Received: 15 July 2021 | Revised: 1 October 2021 | Accepted: 4 October 2021

DOI: 10.1002/hon.2935

ORIGINAL ARTICLE

WILEY

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

Luca Ceriani[1,2] | Lisa Milan[1] | Luciano Cascione[2,3] | Giuseppe Gritti[4] | Federico Dalmasso[5] | Fabiana Esposito[6] | Maria Cristina Pirosa[6] | Sämi Schär[7] | Andrea Bruno[8] | Stephan Dirnhofer[9] | Luca Giovanella[1,10] | Stefanie Hayoz[7] | Christoph Mamot[11] | Alessandro Rambaldi[4,12] | Stephane Chauvie[5] | Emanuele Zucca[2,6,13]

[1]Nuclear Medicine and PET/CT Centre, Imaging Institute of Southern Switzerland, Ente Ospedaliero Cantonale, Bellinzona, Switzerland

[2]Faculty of Biomedical Sciences, Institute of Oncology Research, Università della Svizzera Italiana, Bellinzona, Switzerland

[3]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

[4]Hematology Unit, Azienda Ospedaliera Papa Giovanni XXIII, Bergamo, Italy

[5]Medical Physics Unit, Santa Croce e Carlo Hospital, Cuneo, Italy

[6]Medical Oncology, Oncology Institute of Southern Switzerland, Ente Ospedaliero Cantonale, Bellinzona, Switzerland

[7]Swiss Group for Clinical Cancer Research (SAKK) Coordinating Center, Bern, Switzerland

[8]Department of Nuclear Medicine, Azienda Ospedaliera Papa Giovanni XXIII, Bergamo, Italy

[9]Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Switzerland

[10]Department of Nuclear Medicine, University Hospital Zurich and University of Zurich, Zurich, Switzerland

[11]Division of Oncology, Cantonal Hospital Aarau, Aarau, Switzerland

[12]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy

[13]Department of Medical Oncology, Bern University Hospital and University of Bern, Bern, Switzerland

Correspondence
Luca Ceriani, Nuclear Medicine and PET/CT Centre, Imaging Institute of Southern Switzerland, Ospedale San Giovanni, CH-6500 Bellinzona, Switzerland.
Email: luca.ceriani@eoc.ch

Funding information
Ente Ospedaliero Cantonale, Grant/Award Number: ABREOC 22008-262; Amgen; Oncosuisse, Grant/Award Number: OCS-02270-08-2008

Abstract

Functional parameters from positron emission tomography (PET) seem promising biomarkers in various lymphoma subtypes. This study investigated the prognostic value of PET radiomics in diffuse large B-cell lymphoma (DLBCL) patients treated with R-CHOP given either every 14 (testing set) or 21 days (validation set). Using the PyRadiomics Python package, 107 radiomics features were extracted from baseline PET scans of 133 patients enrolled in the Swiss Group for Clinical Cancer Research 38/07 prospective clinical trial (SAKK 38/07) [ClinicalTrial.gov identifier: NCT00544219]. The international prognostic indices, the main clinical parameters and standard PET metrics, together with 52 radiomics uncorrelated features (selected using the Spearman correlation test) were included in a least absolute shrinkage and selection operator (LASSO) Cox regression to assess their impact on progression-free (PFS), cause-specific (CSS), and overall survival (OS). A linear

Luca Ceriani, Lisa Milan, and Luciano Cascione contributed equally to this work.

combination of the resulting parameters generated a prognostic radiomics score (RS) whose area under the curve (AUC) was calculated by receiver operating characteristic analysis. The RS efficacy was validated in an independent cohort of 107 DLBCL patients. LASSO Cox regression identified four radiomics features predicting PFS in SAKK 38/07. The derived RS showed a significant capability to foresee PFS in both testing (AUC, 0.709; $p < 0.001$) and validation (AUC, 0.706; $p < 0.001$) sets. RS was significantly associated also with CSS and OS in testing (CSS: AUC, 0.721; $p < 0.001$; OS: AUC, 0.740; $p < 0.001$) and validation (CSS: AUC, 0.763; $p < 0.0001$; OS: AUC, 0.703; $p = 0.004$) sets. The RS allowed risk classification of patients with significantly different PFS, CSS, and OS in both cohorts showing better predictive accuracy respect to clinical international indices. PET-derived radiomics may improve the prediction of outcome in DLBCL patients.

KEYWORDS
18FDG-PET/CT, DLBCL, PET metrics, prognostic models, radiomics

1 | INTRODUCTION

The 18F-fluorodeoxyglucose (18FDG) positron emission tomography/computed tomography (PET/CT) represents the current standard procedure for staging and therapeutic response assessment in diffuse large B cell lymphoma, not otherwise specified (DLBCL, NOS), the most common lymphoma subtype.[1–3] The addition of the anti-CD20 monoclonal antibody, rituximab to the standard CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone) chemotherapy regimen, given either every 2 (R-CHOP14) or 3 (R-CHOP21) weeks, has improved the outcome in patients with DLBCL.[4,5] However, a relevant portion of patients (30%–40%) still experiences front-line treatment failure and has a very poor prognosis.[1] Three prognostic indices incorporating simple clinical features (age, lactate dehydrogenase, number/sites of involvement, stage, and performance status) are widely used for DLBCL: the International Prognostic Index (IPI), the revised IPI (R-IPI), and the National Comprehensive Cancer Network IPI (NCCN-IPI).[6] In the rituximab era, however, none of these clinical risk scores have been able to consistently identify the patients with very short survival[6]: the best predictor of DLBCL outcome is the metabolic response on the end-of-treatment PET/CT scans.[2] Hence, the precise and early identification of high-risk patients who could benefit from different therapeutic approaches remains an unmet need.

Functional parameters derived from baseline PET/CT—including the maximum standardized uptake value (SUVmax), the total lesion glycolysis (TLG), and the metabolic tumor volume (MTV)—seem promising biomarkers in various lymphoma subtypes,[7–11] including DLBCL.[12–15] There are also emerging data suggesting that both the metabolic heterogeneity (MH) and the maximum distance between 18FDG-avid lesions (D_{max}) on PET/CT baseline scans may be associated with lymphoma survival.[16–19] We and others have previously shown that models based on MTV at baseline enable outcome prediction in DLBCL and that the combination of baseline PET metrics and early response on interim PET/CT as well as the integration of quantitative PET parameters with clinical and biological features might lead to improve prognostic models.[10,15,20–26] However, the lack of standardized methods for segmentation, acquisition, and processing have thus far prevented the quantitative applications of PET/CT in the clinical setting and further study and validation will be necessary before they can be integrated into risk stratification systems for either clinical trials or daily clinical practice.[3]

Radiomics encompasses the high-throughput extraction of a wide variety of quantitative data from digital medical images.[27] In fact, several additional features beyond MTV and TLG may originate from PET/CT scans, describing the lesion textural pattern, the radiotracer intensity, tumor heterogeneity, and shape and may reflect biological characteristics, such as proliferation rate, hypoxia, necrosis, and angiogenesis.[28,29]

The main aim of the present post-hoc investigation is to explore whether PET radiomics may improve prediction of progression-free survival (PFS) and risk stratification in DLBCL patients treated with R-CHOP.

2 | METHODS

2.1 | Ethics approval

The SAKK 38/07 clinical trial of the Swiss Group for Clinical Cancer Research (SAKK) trial was conducted in accordance with the precepts of the 1964 Helsinki Declaration and its later amendments. The study was authorized at the Oncology Institute of Southern Switzerland by the local Cantonal Ethics Committee (CE-TI-1981) and approved by the Institutional Review Board/Ethics Committee of each participating center. This trial was registered at www.clinicaltrials.gov as #NCT00544219.

2.2 | Testing set

Among the 156 patients with any stage of untreated DLBCL prospectively enrolled in the SAKK 38/07 trial,[30] 133 with baseline 18FDG-PET/CT scans suitable for radiomics analysis and with a complete clinical follow-up were eligible for this study. Central pathology review confirmed DLBCL diagnoses; the presence of cases with BCL2 and/or BCL6 and C-MYC rearrangement (now called "High-grade B-cell lymphoma with double/triple hit") was ruled out by in situ hybridization analyses.[31] All patients received six cycles of R-CHOP14 (rituximab 375 mg/m^2, cyclophosphamide 750 mg/m^2, doxorubicin 50 mg/m^2, vincristine 1.4 mg/m^2 on day 1, and prednisone 100 mg/m^2 for 5 days, every 14 days) with growth factor support, followed by two cycles of rituximab. Based on local guidelines, 20 (15%) patients had consolidation radiotherapy.[30]

2.3 | Validation set

The validation set comprised 107 patients with any stage of DLBCL previously enrolled in a retrospective PET survey (ABREOC 02-008 cohort). Seventy-one patients were obtained from the database of the Oncology Institute of Southern Switzerland, Bellinzona, Switzerland, and 36 from that of the Hematology Unit of the Ospedale Papa Giovanni XXIII, Bergamo, Italy. Age >18 years, 18FDG-PET/CT baseline scan suitable for radiomics analysis, front-line treatment with six cycles of R-CHOP21 (same doses as R-CHOP14, given every 21 days), and a clinical follow-up of at least 18 months were the criteria of selection. Thirteen patients (12%) received consolidation radiotherapy. More detailed descriptions of these patient cohorts have been previously published.[20,30]

2.4 | Clinical prognostic factors

The clinical parameters included in the analysis were age, performance status (PS), extranodal site number (EN), level of lactate dehydrogenase (LDH), stage of the disease, and the cell of origin (COO). Moreover, the International Prognostic Index (IPI), revised-IPI (R-IPI) and National Comprehensive Cancer Center Network enhanced-IPI (NCCN-IPI) were calculated for each patient.[6] The clinical information was retrieved from the SAKK 38/07 study database or from their hospital records, as appropriate. The COO was centrally determined by immunohistochemistry using the Hans algorithm.[32]

2.5 | PET/CT images analysis and radiomics features extraction

PET/CT scan protocol details have been previously reported.[20,21] All scans were centrally evaluated with a standard protocol and a dedicated imaging software (PET-Encore, MIM software Inc. version 6.7.10) workstation. All the lesions of each patient were automatically segmented by fixing a SUV value of 4 as threshold for MTV estimation[33,34]; the segmentation results were visually checked by an expert nuclear physician and, if necessary, manually corrected. A 3D contoured map including all the segmented lesions was generated for each patient and exported for the further radiomics analysis of the whole tumor volume.

The mining of radiomics features (RF) from each PET/CT study was performed by the PyRadiomics open-source Python package (version 2.2.0).[35] RFs were extracted from all the segmented lesions included in the contour map. Patients with a total MTV <5 ml (irrespective of the number of lesions) were excluded from the analysis to avoid undersampling and minimize the statistical fluctuation related to the poor information achievable from small volumes.[36] To standardize this process, gray-level intensities and voxel dimensions of the original images were preliminary resampled following the Image Biomarkers Standardization Initiative (IBSI)[37] and 107 3D standardized features evaluating different characteristics of the tumor lesions were initially extracted. The details of the extraction settings as well as the complete list of features is reported in Supplementary materials (Tables S1 and S2). In addition to the calculated variables, we have also evaluated the maximum distance between two points belonging to the MTV normalized by the body surface area (Shape_Maximum3DDiameter/BSA) since it was demonstrated to be a promising prognostic parameter in DLBCL.[17]

2.6 | Radiomics features analysis

Each RF was first standardized using the Z-score (with $Z = (\mu - \bar{\mu})/\sigma$, where μ, $\bar{\mu}$, and σ are the value, the mean, and the standard deviation, respectively).[38] Considering RF redundancy, we used Spearman correlation test ($p < 0.05$) with false discovery rate (FDR) correction in order to reduce the overfitting problem and to discard features highly correlated to SUVmax and MTV (correlation coefficient > 0.7).[39]

We use the least absolute shrinkage and selection operator (LASSO) Cox regression (with 10-fold cross validation) to select the most informative prognostic parameters of PFS in the testing cohort[40] (Figure S2). In this analysis, the main clinical parameters (PS, EN, age, LDH, COO, stage) and prognostic indices (IPI, R-IPI, NCCN-IPI) as well as TLG (being the other main PET metrics, namely SUVmax, SUVmean and MTV comprised among the first-order and shape Pyradiomics parameters) were added to RFs.

2.7 | Statistics

Continuous variables were expressed as median and interquartile range (IQR) and their distribution between groups was compared using the Mann–Whitney *U*-test. Differences between the frequencies of categorical data were assessed with the chi-square test or the Fisher exact test, as applicable.

Receiver operating characteristic (ROC) curves for PFS and for cause-specific survival (CSS) and overall survival (OS) were generated on both the testing and validation datasets to define the area under the curve (AUC) of RS.

PFS was defined as the time from study entry until lymphoma progression or death from any cause; CSS as the time from entry to death from lymphoma or its treatment and OS from entry to death from any cause.[41] Survival endpoints were estimated by Kaplan–Meier analysis and survival curves compared by log-rank testing.

The predictive accuracy of the different prognostic was compared by a concordance probability estimate (CPE), using the Harrell's C statistic, which indicates the probability that the patient with the shorter survival time also has the higher predicted risk.[42] The indices' relative quality was further assessed using the Akaike information criterion (AIC), an in-sample fit approach to model selection.[43]

A p-value <0.05 was considered statistically significant. Statistical analyses were conducted using the STATA 16.1 statistical software: (StataCorp) and the RStudio 1.2.1335 software package (RStudio, PBC) with R version 4.0.3,[44] as appropriate.

3 | RESULTS

Detailed clinical features and outcome of the patients enrolled in the SAKK 38/07 trial and in the validation cohort (ABREOC 02-08) have been published previously.[20,30] Table 1 summarizes the main baseline demographic, clinical characteristics, and PET metrics of patients included in the present analysis. At a median follow-up of 64 months (IQR, 60–67 months), 28 progressions of disease and 21 deaths (of which 18 were due to lymphoma) were recorded in the testing set. The estimated 5-year PFS, CSS, and OS rates were 79% (95% CI, 71%–85%), 87% (95% CI, 80%–92%) and 85% (95% CI, 78%–90%), respectively (see Figure S1).

Compared with those of the testing set, the patients in the validation set showed significantly older age, a lower proportion of patients with good-risk according to the IPI and NCCN-IPI (Table 1) and a shorter follow-up time (36 months; IQR, 27–55). Twenty-one patients had disease progressions and 19 died (of which 14 due to lymphoma) in the validation cohort. Their estimated 5-year PFS, CSS, and OS rates were 75% (95% CI, 63%–84%), 80% (95% CI, 66%–89%) and 74% (95% CI, 60%–83%), respectively (see Figure S1).

Figure 1 describes the different steps in our study summarizing the procedures leading to the construction of the radiomics signature standard score (RS) and the subsequent development of a radiomics-based prognostic model (RS index).

Among the 108 RFs originally extracted, 56 resulted to be highly correlated to SUVmax or MTV in the testing set and therefore we removed them from subsequent analysis (detailed list in the Table S2). The LASSO Cox regression (including the remaining 52 RFs, TLG and the main clinical features and prognostic indices) identified four non-redundant predictors of PFS (Table 2). All these selected variables are RFs. The Shape_Maximum3DDiameter/BSA is a geometric index of disease dissemination: it measures the distance between the two farthest lesions (or the maximum diameter when disease is characterized by a single localization). The other three RFs (GLCM_SumOfSquares, GLSZM_GrayLevelNonUniformityNormalized and GLDM_GrayLevelVariance) are descriptors of the lesion heterogeneity. They characterize the spatial distribution of image intensities, namely the variation of SUV, using different mathematical operators (see supporting information S1).

The regression coefficients of the selected RFs are reported in Table 2. LASSO regression discarded all clinical variables, prognostic indexes and standard PET metrics.

In the testing set, the continuous RS obtained from the linear combination of the 4 LASSO regression-selected RFs showed a significant ability to predict PFS (AUC, 0.709; 95% CI, 0.623–0.784; $p < 0.001$) as well as CSS (AUC, 0.721; 95% CI, 0.636–0.795; $p < 0.001$) and OS (AUC, 0.740; 95% CI, 0.657–0.812; $p < 0.001$).

The distribution of RS values is shown in Figure 2A. The stratification of patients according to quartiles identified subgroups of patients with a progressively shorter PFS (Figure 2B). RS was poorly correlated with MTV ($r = 0.39$) and maintained its prognostic value (HR, 2.1; 95% CI, 1.4–3.2) after controlling for MTV in multivariable Cox regression. Figure 2C provides some clinical examples of the lack of correlation between RS and MTV.

Kaplan–Meier analysis demonstrated very similar outcomes for the patients in the first and second quartiles, who had a significantly better survival than those with RS values above the median. The patients in the third and fourth quartile displayed seemingly different PFS curves (Figure 2B). On this basis, a RS index could be built separating three groups of patients with significantly ($p = 0.0002$) different outcome: low risk (1st and 2nd quartile) intermediate risk (third quartile) and high-risk group (fourth quartile) (Figure 3A). This index was also able to discriminate patients with different CSS and OS ($p = 0.0020$ and $p = 0.0003$, respectively; Figure 3C,E).

In the validation cohort, the ROC analysis confirmed the validity of RS in predicting PFS (AUC, 0.706; 95% CI, 0.610–0.790; $p < 0.001$), CSS (AUC, 0.763; 95% CI, 0.671–0.840; $p < 0.0001$) and OS (AUC, 0.703; 95% CI, 0.607–0.788; $p = 0.004$). Kaplan–Meyer curves illustrates the capacity of the RS index to discriminate groups of patients with different outcomes also in the validation set for either PFS ($p = 0.0046$, Figure 3B), CSS ($p = 0.0022$, Figure 3D), or OS ($p = 0.0056$, Figure 3F).

The predictive accuracy and discriminatory power of the RS index was superior, in both the testing and the validation sets, in comparison to the ones of the international clinical indices when assessed by either the Akaike information criterion or the Harrell C statistic (Table 3).

4 | DISCUSSION

This study showed that a radiomics signature derived from the staging PET/CT can have a significant role in the prediction of PFS and OS in DLBCL patients treated with conventional

TABLE 1 Comparison of baseline patient characteristics and functional PET parameters

Patient characteristics Clinical features	Testing set, *N* = 133 (SAKK 38/07) *n* (%)	Validation set, *N* = 107 (ABREOC 08-002) *n* (%)	*p*-value*
Sex			0.882
Male	67 (50%)	52 (49%)	
Female	66 (50%)	55 (51%)	
Age			
Median; IQR	59 years; 49–68	70 years; 57–76	*<0.0001*
>60 years	64 (48%)	75 (70%)	*0.0006*
Ann Arbor stage			0.399
1	14 (10%)	14 (13%)	
2	42 (32%)	34 (32%)	
3	31 (23%)	16 (15%)	
4	46 (35%)	43 (40%)	
Bulky disease			
>7.5 cm	66 (50%)	49 (46%)	0.538
>10 cm	39 (29%)	25 (23%)	0.295
Elevated LDH	63 (47%)	52 (49%)	0.758
ECOG PS > 1	10 (7%)	10 (9%)	0.569
Extranodal sites >1	34 (26%)	23 (21%)	0.367
COO (Hans algorithm)	*N* = 107	*N* = 93	*<0.0001*
GCB	28 (26%)	54 (58%)	
Non-GCB	79 (74%)	39 (42%)	
IPI group			*0.044*
Low risk	61 (46%)	33 (31%)	
Intermediate–low risk	30 (23%)	39 (36%)	
Intermediate–high risk	25 (19%)	24 (22%)	
High risk	17 (13%)	11 (10%)	
R-IPI group			0.079
Low risk	21 (16%)	7 (6%)	
Intermediate risk	70 (53%)	65 (61%)	
High risk	42 (32%)	35 (33%)	
NCCN-IPI group			*0.006*
Low risk	21 (16%)	6 (6%)	
Intermediate–low risk	62 (47%)	40 (38%)	
Intermediate–high risk	38 (29%)	51 (48%)	
High risk	12 (9%)	10 (9%)	
PET parameters	Median (IQR)	Median (IQR)	
SUVmax	20.1 (15.5–28.3)	21.2 (13.8–28.0)	0.963
SUVmean	8.1 (6.4–10.5)	8.0 (6.2–9.8)	0.452
MTV	243 (72–757)	201 (44–446)	0.086
TLG	2176 (580–7291)	1977 (322–3333)	0.062

Note: Statistically significant *P*-values are in italics.

Abbreviation: IQR, interquartile range.

*Chi square test for comparison of categorical data and Mann–Whitney test for comparison of continuous data.

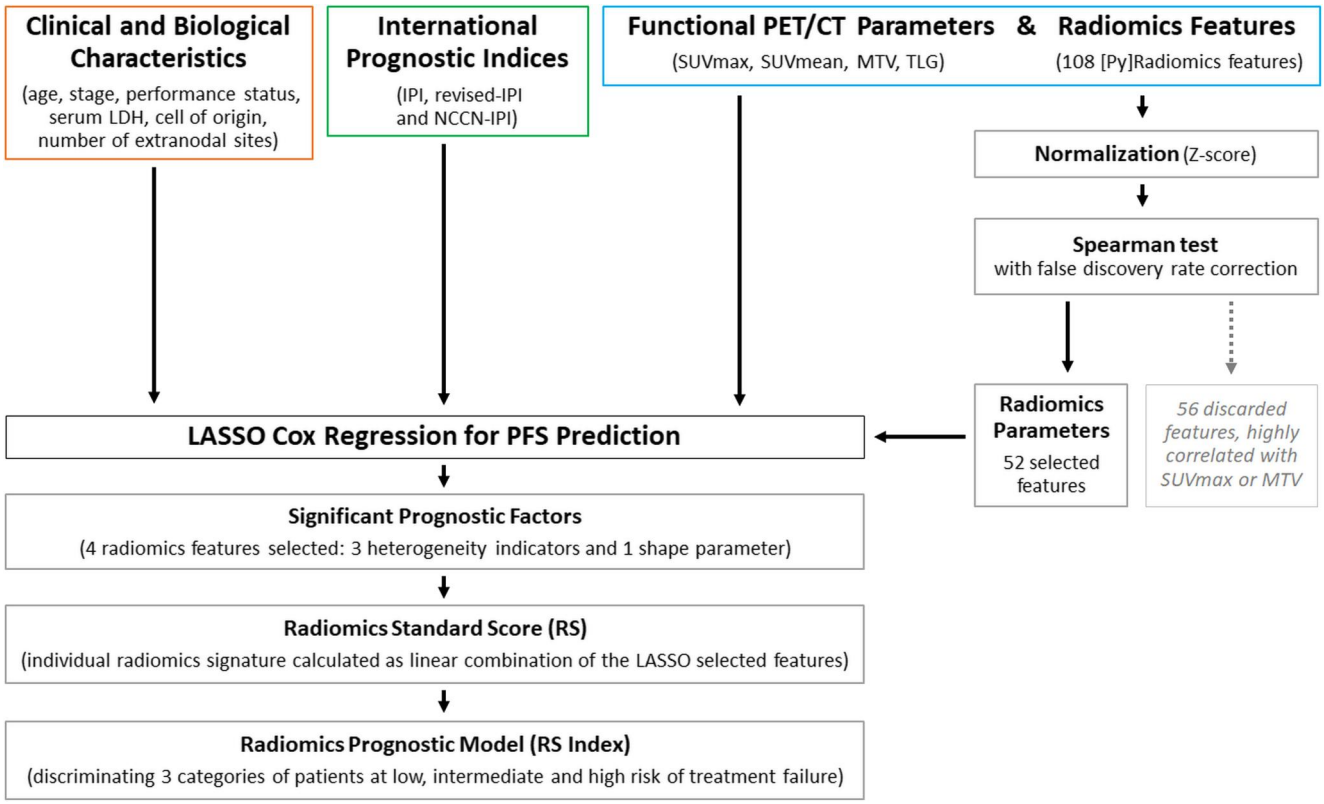

FIGURE 1 Flow-chart of the procedures leading to the construction of the radiomics signature standard score and the subsequent development of a radiomics-based prognostic model

TABLE 2 Selected features from LASSO Cox regression analysis for PFS prediction

Feature name	Regression coefficients
GLCM_SumSquares	−0.2628
Shape_Maximum3DDiameter/BSA	0.1111
GLDM_GrayLevelVariance	−0.0358
GLSZM_GrayLevelNonUniformityNormalized	0.0228

Note: Selected features are listed from the most predictive feature in descending order. The shape_Maximum3DDiameter/BSA is an index of geometric dissemination of the disease, taking into account the patient's body surface area. The other three features are descriptors of the lesion heterogeneity The GLCM_SumOfSquares measures the dispersion of the gray level distribution with regard to the mean intensity value. GLDM_GrayLevelVariance measures the variance in gray level in the image. GLSZM_GrayLevelNonUniformityNormalized measures the variability of gray-level intensity values in the image, with a lower value indicating a greater similarity in intensity values.

immunochemotherapy (R-CHOP regimen given either every 14 or 21 days) and substantiated the clinical relevance of metabolic tumor heterogeneity.

This study provides novel solid evidence supporting the prognostic value of PET-derived radiomics in DLBCL. To our knowledge, only very few preliminary studies have thus far addressed this issue, which used different RF extraction methods and did not have univocal results.[45–47] Parvez et al. tested the prognostic role of the radiomics features in a single-center series (66 DLBCL patients) after partial sampling of the disease including only 1–3 lesions with the highest uptake; their retrospective analysis failed to identify a radiomics signature as prognosticator of outcome although single different RFs showed a relationship with disease-free and overall survival.[45] Aide and colleagues, in another retrospective study of 132 DLBCL patients with a small internal validation set (27 patients), hypothesized the value of radiomics-defined MH of the largest tumor lesion in predicting event-free survival.[46] Lue et al., analyzed a small patient cohort, randomly divided between a training set (58 patients) and a smaller internal validation set (25 patients); their exploratory study used a whole-tumor image analysis and found a single heterogeneity-related radiomics feature to be associated with PFS and OS.[47]

A major strength of our study is that the results generated in the selected patients treated with R-CHOP14 in a multi-center clinical trial were fully validated in a large independent cohort of real-world DLBCL patients treated with standard R-CHOP21. The different treatment schedule in the two cohorts further enhances the prognostic reliability and reproducibility of the RS models.

Our results are in keeping with the mentioned preliminary radiomics observations[45–47] and with prior studies of volumetric PET parameters[18,20] and provide further and robust evidence on the potential prognostic importance of MH in DLBCL. In advanced-stage disease, biological and clinical characteristics may vary at the different tumor sites. We evaluated MH of the whole tumor volume

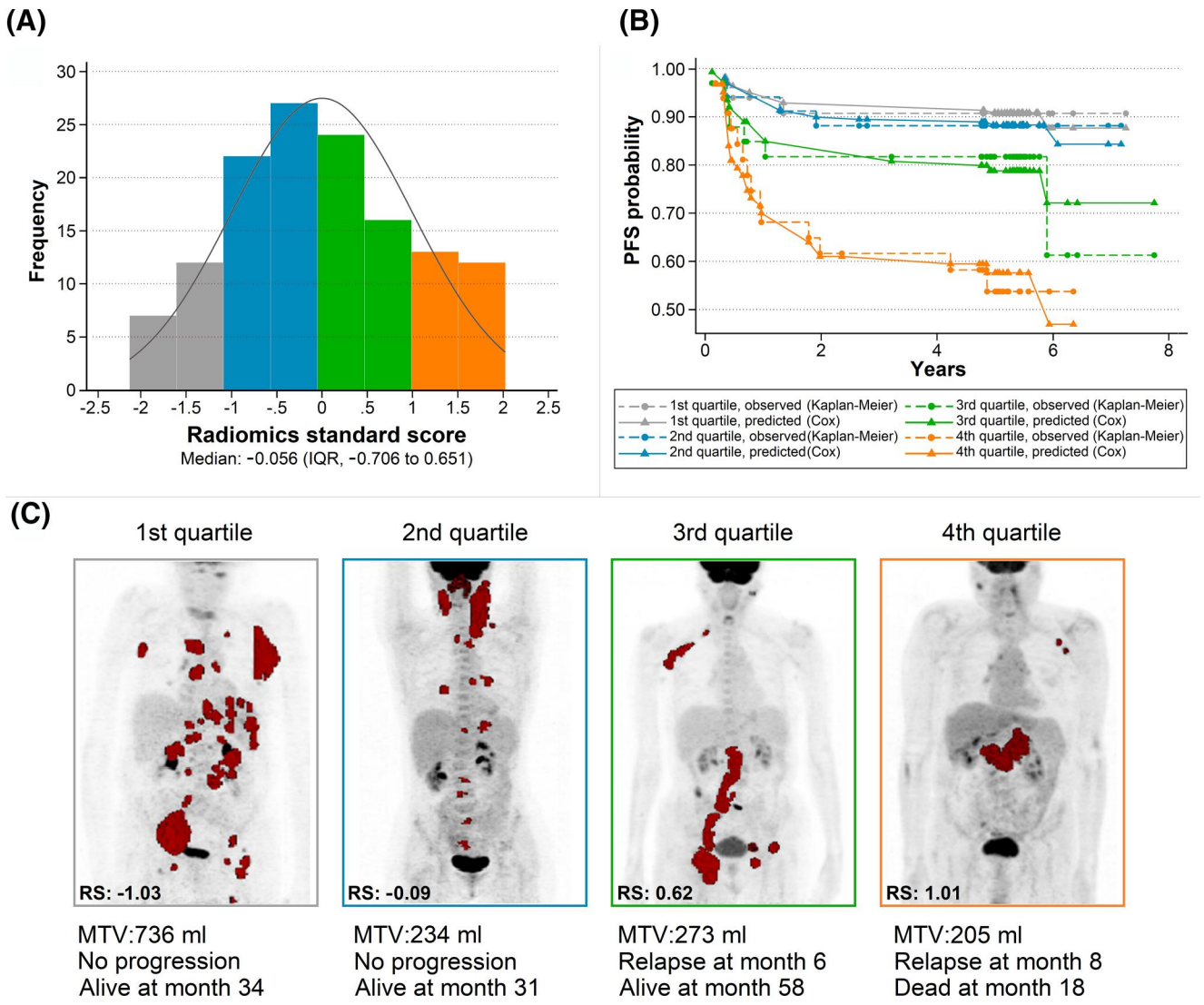

FIGURE 2 Distribution of the radiomics signature standard score (RS), with normal density described by the black curve over the histogram (Panel A). Plotted Kaplan–Meier observed (dashed line with circles) and Cox predicted (solid line with triangles) progression-free survival by RS quartiles in the testing set (Panel B). Panel C provides some examples of the absence of correlation between RS and MTV (segmented volume in red), showing a case (A) with low RS and very high MTV, and three patients (B–D) with similar MTV belonging to different quartiles (and different risk group). Different colors identify the quartiles (first quartile, gray; second, blue; third, green; fourth, orange) in histogram bins (Panel A), survival curves (Panel B) and PET scans (Panel C)

to reduce the potential biases of target lesion selection, which occur when the analysis is restricted to the largest or most metabolically active lesion. Hence, our RS describes the biological heterogeneity of the entire neoplastic population overcoming the limitation of partial tumor sampling. Moreover, our study confirms the very recent findings of Cottereau et al. that anatomic dissemination features such as the maximum distance between 18FDG-avid lesions, may contribute to predict DLBCL outcome.[16,17]

Somewhat surprisingly, LASSO regression discarded the functional quantitative PET parameters (such as SUVmax, MTV, and TLG), which have been previously shown to be significant outcome predictors in our series[20,21] and in several reports in the literature.[10,12,13,15,25,26,48] The results of the present study may appear divergent from our previous report, which showed, in the same cohorts, that the MH of the "hottest" lesion could help identify, among patients with high baseline MTV, a subset of very high-risk patients with shorter PFS and OS.[20] In the current study, instead, we estimated the MH in the entire disease volume and the RS allowed risk stratification of patients irrespective of MTV. This suggests a great potential of radiomics for a better fine-tuning of image-based prognostic models, which deserves further exploration.

Since PET/CT imaging has a fundamental role in the management of lymphomas, radiomics may become a non-invasive and inexpensive tool to produce independent biomarkers for personalized medicine. There are, however significant methodological limitations that may impede the clinical translation, being reproducibility and validation of radiomics analysis a major challenge. In fact, radiomics analysis may be influenced by extraction platform choice, lesion segmentation,

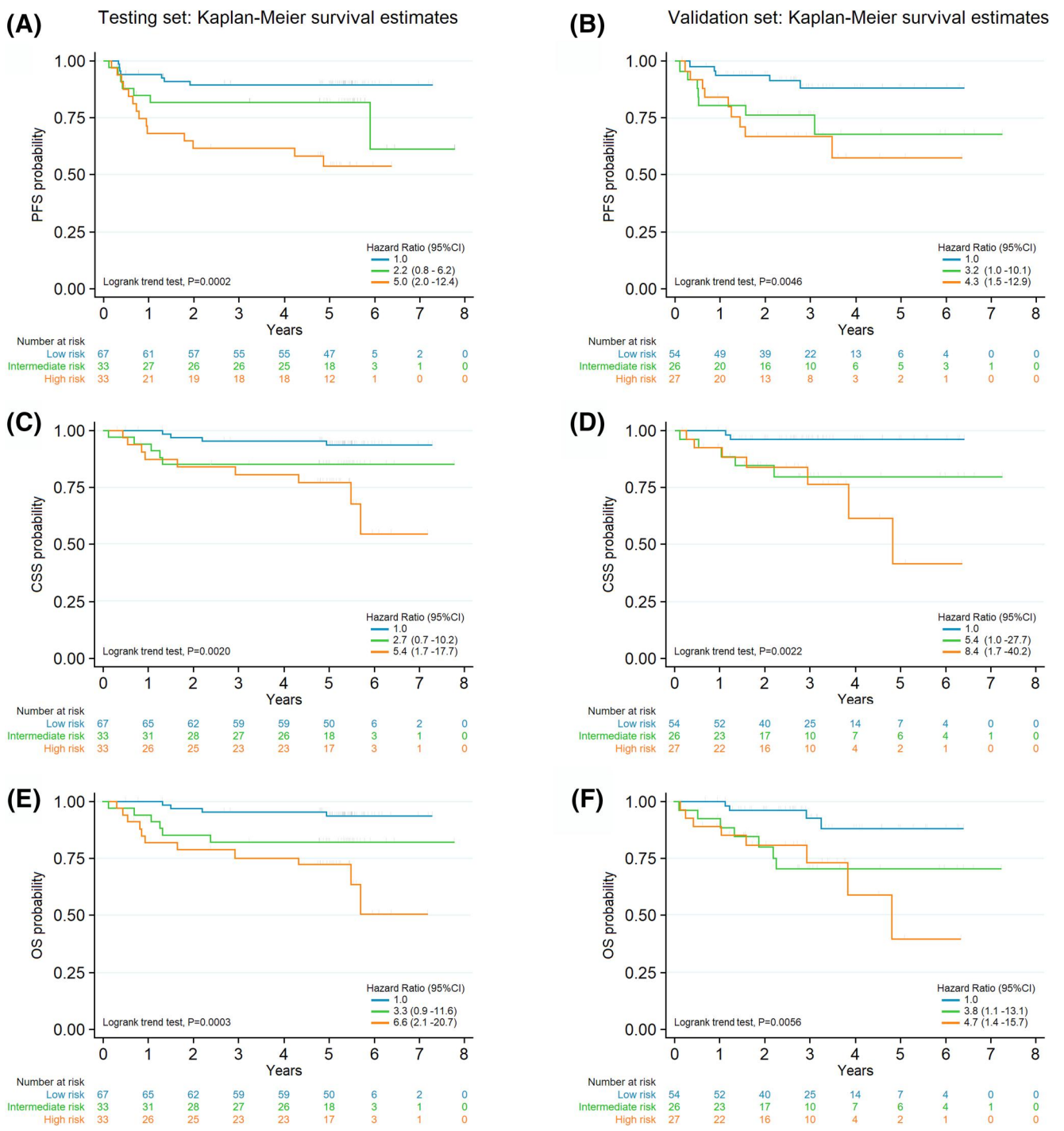

FIGURE 3 Kaplan–Meier estimates of progression free survival (PFS, chart A), cause-specific survival (CSS, chart C), and overall survival (OS, chart E), according to a radiomics-based prognostic model (RS index) in the testing set. Kaplan–Meier estimates of progression-free survival (PFS, chart B), cause-specific survival (CSS, chart D), and overall survival (OS, chart F), according to a radiomics-based prognostic model (RS index) in the validation set. RS values below the median define the low risk group; the intermediate-risk group includes patients with RS between the 50th and 75th percentile and a score over the 75th percentile characterizes the high-risk group

scanner acquisition, and reconstruction parameters.[48,49] These limitations influence model performance limiting the applicability on a multicenter scale. In this perspective, our results are highly encouraging, since both the testing and validation sets are multi-center cohorts.

Notably, LASSO regression also discarded all the clinical indices; hence, it may be not unexpected that in the testing set, both, the information test and the concordance probability estimate showed superior performance of the RS index in comparison to the international indices. Nevertheless, its better predictive accuracy and

TABLE 3 Comparison of the main international prognostic indices with the radiomics signature index by either the concordance probability estimate (CPE) using the Harrell's C statistic or the Akaike information criterion (AIC)

	PFS		CSS		OS	
	AIC	CPE (Harrell's C)	AIC	CPE (Harrell's C)	AIC	CPE (Harrell's C)
Testing set						
RS index	250.6	0.672	159.4	0.676	184.9	0.699
IPI	259.5	0.609	164.7	0.608	192.5	0.624
R-IPI	258.6	0.605	163.9	0.623	191.1	0.634
NCCN-IPI	258.5	0.602	162.5	0.628	187.8	0.662
Validation set						
RS index	179.2	0.666	113.3	0.689	156.5	0.676
IPI	182.7	0.607	119.5	0.605	160.4	0.617
R-IPI	185.2	0.556	120.6	0.556	161.7	0.575
NCCN-IPI	184.9	0.564	120.6	0.549	160.3	0.595

Note: Harrell's C statistic indicates the proportion of observations that the model can order correctly in terms of survival times; C values range from 0.5 (random concordance) to 1 (perfect concordance), higher values indicate better discrimination. AIC estimates the relative amount of information lost by a given prognostic model to assess the likelihood to correctly predict future outcomes. The less information a model loses, the higher its quality; hence, the best model is the one with the lowest AIC.

discriminatory power in validation set provide further evidence of its robustness.

5 | CONCLUSIONS

Given that PET/CT is the gold standard for DLBCL staging,[2] the RS index may help improve risk stratification, for example, for patients entering in clinical trials. Further studies, however, are warranted to clarify potential drawbacks, since the results of texture analysis are strictly dependent on functional volume calculation.[49,50] Hence, our study provides proof of principle for the application of radiomic models, however, the index we generated may require validation with different segmentation methods and in different patient cohorts before being applied in the clinical setting. Notably, our model was developed using a widely available open-source platform, which may facilitate further investigations.

ACKNOWLEDGMENTS

We thank investigators, research nurses and data managers of the SAKK 38/07 trial at each study center, as well as the central study team at the SAKK coordinating center for their support in data collection and study conduction. Special thanks go to Sarah Ortelli Giannakis for her editorial support and to Rita Gianascio Gianocca for the secretarial help. The SAKK 38/07 trial was subsidized by a grant from Amgen (Switzerland) and a grant (OCS 02270-08-2008) from Oncosuisse (Switzerland). The present analysis was partly supported by a grant (ABREOC 22008-262) from the Ente Ospedaliero Cantonale, Bellinzona (Switzerland). The SAKK organization is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI).

CONFLICT OF INTEREST

All authors declare no competing interests that are related, directly or indirectly, to the present study.

AUTHOR CONTRIBUTIONS

Luca Ceriani, Lisa Milan, and Emanuele Zucca designed the study, performed research, analyzed the data, and wrote the paper. Luciano Cascione participated to the study design data analysis and manuscript writing. Federico Dalmasso and Stephane Chauvie contributed to the extraction of radiomics features. Stefanie Hayoz and Sämi Schär reviewed the statistical analysis. Stephan Dirnhofer performed the central pathology review. All authors contributed to data collection, reviewed and approved the manuscript, and shared final responsibility for the decision to submit.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are not publicly available due to legal restrictions. Deidentified individual data used for the reported analysis are currently being shared with the PETRA (PET re-analysis; https://www.petralymphoma.org) consortium. Further data sharing could be made available on reasonable request, which should be addressed to the SAKK—Swiss Group for Clinical Cancer Research (https://www.sakk.ch/en/contact).

ORCID

Luca Ceriani https://orcid.org/0000-0002-6371-097X

Emanuele Zucca https://orcid.org/0000-0002-5522-6109

TRANSPARENT PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/hon.2935.

REFERENCES

1. Sehn LH, Salles G. Diffuse large B-cell lymphoma. *N Engl J Med*. 2021;384(9):842-858.
2. 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):3059-3067.
3. 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.
4. Delarue R, Tilly H, Mounier N, et al. Dose-dense rituximab-CHOP compared with standard rituximab-CHOP in elderly patients with diffuse large B-cell lymphoma (the LNH03-6B study): a randomised phase 3 trial. *Lancet Oncol*. 2013;14(6):525-533.
5. Cunningham D, Hawkes EA, Jack A, et al. Rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisolone in patients with newly diagnosed diffuse large B-cell non-Hodgkin lymphoma: a phase 3 comparison of dose intensification with 14-day versus 21-day cycles. *Lancet (London, England)*. 2013;381(9880):1817-1826.
6. 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.
7. Kim CY, Hong CM, Kim DH, et al. Prognostic value of whole-body metabolic tumour volume and total lesion glycolysis measured on (1)(8)F-FDG PET/CT in patients with extranodal NK/T-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2013;40(9):1321-1329.
8. Kanoun S, Rossi C, Berriolo-Riedinger A, et al. Baseline metabolic tumour volume is an independent prognostic factor in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41(9):1735-1743.
9. Ceriani L, Martelli M, Zinzani PL, et al. Utility of baseline 18FDG-PET/CT functional parameters in defining prognosis of primary mediastinal (thymic) large B-cell lymphoma. *Blood*. 2015;126(8):950-956.
10. 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.
11. Zhou M, Chen Y, Huang H, Zhou X, Liu J, Huang G. Prognostic value of total lesion glycolysis of baseline 18F-fluorodeoxyglucose positron emission tomography/computed tomography in diffuse large B-cell lymphoma. *Oncotarget*. 2016;7(50):83544-83553.
12. 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.
13. Chang CC, Cho SF, Chuang YW, et al. Prognostic significance of total metabolic tumor volume on (18)F-fluorodeoxyglucose positron emission tomography/computed tomography in patients with diffuse large B-cell lymphoma receiving rituximab-containing chemotherapy. *Oncotarget*. 2017;8(59):99587-99600.
14. Kim J, Hong J, Kim SG, et al. Prognostic value of metabolic tumor volume estimated by (18) F-FDG positron emission tomography/computed tomography in patients with diffuse large B-cell lymphoma of stage II or III disease. *Nuclear Medicine and Molecular Imaging*. 2014;48(3):187-195.
15. 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.
16. Cottereau AS, Nioche C, Dirand AS, et al. F-FDG PET dissemination features in diffuse large B-cell lymphoma are predictive of outcome. *J Nucl Med*. 2020;61(1):40-45.
17. Cottereau AS, Meignan M, Nioche C, et al. Risk stratification in diffuse large B-cell lymphoma using lesion dissemination and metabolic tumor burden calculated from baseline PET/CT(†). *Ann Oncol*. 2021;32(3):404-411.
18. 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.
19. Ceriani L, Milan L, Martelli M, et al. Metabolic heterogeneity on baseline 18FDG-PET/CT scan is a predictor of outcome in primary mediastinal B-cell lymphoma. *Blood*. 2018;132(2):179-186.
20. Ceriani L, Gritti G, Cascione L, et al. SAKK38/07 study: integration of baseline metabolic heterogeneity and metabolic tumor volume in DLBCL prognostic models. *Blood Advances*. 2020;4(6):1082-1092.
21. Zucca E, Cascione L, Ruberto T, et al. Prognostic models integrating quantitative parameters from baseline and interim positron emission computed tomography in patients with diffuse large B-cell lymphoma: post-hoc analysis from the SAKK38/07 clinical trial. *Hematol Oncol*. 2020;38(5):715-725.
22. Zhang YY, Song L, Zhao MX, Hu K. A better prediction of progression-free survival in diffuse large B-cell lymphoma by a prognostic model consisting of baseline TLG and %DeltaSUVmax. *Cancer medicine*. 2019;8(11):5137-5147.
23. 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.
24. Toledano MN, Desbordes P, Banjar A, et al. Combination of baseline FDG PET/CT total metabolic tumour volume and gene expression profile have a robust predictive value in patients with diffuse large B-cell lymphoma. *Eur J Nucl Med Mol imaging*. 2018;45(5):680-688.
25. 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.
26. 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.
27. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762.
28. Cook GJR, Azad G, Owczarczyk K, Siddique M, Goh V. Challenges and promises of PET radiomics. *Int J Radiat Oncol Biol Phys*. 2018;102(4):1083-1089.
29. Milgrom SA, Elhalawani H, Lee J, et al. A PET radiomics model to predict refractory mediastinal Hodgkin lymphoma. *Sci Rep*. 2019;9(1):1322.
30. 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 official J Am Soc Clin Oncol*. 2015;33(23):2523-2529.
31. Tzankov A, Leu N, Muenst S, et al. Multiparameter analysis of homogeneously R-CHOP-treated diffuse large B cell lymphomas identifies CD5 and FOXP1 as relevant prognostic biomarkers: report of the prospective SAKK 38/07 study. *J Hematol Oncol*. 2015;8:70.
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. 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*. 2020.

34. 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.
35. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104-e107.
36. Soussan M, Orlhac F, Boubaya M, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. *PLoS One*. 2014;9(4):e94017.
37. Zwanenburg A, Vallières 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.
38. Abdi H. Z-scores. In: Salkind NJ. ed. *Encyclopedia of measurement and statistics*. Vol 3. SAGE Publications, Inc.; 2007:1055-1058.
39. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24(3):69-71.
40. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385-395.
41. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol official J Am Soc Clin Oncol*. 2007;25(5):579-586.
42. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.
43. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol*. 2004;53(5):793-808.
44. R Core Team. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020.
45. Parvez A, Tau N, Hussey D, Maganti M, Metser U. 18 F-FDG PET/CT metabolic tumor parameters and radiomics features in aggressive non-Hodgkin's lymphoma as predictors of treatment outcome and survival. *Ann Nucl Med*. 2018;32(6):410-416.
46. 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.
47. Lue K-H, Wu Y-F, Lin H-H, et al. Prognostic value of baseline radiomic features of 18F-FDG PET in patients with diffuse large B-cell lymphoma. *Diagnostics*. 2021;11(1):36.
48. 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.
49. Fornacon-Wood I, Mistry H, Ackermann CJ, et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur Radiol*. 2020;30(11):6241-6250.
50. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present any future? *Eur J Nucl Med Mol imaging*. 2017;44(1):151-165.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: 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*. 2021;1-11. https://doi.org/10.1002/hon.2935