Computational and Structural Biotechnology Journal 21 (2023) 1102–1114

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Semi-automated ^{18}F-FDG PET segmentation methods for tumor volume determination in Non-Hodgkin lymphoma patients: a literature review, implementation and multi-threshold evaluation

Kylie Keijzer [a,b], Anne G.H. Niezink [b], Janneke W. de Boer [a], Jaap A. van Doesum [a], Walter Noordzij [c], Tom van Meerten [a], Lisanne V. van Dijk [b,*]

[a] Department of Hematology, University Medical Center Groningen, Hanzeplein 1, 9713GZ Groningen, the Netherlands
[b] Department of Radiation Oncology, University Medical Center Groningen, Hanzeplein 1, 9713GZ Groningen, the Netherlands
[c] Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, Hanzeplein 1, 9713GZ Groningen, the Netherlands

ARTICLE INFO

Article history:
Received 16 November 2022
Received in revised form 18 January 2023
Accepted 18 January 2023
Available online 20 January 2023

Keywords:
^{18}F-FDG PET
SUV thresholding
PET segmentation
Segmentation software
Metabolic tumor volume
Non-Hodgkin lymphoma

ABSTRACT

In the treatment of Non-Hodgkin lymphoma (NHL), multiple therapeutic options are available. Improving outcome predictions are essential to optimize treatment. The metabolic active tumor volume (MATV) has shown to be a prognostic factor in NHL. It is usually retrieved using semi-automated thresholding methods based on standardized uptake values (SUV), calculated from ^{18}F-Fluorodeoxyglucose Positron Emission Tomography (^{18}F-FDG PET) images. However, there is currently no consensus method for NHL. The aim of this study was to review literature on different segmentation methods used, and to evaluate selected methods by using an in house created software tool.

A software tool, **MU**ltiple **S**UV **T**hreshold (MUST)-segmenter was developed where tumor locations are identified by placing seed-points on the PET images, followed by subsequent region growing. Based on a literature review, 9 SUV thresholding methods were selected and MATVs were extracted. The MUST-segmenter was utilized in a cohort of 68 patients with NHL. Differences in MATVs were assessed with paired t-tests, and correlations and distributions figures.

High variability and significant differences between the MATVs based on different segmentation methods ($p < 0.05$) were observed in the NHL patients. Median MATVs ranged from 35 to 211 cc.

No consensus for determining MATV is available based on the literature. Using the MUST-segmenter with 9 selected SUV thresholding methods, we demonstrated a large and significant variation in MATVs. Identifying the most optimal segmentation method for patients with NHL is essential to further improve predictions of toxicity, response, and treatment outcomes, which can be facilitated by the MUST-segmenter.

Abbreviations: AT, adaptive thresholding methods; CAR, chimeric antigen receptor; cc, cubic centimeter; CT, computed tomography; DICOM, Digital Imaging and Communications in Medicine; DLBCL, Diffuse large B-cell lymphoma; EANM, European Association of Nuclear Medicine; EARL, EANM Research Ltd.; FDG, fluorodeoxyglucose; HL, Hodgkin lymphoma; IMG, robustness across image reconstruction methods; IQR, interquartile range; LBCL, Large B-cell lymphoma; LDH, lactate dehydrogenase; MAN, clinician based evaluation using manual segmentations; MATV, Metabolic active tumor volume; MIP, Maximum Intensity Projection; MUST, Multiple SUV Thresholding; NHL, Non-Hodgkin lymphoma; OBS, robustness across observers; OS, overall survival; PD-L1, programmed cell death ligand-1; PET, positron emission tomography; PFS, progression free survival; PROG, progression vs non-progression; PTCL, Peripheral T-cell lymphoma; PTLD, Post-transplant lymphoproliferative disorder; QS, quality scores; SOFT, robustness across software; SUV, standardized uptake value; TCL, T-cell lymphoma; UMCG, University Medical Center Groningen; VOI, volume of interest

* Corresponding author.
E-mail addresses: k.keijzer@umcg.nl (K. Keijzer), a.g.h.niezink@umcg.nl (A.G.H. Niezink), j.w.de.boer@umcg.nl (J.W. de Boer), j.a.van.doesum@umcg.nl (J.A. van Doesum), w.noordzij@umcg.nl (W. Noordzij), t.van.meerten@umcg.nl (T. van Meerten), l.v.van.dijk@umcg.nl (L.V. van Dijk).

https://doi.org/10.1016/j.csbj.2023.01.023

1. Introduction

^{18}F-Fluorodeoxyglucose Positron Emission Tomography/Computerized Tomography (^{18}F-FDG PET/CT) scans are widely used in clinical practice for disease staging and evaluation of response to treatment in patients with Non-Hodgkin lymphoma (NHL). Performing baseline, interim and end of treatment ^{18}F-FDG PET/CT imaging has led to risk-adapted treatment approaches to improve treatment response and limit toxicities and adverse events, hence accurate interpretation of ^{18}F-FDG PET/CT is of great importance [1,2]. Response assessment is nowadays performed using the Deauville score, a 5-point grading scale that compares the FDG uptake of diseased areas to uptake in the mediastinum and liver [3]. However, this is a coarse assessment showing a poor inter-reader reliability [4]. In addition, visual assessment of FDG uptake can be biased due to differences in background levels (contrast illusion) [5].

Alternative methods to assess a patient's disease stage using ^{18}FDG-PET image biomarkers have been proposed. Most common is the use of metabolic active tumor volume (MATV), which already showed to be a good prognostic biomarker for treatment outcome predictions [6–13]. MATV is a metric that represents the volume of the tumor tissue segmentations with high FDG uptake (FDG-tumor). Defining this FDG-tumor can be performed manually, but (semi-)automatic segmentation by thresholding on a specific PET image intensity value (i.e., standardized uptake value (SUV)), is a more objective and normalized method. In this procedure, volumes above a certain SUV threshold are classified as tumor. This is either based on fixed (e.g., SUV 2.5) [14] or relative (e.g., 41% of the maximum SUV value in the tumor region) [15] thresholds, and are generally initialized by placing seed points in, or bounding boxes around the identified tumors [16].

NHL is a disease that is often widespread through the body with nodal and extranodal involvement, and with a wide range in tumor volumes. (Fig. 1). Compared to solid tumors, it is therefore more challenging to investigate the prognostic value of different FDG-tumor definitions as this is in the current state very time-consuming [17]. Besides, there is currently no consensus SUV thresholding method [18], resulting in a variety of methods used in clinical studies [19–24].

Thresholding methods incorporated in currently available segmentation software are variable and can only be utilized per single thresholding method at a time; evaluating multi-threshold PET segmentation methods for NHL patients is therefore challenging. There is an unmet need of an easy-to-use PET segmentation tool that can extract MATV of various tumor lesions, using multiple segmentation methods simultaneously. This is needed to facilitate research on the optimal segmentation method for the prediction of treatment outcomes in patients with NHL.

The aim of this study was to review the literature on the different segmentation methods used and to evaluate the methods by using an in-house created software tool that semi-automatically segments NHL lesions on FDG PET scans. This was pursued with the following objectives:

A. To identify promising semi-automated FDG PET NHL segmentation methods by reviewing related segmentation evaluation publications.
B. To develop a segmentation tool that generates FDG-tumor and MATV for different SUV thresholds simultaneously.
C. To evaluate variance in MATVs based on different FDG-tumor threshold in an exemplar patient cohort.

2. Materials and methods

2.1. Literature review

A literature search was performed in PubMed to retrieve publications that evaluated different semi-automated segmentation methods for NHL. The following search query was created: (lymphoma[tiab] OR NHL[tiab] OR DLBCL[tiab] OR MCL[tiab]) AND (positron emission tomography[tiab] OR FDG[tiab] OR PET[tiab] OR SUV [tiab] OR standardised uptake value[tiab] OR standardized uptake

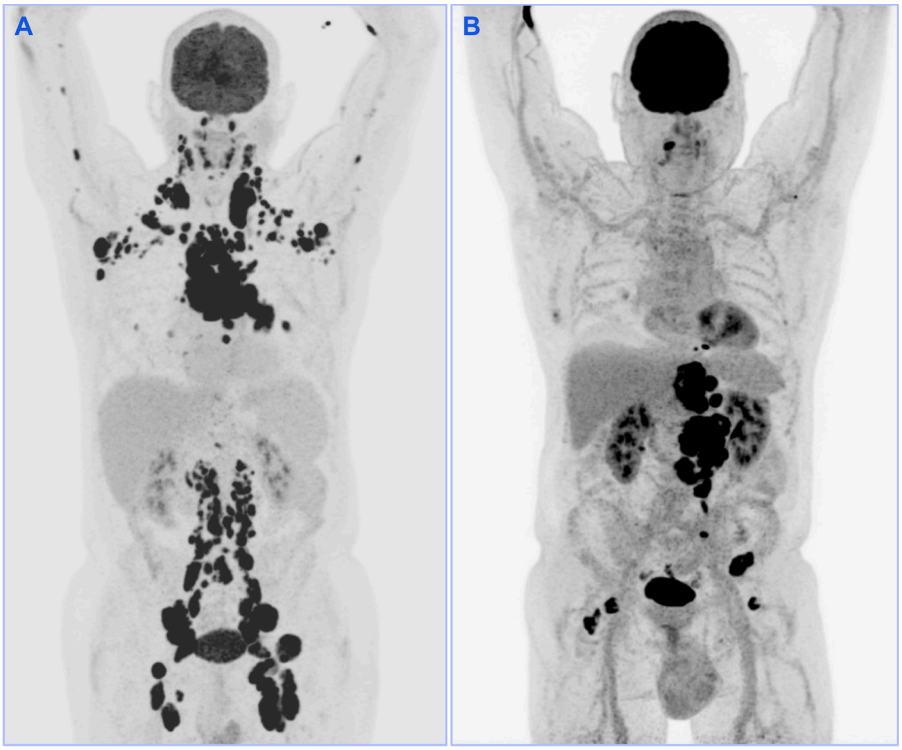

Fig. 1. Maximum Intensity Projections (MIP) of two patients with NHL showing **A**, a large dissemination of tumors and **B**, less dissemination of tumors.

value[tiab]) AND (segment*[tiab] OR deline*[tiab] OR contour*[tiab] OR measurement method[tiab] OR methodolog*[tiab] OR volume measur*[tiab] OR volume calculat*[tiab]) AND (feature*[tiab] OR characteristic*[tiab] OR tumor volume[tiab] OR tumour volume[tiab] OR tumor burden[tiab] OR tumour burden[tiab] OR MTV[tiab] OR MATV[tiab]). The search was completed on August 1st, 2022. Only studies from 2010 till 2022 that had a full-text version available, were considered.

All studies had to fulfill the following inclusion criteria: a) clinical studies b) including patients with NHL c) disease evaluation reported with FDG PET scans d) segmentation performed with multiple semi-automated SUV thresholding methods e) evaluation/comparisons made between the selected segmentation methods.

The selection of studies was performed based on title and abstract, and full text was evaluated in the case of any doubt. Resulting studies were analyzed on methodology, important conclusions, utilized segmentation methods, best segmentation method(s), number of patients and patient disease.

Based on the literature overview, a final set of semi-automated segmentation methods was selected for implementation in our software.

2.2. MUST-segmenter development

The software tool MUST (**MU**ltiple **S**UV **T**hreshold)-segmenter was developed to extract segmentations of multiple tumor lesions with multiple SUV thresholding methods simultaneously, i.e., without having to re-segment for every threshold method. This will allow to perform efficient research on the optimal SUV thresholding method for the prediction of treatment outcomes in NHL patients. The MUST-segmenter is readily available within 3D Slicer [25] as a Python-based [26] plugin and can be used without coding experience. The source code of the tool is open-access and available at https://github.com/kyliekeijzer/Slicer-PET-MUST-segmenter.

The development workflow of the MUST-segmenter was designed as follows: 1) opening and displaying PET imaging data together with avoidance regions; 2) user-detection of NHL lesions with seed points and/or bounding boxes; 3) semi-automatic segmentation using seed points and/or bounding boxes; 4) displaying and exporting segmentation results; 5) calculation of FDG-tumor MATV. A schematic overview of the process is displayed in Fig. 2.

2.2.1. Opening PET/CT imaging and avoidance regions

The input required for the MUST-segmenter are PET (which are then converted to SUV maps) and CT images (in Digital Imaging and Communications in Medicine (DICOM) format). Users can navigate between slices in sagittal, transversal, and coronal views (functionality of 3D Slicer). Optionally, pre-defined high FDG uptake avoidance volumes of interests (VOIs), e.g., in the brain, bladder or kidneys, can be loaded. These VOIs can be excluded from the PET images, to prevent healthy tissue to be included into the tumor segmentation.

2.2.2. Lymphoma lesions identification with seed points

The user is required to identify the NHL lesions on the PET/CT by clicking on a lesion (preferably in the center). Since NHL is typically a disseminated disease, NHL lesion detection needs to commonly be performed for multiple tumor locations. Seed points were defined with the users' mouse clicks at high uptake areas of all individual tumors. Rectangle shaped bounding boxes are generated around the individual tumors, which can be adjusted by the user.

2.2.3. Semi-automatic lesion segmentation with different PET intensity thresholds

The next step is the segmentation process that is performed with a region-growing algorithm using a 6-connected voxel neighborhood. Specifically, the algorithm starts from the placed seeds and spatially keeps adding neighboring tumor voxels that meet the SUV criterium, until no adjacent parts meet the criterium anymore.

Based on the literature review, the following SUV thresholding methods were established for the MUST-segmenter:

Type 1 - fixed SUV threshold

For method type 1, the fixed SUV thresholds, the VOI is defined from the seed point to all connected voxels which meet the absolute threshold of:

1.1 SUV ≥ 2.5 (SUV2.5)
1.2 SUV ≥ 3.0 (SUV3.0)
1.3 SUV ≥ 4.0 (SUV4.0)

Type 2 - relative threshold

For the relative threshold methods, the VOI is defined from the seed point to all connected voxels which meet the relative threshold of:

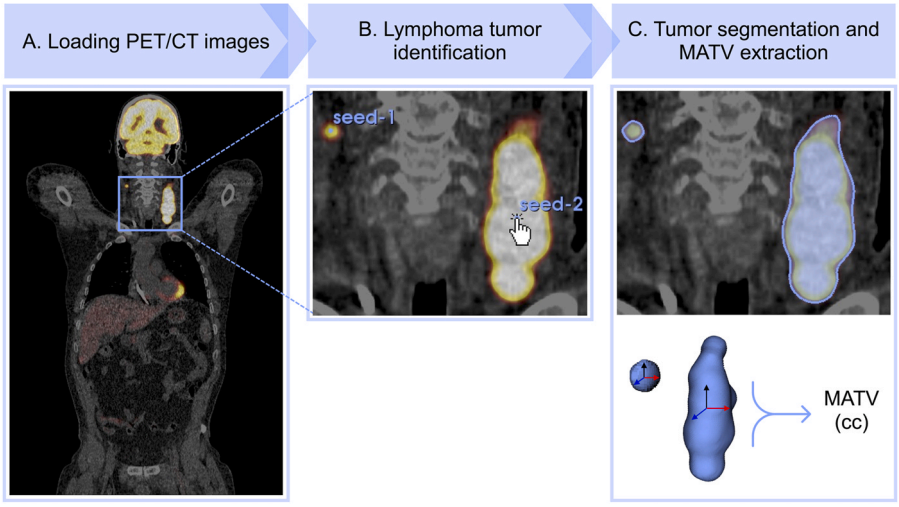

Fig. 2. Development workflow of the MUST-segmenter. **A** loading PET/CT imaging data into 3D Slicer. **B** lymphoma tumor identification by placing seed points, and **C** performing segmentation, displaying the segmentation results, and calculating the metabolic active tumor volume (MATV) of the FDG-tumor.

2.1 SUV ≥ 41% of the maximum SUV of bounding boxes (41suvMax);
Bounding boxes are areas around the lesions based on the seed points, 41% of the maximum SUV value determined in all boxes defines the threshold.
2.2 SUV ≥ the maximum SUV value in a user-defined liver area (liverSuvMax);
The maximum SUV value in a 1.2 cm diameter sphere segment in the right lobe of the liver. The sphere is created directly from a user-defined seed.
2.3 SUV ≥ PERCIST SUV [27] (PERSIST);
Based on the same SUV vales in the sphere area described at 2.2., the PERCIST SUV is defined as: 1.5 x average liver SUV + 2 x Standard Deviation.
2.4 SUV ≥ 41% of the average brain SUV + Age Correction (SUVbrain);

Where,

$$AgeCorrection = (PatientAge - 20) \times \left(\frac{0.125}{70 - 20}\right)$$

Brain SUV values are calculated from uploaded high uptake regions as described before. This threshold was added as it is known from literature that the brain activity is consistent among individuals; however, the activity decreases with age [28–31].

Five additional thresholds based on brain SUV values can be generated, which are described in **Supplementary file A**.

Type 3 - majority voting

Majority voting is based on uniformity between the SUV2.5, SUV4.0, 41suvMax, liverSuvMax and PERCIST methods:

3.1 two or more methods (MV2)
3.2 three or more methods (MV3)

2.2.4. Visualizing segmentations and exporting MATV

The final step in the MUST-segmenter software is the visualization of FDG-tumor defined by the different SUV thresholding methods. These contours are converted to visual representations on the PET images and are paired with 3D representations. Users can make final edits to the FDG-tumor segmentation if necessary, followed by saving them in the desired format using 3D Slicer functionality. The MATV can be extracted and is determined as the number of voxels in the FDG-tumor segmentation, times the voxel resolution. In addition, yet outside the scope of this paper, other image biomarkers can be extracted, including radiomics features [32].

2.3. MATV assessment

2.3.1. Patient cohort

To evaluate the MUST-segmenter and the variance in MATV output, a use-case cohort of 68 patients with Large B-cell lymphoma (LBCL), the most common NHL subtype, treated at the University Medical Center Groningen (UMCG), The Netherlands, was collected. This LBCL cohort was chosen since this is a very diverse disease with commonly multiple tumor sites, and thus appropriate for evaluation of the MUST-segmenter. Patients received CD19-directed chimeric antigen receptor (CAR) T-cell therapy as a third line therapy. PET/CT scans were acquired at two time points: approximately twenty-eight days and six days before CAR T-cell infusion. ^{18}F-FDG PET scans were performed according to The European Association of Nuclear Medicine (EANM) on either a Siemens Biograph mCT (40 or 64), a Siemens Biograph Vision 600 or a Siemens Biograph Vision Quadra, with EARL1 accreditation [15,33].

2.3.2. MUST-segmenter application on patient data

PET/CT scans were loaded into 3D Slicer, together with avoidance regions, which include the heart, spleen, kidneys, and liver and brain contours, that were obtained with Mirada's Embrace Atlas Contouring modality [34]. The MUST-segmenter was utilized to segment PET scans and retrieve the MATV with all SUV methods as described before.

2.3.3. Statistics

Distributions of MATV were visualized by creating boxplots (5–95 percentile whiskers). Significant differences between MATV results were assessed with paired t-tests, adjusted for multiple testing (Benjamini-Hochberg). Correlations and distributions figures of log transformed MATVs among the different segmentation methods were created. Low- and high-MATV patient groups were identified using cut-offs found in literature, and were evaluated using Cochran's Q test and McNemar's tests. Statistical analysis was performed using R v4.2.1.

3. Results

3.1. Literature review

The literature search retrieved 123 results and after the selection procedure, 12 studies were included for the review. The flowchart of the selection process is shown in Fig. 3. An overview of the included 12 studies is given in Table 1.

3.1.1. Patient and PET/CT image characteristics

Of the 12 included studies, patient cohorts ranged from 12 to 239 (mean = 91, median = 95). All studies included patients with NHL ($n = 12$; [35–46]) but some studies additionally included patients with Hodgkin lymphoma (HL) ($n = 3$; [35,40,46]). Six studies [35,36,38,39,45,46] stated whether they included EANM Research Ltd. (EARL) accredited PET images, whereof 2 studies included both EARL accredited and non-accredited images [36,38].

3.1.2. Segmentation software and approach

Studies relied on different segmentation software and approaches. Two studies used seed-based segmentation ($n = 2$) [38,42,44], while the majority deployed user-defined bounding boxes ($n = 6$) [35,39–41,45,46]. Moreover, different bounding box shapes were implemented: cubic shapes [35,45,46], spherical shapes [41,46] and irregular shapes [39,40]. Four studies used auto-segmentation of all high uptake regions without deploying bounding boxes, yet by manually removing all non-tumor regions [35,36,38,43]. Manual removal of non-tumor regions was also performed in 1 other study [44] that applied a semi-automatic approach.

3.1.3. SUV segmentation thresholds

Nearly all studies included fixed SUV thresholds of SUV2.5 ($n = 9$) [35–39,41–44] and SUV4.0 ($n = 5$) [35–38,42] in their segmentation method evaluation. The most common relative SUV threshold method was the 41suvMax method ($n = 10$) [35–39,42–46]; this threshold has been recommended by the EANM for solid tumors [15]. Thresholds based on liver SUV values were less investigated ($n = 1$ for liverSuvMax [39], $n = 3$ for PERCIST [39,41,44]). Methods based on majority voting were included in 5 studies [35–38,42]. Several studies included more complex thresholds based on adaptive threshold (AT) algorithms ($n = 8$) [35–40,42,45], such as thresholds based on tumor background intensities. One study [40] also compared SUV thresholding methods to deep learning algorithms.

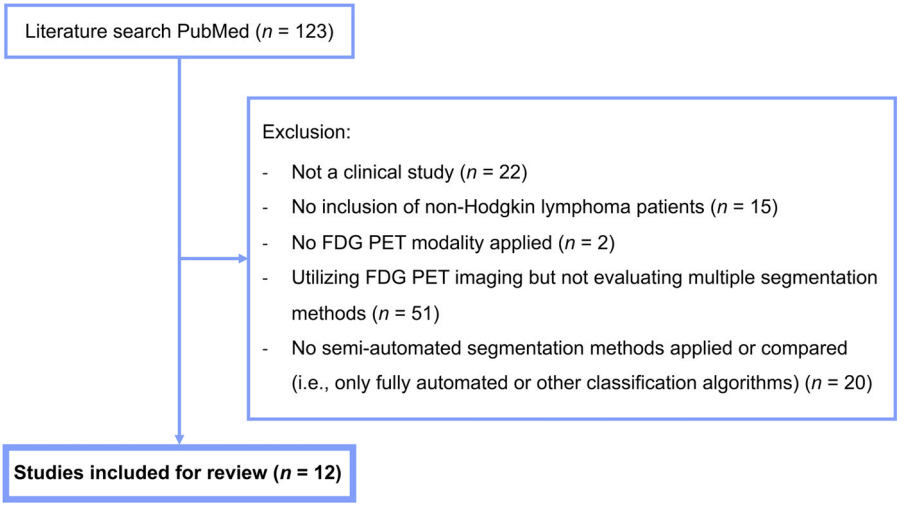

Fig. 3. Flowchart of literature search, resulting in a final selection of 12 studies that were included in the review.

3.1.4. Segmentation methods evaluation and conclusions

Different analyses were used to identify the preferred PET segmentation approach, these were based on 1) predictive performance, 2) robustness, and 3) clinician-based evaluation.

The performance of Overall Survival (OS) and/or Progression Free Survival (PFS) prediction was assessed for MATVs defined with different SUV threshold methods in 5 studies [39,41,43–45]. Significant association of MATV with OS using all segmentation methods (i.e., SUV2.5, 41suvMax, liverSuvMax, PERCIST) was found in 2 studies [39,44], whereas 1 study [43] showed no significance association with the 2 methods investigated (i.e., SUV2.5, 41suvMax). MATV was significantly associated with PFS using all methods in 3 studies [39,44,45], although no significant relation with PFS was found in 1 study [43].

For the prediction of OS and PFS, 3 studies [41,43,44] found that the best predictive performance measures of MATV were achieved with SUV2.5. In contrast, 1 study found the best predictive performance for PFS with 41suvMax [39].

The optimal cut-off points to differentiate low- and high-MATV patient groups were analyzed in 4 studies [39,43–45], and determined for 4 segmentation methods (Table 2). The cut-offs varied between studies, e.g., 396 – 552 cc for PFS using SUV2.5, and between segmentation methods, e.g., 295 – 552 cc for PFS found by Eude et al. [39].

Prediction of disease progression (PROG) was performed by 2 studies [36,38]. One of these studies [36] evaluated prediction models that included multiple PET variables and found SUV4.0 the best preforming method. The other study [38] found MV2 to perform the best.

Although these studies found differences in predictive performance, only 1 study [41] based their preferred method recommendations on the differences in outcome prediction performance.

Robustness across different observers (OBS) was evaluated in 4 studies [39,42–44], whereof 2 studies [39,44] recommended SUV2.5, 1 study [42] the SUV4.0 method, and one study [39] liverSuvMax and PERCIST.

Robustness of the threshold approaches was evaluated for different PET image reconstruction methods (IMG) ($n = 2$; [35,38]). One study [35] compared tumor volume ratios of two types of EARL- and high-resolution-reconstructed images, and evaluated the effect of these reconstruction methods on different SUV thresholding methods. SUV4.0 showed the best alignment between the image reconstruction techniques and results of MV2 were comparable, indicating these two methods were the most robust to image reconstruction. Another study [38] did not find a significant difference in quality scores determined on EARL and non-EARL reconstructed images.

Additionally, 1 study [44] evaluated robustness between software (SOFT) and recommended SUV2.5.

For clinician-based evaluation, segmentation quality scores (QS) have been proposed in 3 studies [37,38,42] as a technique to determine the most optimal method where the quality score is assigned to the segmentation result by a clinician, yet quality scoring criteria differed somewhat between studies. Based on the QS, all 3 studies recommend SUV4.0, 1 study [38] also recommends MV2, and 1 study [37] MV3.

Other clinician-based analyses were performed by comparing manual segmentations (MAN) with different threshold approaches. Where 1 study [46] recommended the 41suvMax over other SUVmax thresholding methods, while 2 other studies [40,43] did not provide a recommendation.

As an additional analysis, 3 studies [37,38,44] investigated the correlation between the segmentation thresholds approaches, and overall found the thresholding methods linearly correlated.

3.2. MUST-segmenter implementation

The MUST-segmenter loaded as a plugin tool into 3D Slicer is shown in Fig. 4. A patient's PET/CT image is displayed in axial, coronal and sagittal representations. The MUST-segmenter functionalities are available on the left side, including options to perform segmentation with fixed, relative, and majority voting-based methods, followed by calculation of MATVs. An example segmentation result retrieved with SUV4.0 is displayed over the PET/CT images (green) and in 3D view (upper right), with the corresponding seed points.

A complete instruction on how to use the MUST-segmenter is available at the repository.

To get an indication of the time to perform segmentation, 1 patient with 12 lesions was segmented with 2 fixed thresholds and 1 relative threshold, using the MUST-segmenter and using ACCURATE [47], a tool utilized in the most recent studies [35–38]. The total segmentation time using MUST-segmenter was *5 minutes and 56 seconds*, compared to *28 minutes and 56 seconds* using ACCURATE.

Table 1
Overview of studies included for literature review.

Author, Year	Lymphoma Type	No. Patients	Software Used	EANM Accredited	SUV Segmentation Thresholds	No. Observers	Evaluation Criteria	Recommended Segmentation Method
Ferrández, 2022 [35]	DLBCL, HL, TCL, PTLD	19	ACCURATE	Yes	SUV2.5, SUV4.0, 41suvMax, AT, MV2, MV3	1	IMG*	SUV4.0 and MV2
Eertink, 2022 [36]	DLBCL	100	ACCURATE	Yes	SUV2.5, SUV4.0, 41suvMax, AT, MV2, MV3	1	PROG	-
Zwezerijnen, 2021 [37]	DLBCL	45	ACCURATE	No	SUV2.5, SUV4.0, 41suvMax, AT, MV2, MV3	3	QS*	MV3 and SUV4.0
Barrington, 2021 [38]	DLBCL	138	ACCURATE	Yes	SUV2.5, SUV4.0, 41suvMax, AT, MV2, MV3	2	PROG, QS*	SUV4.0 and MV2
Eude, 2021 [39]	DLBCL	239	PLANET Onco	Yes	SUV2.5, 41suvMax, liverSuvMax, PERCIST, AT	2	OS, PFS, OBS*	SUV2.5, liverSuvMax, PERCIST
Weisman, 2020 [40]	DLBCL, HL	90	MIM software, MATLAB	No	40suvMax, 50suvMax, AT, active contouring, clustering, deep learning	2	MAN	-
Guzmán Ortiz, 2020 [41]	DLBCL	34	Syngo.Via	No	SUV2.5, 40suvMax, PERCIST	2	OS* , PFS*	SUV2.5
Burggraaff, 2020 [42]	DLBCL	12	ACCURATE	No	SUV2.5, SUV4.0, 41suvMax, AT, MV2, MV3, SUV4.0 and volume ≥ 3cc (fully automated), SUV4.0 + volume ≥ 3cc + manual tumor additions	3	QS* , OBS*	Fully automated segmentation with SUV4.0 and tumor volume ≥ 3cc
Gormsen, 2019 [43]	DLBCL	118	Hermes	No	SUV2.5, 41suvMax, visual assessment	3	OS, PFS, OBS, MAN	-
Ilyas, 2018 [44]	DLBCL	147	Hermes, PETTRA	No	SUV2.5, 41suvMax, PERCIST	2	OS, PFS, OBS* , SOFT*	SUV2.5
Cottereau, 2017 [45]	PTCL	106	PLANET Onco	Yes	41suvMax, AT	1	OS, PFS, REF	-
Meignan, 2014 [46]	DLBCL, HL	40	PETVCAR, Imagys	Yes	25suvMax, 35suvMax, 40suvMax, 41suvMax, 42suvMax, 45suvMax, 50suvMax, 60suvMax	2	MAN*	41suvMax

AT = Adaptive thresholding methods, DLBCL = Diffuse large B-cell lymphoma, HL = Hodgkin lymphoma, IMG = Robustness across image reconstruction methods, MAN = Clinician based evaluation using manual segmentations, OBS = Robustness across observers, OS = Overall survival, PFS = Progression free survival, PROG = Progression vs non-progression, PTCL = Peripheral T-cell lymphoma, PTLD = Post-transplant lymphoproliferative disorder, QS = Quality scores, SOFT = Robustness across software, TCL = T-cell lymphoma
* Recommended segmentation method is based on this criterion

Table 2
MATV (cc) cut-offs per segmentation method to determine low- and high-MATV groups for progression free survival (PFS) and overall survival (OS) predictions, found in the literature.

Study	SUV2.5		41suvMax		LiverSUVmax	PERCIST	
	PFS	OS	PFS	OS	PFS	PFS	OS
Eude [39]	552		295		487	486	
Gormsen [43]	542	585	147	105			
Ilyas [44]	396, 401	458, 401	166	189		327	670
Cottereau [45]			230	260			

3.3. MATV assessment

3.3.1. Patient cohort

In the use-case cohort 68 DLBCL patients were included (Table 3). Pre-CAR T-cell therapy scans from two timepoints were selected, resulting in a total of 110 PET scans available for MATV analysis.

Table 3
Patient Characteristics.

	Total (n = 68)
Age, median (range)	61 (20 – 79)
Sex, *n* (%)	
Male	47 (69.1)
Female	21 (30.9)
Ann Arbor stage, *n* (%)	
I	4 (5.9)
II	8 (11.8)
III	16 (23.5)
IV	40 (58.8)
Extranodal localizations, *n* (%)	
Yes	49 (72.1)
No	19 (27.9)

3.3.2. MATV for the different threshold methods

In Fig. 5, the MATVs are plotted per segmentation method. The median MATV ranged between 35 and 211 cc (**Supplementary file B**). Largest variability was observed for SUVbrain (IQR [64.56 – 501.55 cc]) and smallest variability was retrieved with 41suvMax (IQR [10.62 – 86.72 cc]). MATVs of fixed thresholds decrease with increase of SUV threshold value (i.e., from SUV2.5 to SUV4.0), which translates to the majority voting methods. There is a large amount of variability between all relative methods, which was also observed for the fixed, relative, and majority voting methods.

3.3.3. MATV correlations and differences

MATVs were log-transformed to establish a normal distribution (Fig. 6). The vast majority of MATVs were significant different between segmentation methods (paired t-test $p < 0.05$). Only, the MATVs obtained with SUV4.0 and PERCIST methods were not significantly different from each other, as were those with liverSuvMax and MV2. Evaluation of the correlation plots reveals that the 41suvMax differed the most compared to the other methods. The fixed methods were highly correlated, while the data points were placed further under the diagonal when the threshold increases, which was also observed in Fig. 5. Some MATV points were zero for the SUV3.0 and SUV4.0 methods, as no lesions had SUVs above the fixed threshold.

3.3.4. Low- and high-MATV groups identification

Two main MATV cut-off trends were found in the literature (Table 2), which were 200 and 500 cc. Percentages of patients classified as low- or high-MATV using these two cut-offs (n = 110) are displayed in Table 4. The least high-MATV patients are found with 41suvMax, only 5% and 1%, whereas the other high-MATV patients percentages range from 33 – 51 (cut-off 1) and 11 – 25 (cut-off 2).

For both cut-offs, there was a significant difference in low- and high-MATV patient proportions between all segmentation methods (Cochran's Q p-value < 0.0001). Pairwise comparisons showed significant differences between nearly all methods (Fig. 7). The

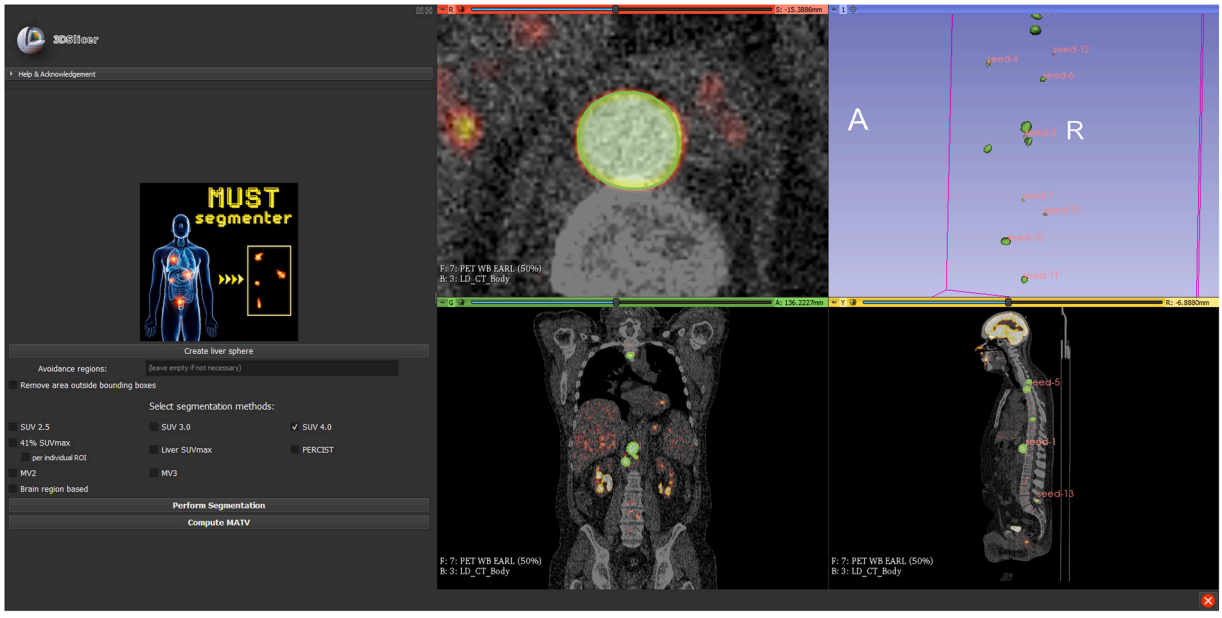

Fig. 4. MUST-segmenter loaded into 3D Slicer. One patient's PET/CT images are shown with an example segmentation result using SUV4.0 (green). On the left side MUST-segmenter functionalities are shown, including the option to perform segmentation with different methods and extracting the corresponding MATVs.

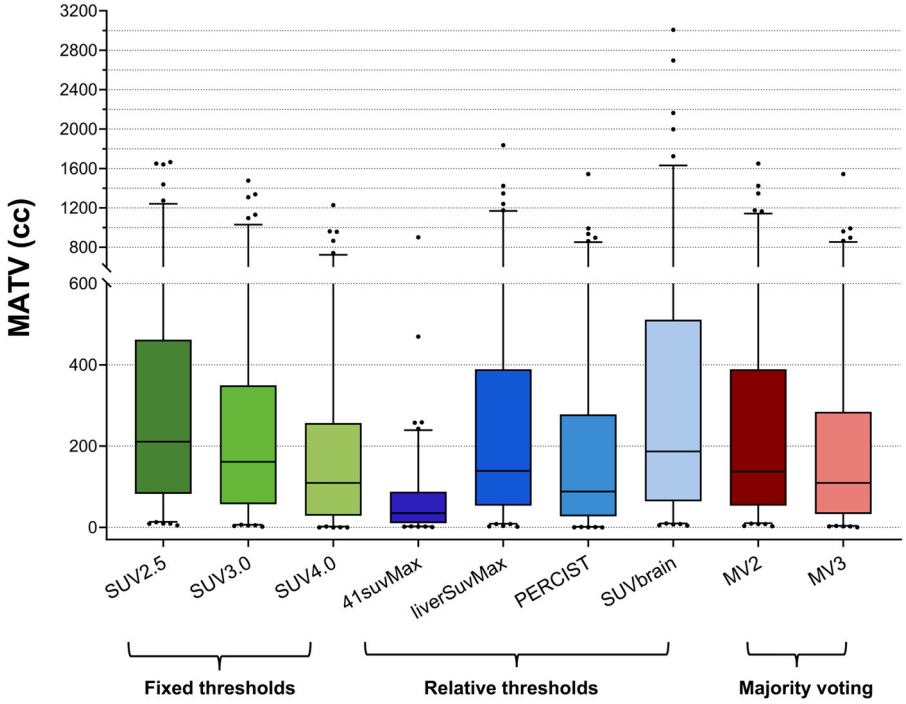

Fig. 5. Boxplots displaying the median and lower- and upper-quartiles for metabolic active tumor volumes (MATVs) retrieved with every segmentation method ($n = 110$ per boxplot).

41suvMax differs the most from all other methods, which is in concordance with the findings in Figs. 5 and 6.

3.3.5. Segmentation results evaluation

Example segmentation results created with the MUST-segmenter are shown in Fig. 8. Four tumors represented as maximum intensity projected SUV images (scale 0–10) with the corresponding FDG-tumor per SUV segmentation threshold. Tumor **A** shows consistent FDG uptake, tumor **B** shows high and low uptake, tumor **C** shows necrosis, thus having a central part with almost no uptake. Tumor **D** shows fluctuations in uptake.

4. Discussion

During the treatment of NHL patients, a wide range of options are available, especially in case of refractory or relapsed disease. Improving response and outcome predictions is therefore essential to further optimize and individualize treatment. The MATV metric has shown to be a prognostic factor in NHL [6–12,48]. However, there is currently no optimal method to determine MATV by segmentation of ^{18}F-FDG PET images [18], resulting in utilization of a large variety of methods in clinical studies [19–24]. As NHL is often widespread, MATV extraction and research on semi-automated segmentation methods are labor-intensive processes.

Several of the reviewed studies advised various thresholding methods to utilize in future research [35,37–39,41,42,44,46], basing their recommendations on different evaluation criteria. Studies that recommended SUV2.5 over relative thresholding methods [39,41,44] did not include SUV4.0 in their evaluation, which if included, was often preferred [35,37,38,42]. Interestingly, the 41suvMax method recommended by the EANM for solid tumors [15] was included in all studies, but only the preferred method in a single study, and was in that study based on comparisons to manual segmentations by a clinician [46]. Other potentially promising relative thresholding methods, such as the liver-based threshold methods (liverSuvMax, PRECIST), were limitedly investigated, yet recommended by one study [39].

The predictive performance of survival or disease progression was evaluated in relatively small patient cohorts for modelling (median = 118) [36,38,39,41,43–45], which might make the performance outcomes nongeneralizable. Hence, evaluation of segmentation methods for the predictive performance needs to be investigated in larger NHL patient cohorts in the future.

Optimal cut-offs to determine high-MATV patient groups differed between segmentation methods in 4 studies [39,43–45]. These cut-offs were only applied to their corresponding methods, resulting in no inherent value since there is a need for a generic cut-off independent of the method. This indicates that determination of high-MATV groups is still dependent on the segmentation methodology and suggests the need for a standardized method to measure MATV.

Ferrández et al. [35] recommended SUV4.0 based on robustness among image reconstruction methods, yet they included patients with four different lymphoma types, which may influence the conclusion. Barrington et al. [38] observed no differences based on image reconstruction. Ilyas et al. [44] evaluated robustness across software platforms, which resulted in the expected conclusion that segmentation algorithms are similar among software.

Recommendations based on quality scores [37,38,42], subjective observations from clinicians, suggested that the SUV4.0 was visually the best performing method. Nevertheless, the studies also showed that these measurements are sensitive to inter-observer variability; moreover, the visual best method may not always have the best predictive performance, as indicated by Eertink et al. [36].

In conclusion, there is no consistent recommended thresholding method based on the literature of PET-based segmentation method comparisons in predictive performance, robustness, or clinician-based evaluation. The best method may also depend on the specific purpose of the segmentation utilization (e.g., whether this is for prediction or segmentation for radiation). Thus, the need for a standardized FDG-PET segmentation method remains. Although there are several segmentation software available, they do not contain all the threshold methods discovered in the literature review. The MUST-segmenter presented in the current study includes

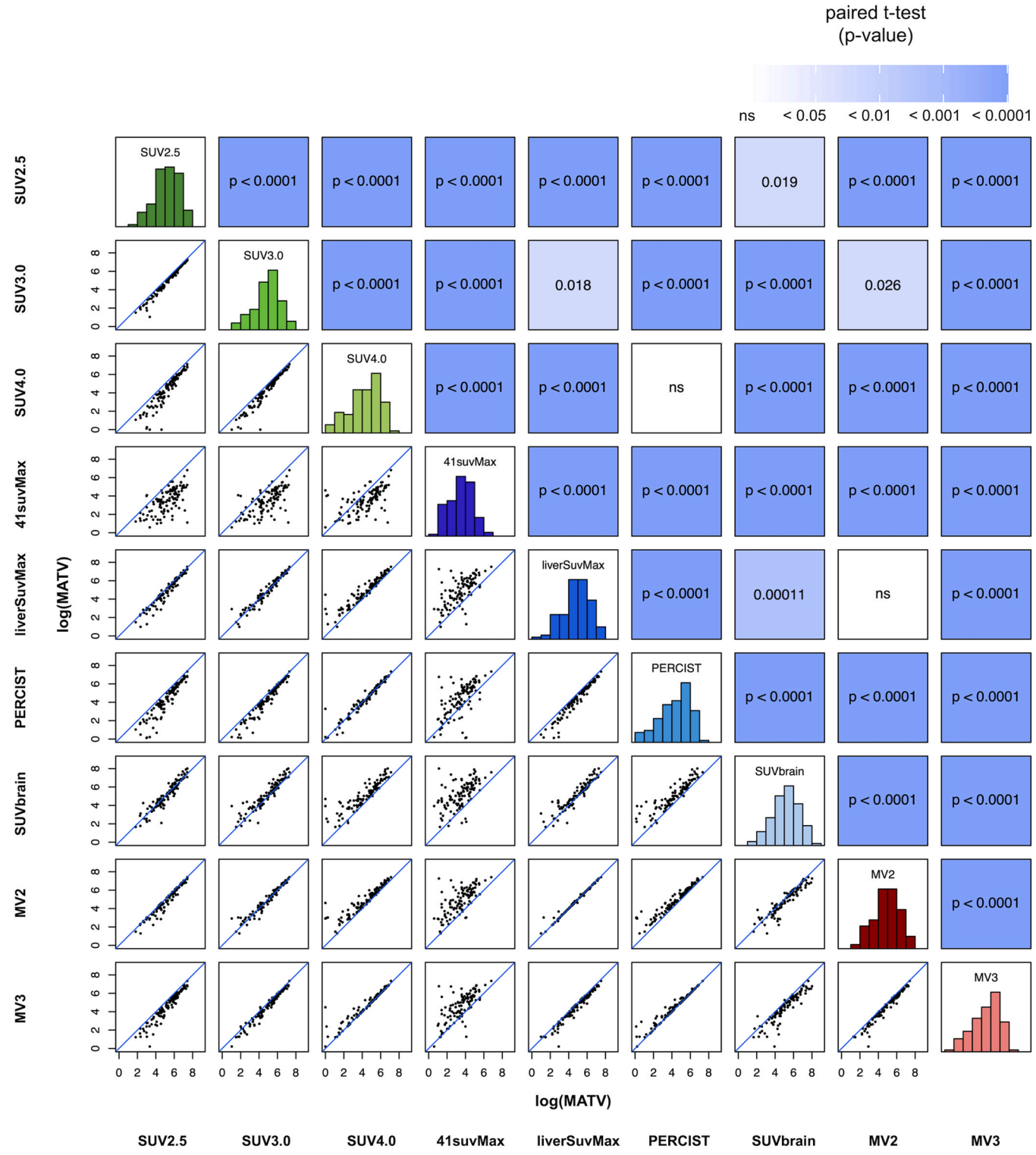

Fig. 6. Distributions, correlations, and paired t-test results (*p*-values) of log-transformed metabolic active tumor volumes (MATVs) retrieved with MUST-segmenter using different segmentation thresholds.

these methods with the ability to apply them at once, hence research to find an optimal segmentation method can be facilitated.

The MUST-segmenter is implemented as an open-source extension for 3D Slicer, where users can extract MATV (and other image biomarkers) based on multiple SUV thresholding methods by a one-click approach. Avoidance regions can be excluded from the segmentation results, without the need to manually remove physiological uptake areas, which was frequently required in our reviewed studies [35,36,38,43,44]. The first user-experience of the MUST-segmenter indicated that it is an efficient and user-friendly segmentation tool, as no manual alterations were needed, and multi-threshold segmentation can be applied to larger patient cohorts with multiple tumor localizations.

Although the MUST-segmenter is a free, accessible application, some limitations considering time-efficiency come with it. Most important is that the MUST-segmenter is implemented in 3D Slicer,

Table 4
Percentages of patients classified as low- and high- metabolic active tumor volume (MATV) (total $n = 110$) identified with two MATV cut-offs (200 and 500 cc).

	***n* = 110**			
	Cut-off 1 200 cc		**Cut-off 2 500 cc**	
	Low-MATV	**High-MATV**	**Low-MATV**	**High-MATV**
SUV2.5	49	51	77	23
SUV3.0	54	46	84	16
SUV4.0	65	35	89	11
41suvMax	95	5	99	1
liverSuvMax	55	45	81	19
PERCIST	77	33	89	11
SUVbrain	52	48	75	25
MV2	54	46	82	18
MV3	64	36	88	12

requiring a learning curve before the ability to use the actual tool. The segmentation process itself can be cumbersome in the case of widespread disease, as it is necessary to place seeds on all the identified tumors; however, this is a general drawback in seed-based segmentation. As found by our indication of segmentation process time, for a patient with 12 lesions it takes approximately *6 minutes* to create segmentations and extract MATV using 3 SUV thresholds, compared to *29 minutes* using ACCURATE.

No inter-observer variability was included in the current study, which is a well-known problem in manual delineations due to the bias in visual assessment of FDG uptake compared to differences in background levels, also referred to as contrast illusion [5]. However, this less of an issue for objective segmentation methods.

The MATVs determined in the cohort of NHL patients significantly differed among almost all methods, demonstrating the differences in SUV thresholding methods. The 41suvMax method showed the largest difference compared to other methods and produced the smallest MATVs (median 35 cc), which was also stated by Barrington et al. [38] in a DLBCL patient cohort. The comparisons of proportions of low- and high-MATV patient groups determined with two cut-offs, 200 and 500 cc, showed similar results. Nearly all proportions differed from each other meaning that the different approaches of determining MATV have a clinical impact. Herewith, the 41suvMax differed the most from the other methods; due to this method creating the lowest MATVs, patients included in the high-MATV group were minimal. A study by Driessen et al. [49] found a largest MATV retrieved with 41suvMax compared to other thresholding methods, explained by lower SUV values in their HL patient cohort resulting into a lower threshold. Due to common tumor heterogeneity in NHL, the 41suvMax method may not be very applicable in NHL.

Boxplots of MATVs (Fig. 5) demonstrated that distributions of liverSuvMax and MV2 are very similar, which also holds for SUV4.0, PERCIST and MV3. Similarity between SUV4.0 and MV3 can be explained by the fact that the majority voting methods, that are based on both SUV2.5 and SUV4.0, will always include all voxels identified by SUV4.0 as they are always included in SUV2.5 tumor segmentation [35–38,42]. Furthermore, the MATVs showed visually high correlation among the different threshold methods (Fig. 6), as was found in 3 of the reviewed studies [37,38,44]. Both fixed thresholding methods and majority voting methods were strongly correlated, as expected.

The MUST-segmenter architecture allows for additions and adjustment, making it adaptable to future unforeseen needs. For example, the MUST-segmenter could be further optimized by including CT images in tumor segmentation. SUV thresholding segmentation causes inclusion of only high-uptake tumor parts, while it may miss tumor regions with low FDG-PET uptake, such as necrotic regions (Fig. 8). Depending on the application of the segmentation, this can be important to include. More research is needed to determine whether exclusion of low FDG-avid tumor regions matter for NHL outcome predictions [50]. To incorporate CT information in the segmentation, deep learning approaches are considered promising [51]. Additionally, depending on the outcome, different segmentation methods could be optimal [40], but also alternatives to the

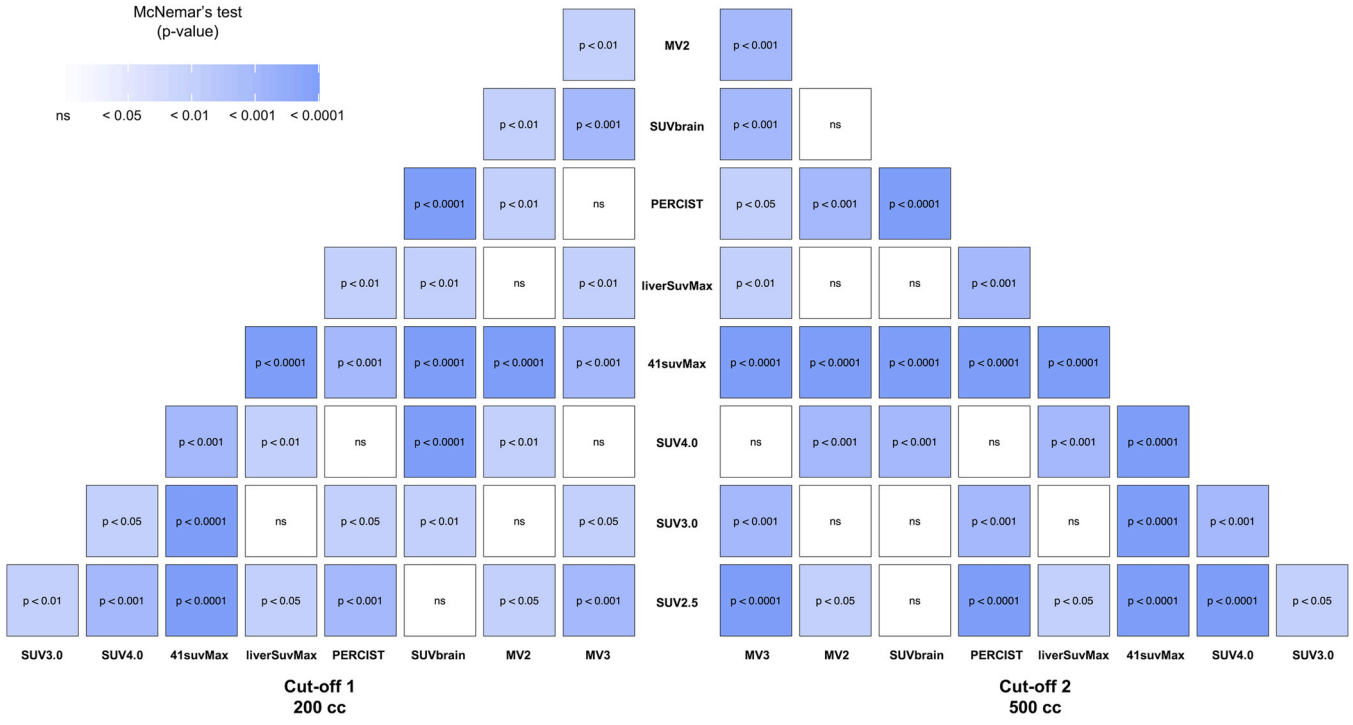

Fig. 7. Significance (McNemar's test) heatmap of differences between low- and high-MATV patient proportions determined with all segmentation methods, using two cut-offs (200 and 500 cc).

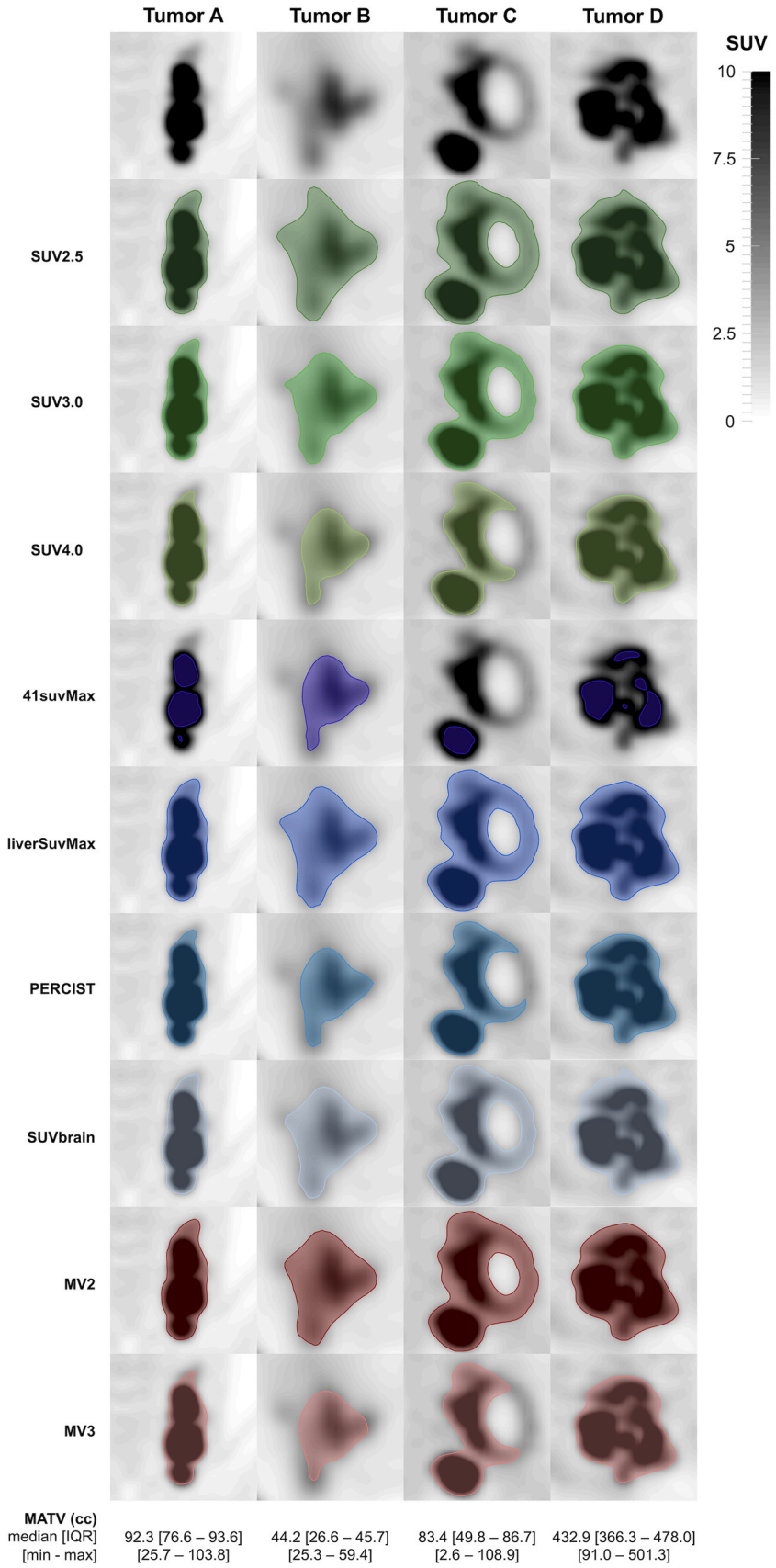

Fig. 8. Example tumor segmentation results for multiple SUV thresholds, retrieved with the MUST-segmenter. Four tumors represented as maximum intensity projected SUV images (scale 0–10) with corresponding segmentation results as a contour.

MATV could be of interest. Other ^{18}F-FDG PET features for lymphoma are reported for outcome predictions and the segmentation method may also have an impact on these outcomes [52,53]. As already indicated by the MATV cut-off analysis, the segmentation method influences the proportions of patients classified in the low- and high-MATV groups. More features can be extracted via the MUST-segmenter and it has the functionality to add more.

5. Conclusions

Based on the systematic literature review, there is no preferred segmentation method for patients with NHL available. The MUST-segmenter makes it possible to extract MATVs using different thresholding methods. Significant differences in MATVs between the segmentation methods were observed, suggesting differences in outcome prediction of target delineation results. To further optimize treatment in NHL patients, research on finding an optimal segmentation method to predict response and treatment outcomes needs to be continued, and can be more readily researched with the resulting MUST-segmenter.

CRediT authorship contribution statement

Kylie Keijzer: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Anne G.H. Niezink:** Conceptualization, Methodology, Resources, Writing – review & editing, Project administration, Supervision. **Janneke W. de Boer:** Resources, Data curation, Writing – review & editing. **Jaap A. van Doesum:** Resources, Data curation. **Walter Noordzij:** Resources, Writing – review & editing. **Tom van Meerten:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Lisanne V. van Dijk:** Conceptualization, Methodology, Formal analysis, Resources, Writing – review & editing, Project administration, Supervision.

Conflict of Interest

The Department of Radiation Oncology has research collaborations with Elekta, IBA, RaySearch, Siemens, Mirada, Bergoz Instrumentation and Medical Data Works, Genentech.

Dr. van Meerten reported research grants from Genentech, Celgene/BMS; personal fees from Kite/Gilead and Janssen (advisory boards); and honoraria from Kite/Gilead and Celgene/BMS (speaker fees) outside the submitted work.

Acknowledgements

Dr. van Dijk, received/receives funding and salary support unrelated to this project from NWO ZonMw via the VENI (NWO-09150162010173) Individual career development grant and KWF Dutch Cancer Society Young Investigator Grant (KWF-13529).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.01.023.

References

[1] Ceriani L, Gritti G, Cascione L, Pirosa MC, Polino A, Ruberto T, et al. SAKK38/07 study: integration of baseline metabolic heterogeneity and metabolic tumor volume in DLBCL prognostic model. Blood Adv 2020;4:1082–92. https://doi.org/10.1182/BLOODADVANCES.2019001201

[2] Friedberg JW. Relapsed/refractory diffuse large B-cell lymphoma. Hematology 2011;2011:498–505. https://doi.org/10.1182/ASHEDUCATION-2011.1.498

[3] Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET scan in lymphoma. Https://DoiOrg/101080/10428190903040048 2009;50:1257–60. https://doi.org/10.1080/10428190903040048

[4] Kluge R, Chavdarova L, Hoffmann M, Kobe C, Malkowski B, Montravers F, et al. Inter-reader reliability of early FDG-PET/CT response assessment using the deauville scale after 2 cycles of intensive chemotherapy (OEPA) in Hodgkin's Lymphoma. PLoS One 2016;11:e0149072https://doi.org/10.1371/JOURNAL.PONE.0149072

[5] Hasenclever D, Kurch L, Mauz-Körholz C, Elsner A, Georgi T, Wallace H, et al. qPET - a quantitative extension of the Deauville scale to assess response in interim FDG-PET scans in lymphoma. Eur J Nucl Med Mol Imaging 2014;41:1301–8. https://doi.org/10.1007/S00259-014-2715-9/FIGURES/5

[6] Burggraaff CN, Eertink JJ, Lugtenburg PJ, Hoekstra OS, Arens AIJ, Keizer B de, et al. 18F-FDG PET improves baseline clinical predictors of response in diffuse large B-cell lymphoma: The HOVON-84 study. (jnumed). J Nucl Med 2021;121:262205https://doi.org/10.2967/JNUMED.121.262205

[7] Mikhaeel NG, Smith D, Dunn JT, Phillips M, Møller H, Fields PA, 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;43:1209. https://doi.org/10.1007/S00259-016-3315-7

[8] Reinert CP, Perl RM, Faul C, Lengerke C, Nikolaou K, Dittmann H, et al. Value of CT-textural features and volume-based PET parameters in comparison to serologic markers for response prediction in patients with diffuse large B-cell lymphoma undergoing CD19-CAR-T cell therapy. Vol 11, Page 1522 2022 J Clin Med 2022;11:1522. https://doi.org/10.3390/JCM11061522

[9] Schmitz C, Hüttmann A, Müller SP, Hanoun M, Boellaard R, Brinkmann M, 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. https://doi.org/10.1016/J.EJCA.2019.09.027

[10] Aide N, Fruchart C, Nganoa C, Gac AC, Lasnon C. Baseline 18F-FDG PET radiomic features as predictors of 2-year event-free survival in diffuse large B cell lymphomas treated with immunochemotherapy. Eur Radio 2020;30:4623–32. https://doi.org/10.1007/S00330-020-06815-8/FIGURES/3

[11] Cottereau, Becker S, Broussais F, Casasnovas O, Kanoun S, Roques M, et al. Prognostic value of baseline total metabolic tumor volume (TMTV0) measured on FDG-PET/CT in patients with peripheral T-cell lymphoma (PTCL. Ann Oncol 2016;27:719–24. https://doi.org/10.1093/ANNONC/MDW011

[12] Meignan M, Cottereau AS, Versari A, Chartier L, Dupuis J, Boussetta S, et al. Baseline metabolic tumor volume predicts outcome in high-tumor-burden follicular lymphoma: a pooled analysis of three multicenter studies. J Clin Oncol 2016:34. https://doi.org/10.1200/JCO.2016.66.9440

[13] Mikhaeel NG, Heymans MW, Eertink JJ, de Vet HCW, Boellaard R, Dührsen U, et al. Proposed new dynamic prognostic index for diffuse large B-cell lymphoma: international metabolic prognostic index. J Clin Oncol 2022;40:2352. https://doi.org/10.1200/JCO.21.02063

[14] Paulino AC, Koshy M, Howell R, Schuster D, Davis LW. Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer. Int J Radiat Oncol*Biol*Phys 2005;61:1385–92. https://doi.org/10.1016/J.IJROBP.2004.08.037

[15] Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging 42. 2015. p. 328–54. https://doi.org/10.1007/S00259-014-2961-X/TABLES/2

[16] Mičušík B, Hanbury A. Automatic image segmentation by positioning a seed. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 3952 LNCS 2006:468–80. https://doi.org/10.1007/11744047_36/COVER

[17] Ilyas H, Mikhaeel NG, Dunn JT, Rahman F, Møller H, Smith D, 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;45:1142–54. https://doi.org/10.1007/S00259-018-3953-Z/TABLES/3

[18] Barrington SF, Meignan M. Time to prepare for risk adaptation in lymphoma by standardizing measurement OF METABOLIC TUMOr Burden. J Nucl Med 2019;60:1096–102. https://doi.org/10.2967/jnumed.119.227249

[19] Mikhaeel NG, Smith D, Dunn JT, Phillips M, Møller H, Fields PA, 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;43:1209. https://doi.org/10.1007/S00259-016-3315-7

[20] Kurtz DM, Green MR, Bratman S v, Scherer F, Liu CL, Kunder CA, et al. Lymphoid neoplasia: noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. Blood 2015;125:3679. https://doi.org/10.1182/BLOOD-2015-03-635169

[21] Kostakoglu L, Martelli M, Sehn LH, Belada D, Carella A-M, Chua N, et al. Baseline PET-derived metabolic tumor volume metrics predict progression-free and overall survival in DLBCL after first-line treatment: results from the phase 3 GOYA Study. Blood 2017;130:824. https://doi.org/10.1182/BLOOD.V130.SUPPL_1.824.824

[22] Cottereau AS, Lanic H, Mareschal S, Meignan M, Vera P, Tilly H, 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;22:3801–9. https://doi.org/10.1158/1078-0432.CCR-15-2825

[23] Ceriani L, Milan L, Martelli M, Ferreri AJM, Cascione L, Zinzani PL, et al. Metabolic heterogeneity on baseline 18FDG-PET/CT scan is a predictor of outcome in primary mediastinal B-cell lymphoma. Blood 2018;132:179–86. https://doi.org/10.1182/BLOOD-2018-01-826958

[24] Cottereau, Becker S, Broussais F, Casasnovas O, Kanoun S, Roques M, et al. Prognostic value of baseline total metabolic tumor volume (TMTV0) measured on FDG-PET/CT in patients with peripheral T-cell lymphoma (PTCL. Ann Oncol 2016;27:719–24. https://doi.org/10.1093/ANNONC/MDW011
[25] Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 2012;30:1323–41. https://doi.org/10.1016/J.MRI.2012.05.001
[26] van Rossum G, Drake FL. Python 3 Reference Manual; CreateSpace. Scotts Val, CA 2009:242.
[27] 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;50:122S. https://doi.org/10.2967/JNUMED.108.057307
[28] Moeller JR, Ishikawa T, Dhawan V, Spetsieris P, Mandel F, Alexander GE, et al. The metabolic topography of normal aging. J Cereb Blood Flow Metab 1996;16:385–98. https://doi.org/10.1097/00004647-199605000-00005
[29] Bentourkia M, Bol A, Ivanoiu A, Labar D, Sibomana M, Coppens A, et al. Comparison of regional cerebral blood flow and glucose metabolism in the normal brain: effect of aging. J Neurol Sci 2000;181:19–28. https://doi.org/10.1016/S0022-510X(00)00396-8
[30] Willis MW, Ketter TA, Kimbrell TA, George MS, Herscovitch P, Danielson AL, et al. Age, sex and laterality effects on cerebral glucose metabolism in healthy adults. Psychiatry Res Neuroimaging 2002;114:23–37. https://doi.org/10.1016/S0925-4927(01)00126-3
[31] Waxman A.D., Herholz K., Lewis D.H., Herscovitch P., Minoshima S., Ichise M., et al. Society of Nuclear Medicine Procedure Guideline for FDG PET Brain Imaging n.d.
[32] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014 5:1 2014;5:1–9. https://doi.org/10.1038/ncomms5006
[33] Boellaard R, Oyen WJG, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging 2008;35:2320–33. https://doi.org/10.1007/S00259-008-0874-2/TABLES/6
[34] Multi Atlas M.R. contouring automation | Embrace MR n.d. ⟨https://mirada-medical.com/product/embrace-mr-multi-atlas-contouring/⟩.
[35] Ferrández MC, Eertink JJ, Golla SSV, Wiegers SE, Zwezerijnen GJC, Pieplenbosch S, et al. Combatting the effect of image reconstruction settings on lymphoma [18F]FDG PET metabolic tumor volume assessment using various segmentation methods. EJNMMI Res 2022;12:44. https://doi.org/10.1186/S13550-022-00916-9
[36] Eertink JJ, Pfaehler EAG, Wiegers SE, van de Brug T, Lugtenburg PJ, Hoekstra OS, et al. Quantitative radiomics features in diffuse large B-cell lymphoma: does segmentation method matter? J Nucl Med 2022;63:389–95. https://doi.org/10.2967/JNUMED.121.262117
[37] Zwezerijnen GJC, Eertink JJ, Burggraaff CN, Wiegers SE. Shaban EAIN, Pieplenbosch S, et al. Interobserver agreement on automated metabolic tumor volume measurements of deauville score 4 and 5 Lesions at Interim 18F-FDG PET in diffuse large B-Cell Lymphoma. J Nucl Med 2021;62:1531. https://doi.org/10.2967/JNUMED.120.258673
[38] Barrington SF, Zwezerijnen BGJC, de Vet HCW, Heymans MW, Mikhaeel NG, Burggraaff CN, 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;62:332. https://doi.org/10.2967/JNUMED.119.238923
[39] Eude F, Toledano MN, Vera P, Tilly H, Mihailescu SD, Becke S. Reproducibility of baseline tumour metabolic volume measurements in diffuse large b-cell lymphoma: is there a superior method? Metabolites 2021;11:1–16. https://doi.org/10.3390/METABO11020072
[40] Weisman AJ, Kieler MW, Perlman S, Hutchings M, Jeraj R, Kostakoglu L, et al. Comparison of 11 automated PET segmentation methods in lymphoma. Phys Med Biol 2020;65:235019https://doi.org/10.1088/1361-6560/ABB6BD
[41] Guzmán Ortiz S, Mucientes Rasilla J, Vargas Núñez JA, Royuela A, Navarro Matilla B, Mitjavila Casanovas M. Evaluation of the prognostic value of different methods of calculating the tumour metabolic volume with 18F-FDG PET/CT, in patients with diffuse large cell B-cell lymphoma. Rev Esp De Med Nucl e Imagen Mol (Engl Ed) 2020;39:340–6. https://doi.org/10.1016/J.REMNIE.2020.09.003
[42] Burggraaff CN, Rahman F, Kaßner I, Pieplenbosch S, Barrington SF, Jauw YWS, et al. Optimizing workflows for fast and reliable metabolic tumor volume measurements in diffuse large B cell lymphoma. Mol Imaging Biol 2020;22:1102. https://doi.org/10.1007/S11307-020-01474-Z
[43] Gormsen LC, Vendelbo MH, Pedersen MA, Haraldsen A, Hjorthaug K, Bogsrud TV, et al. A comparative study of standardized quantitative and visual assessment for predicting tumor volume and outcome in newly diagnosed diffuse large B-cell lymphoma staged with 18F-FDG PET/CT. EJNMMI Res 2019:9. https://doi.org/10.1186/S13550-019-0503-Z
[44] Ilyas H, Mikhaeel NG, Dunn JT, Rahman F, Møller H, Smith D, 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. https://doi.org/10.1007/S00259-018-3953-Z
[45] Cottereau AS, Hapdey S, Chartier L, Modzelewski R, Casasnovas O, Itti E, et al. Baseline total metabolic tumor volume measured with fixed or different adaptive thresholding methods equally predicts outcome in peripheral T cell lymphoma. 2017;58:276–81. https://doi.org/10.2967/JNUMED.116.180406
[46] Meignan M, Sasanelli M, Casasnovas RO, Luminari S, Fioroni F, Coriani C, et al. Metabolic tumour volumes measured at staging in lymphoma: Methodological evaluation on phantom experiments and patients. Eur J Nucl Med Mol Imaging 2014:41. https://doi.org/10.1007/S00259-014-2705-Y/FIGURES/5
[47] Boellaard R. Quantitative oncology molecular analysis suite: ACCURATE. J Nucl Med 2018;59. 1753–1753.
[48] Kanoun S, Rossi C, Berriolo-Riedinger A, Dygai-Cochet I, Cochet A, Humbert O, et al. Baseline metabolic tumour volume is an independent prognostic factor in Hodgkin lymphoma. Eur J Nucl Med Mol Imaging 2014;41:1735–43. https://doi.org/10.1007/S00259-014-2783-X/FIGURES/4
[49] Driessen J, Zwezerijnen GJC, Schöder H, Drees EEE, Kersten MJ, Moskowitz AJ, et al. The impact of semi-automatic segmentation methods on metabolic tumor volume, intensity and dissemination radiomics in 18F-FDG PET scans of patients with classical Hodgkin lymphoma. J Nucl Med 2022;63:1424–30. https://doi.org/10.2967/JNUMED.121.263067
[50] Song Q, Bai J, Han D, Bhatia S, Sun W, Rockey W, et al. Optimal Co-segmentation of tumor in PET-CT images with context information. IEEE Trans Med Imaging 2013;32:1685–97. https://doi.org/10.1109/TMI.2013.2263388
[51] Blanc-Durand P, Jégou S, Kanoun S, Berriolo-Riedinger A, Bodet-Milin C, Kraeber-Bodéré F, et al. Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. Eur J Nucl Med Mol Imaging 2021;48:1362–70. https://doi.org/10.1007/S00259-020-05080-7/FIGURES/3
[52] Rizzo A, Triumbari EKA, Gatta R, Boldrini L, Racca M, Mayerhoefer M, et al. The role of 18F-FDG PET/CT radiomics in lymphoma. Clin Transl Imaging 2021;9:589–98. https://doi.org/10.1007/S40336-021-00451-Y/TABLES/2
[53] Pfaehler E, Zhovannik I, Wei L, Boellaard R, Dekker A, Monshouwer R, et al. A systematic review and quality of reporting checklist for repeatability and reproducibility of radiomic features. Phys Imaging Radiat Oncol 2021;20:69–75. https://doi.org/10.1016/j.phro.2021.10.007