of 68.8% (95% CI, 59.2-80.0) (supplemental Figure 1), compared with a much lower 2-year TTP of 50.0% (95% CI, 39.3-63.6) for the high-risk patients identified with the radiomics + *MYC* model. High-risk patients according to the radiomics + *MYC* model had a 2-year PFS of 50.6% (95% CI, 40.7-63.0) and a 2-year OS of 57.4% (95% CI, 45.6-72.2). Two-year TTP for high-risk patients identified with the radiomics model was 51.5% (95% CI, 40.8-65.1). Survival rates for other prediction models using 2-year PFS and 2-year OS as outcome parameters are presented in supplemental Table 6.

Table 3. Model performances of all models

	CV-AUC ± standard deviation (95% CI)	AIC
IPI	0.65 ± 0.07 (0.50-0.78)	191.1
MYC	0.58 ± 0.08 (0.43-0.72)	197.3
IPI + *MYC*	0.68 ± 0.08 (0.52-0.83)	186.7
IPI + *MYC* + MTV	0.74 ± 0.07 (0.59-0.87)	180.1
Radiomics	0.77 ± 0.07 (0.62-0.89)	175.0
Radiomics + *MYC*	0.77 ± 0.07 (0.63-0.90)	173.1

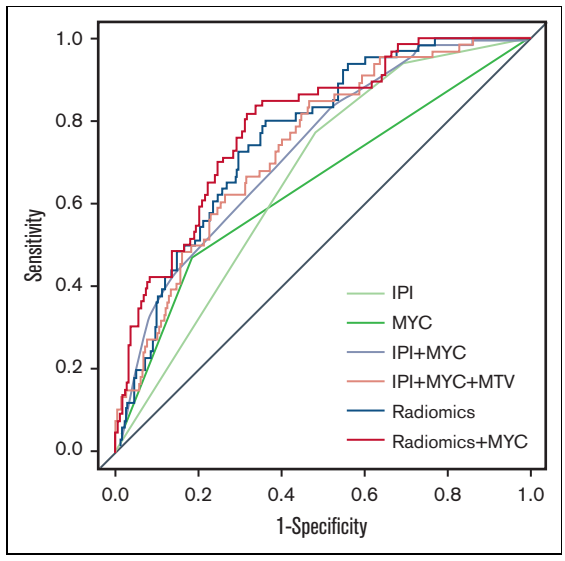

Figure 1. Receiver operating characteristic curves for 2-year TTP for IPI, *MYC*, IPI + *MYC*, IPI + *MYC* + MTV, radiomics, and radiomics + *MYC* prediction models.

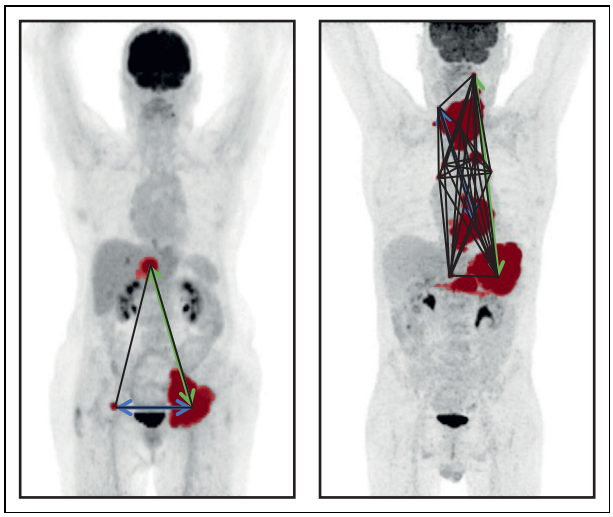

Figure 2. Maximum intensity projections of 2 patients. MTV is indicated in red, $Dmax_{bulk}$ with a green arrow, $DSUVpeak_{patient}$ with a blue arrow, and $Spread_{patient}$ with black lines.

Discussion

Our study shows that baseline PET radiomics features can identify high-risk patients with aggressive B-cell lymphoma, and that a prediction model based only on radiomics features can select high-risk patients more accurately than a model that combines IPI and *MYC*-R status. Moreover, adding dissemination and intensity features to MTV improves the predictive value and diagnostic accuracy of our prediction model. Better selection of high-risk patients is clinically relevant because it offers these patients a timely switch to innovative new treatment options, as well as including these patients in clinical trials offering them chimeric antigen receptor T-cell or bispecific monoclonal therapy.

Our results show that MTV values of patients with *MYC*-SH DLBCL are comparable to those with *MYC*-WT DLBCL. In addition, SUV metrics are significantly lower too. Patients with DH/TH had higher MTVs, higher SUVs, and larger dissemination at baseline than patients with *MYC*-SH DLBCL and *MYC*-WT DLBCL. To the best of our knowledge, no other studies compared either MTV, SUV, or dissemination features stratified for *MYC* and *BCL2* and/or *BCL6* rearrangement status in aggressive B-cell lymphoma. The higher intensity values and MTV of patients with DH/TH could be explained by the different pathological behavior (higher cell metabolism). We previously showed that dissemination expressed as distance does not correlate with MTV.[16] This current study shows that SUV metrics and dissemination in volume and intensity, respectively, do not correlate with MTV and are independent predictors of the outcome.

Several studies have shown that radiomics features extracted from baseline 18F-FDG PET scans are predictive of outcome in DLBCL[16,17,20,34] and the independent predictive value of both MTV and dissemination is expressed as distance.[16,28,35] In this study, we showed that both features were retained in the prediction model when adding *MYC*-R status and adding new dissemination features. Cottereau et al[36] showed that double expressor and patients with *MYC*-positive DLBCL using complementary DNA–mediated annealing, selection, ligation, and extension technology had an increased risk of relapse or progression, regardless of their MTV. Complementary DNA–mediated annealing, selection, ligation, and extension provides an expression profile, therefore, this signature does not capture *MYC* DH/TH translocation status. To the best of our knowledge, no studies incorporated both molecular genotypes and radiomics features extracted from 18F-FDG PET/CT scans.

Our model that only included radiomics features showed almost identical model performance compared with the model that included radiomics features and *MYC*-R related to the CV-AUC, standard deviation, 95% CI, and the AIC index. Moreover, the PPV and progression rates of the high-risk group identified with only radiomics features were very comparable to the PPV and progression and PFS rates of the model that combined radiomics features and *MYC*-R status. However, it should be noted that the model that included radiomics features and *MYC*-R status showed a steeper initial decline in the high-risk group using 2-year TTP and is superior in identifying primary refractory patients. For OS, 2-year survival rates dropped by an additional 5% when *MYC*-R status was added. Both models show the high predictive value of baseline radiomics features in patients with aggressive B-cell lymphoma. Furthermore, the survival rate of high-risk patients identified by the radiomics + *MYC* model dropped by 20% using 2-year TTP, by 14.3% using 2-year PFS, and by 11.3% using 2-year OS as an outcome parameter compared with high-risk IPI patients. Furthermore, our model that included both *MYC*-R status and radiomics features correctly identified 15 patients with DH/TH (25% of the population) as low risk. Moreover, compared with the IPI and *MYC*-R status, another advantage of our radiomics + *MYC* prediction model is the fact that it allows individual risk prediction per patient. Individual patients with poorer outcomes in need of treatment escalation can be identified and the optimal cutoff of the model can be selected based on the clinical context.

This study showed that when adding radiomics features extracted from baseline 18F-FDG PET/CT scans to *MYC*-R status, the selection of high-risk patients became more accurate with a higher PPV and CV-AUC. These data are important to place in the context

Table 4. Diagnostic measures of prediction models

	Sensitivity	Specificity	PPV	NPV
IPI	31.8 (20.9-44.4)	80.5 (75.2-85.2)	29.6 (21.4-39.3)	82.1 (79.4-84.6)
MYC	34.9 (23.5-47.6)	87.9 (83.3-91.7)	42.6 (31.8-54.2)	84.0 (81.4-86.3)
IPI + *MYC*	60.6 (47.8-72.4)	77.0 (71.4-82.0)	40.4 (33.5-47.7)	88.4 (84.9-91.2)
IPI + *MYC* + MTV	40.9 (29.0-53.7)	84.8 (79.8-89.0)	40.9 (31.5-51.0)	84.8 (82.0-87.3)
Radiomics	48.5 (36.0-61.1)	86.8 (82.0-90.7)	48.5 (38.7-58.4)	86.8 (83.8-89.3)
Radiomics + *MYC*	50.0 (37.4-62.6)	87.2 (82.4-91.0)	50.0 (40.1-59.9)	87.2 (84.2-89.7)

Table 5. Survival rates of high-risk and low-risk patients according to prediction models

	TTP (95% CI)
IPI	
Low	95.2 (90.7-99.9)
Low-intermediate	83.1 (74.4-92.7)
High-intermediate	71.2 (63.0-80.4)
High*	70.4 (60.6-81.9)
MYC	
MYC-WT	85.7 (81.4-90.2)
MYC-SH	66.7 (50.2-88.5)
DH/TH	57.4 (45.6-72.2)
IPI + *MYC*†	
Low	88.4 (84.3-92.7)
High	59.6 (50.7-70.1)
IPI + *MYC* + MTV	
Low	84.8 (80.5-89.3)
High	59.1 (48.3-72.2)
Radiomics	
Low	86.8 (82.7-91.0)
High	51.5 (40.8-65.1)
Radiomics + *MYC*	
Low	87.2 (83.2-91.3)
High	50.0 (39.3-63.6)

*n = 72 patients as high risk.
†n= 99 patients as high risk, all other models included n = 66 patients as high risk.

of modifications to frontline therapy. Several approaches are being explored to improve outcomes in high-risk subgroups. However, so far, no alternative frontline therapy has improved survival rates.[24,25,37,38] Widespread adoption of alternative treatments for biologically high-risk patients can increase treatment toxicity and health care expenditure. Therefore, a high PPV or upfront diagnosis is important to avoid unnecessary intensive therapy in a subset of patients with relatively favorable outcomes.[39,40]

MYC-R status is not always available before treatment and patients frequently receive 1 cycle of R-CHOP before they shift to intensified chemotherapy. The radiomics features that were extracted in this study can be calculated easily from the baseline PET without treatment delay allowing rapid stratification of patients starting with frontline therapy. Multiple vendors of PET/CT systems have implemented algorithms to calculate MTV in their clinical software. If the workflow is optimized, MTV can be calculated in 3 to 6 minutes, with complex cases taking up to 10 to 20 minutes.[41] Dissemination features are currently only extracted in research settings. However, these features are also relatively simple to calculate and relatively insensitive to differences in acquisition, reconstruction, and delineation methods.[42,43] Therefore, the implementation of the calculation of these features should be feasible in a reproducible manner in most clinical PET centers. We expect and hope that vendors implement the calculation of radiomics features in their software in the foreseeable future once more evidence of their clinical value becomes apparent. In the meantime, our image analysis tool, ACCURATE, is provided as an open tool to facilitate research.

To the best of our knowledge, our study is the first to incorporate both quantitative PET metrics and genetic markers with relatively large subsets of patients available with *MYC*-SH and DH/TH. The

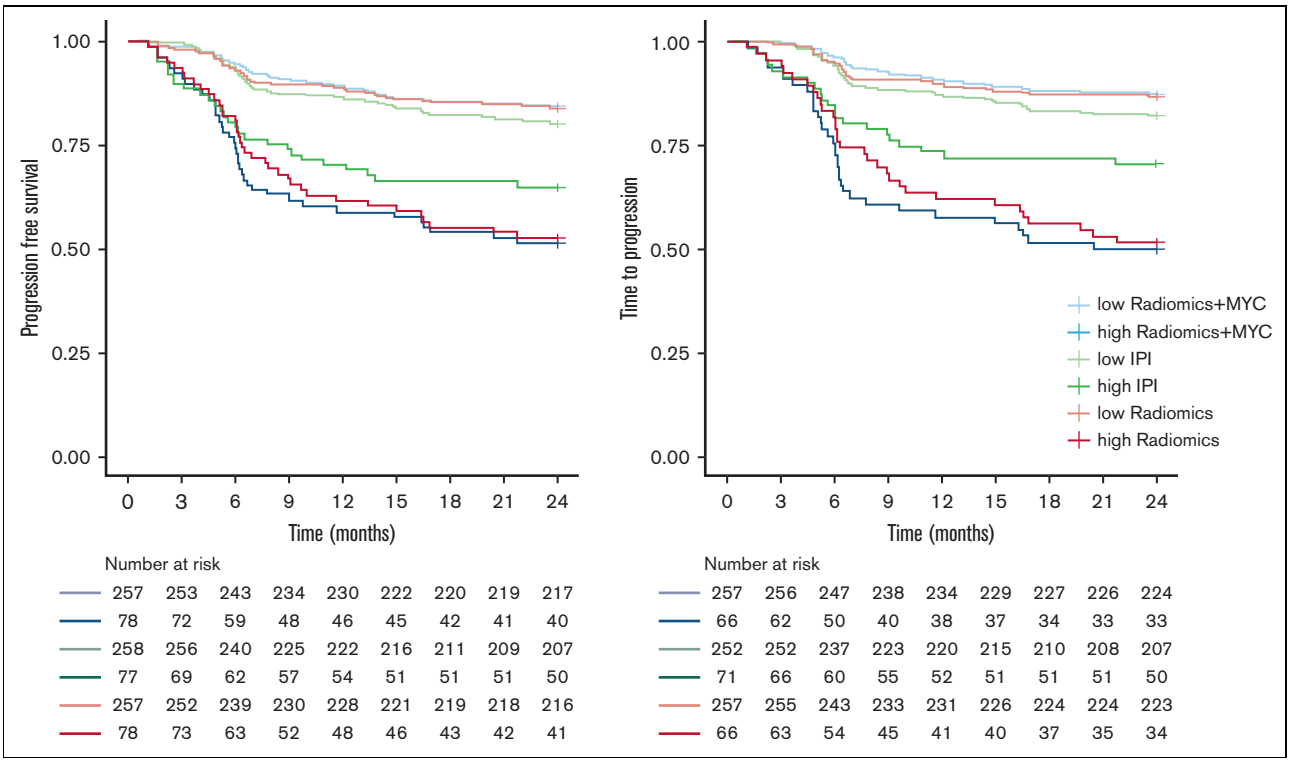

Figure 3. Kaplan-Meier survival curves of high- and low-risk groups for 2-year TTP and 2-year PFS.

uniform analysis of baseline 18F-FDG PET/CT scans and uniform FISH analysis in this study resulted in high-quality data. Nevertheless, several limitations of the current study should be noted. First, due to the retrospective nature of this study, treatment subgroups were heterogeneous. Patients with *MYC*-WT DLBCL were almost exclusively treated with R-CHOP regimens; whereas, one-third of the *MYC*-SH patients and 25% of patients with DH/TH received R-CHOP–based treatment. However, as we sampled patients for each fold based on *MYC*-R, only 5% of the patients in each fold were not treated according to current standards. Consequently, the treatment effect in our study was likely limited; yet an effect cannot be precluded. Moreover, not all patients in the prospective clinical trials had a baseline PET/CT scan available and/or sufficient biopsy material to assess *MYC*, *BCL2*, and *BCL6* rearrangement status. As a result, not all enrolled patients could be included in our analysis, possibly resulting in patient selection bias. Finally, our results regarding differences in radiomics features between patients with *MYC*-SH DLBCL and DH/TH could be suffering from relatively small sample sizes and should be validated in a larger cohort.

We chose to increase the internal validity of our model instead of leaving out 1 of the 3 trials or selecting a holdout set a priori. Even though the sample size of this study is large for a PET study, from a statistical perspective it was rather small. Small internal or external data sets suffer from large uncertainties when predicting outcomes, therefore, appropriate internal validation approaches using the full training data set are preferred over a small external data set or a holdout set, which is essentially the same as onefold in the cross-validation as the patient characteristics of the train and test set are identical if you leave out 1 part of the data.[33,44-46]

In summary, robust and easy-to-use biomarkers for the early identification of poor responders in this patient group are essential. We showed that radiomics features extracted from baseline 18F-FDG PET/CT scans accurately predict outcomes in aggressive B-cell lymphoma and an integrative approach with both molecular data and quantitative PET metrics could improve the prediction of prognosis and guide the choice of therapies.

Acknowledgments

This work was financially supported by the Dutch Cancer Society (VU 2018–11648). The PETAL trial was supported by grants from the Deutsche Krebshilfe (107592 and 110515).

Authorship

Contribution: J.J.E., G.J.C.Z., Y.W.S.J., O.S.H., H.C.W.d.V., R.B., and J.M.Z. contributed to the concept and design of this study; P.J.L., M.E.D.C., U.D., and A.H. were responsible for acquiring the data; J.J.E., S.E.W., S.P., C.H., and G.J.C.Z. performed positron emission tomography/computed tomography analyses; D.d.J., B.Y., M.M., J.R., and W.K. performed fluorescence in situ hybridization analyses; and all authors contributed to the interpretation of the data and critically reviewed and approved the manuscript.

Conflict-of-interest disclosure: M.E.D.C. received financial support for clinical trials from Celgene, Bristol-Myers Squibb, and Gilead. J.M.Z. achieved financial support for clinical trials from Roche, Gilead, and Takeda. The remaining authors declare no competing financial interests.

A complete list of the members of the PETRA consortium appears in "Appendix."

ORCID profiles: J.J.E., 0000-0002-6094-0016; G.J.C.Z., 0000-0002-9571-9362; S.E.W., 0000-0003-3698-8121; S.P., 0000-0002-7897-6383; P.J.L., 0000-0002-6735-8651; D.d.J., 0000-0002-9725-4060; B.Y., 0000-0001-9479-3010; M.M., 0000-0002-6130-1785; A.H., 0000-0003-2230-3873; W.K., 0000-0001-7208-4117; H.C.W.d.V., 0000-0002-5454-2804; J.M.Z., 0000-0003-1074-5922.

Correspondence: Jakoba Johanna Eertink, Department of Hematology, Amsterdam University Medical Center (UMC), Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands; email: j.eertink@amsterdamumc.nl.

Appendix

The members of the PETRA consortium are Josée Zijlstra, Riekie de Vet, Otto Hoekstra, Ronald Boellaard, Corinne Eertink, Coreline Burggraaff, Sanne Wiegers, Simone Pieplenbosch, Maria Ferrandez Ferrandez, Sandeep Golla, Ben Zwezerijnen, Annelies Bes, Martijn Heijmans, Yvonne Jauw, Elly Lugtenburg, Martine Chamuleau, Sally Barrington, George Mikhaeel, Ulrich Dührsen, Andreas Hüttmann, Lars Kurch, Christine Hanoun, Emanuele Zucca, Luca Ceriani, Robert Carr, Tamás Györke, Sándor Czibor, Stefano Fanti, Lale Kostakoglu, Annika Loft, Martin Hutchings, and Sze Ting Lee.

References

1. 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.
2. Savage KJ, Johnson NA, Ben-Neriah S, et al. MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood*. 2009;114(17):3533-3537.
3. Aukema SM, Kreuz M, Kohler CW, et al. Biological characterization of adult MYC-translocation-positive mature B-cell lymphomas other than molecular Burkitt lymphoma. *Haematologica*. 2014;99(4):726-735.
4. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375-2390.
5. Barrans S, Crouch S, Smith A, et al. Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab. *J Clin Oncol*. 2010;28(20):3360-3365.

6. Landsburg DJ, Falkiewicz MK, Petrich AM, et al. Sole rearrangement but not amplification of MYC is associated with a poor prognosis in patients with diffuse large B cell lymphoma and B cell lymphoma unclassifiable. *Br J Haematol*. 2016;175(4):631-640.
7. Davies A. Tailoring front-line therapy in diffuse large B-cell lymphoma: who should we treat differently? *Hematology Am Soc Hematol Educ Program*. 2017;2017(1):284-294.
8. National Comprehensive Cancer Network. B-Cell lymphomas (version 2.2022);2022.
9. 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.
10. 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.
11. 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 Imag Biol*. 2020;22(4):1102-1110.
12. 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.
13. 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.
14. Schmitz C, Huttmann A, Muller 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.
15. 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.
16. 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 Imag*. 2022;49(3):932-942.
17. Senjo H, Hirata K, Izumiyama K, et al. High metabolic heterogeneity on baseline 18FDG-PET/CT scan as a poor prognostic factor for newly diagnosed diffuse large B-cell lymphoma. *Blood Adv*. 2020;4(10):2286-2296.
18. 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.
19. Aide N, Fruchart C, Nganoa C, Gac AC, Lasnon C. Baseline (18)F-FDG PET radiomic features as predictors of 2-year event-free survival in diffuse large B cell lymphomas treated with immunochemotherapy. *Eur Radiol*. 2020;30(8):4623-4632.
20. Frood R, Burton C, Tsoumpas C, et al. Baseline PET/CT imaging parameters for prediction of treatment outcome in Hodgkin and diffuse large B cell lymphoma: a systematic review. *Eur J Nucl Med Mol Imag*. 2021;48(10):3198-3220.
21. Scott DW, Wright GW, Williams PM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood*. 2014;123(8):1214-1217.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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 Imag*. 2015;42(2):328-354.
27. Barrington SF, Zwezerijnen BG, de Vet HC, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method is most successful. *J Nucl Med*. 2021;62(3):332-337.
28. 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.
29. Pfaehler E, Zwanenburg A, de Jong JR, RaCaT Boellaard R. An open source and easy to use radiomics calculator tool. *PLoS One*. 2019;14(2):e0212223.
30. Zwanenburg A, Vallieres M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338.
31. 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.
32. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275-282.
33. Steyerberg EW. *Clinical Prediction Models*. Springer; 2009.

34. Ceriani L, Milan L, Cascione L, et al. Generation and validation of a PET radiomics model that predicts survival in diffuse large B cell lymphoma treated with R-CHOP14: a SAKK 38/07 trial post-hoc analysis. *Hematol Oncol*. 2022;40(1):11-21.

35. 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.

36. Cottereau AS, Lanic H, Mareschal S, et al. Molecular profile and FDG-PET/CT total metabolic tumor volume improve risk classification at diagnosis for patients with diffuse large B-cell lymphoma. *Clin Cancer Res*. 2016;22(15):3801-3809.

37. Bartlett NL, Wilson WH, Jung SH, et al. Dose-adjusted EPOCH-R compared with R-CHOP as frontline therapy for diffuse large B-cell lymphoma: clinical outcomes of the phase III intergroup trial alliance/CALGB 50303. *J Clin Oncol*. 2019;37(21):1790-1799.

38. Younes A, Sehn LH, Johnson P, et al. Randomized phase III trial of ibrutinib and rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in non-germinal center B-cell diffuse large B-cell lymphoma. *J Clin Oncol*. 2019;37(15):1285-1295.

39. Wilson WH, Sin-Ho J, Pitcher BN, et al. Phase III randomized study of R-CHOP versus DA-EPOCH-R and molecular analysis of untreated diffuse large b-cell lymphoma: CALGB/Alliance 50303. *Blood*. 2016;128(22):469.

40. Villa D, Sehn LH. Double hit lymphoma: do we need a 'double hit' of intensive therapy? *Leuk Lymphoma*. 2018;59(8):1767-1768.

41. Ilyas H, Mikhaeel NG, Dunn JT, et al. Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imag*. 2018;45(7):1142-1154.

42. Pfaehler E, van Sluis J, Merema BBJ, et al. Experimental multicenter and multivendor evaluation of the performance of PET radiomic features using 3-dimensionally printed phantom inserts. *J Nucl Med*. 2020;61(3):469-476.

43. Eertink JJ, Pfaehler EAG, Wiegers SE, et al. Quantitative radiomics features in diffuse large B-cell lymphoma: does segmentation method matter? *J Nucl Med*. 2022;63(3):389-395.

44. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247.

45. Smith GC, Seaman SR, Wood AM, Royston P, White IR. Correcting for optimistic prediction in small data sets. *Am J Epidemiol*. 2014;180(3):318-324.

46. Eertink JJ, Heymans MW, Zwezerijnen GJC, Zijlstra JM, de Vet HCW, Boellaard R. External validation: a simulation study to compare cross-validation versus holdout or external testing to assess the performance of clinical prediction models using PET data from DLBCL patients. *EJNMMI Res*. 2022;12(1):58.