

EUR Research Information Portal

Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma

Published in:
Blood

Publication status and date:
Published: 22/06/2023

DOI (link to publisher):
[10.1182/blood.2022018558](https://doi.org/10.1182/blood.2022018558)

Document Version
Publisher's PDF, also known as Version of record

Document License/Available under:
CC BY-NC-ND

Citation for the published version (APA):

Eertink, J. J., Zwezerijnen, G. J. C., PETRA consortium, Heymans, M. W., Pieplenbosch, S., Wieggers, S. E., Dührsen, U., Hüttmann, A., Kurch, L., Hanoun, C., Lugtenburg, P. J., Barrington, S. F., Mikhaeel, N. G., Ceriani, L., Zucca, E., Czibor, S., Györke, T., Chamuleau, M. E. D., Hoekstra, O. S., ... Zijlstra, J. M. (2023). Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma. *Blood*, *141*(25), 3055-3064.
<https://doi.org/10.1182/blood.2022018558>

[Link to publication on the EUR Research Information Portal](#)

Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

LYMPHOID NEOPLASIA

Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma

J. J. Eertink,^{1,2} G. J. C. Zwezerijnen,^{2,3} M. W. Heymans,^{4,5} S. Pieplenbosch,^{2,3} S. E. Wiegers,^{2,3} U. Dührsen,⁶ A. Hüttmann,⁶ L. Kurch,⁷ C. Hanoun,⁶ P. J. Lugtenburg,⁸ S. F. Barrington,⁹ N. G. Mikhaeel,¹⁰ L. Ceriani,^{11,12} E. Zucca,^{12,13} S. Czibor,¹⁴ T. Györke,¹⁴ M. E. D. Chamuleau,^{1,2} O. S. Hoekstra,^{2,3} H. C. W. de Vet,^{4,5} R. Boellaard,^{2,3} and J. M. Zijlstra,^{1,2} on behalf of the PETRA Consortium

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

KEY POINTS

- Baseline ¹⁸F-FDG-PET radiomics features can select patients at high risk more accurately than the IPI risk score.
- The clinical PET model that was developed in the HOVON-84 data set remained predictive of the outcome in 6 independent studies.

The objective of this study is to externally validate the clinical positron emission tomography (PET) model developed in the HOVON-84 trial and to compare the model performance of our clinical PET model using the international prognostic index (IPI). In total, 1195 patients with diffuse large B-cell lymphoma (DLBCL) were included in the study. Data of 887 patients from 6 studies were used as external validation data sets. The primary outcomes were 2-year progression-free survival (PFS) and 2-year time to progression (TTP). The metabolic tumor volume (MTV), maximum distance between the largest lesion and another lesion ($D_{max_{bulk}}$), and peak standardized uptake value (SUV_{peak}) were extracted. The predictive values of the IPI and clinical PET model (MTV, $D_{max_{bulk}}$, SUV_{peak} , performance status, and age) were tested. Model performance was assessed using the area under the curve (AUC), and diagnostic performance, using the positive predictive value (PPV). The IPI yielded an AUC of 0.62. The clinical PET model yielded a significantly higher AUC of 0.71 ($P < .001$). Patients with high-risk IPI had a 2-year PFS of 61.4% vs 51.9% for those with high-risk clinical PET, with an increase in PPV from 35.5% to 49.1%, respectively. A total of 66.4% of patients with high-risk IPI were free from progression or relapse vs 55.5% of patients with high-risk clinical PET scores, with an increased PPV from 33.7% to 44.6%, respectively. The clinical PET model remained predictive of outcome in 6 independent first-line DLBCL studies, and had higher model performance than the currently used IPI in all studies.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of aggressive non-Hodgkin lymphoma in adults with large variations in outcomes. Approximately 20% to 50% of patients with DLBCL are refractory to standard chemotherapy or relapse after achieving complete response.¹ With more available innovative treatment options (such as chimeric antigen T-cell and bispecific monoclonal therapy), better selection of patients at high risk is highly

relevant to potentially offer these patients a timely switch to these new treatment options.

Thirty years after its development, the international prognostic index (IPI)² is still the most widely used prognostic index for DLBCL. The addition of rituximab has significantly increased the cure rate.³ The ability to identify patients at high risk with a long-term survival of <50% using the IPI, revised IPI, and National Comprehensive Cancer Network IPI is limited.^{4,5} Therefore, more accurate prognostic markers are essential to

identify patients at high risk of progression or relapse. In recent years, several studies have explored the potential of the baseline metabolic tumor volume (MTV) extracted from ^{18}F -fluorodeoxyglucose positron emission tomography-computed tomography (^{18}F -FDG-PET/CT) scans to predict the DLBCL outcome. The results consistently showed that MTV is inversely related to overall survival and progression-free survival (PFS).⁶⁻¹¹ Recently, a new international prognostic index (IMPI) incorporating MTV, age, and Ann Arbor stage was developed, thereby allowing improved individual outcome prediction.¹²

MTV reflects the ^{18}F -FDG-avid tumor burden but does not include phenotypical aspects such as the spatial distribution, heterogeneity, and shape of lesions. Recently developed quantitative ^{18}F - ^{18}F -FDG-PET/CT features, also referred to as radiomics, reveal the biological characteristics of the disease and could help to improve outcome prediction. Adding ^{18}F -FDG-PET radiomics features to the currently used predictors may improve the identification of patients with poor prognosis. Features quantifying dissemination, in particular, have shown high predictive value independent from MTV in DLBCL.^{11,13} Therefore, we previously developed a prediction model that incorporated MTV, the peak of the standardized uptake value (SUV_{peak}), the maximum distance between the largest lesion and any other lesion ($\text{D}_{\text{max,bulk}}$), World Health Organization (WHO) performance status, and age using data of the HOVON-84 trial.¹¹ The advantage of this model over other models using dichotomous cutoffs is that it allows for individual patient risk prediction and is less sensitive to data-driven cutoffs.

The objective of this study is to externally validate the clinical positron emission tomography (PET) model developed in the HOVON-84 trial¹¹ using 887 patients from the PETRA database and to compare the model performance of our clinical PET model with the currently used IPI.

Methods

Study population

Adult patients with de novo DLBCL ($n = 1466$) with a baseline ^{18}F -FDG-PET scan and 2-year follow-up data were included. Clinical data and [^{18}F]FDG-PET scans were collated and harmonized by the PETRA consortium.¹⁴ Patients were originally included in 7 individual studies: GSTT15,⁷ HOVON-84,¹⁵ HOVON-130,¹⁶ IAEA,¹⁷ NCRI,¹⁸ PETAL,¹⁹ and SAKK 38/07²⁰ (hereafter referred to as SAKK). Individual trials were approved by the institutional review board and all patients provided written informed consent. The use of all data within the PETRA imaging database was approved by the institutional review board of VU University Medical Center (JR/20140414).

^{18}F -FDG-PET/CT analysis

Scans did not pass quality control if (1) whole body ^{18}F -FDG-PET/CT scans were incomplete, (2) essential Digital Imaging and Communications in Medicine (DICOM) information was missing, (3) no FDG-avid lesions were present, and (4) plasma glucose levels and hepatic SUV_{mean} were outside the suggested ranges of the European Association of Nuclear Medicine.²¹ Scans were included when the hepatic SUV_{mean} was outside

the suggested ranges, but the total image activity was between 50% and 80% of the total injected activity.

Quantitative analysis of all ^{18}F -FDG-PET scans that passed quality control was performed using the ACCURATE tool.²² Lesions were delineated at baseline using a fully automated preselection defined by $\text{SUV} \geq 4.0$, and a volume threshold ≥ 3 mL.²³ Previous studies showed that an SUV threshold of 4.0 and a volume threshold of ≥ 3 mL resulted in the highest success rate and interobserver variability.^{23,24} Physiological uptake was deleted, and lymphoma lesions < 3 mL were added with single mouse clicks. The physiological uptake (eg, bladder and kidneys) adjacent to the tumor regions was removed manually. All scans were reviewed by a nuclear medicine physician who was blinded to the outcome. Delineations were performed by a nuclear medicine physician (GSTT15 and IAEA) or under the supervision of a nuclear medicine physician by trained researchers (with > 5 years of experience; HOVON-84, HOVON-130, PETAL, NCRI, and SAKK). We assessed the concordance of MTV between a nuclear medicine physician and a trained researcher for the SAKK study, and observed a correlation of 0.99.¹² To further harmonize quantitative ^{18}F -FDG-PET analysis between studies, all segmentations were visually checked for missed lesions or missed physiological uptake by a trained researcher before calculating the radiomics features. Based on these delineations, the MTV, SUV_{peak} ,²⁵ and $\text{D}_{\text{max,bulk}}$ were extracted for all patients. During model development using the HOVON-84 trial, we choose SUV_{peak} instead of SUV_{max} because the SUV_{peak} is relatively less sensitive to noise.²⁶ All image-processing and feature calculations were performed using RaCaT software,²⁷ which is in compliance with the imaging biomarker standardization initiative criteria.²⁸

Statistical analysis

Prediction models Multivariable logistic regression with backward feature selection was used to predict the risk of progression, relapse, or death after 2 years (2-year PFS) and the risk of progression or relapse after 2 years (2-year time to progression [TTP]). Follow-up started at the time of baseline [^{18}F]FDG-PET/CT scan. Patients who died within 2 years without signs of progression or relapse were excluded from the TTP prediction model.

We tested the predictive value of the following models:

1. IPI: the IPI risk score using low, low-intermediate, high-intermediate and high-risk groups.²
2. Clinical PET model as developed in the HOVON-84 trial: the natural logarithms of MTV and SUV_{peak} , the maximum distance between the largest lesion and any other lesion ($\text{D}_{\text{max,bulk}}$), WHO performance status, and age.¹¹

For the clinical PET model, the sum of individual predictors, weighted based on regression coefficients, together with the intercept of the model, were used to derive the predicted probability of an event for each patient. The model performance was assessed using the area under the curve (AUC) of the receiver operating characteristic curve. Differences between the model performances of prediction models, expressed as AUC, were assessed using the two-sided DeLong test.²⁹

Updating the model Ideally, a prediction model provides valid predictions of the outcome for individual patients in a setting other than that in which the model was developed. Recalibration methods for reestimating the coefficients of a model are attractive because of their stability. The validity of the model predictions can be assessed by comparing the observed outcomes and predictions when empirical data from this external setting are available,³⁰ which is the case now that we have 887 patients available from 6 external studies. We updated the model using all available data within the PETRA using logistic calibration. The intercept was updated to make the average predicted probability equal to the observed overall event rate (so-called calibration-in-the-large), and individual coefficients were reestimated.³⁰ Detection of calibration-in-the-large problems avoids miscalibration of the model and, consequently, wrong decision making.³⁰

Sensitivity analysis We assessed model performances among patients exclusively treated with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP). Secondly, we investigated the added value of the cell of origin (COO) to our clinical PET prediction model in a subset of patients with available COO information.

Furthermore, to compare the model performance of our clinical PET model with that of the IMPI model¹² and a model that combined MTV and WHO performance status (MTV/ECOG),³¹ we applied Cox regression models with a 2-year PFS as the outcome and assessed model performance, using the C-index and the Akaike information criteria.

Diagnostic performance To calculate the diagnostic performance of the models, high- and low-risk groups were defined. For the IPI prediction model, patients with 4 or 5

adverse factors were considered as high risk. For the clinical PET model, patients with the highest predicted probabilities were used to define the high-risk group. To allow comparison of the high-risk groups of the IPI and clinical PET models, the high-risk patient group for the clinical PET model was of equal size to the high-risk IPI group. The diagnostic performance of the prediction models was assessed using sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). For the Cox regression models, high-risk groups for the IMPI and clinical PET models were of equal size as the high-risk IPI group and the MTV/ECOG group with 2 risk points. Survival curves were obtained with Kaplan-Meier analyses, using the probabilities of the Cox regression models to create risk groups.

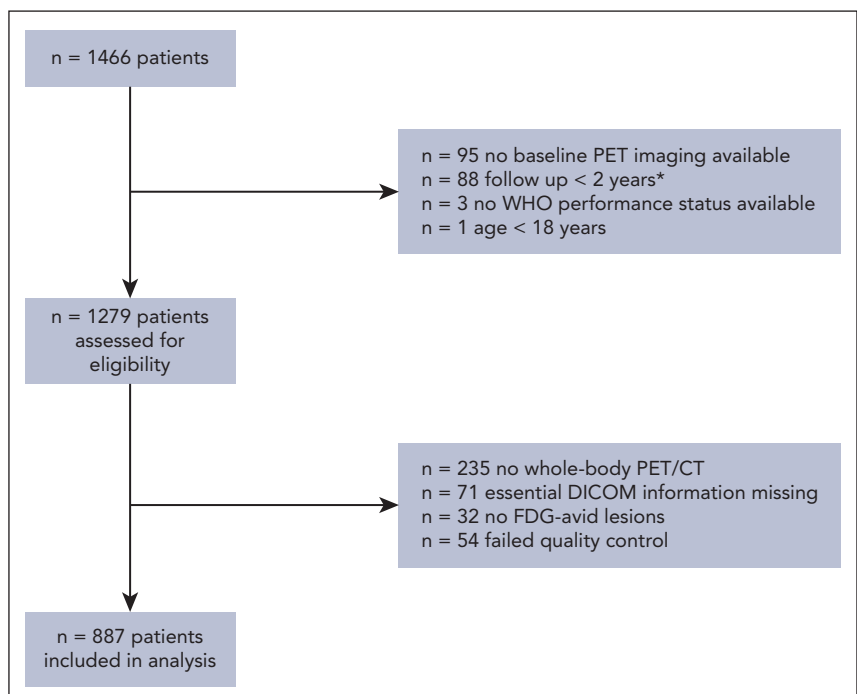
Statistical analysis was performed using R (version 4.2.1). $P < .05$ was considered statistically significant.

Results

Patient characteristics

A total 1466 eligible patients with de novo DLBCL from studies other than the HOVON-84 study were available in the PETRA database, of whom 887 were included in this analysis (Figure 1). Patients with no baseline ¹⁸F-FDG-PET imaging available (n = 95), who were lost to follow-up within 2 years and did not show any signs of progression (n = 88), aged <18 years (n = 1), and with missing WHO performance status (n = 3) were ineligible for this study. ¹⁸F-FDG-PET quality control led to the exclusion of patients with incomplete ¹⁸F-FDG-PET/CT scans (n = 235), missing essential DICOM information (n = 71), no ¹⁸F-FDG-avid lesions (n = 32), and scans outside the quality control range (n = 54). For the Cox regression models, patients who had a follow-up shorter than 2 years and an ¹⁸F-FDG-PET/CT scan that was within our quality control were included (n = 58).

Figure 1. CONSORT diagram of included patients for external validation. *Patients who were not included in the logistic regression model but were included in the Cox regression model.



Together with 308 patients from the HOVON-84 study, a total of 1195 patients were included in this analysis. Descriptive statistics of the baseline characteristics of all included patients stratified per the study are presented in Table 1. Two hundred and forty-one patients developed progression or relapse within 2 years after baseline ^{18}F -FDG-PET/CT, and 50 patients died within 2 years after baseline ^{18}F -FDG-PET/CT. The median baseline MTV of all patients was 324.4 mL (interquartile range [IQR], 81.7-828.8), with a median SUV_{peak} of 17.6 (IQR, 12.1-24.4) and a median $\text{D}_{\text{maxbulk}}$ of 22.2 cm (4.8-41.2; supplemental Table 1, available on the *Blood* website).

Prediction model

Using a 2-year PFS as the outcome, the AUC of the HOVON-84 trial was 0.67 for the IPI model and 0.75 for the clinical PET

model.¹¹ The IPI model yielded an AUC of 0.62 using all patients (Table 2; Figure 2). Within individual studies, the AUC of the IPI model ranged from 0.51 for the SAKK study to 0.65 for the PETAL study. The clinical PET model yielded an AUC of 0.71, which was significantly higher than that of the IPI model ($P < .001$). The AUC of the clinical PET model ranged between 0.59 for the HOVON-130 study to 0.75 for the PETAL study. For all individual studies, the AUC of the clinical PET model was higher than that of the IPI model, especially for the IAEA and SAKK studies.

Comparable results were obtained using a 2-year TTP as the outcome. The AUC of the HOVON-84 trial for IPI was 0.69, vs 0.79 for the clinical PET model. The IPI model yielded an AUC of 0.62, and the clinical PET model yielded an AUC of 0.71,

Table 1. Characteristics of included patients

	Total (n = 1195)	GSTT15 ⁷ (n = 97)	HOVON-130 ¹⁶ (n = 65)	HOVON-84 ¹⁵ (n = 308)	IAEA ¹⁷ (n = 104)	NCRI ¹⁸ (n = 133)	PETAL ¹⁹ (n = 368)	SAKK ²⁰ (n = 120)
Age (median, IQR)	62 (51-70)	61 (49-70)	63 (54-72)	65 (56-72)	57 (43-65)	61 (49-68)	61 (51-70)	59 (49-68)
>60 y	547 (46)	47 (48)	30 (46)	100 (32)	63 (61)	63 (47)	179 (49)	65 (54)
≤60 y	648 (54)	50 (52)	35 (54)	208 (68)	41 (39)	70 (53)	189 (51)	55 (46)
Ann Arbor stage								
I	108 (9)	9 (9)	0	0	11 (11)	8 (6)	66 (18)	14 (12)
II	284 (24)	20 (21)	7 (11)	55 (18)	25 (24)	51 (38)	80 (22)	42 (35)
III	269 (23)	11 (11)	8 (12)	70 (23)	23 (22)	35 (26)	75 (20)	26 (22)
IV	534 (45)	57 (59)	50 (77)	183 (59)	45 (43)	39 (29)	147 (40)	38 (32)
WHO performance status								
0	590 (49)	32 (33)	38 (58)	175 (57)	36 (35)	75 (56)	166 (45)	68 (57)
1	449 (38)	35 (36)	23 (35)	94 (31)	44 (42)	44 (33)	165 (45)	44 (37)
2	124 (10)	18 (19)	3 (5)	39 (13)	15 (14)	14 (11)	27 (7)	8 (7)
3	30 (3)	12 (12)	1 (2)	0	7 (7)	0	10 (3)	0
4	2	0	0	0	2 (2)	0	0	0
LDH								
≤ Normal	478 (40)	35 (36)	16 (25)	100 (32)	54 (52)	51 (38)	154 (42)	62 (52)
Normal	713 (60)	62 (64)	45 (69)	208 (68)	50 (48)	82 (62)	214 (58)	58 (48)
Missing	4		4 (6)					
Extranodal involvement								
≥1	773 (65)	47 (48)	30 (46)	182 (59)	67 (64)	106 (80)	249 (68)	92 (77)
<1	422 (35)	50 (52)	35 (54)	126 (41)	37 (36)	27 (20)	119 (32)	28 (23)
IPI low	368 (31)	26 (27)	9 (14)	51 (17)	44 (42)	52 (39)	125 (34)	61 (51)
Low-intermediate	264 (22)	10 (10)	14 (22)	75 (24)	16 (15)	28 (31)	97 (26)	24 (20)
High-intermediate	331 (28)	30 (31)	29 (45)	106 (34)	22 (21)	35 (26)	89 (24)	20 (17)
High	232 (19)	31 (32)	13 (20)	76 (25)	22 (21)	18 (14)	57 (15)	15 (13)

LDH, lactate dehydrogenase.

Table 2. AUCs of the IPI prediction model and clinical PET prediction models for all individual studies and all patients using 2-year PFS and 2-year TTP as the outcomes

Study name	2-y PFS		2-y TTP	
	IPI	Clinical PET	IPI	Clinical PET
HOVON-84 (test)	0.67	0.75	0.69	0.79
All patients	0.62	0.71	0.62	0.71
GSTT15	0.63	0.72	0.62	0.71
HOVON-130	0.53	0.59	0.53	0.60
IAEA	0.56	0.65	0.56	0.66
NCRI	0.56	0.71	0.59	0.70
PETAL	0.65	0.75	0.62	0.75
SAKK	0.51	0.71	0.51	0.70

when using all patients ($P < .001$). Again, for all individual studies, the AUCs of the clinical PET models were consistently higher than the AUCs of the IPI model.

Diagnostic performance

Patients at high risk according to the IPI model had a 2-year PFS probability of 61.4% (95% confidence interval [CI], 55.5-67.9; [Figure 3](#)). Patients at high risk according to the clinical PET model had a probability for 2-year PFS of 51.9% (95% CI, 45.9-58.7). The sensitivity, specificity, PPV, and NPV were higher for the clinical PET model than for the IPI model ([Table 3](#)). Specificity and NPV showed a small increase, but sensitivity increased from 29.5% to 39.0%, and PPV increased from 35.5% in the IPI model to 49.1% in the clinical PET model.

For 2-year TTP as the outcome, patients with high-risk IPI scores had a survival rate of 66.4% (95% CI, 60.3-73.0). Patients with

high-risk clinical PET scores had a survival rate of 55.5% (95% CI, 49.1-62.6). Again, sensitivity, specificity, PPV, and NPV were higher for the clinical PET than for the IPI model. The PPV increased from 33.7% to 44.6% in the clinical PET model compared with that in the IPI model.

Patients with 2 risk points in the MTV/ECOG model had a 2-year PFS of 62.8% (95% CI, 55.0-71.6; [Figure 4](#)), whereas patients at high risk according to the IMPI scores had a 2-year PFS of 59.1% (95% CI, 53.2-65.7). Patients at high risk according to the clinical PET model had the lowest survival rate, with a 2-year PFS of 51.9% (95% CI, 45.9-58.7). When using the same group sizes for the high-risk group as those of the patients with 2 risk points in the MTV/ECOG model, the 2-year PFS rates of the patients at high risk according to the IMPI scores were 55.2% (95% CI, 47.4-64.4) and 48.6% (95% CI, 40.8-57.9) using the clinical PET model, showing a clear superiority of both the IMPI and clinical

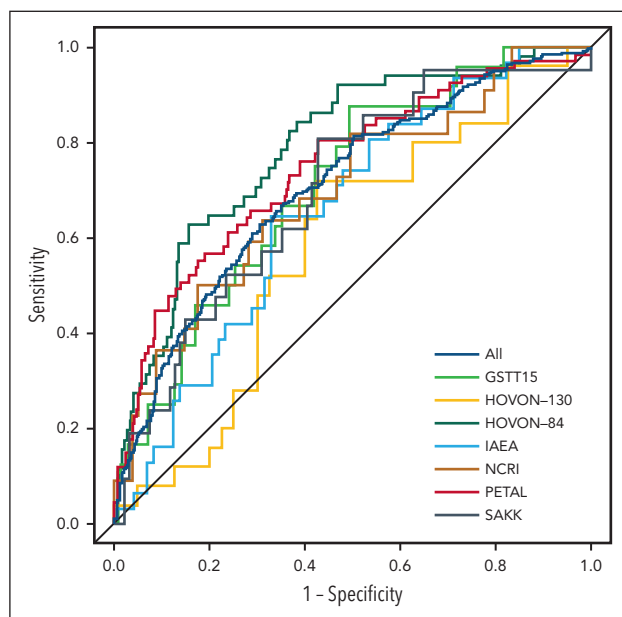


Figure 2. Receiver operating characteristic curves for 2-year PFS for all included patients and separate studies.

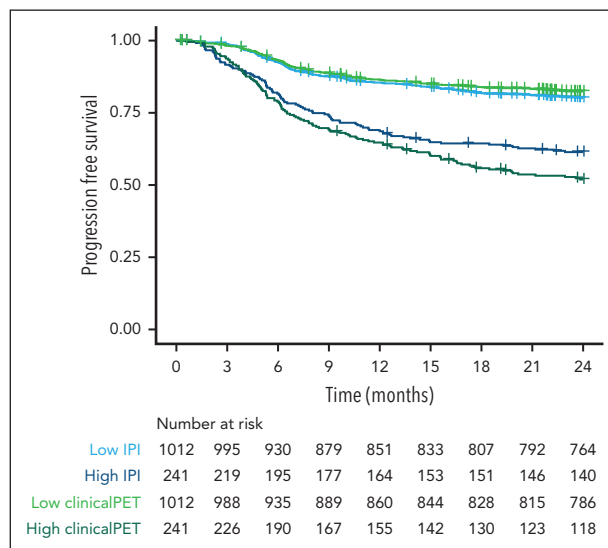


Figure 3. Survival curves of patients at high and low risk, as identified with IPI and clinical PET models, using 2-year PFS as the outcome.

Table 3. Diagnostic performance of the IPI and clinical PET models

		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
PFS	IPI	27.90 (22.69-33.59)	84.51 (81.99-86.81)	35.48 (30.13-41.23)	79.34 (78.02-80.59)
	Clinical PET	39.18 (33.53-45.04)	86.95 (84.57-89.08)	49.14 (43.65-54.65)	81.62 (80.14-83.01)
TTP	IPI	29.46 (23.78-35.65)	84.51 (81.99-86.81)	33.65 (28.36-39.38)	81.80 (80.48-83.05)
	Clinical PET	39.00 (32.81-45.47)	87.06 (84.69-89.18)	44.55 (38.93-50.31)	84.26 (82.83-85.59)

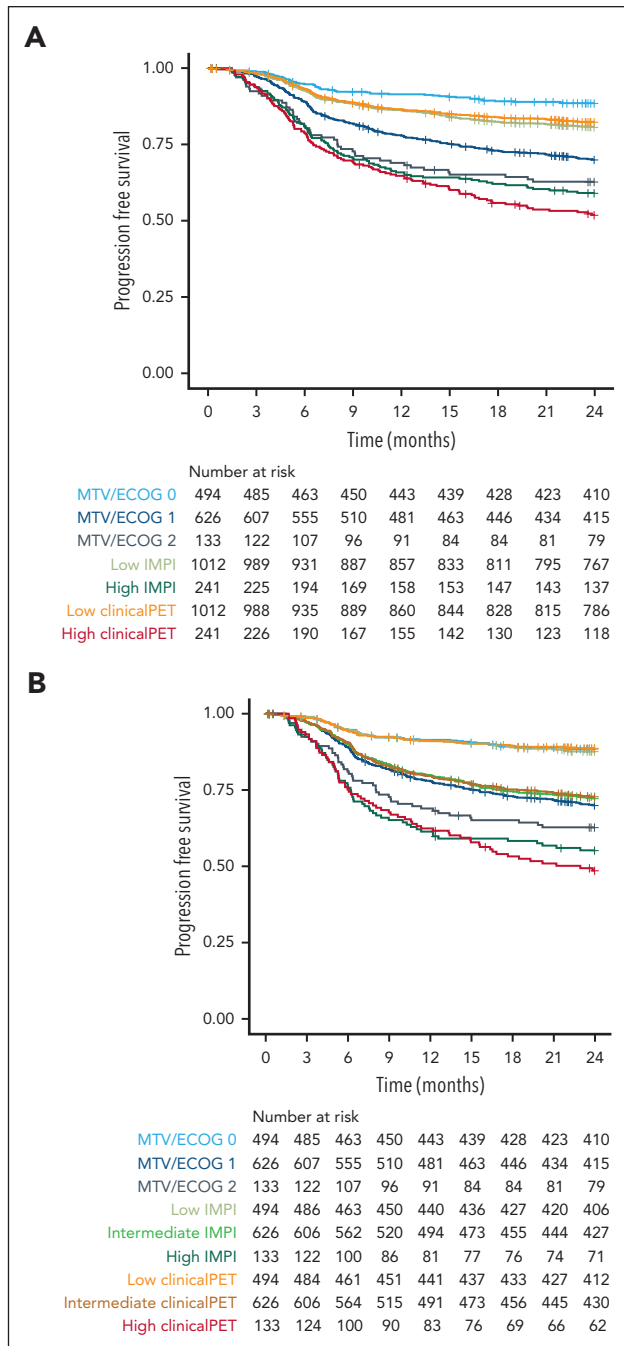


Figure 4. Survival curves of patients at high and low risk as identified with MTV/ECOG, IMPI, and clinical PET models using 2-year progression-free survival as outcome. (A) Risk groups of the MTV/ECOG as defined in the original publication³² and high-risk groups of the IMPI and clinical PET models are of equal size as the high-risk IPI group. (B) Risk groups for all models are of equal size as the MTV/ECOG groups.

PET model, with the best selection of patients at high risk by the clinical PET model, which is in line with the C-index and AIC values of the models.

Updating the model

After updating the model, its model performance (supplemental Table 2) and diagnostic performance (supplemental Table 3) were comparable with those of the original HOVON-84 model. For the GSTT, PETAL, and NCRI studies, the model performance slightly improved after calibration, whereas it decreased for the HOVON-130, IAEA, and SAKK studies. The diagnostic performance was slightly higher after model recalibration.

Sensitivity analysis

Similar results were obtained when only patients treated with R-CHOP were included (n = 1157 patients). The performance of the clinical PET model increased for the GSTT15, IAEA, and PETAL studies (supplemental Table 2). For both 2-year PFS and 2-year TTP, the AUC of IPI was 0.62, and that our clinical PET model was 0.71. A total of 493 patients had COO information available. In this subset, the COO was not a significant predictor of outcome after backward feature selection.

Furthermore, Cox regression modeling showed that model performance was highest for the clinical PET model (C-index, 0.69) and lowest for the MTV/ECOG model (C-index, 0.63); IMPI had a C-index of 0.66. Similar results were observed for the AIC (supplemental Table 4).

Discussion

Our study shows that the clinical PET model that was developed in the HOVON-84 trial remained predictive of outcome in 6 independent studies and had better model performance than the currently used IPI in all studies. Baseline ¹⁸F-FDG-PET clinical PET features were superior to IPI in identifying patients with high-risk DLBCL, with a relatively better model performance and higher PPV.

Several other studies have evaluated the predictive value of baseline radiomics features in DLBCL.^{11,32-38} Because of the different (numbers of) features that were extracted, it is hard to compare these studies directly. In general, all studies confirm that radiomics features are predictive of outcome. Moreover, previous studies showed that dissemination is a predictor of outcome independent of MTV.^{13,32} A recent study compared the 3 IPI variants in 2124 patients; according to the original IPI, patients had a 2-year PFS of almost 60%,⁵ which is comparable to the IPI performance in our study.

Cottreau et al³² published a risk stratification model that included the maximum distance between 2 lesions normalized for the body surface area (SDmax) and MTV in 301 patients. They showed that patients with both high MTV and SDmax had significantly lower survival rates, with a 2-year PFS of ~50%. These results are comparable with our results, given that we reported a 2-year PFS of 51.9% in the high-risk group. Both high-risk groups included ~20% of the patients. However, it should be noted that they applied a different segmentation method to delineate lesions, which could probably explain the lower median MTV (253 mL vs 324.4 mL) and hampers direct comparison of their model to ours, because multiple studies have shown large differences in extracted MTVs using the SUV4.0 or 41% max segmentation methods.^{6,24,39} Previous analysis in the HOVON-84 study showed that correction of $D_{max_{bulk}}$ for height did not influence our model performance.¹¹ Moreover, the advantage of our clinical PET model is that it allows individual patient risk prediction because MTV and $D_{max_{bulk}}$ are included as continuous variables. Therefore, it is less influenced by data-driven optimal cutoffs. A dichotomous cutoff results in different survival estimates for MTV and SDmax values that are close to the cutoffs, whereas the actual survival is similar and more accurately predicted with our clinical PET model.

Kostakoglu et al⁴⁰ recently published a radiomics prediction model based on 1263 patients from the GOYA trial. Patient characteristics were comparable, although their study included patients with slightly more advanced-stage diseases (84% vs 68%, respectively), and our study included more patients with high-risk IPI (15% vs 19%, respectively). Although their model performance was lower (AUC 0.64), the patients at high risk (33% of the total population), which their random forest prediction model identified, had a 2-year PFS of ~50%. In this study, 42 radiomics features were used. In addition to the MTV, 7 textural features were included in the final random forest model. Textural features are sensitive to different acquisition, reconstruction, and segmentation methods,^{39,41,42} leading to limited reproducibility in multicenter, multivendor studies, which was the case for 5 out of the 7 textural features included in their prediction model.⁴² Moreover, interpretation of these textural features is complex. Contrary to textural radiomics features, dissemination features are easy to interpret because they quantitatively reflect what can be visualized using ¹⁸F-FDG-PET/CT scans. They are also relatively simple to calculate and are relatively insensitive to scan protocol differences.

The recently published IMPI included Ann Arbor stage, age, and MTV.¹² In our clinical PET model, Ann Arbor stage is replaced by $D_{max_{bulk}}$ and WHO performance status. Both IMPI and clinical PET models allow individual risk prediction. Looking at the 2-year PFS rates, the clinical PET model outperformed both IMPI and MTV/ECOG prediction models.

None of the previously described prognostic models reported the PPV, NPV, sensitivity, and specificity; therefore, we cannot compare the diagnostic measures of these radiomics models with those of our clinical PET model. The high-risk groups in all the mentioned prediction models and our clinical PET model had a survival rate of ~50%, indicating that none of the indices identified a truly high-risk group. There is an unmet need to

identify patients with high-risk DLBCL shortly after diagnosis. Therefore, the identification of robust and easy-to-use biomarkers for the early identification of patients at high risk in this patient group is essential. Although not perfect, the clinical PET model is the best we have to select patients at high risk with limited additional costs and limited additional time because, on an average, MTV can be calculated for patients within 3 to 6 minutes, taking up to 10 to 20 minutes for complex cases.⁴³

The focus of a validation study should not be on the statistical testing of differences in performance but on the generalizability of the model in other settings.^{44,45} A prediction model ideally provides valid predictions of outcomes for individual patients in real life. Our study showed that our clinical PET model was generalizable because it remained predictive of outcome in all external studies, which were clinical cohorts of unselected patients that can represent real-life settings. After updating the model (ie, recalibration of the intercept and coefficients), comparable model and diagnostic performances were confirmed. However, case-mix differences between individual studies were present regarding patient characteristics, outcome, treatment, and ¹⁸F-FDG-PET parameters. This led to different model performances between studies for both IPI and clinical PET model. This is most prominent in HOVON-130, a study with most aberrant patient and ¹⁸F-FDG-PET characteristics, compared with other studies, because it only included patients with *MYC* gene rearrangements, and a subgroup of these patients showed poor survival rates irrespective of disease burden quantified based on radiomics features.⁴⁶ The SAKK study mainly included patients at low risk, which led to poor performance of the IPI risk score. However, our clinical PET model was still able to accurately predict the outcome for these patients at low risk. The patient characteristics in Table 1 show that the NCRI and SAKK studies included relatively more patients at limited stages, whereas the HOVON-130, HOVON-84, and GSTT15 studies included more patients at advanced stages. These differences were also visible in the IPI score. These case-mix differences are more pronounced when the sample sizes are relatively small, which is the case for the GSTT15, HOVON-130, IAEA, NCRI, and SAKK studies. The uncertainty of the model increases, leading to a large range of CIs,⁴⁷ possibly explaining the large variation in model performance. Regardless of these case-mix differences, the model performances of the clinical PET model always outperformed those of the IPI model. This led to a more accurate selection of patients at high risk, as shown by the decrease of 10% (IPI, 61.4% vs 51.9% for clinical PET model) in the survival for the high-risk group and an increase of 14% (35.5 vs 49.1 respectively) for the PPV (compared with the IPI model).

Significant efforts have been made to standardize ¹⁸F-FDG-PET scanning, including initiatives by the European Association for Nuclear Medicine Research Limited and the US Society of Nuclear Medicine.^{48,49} However, the absence of a standardized methodology has hampered the use of quantitative PET parameters in clinical practice. However, multiple vendors of ¹⁸F-FDG-PET systems have implemented algorithms to calculate the MTV. Currently, dissemination features are included only in the context of the research. However, these features are relatively insensitive to differences in segmentation methods, acquisition, and reconstruction^{39,42} and are relatively simple to calculate. Therefore, implementation of the calculation of these

radiomics features should be feasible in a reproducible manner in most clinical PET centers. We expect and hope that vendors will implement the calculation of radiomics features in their software in the foreseeable future, once more evidence on their clinical value becomes apparent. In the meantime, our image analysis tool, ACCURATE, is provided as an open tool to facilitate research use.

This study has several strengths. By applying 2 risk scores to the same individual patient data from high-quality studies, this analysis allowed for the direct comparison of risk indices. Furthermore, the applied PET quality control criteria and uniform analysis of the baseline ^{18}F -FDG-PET/CT scans resulted in the inclusion of high-quality PET data. Moreover, survival data were harmonized by recalculating the follow-up between the original studies. We decided to truncate survival at 2 years because the most clinically relevant events occurred during this period. An individual patient data analysis reported that patients who are alive without progression at 2 years have similar survival rates as the age-, sex-, and country-matched population 7 years after this time.⁵⁰ A limitation of our study was that for some patients included in the PETRA database, the baseline ^{18}F -FDG-PET/CT scan was either not performed or performed on a PET-only system (235 out of 392). Therefore, not all patients were included in the post hoc analysis. However, we believe that for prospective trials, fewer patients will be excluded because of insufficient PET quality, given that there is increased awareness of scanning and anonymization procedures compared with the timeframe when prospective clinical trials were performed. Furthermore, we decided to include TTP as an outcome parameter, because PFS and overall survival are affected by aging.⁶ The outcome of older patients is determined not only by lymphoma but also by age-related comorbidities, adverse treatment effects, and limited life expectancy in general. Lastly, although most patients were treated with R-CHOP, differences in treatment regimens between studies existed with regard to the number of cycles and intensification of treatment.

In conclusion, the clinical PET model that was developed in the HOVON-84 data set remained predictive of outcome in 6 independent studies and had a better model performance than the currently used IPI risk score in all studies. Therefore, baseline ^{18}F -FDG-PET radiomics features can be used to select patients at high risk more accurately than the IPI model, given its relatively higher model performance and PPV.

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 ^{18}F -FDG-PET/CT scans for individual studies.

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). S.F.B. 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/EPSCRC Centre for Medical Engineering at King's College London (WT203148/Z/16/Z) and the National Institute for Health and Care Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' National Health Service (NHS) Foundation Trust and King's College London and the NIHR Clinical Research Facility.

The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or the Department of Health and Social Care.

Authorship

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

Conflict-of-interest disclosure: S.F.B. received departmental funding from Amgen, AstraZeneca, BMS, Novartis, Pfizer and Takeda. M.E.D.C. received financial support for the clinical trials from Celgene, BMS and Gilead. J.M.Z. received 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 the supplemental Appendix.

ORCID profiles: J.J.E., [0000-0002-6094-0016](https://orcid.org/0000-0002-6094-0016); G.J.C.Z., [0000-0002-9571-9362](https://orcid.org/0000-0002-9571-9362); M.W.H., [0000-0002-3889-0921](https://orcid.org/0000-0002-3889-0921); S.P., [0000-0002-7897-6383](https://orcid.org/0000-0002-7897-6383); S.E.W., [0000-0003-3698-8121](https://orcid.org/0000-0003-3698-8121); U.D., [0000-0002-4034-9472](https://orcid.org/0000-0002-4034-9472); A.H., [0000-0003-2230-3873](https://orcid.org/0000-0003-2230-3873); L.K., [0000-0002-3396-4880](https://orcid.org/0000-0002-3396-4880); P.J.L., [0000-0002-6735-8651](https://orcid.org/0000-0002-6735-8651); S.F.B., [0000-0002-2516-5288](https://orcid.org/0000-0002-2516-5288); N.G.M., [0000-0003-0359-0328](https://orcid.org/0000-0003-0359-0328); L.C., [0000-0002-6371-097X](https://orcid.org/0000-0002-6371-097X); E.Z., [0000-0002-5522-6109](https://orcid.org/0000-0002-5522-6109); S.C., [0000-0002-7679-3137](https://orcid.org/0000-0002-7679-3137); T.G., [0000-0002-8772-9931](https://orcid.org/0000-0002-8772-9931); M.E.D.C., [0000-0002-0123-9182](https://orcid.org/0000-0002-0123-9182); O.S.H., [0000-0003-0767-2734](https://orcid.org/0000-0003-0767-2734); H.C.W.d.V., [0000-0002-5454-2804](https://orcid.org/0000-0002-5454-2804); R.B., [0000-0002-0313-5686](https://orcid.org/0000-0002-0313-5686); J.M.Z., [0000-0003-1074-5922](https://orcid.org/0000-0003-1074-5922).

Correspondence: J. J. Eertink, Department of Hematology, Amsterdam UMC, location VUmc, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands; email: j.eertink@amsterdamumc.nl.

Footnotes

Submitted 27 September 2022; accepted 27 February 2023; prepublished online on *Blood* First Edition 31 March 2023. <https://doi.org/10.1182/blood.2022018558>.

All data are available on request from the corresponding author, J. J. Eertink (j.eertink@amsterdamumc.nl). Deidentified individual participant data can be requested through the PETRA consortium request platform at <https://petralymphoma.org> (petra@amsterdamumc.nl).

The online version of this article contains a data supplement.

There is a [Blood Commentary](#) on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

REFERENCES

- Crump M, Neelapu SS, Farooq U, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. *Blood*. 2017;130(16):1800-1808.
- 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.
- Habermann TM, Weller EA, Morrison VA, et al. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. *J Clin Oncol*. 2006;24(19):3121-3127.
- Gleeson M, Counsell N, Cunningham D, et al. Prognostic indices in diffuse large B-cell lymphoma in the rituximab era: an analysis of the UK National Cancer Research Institute R-CHOP 14 versus 21 phase 3 trial. *Br J Haematol*. 2021;192(6):1015-1019.
- Ruppert AS, Dixon JG, Salles G, et al. International prognostic indices in diffuse large B-cell lymphoma: a comparison of IPI, R-IPI, and NCCN-IPI. *Blood*. 2020;135(23):2041-2048.
- 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.
- 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.
- 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.
- Sasanelli M, Meignan M, Haioun C, et al. Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41(11):2017-2022.
- Cottreau 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.
- 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.
- 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.
- Cottreau 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Boellaard R. Quantitative oncology molecular analysis suite: ACCURATE. *J Nucl Med*. 2018;59(suppl 1):1753.
- 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.
- 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.
- Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009;50(suppl 1):122S-150S.
- Kaalep A, Burggraaff CN, Pieplenbosch S, et al. Quantitative implications of the updated EARL 2019 PET-CT performance standards. *EJNMMI Phys*. 2019;6(1):28.
- Pfaehler E, Zwanenburg A, de Jong JR, Boellaard R. An open source and easy to use radiomics calculator tool. *PLoS One*. 2019;14(2):e0212223.
- 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.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.
- Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating. Statistics for biology and health*, 2197-5671. Springer; 2019.
- Thieblemont C, Chartier L, Duhrsen U, et al. A tumor volume and performance status model to predict outcome before treatment in diffuse large B-cell lymphoma. *Blood Adv*. 2022;6(23):5995-6004.
- Cottreau 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. *Ann Oncol*. 2021;32(3):404-411.
- 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.
- Senjo H, Hirata K, Izumiya 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.
- 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.
- Frood R, Clark M, Burton C, et al. Discovery of pre-treatment FDG PET/CT-derived radiomics-based models for predicting Outcome in diffuse large B-cell lymphoma. *Cancers (Basel)*. 2022;14(7):1711.
- Jiang C, Li A, Teng Y, et al. Optimal PET-based radiomic signature construction based on the cross-combination method for

- predicting the survival of patients with diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2022;49(8):2902-2916.
38. Zhang X, Chen L, Jiang H, et al. A novel analytic approach for outcome prediction in diffuse large B-cell lymphoma by [(18)F]FDG PET/CT. *Eur J Nucl Med Mol Imaging*. 2022; 49(4):1298-1310.
 39. 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.
 40. Kostakoglu L, Dalmaso F, Berchiolla P, et al. A prognostic model integrating PET-derived metrics and image texture analyses with clinical risk factors from GOYA. *EJHaem*. 2022;3(2):406-414.
 41. Pfaehler E, Beukinga RJ, de Jong JR, et al. Repeatability of (18) F-FDG PET radiomic features: a phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method. *Med Phys*. 2019;46(2):665-678.
 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. 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 Imaging*. 2018;45(7):1142-1154.
 44. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5): 441-447.
 45. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247.
 46. Eertink JJ, Zwezerijnen GJ, Wiegers SE, et al. Baseline radiomics features and MYC rearrangement status predict progression in aggressive B-cell lymphoma. *Blood Adv*. 2023;7(2):214-223.
 47. 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.
 48. Sunderland JJ, Christian PE. Quantitative PET/CT scanner performance characterization based upon the society of nuclear medicine and molecular imaging clinical trials network oncology clinical simulator phantom. *J Nucl Med*. 2015;56(1): 145-152.
 49. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *Eur J Nucl Med Mol Imaging*. 2017;44(suppl 1): 17-31.
 50. Maurer MJ, Habermann TM, Shi Q, et al. Progression-free survival at 24 months (PFS24) and subsequent outcome for patients with diffuse large B-cell lymphoma (DLBCL) enrolled on randomized clinical trials. *Ann Oncol*. 2018;29(8):1822-1827.

© 2023 by The American Society of Hematology. Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.