Inter-reader agreement of quantitative FDG PET/CT biomarkers in lymphoma: a multicentre evaluation of MTV, TLG and Dmax

(article starts on next page)

RESEARCH

Open Access

Inter-reader agreement of quantitative FDG PET/CT biomarkers in lymphoma: a multicentre evaluation of MTV, TLG and Dmax

Elin Trägårdh[1,2*], Malin Lewold[1,2], Jesus Lopez Urdaneta[3], Måns Larsson[4], Olof Enqvist[4,5], Sally F. Barrington[6], Mats Jerkeman[7,8], Lars Edenbrandt[3] and May Sadik[3]

Abstract

Background The Deauville score is a key prognostic factor in Hodgkin lymphoma (HL) and diffuse large B-cell lymphoma (DLBCL) during interim and end-of-treatment PET/CT evaluations. However, additional measurements, particularly at baseline, such as metabolic tumour volume (MTV), total lesion glycolysis (TLG), and the maximum distance between hypermetabolic lymphoma lesions (Dmax) may offer enhanced prognostic value. This study evaluates the inter-reader agreement of these metrics to assess their reliability across different physicians.

Methods This study included 117 patients with untreated HL or DLBCL who had baseline [^{18}F]fluorodeoxyglucose PET/CT scans. Nine nuclear medicine physicians independently segmented lymphoma lesions using the online platform Recomia (www.recomia.org), without specific instructions beyond identifying lymphoma-related lesions. MTV, TLG, and Dmax were calculated from these segmentations. MTV was defined as the summed volume in cm^3, TLG as MTV multiplied by SUVmean and Dmax as the distance between the centroids of the two farthest lesions, measured in the 3D reconstruction. Each patient was segmented by two physicians. Inter-reader agreement was assessed using Spearman correlation coefficients for continuous values and Cohen's kappa coefficient (κ) for dichotomized values (above/below median).

Results The mean age of the 117 patients was 50 years (standard deviation 19), 39% female. Median (± interquartile range) values were 321 (± 597) cm^3 for MTV, 2200 (± 4399) cm^3 for TLG, and 35 (± 50) cm for Dmax. Spearman correlations between readers were 0.97 for MTV, 0.98 for TLG and 0.72 for Dmax (all $p < 0.01$). Agreement on dichotomized values was 95.7% for MTV (κ = 0.91), 97.4% for TLG (κ = 0.95), 83.8% for Dmax (κ = 0.68).

Conclusions MTV and TLG demonstrated good inter-reader reliability, even without standardized segmentation protocols. In contrast, Dmax showed moderate variability. These findings support the robustness of MTV and TLG as quantitative biomarkers. For Dmax to be clinically reliable, clearer segmentation guidelines are essential. Especially,

*Correspondence:
Elin Trägårdh
Elin.tragardh@med.lu.se
Full list of author information is available at the end of the article

Trägårdh *et al. BMC Medical Imaging*   (2025) 25:368

Page 2 of 8

Background

[^{18}F]fluorodeoxyglucose ([^{18}F]FDG) positron emission tomography/computed tomography (PET/CT) has become an indispensable tool in the management of lymphoma, particularly Hodgkin lymphoma (HL) and diffuse large B-cell lymphoma (DLBCL) [1]. Among its many applications, interim and/or end-of-treatment PET/CT scans are routinely used to assess treatment response, with the Deauville five-point scale serving as a widely accepted assessment method [2]. While the Deauville score provides valuable prognostic information and informs treatment escalation or de-escalation strategies [3–7], it remains a relatively simplistic metric that may not fully reflect the underlying disease biology. Moreover, its use is confined to treatment response assessment, despite evidence suggesting that staging PET/CT also provides prognostic information [8, 9].

In recent years, there has been increasing interest in quantitative PET/CT biomarkers that may provide prognostic information at diagnosis and complement the predictive capabilities of the Deauville score. Parameters such as metabolic tumour volume (MTV), total lesion glycolysis (TLG), and the maximum distance between hypermetabolic lesions (Dmax) have potential to refine risk stratification and predict outcomes in both HL and DLBCL [10–18]. These parameters provide a detailed characterization of tumour burden and spatial dissemination, potentially enhancing clinical decision-making at baseline and during follow-up.

However, the clinical adoption of these quantitative metrics depends on their reproducibility across different readers and institutions. Variability in lesion segmentation, particularly in the absence of standardized protocols, can significantly impact the reliability of MTV, TLG, and Dmax measurements. This issue is especially relevant in multicentre trials and real-world clinical settings, where consistency in image interpretation is essential for accurate prognostication and treatment planning.

Despite growing interest, few studies have systematically evaluated the inter-reader agreement of these quantitative PET/CT metrics in a controlled setting. Evaluating the degree of agreement among nuclear medicine physicians in measuring MTV, TLG, and Dmax is essential for validating their robustness as biomarkers and guiding the development of standardized segmentation protocols.

This study aimed to address this gap by investigating the inter-reader agreement of MTV, TLG, and Dmax in a cohort of patients with newly diagnosed HL or DLBCL.

Using manual segmentations performed by multiple experienced nuclear medicine physicians via an online platform, we aimed to quantify the consistency of these measurements.

Methods

Patients

This retrospective study included 117 consecutive, newly diagnosed, and untreated patients who underwent staging with [^{18}F]FDG PET/CT at Sahlgrenska University Hospital, Gothenburg, Sweden, between 2017 and 2022. Of these, 48 patients with biopsy-confirmed HL, examined between 2017 and 2018, had been previously included in a published study [19]. The remaining 69 patients had biopsy-proven DLBCL and were examined between 2019 and 2022.

The study was approved by the Regional Ethics Committee in Gothenburg (#2019 – 01274), and the need for written informed consent was waived. The study was performed according to the principles of the Declaration of Helsinki and its later amendments.

PET/CT imaging

Patients fasted for at least 6 h prior to the intravenous administration of [^{18}F]FDG. Adults received a dose of 4 MBg/kg (maximum 400 MBq), while pediatric doses were administered according to the EANM Dosage Card (version 5.7.2016). After a 60-minute uptake time, imaging was performed using one of the following integrated PET/CT systems: Siemens Biograph 64 TruePoint, GE Healthcare Discovery MI 5R or GE Healthcare Omni Legend 32. Whole-body PET/CT scans were acquired from the base of the skull to the mid-thigh, with acquisition times of 1–3 min per bed position, depending on the camera system. Images were corrected for scatter and attenuation. A low-dose CT scan (64-slice helical, 120 kV, 30 mAs, 512×512 matrix) was obtained and reconstructed with filtered back projection algorithm.

Image analysis

Quantitative PET metrics (MTV, TLG and Dmax) were derived from manual segmentations. Segmentations were performed by nine nuclear medicine physicians (eight specialists and one resident) from eight different hospitals. Readers were blinded to all clinical data except for the diagnosis of newly diagnosed HL or DLBCL. Lesions were segmented according to published recommendations [20], including:

1. Viable lymph nodes with increased FDG uptake;
2. Focal splenic uptake, regardless of spleen size;
3. Focal uptake in the bone marrow or other extra-nodal sites;
4. Diffuse splenic uptake exceeding liver uptake (spleen/liver ratio > 1.5), in the absence of reactive bone marrow uptake (bone marrow/liver ratio < 1.0).

Segmentations were performed using the cloud-based RECOMIA platform (www.recomia.org), which provided access to CT, PET, fused PET/CT, and maximum intensity projection images [21]. Readers could navigate in sagittal, coronal, and axial planes, adjust PET colour scales and SUV thresholds, and modify CT window settings (e.g. soft tissue, lung, bone). The physicians first visually assessed the scans, then performed segmentations in all slices with suspected lymphoma lesions, using the criteria listed above. No preferred segmentation method was suggested to the physicians. The RECOMIA platform supports fully manual segmentation and segmentation using a manually selected SUV threshold; all segmentations were performed slice by slice. Each patient was independently segmented by two randomly assigned readers, referred to as Reading A and Reading B. A single reader could serve as Reading A for one case and Reading B for another. Six of the readers segmented 12 cases each, and the remaining three readers each segmented 28, 53 and 81 cases, respectively.

MTV (cm³) was defined as the total volume of all voxels labelled as lymphoma. TLG (cm³) was calculated by multiplying the mean standardized uptake value (SUVmean) by the MTV for each lesion and summing across all lesions. Dmax (cm) was defined as the Euclidean distance between the centroids of the two most widely separated lesions in three-dimensional space and was calculated based on the manual segmentations. If one or no lesions were identified, Dmax was set to zero.

Statistical analysis
Descriptive statistics (mean, standard deviation (SD) and range, or median and inter-quartile range (IQR)) were used to summarize the data. Spearman's correlation coefficients and 95% confidence intervals (CI) were computed to evaluate the linear relationship between the two manual readings. The mean difference (bias) and 95% limits of agreement ($\pm 1.96 \times$ SD of the differences) between Readings A and B for MTV, TLG, and Dmax were calculated, and Bland-Altman plots were created. To further assess agreement between the readings, the metrics MTV, TLG and Dmax were split by the median and the percentage agreement of values below/above the median were calculated, as well as the Cohen's kappa coefficient (κ). This method was used to illustrate agreement of a metric being classified as above or below a potential cutoff value for the two Readings. All calculations were also performed when excluding lesions smaller than 3 cm³, to evaluate the impact on inclusion/exclusion of small lesions. A p-value < 0.05 was considered statistically significant.

Results
Patient characteristics
In the total cohort ($n = 117$), mean age was 50 years ($\pm$ SD 19, range 7–90), and 39% were female. Among the 48 patients with HL, the mean age was 37 years ($\pm$ SD 19, range 7–75), with 46% female. In the 69 patients with DLBCL, the mean age was 59 years ($\pm$ SD 14, range 17–90), and 35% were female.

Inter-reader agreement for all lesions
The median ($\pm$ IQR) values for MTV were 336 ± 591 cm³ for Reading A and 307 ± 614 cm³ for Reading B. For TLG, the corresponding values were 2235 ± 4274 cm³ and 2133 ± 4488 cm³, respectively. For Dmax, the values were 34 ± 48 cm for Reading A and 35 ± 52 cm for Reading B.

Quantitative PET/CT metrics generally showed high consistency between the two independent readings. Spearman correlation coefficients demonstrated strong agreement for MTV (0.97, 95% CI 0.94–0.98) and TLG (0.98, 95% CI 0.97–0.99), but lower for Dmax (0.72, 95% CI 0.55–0.83), ($p < 0.01$ for all metrics). Figure 1 displays scatter and Bland-Altman plots illustrating the variability for MTV, TLG and Dmax. Overall, inter-reading bias was minimal, particularly for TLG, which is expected since the physicians segmenting a case were randomly assigned as Reading A or Reading B. However, the limits of agreement were notably wider for Dmax (Table 1).

When dichotomizing the metrics based on the median to assess agreement in classifying measurements as above or below the median (the median being used as an example of a cutoff) for the two Readings, the overall agreement was higher for MTV (95.7%, κ = 0.91) and TLG (97.4%, κ = 0.95) than for Dmax (83.8%, κ = 0.68). Table 2 shows the number of patients classified as below or above the median for both Readings (agreement) as well as below the median for one of the Readings and above for the other Reading (non-agreement). The median value was 321 cm³ for MTV, 2200 cm³ for TLG and 35 cm for Dmax. Figure 2 shows a patient example with segmentations performed by two physicians, with large differences in Dmax and smaller differences in MTV and TLG.

Inter-reader agreement when excluding small lesions
When excluding lesions smaller than 3 cm³, the Spearman correlation coefficients for MTV and TLG remained similar to those obtained when all lesions were included (0.96, 95% CI 0.94–0.98 and 0.98, 95% CI 0.97–0.99, respectively), while the coefficient for Dmax increased
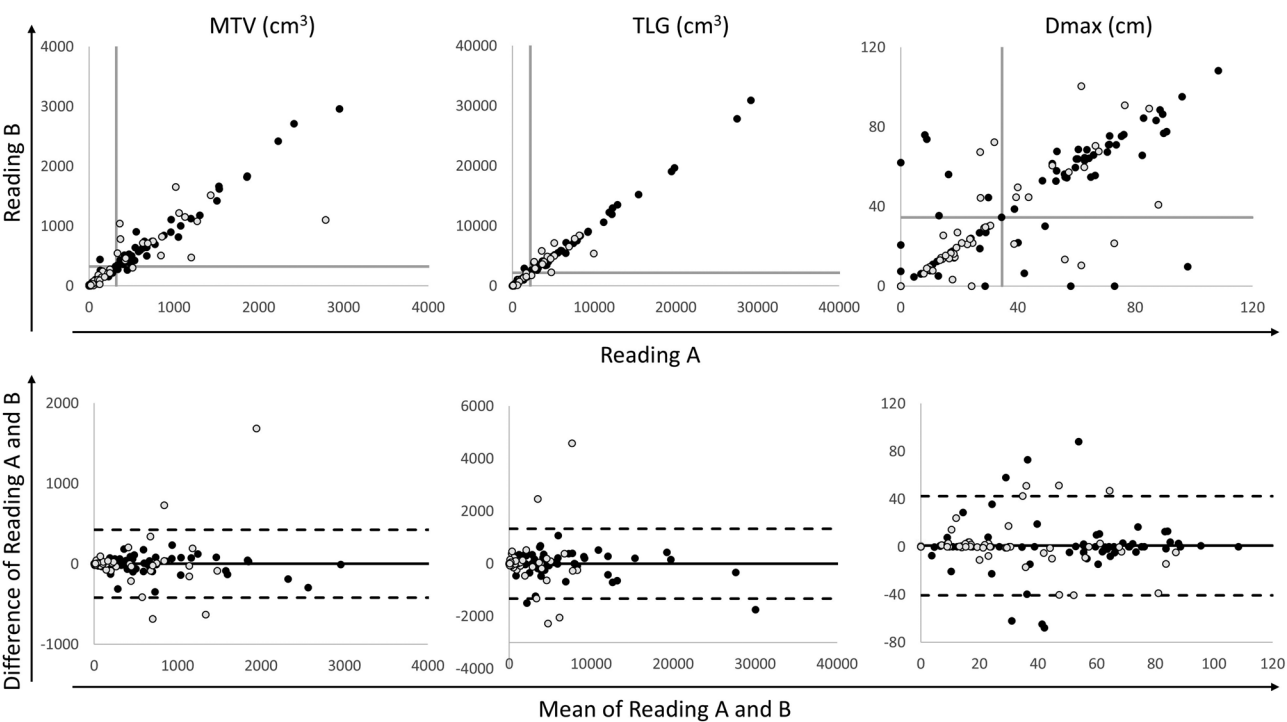
Fig. 1 Inter-reader agreement. Scatter (upper row) and Bland-Altman (lower row) plots illustrating inter-reader agreement for MTV, TLG and Dmax between Reading **A** and Reading **B**. White circles represent patients with HL and black circles represent patients with DLBCL. The grey lines in the scatter plots represent the median value for each respective metrics. The solid line in the Bland-Altman plots represents the bias and the dashed lines represent the limits of agreement

Table 1 Inter-reading bias and limits of agreement for MTV, TLG and Dmax between reading A and reading B

	Bias	Limits of agreement
MTV (cm³)	3.9	±422
TLG (cm³)	-1.6	±1329
Dmax (cm)	0.1	±41

Dmax – maximum distance between hypermetabolic tumour lesions, MTV – metabolic tumour volume, TLG – MTV x SUVmean

Table 2 The number of patients categorized as below/above the median (agreement vs. non-agreement when using the median as cutoff) for MTV, TLG and Dmax for the two readings

MTV		Reading B	
		Below median	Above median
Reading A	Below median	56	2
	Above median	3	56
TLG		Reading B	
		Below median	Above median
Reading A	Below median	57	1
	Above median	2	57
Dmax		Reading B	
		Below median	Above median
Reading A	Below median	49	10
	Above median	9	49

Dmax – maximum distance between hypermetabolic tumour lesions, MTV – metabolic tumour volume, TLG – MTV x SUVmean

(0.83, 95% CI 0.71–0.91). All metrics showed statistically significant correlations ($p < 0.01$). A detailed summary of the median ($\pm$ IQR), bias, and limits of agreement are provided in the Supplementary material.

When dichotomizing the metrics based on the median to assess agreement in classifying measurements as above or below the median for the two Readings, the overall remained unchanged for MTV (95.7%, $\kappa = 0.91$) and TLG (97.4%, $\kappa = 0.95$) compared to the analysis including all lesions. However, the agreement improved for Dmax (90.6%, $\kappa = 0.81$) when small lesions were excluded. Full results are available in the Supplementary material.

Discussion

This study evaluated the inter-reader agreement of the three quantitative FDG PET/CT biomarkers MTV, TLG, and Dmax in patients with newly diagnosed HL or DLBCL. The findings demonstrate high inter-reader agreement for MTV and TLG. In contrast, Dmax showed greater variability. Excluding small lesions led to improved agreement for Dmax; however, it remained lower than that observed for MTV and TLG. These results suggest that MTV and TLG are more robust and reproducible across readers than Dmax, even in the absence of standardized segmentation protocols. Since it can be anticipated that future use of the metrics in prognostic lymphoma models will be based on cutoff values,
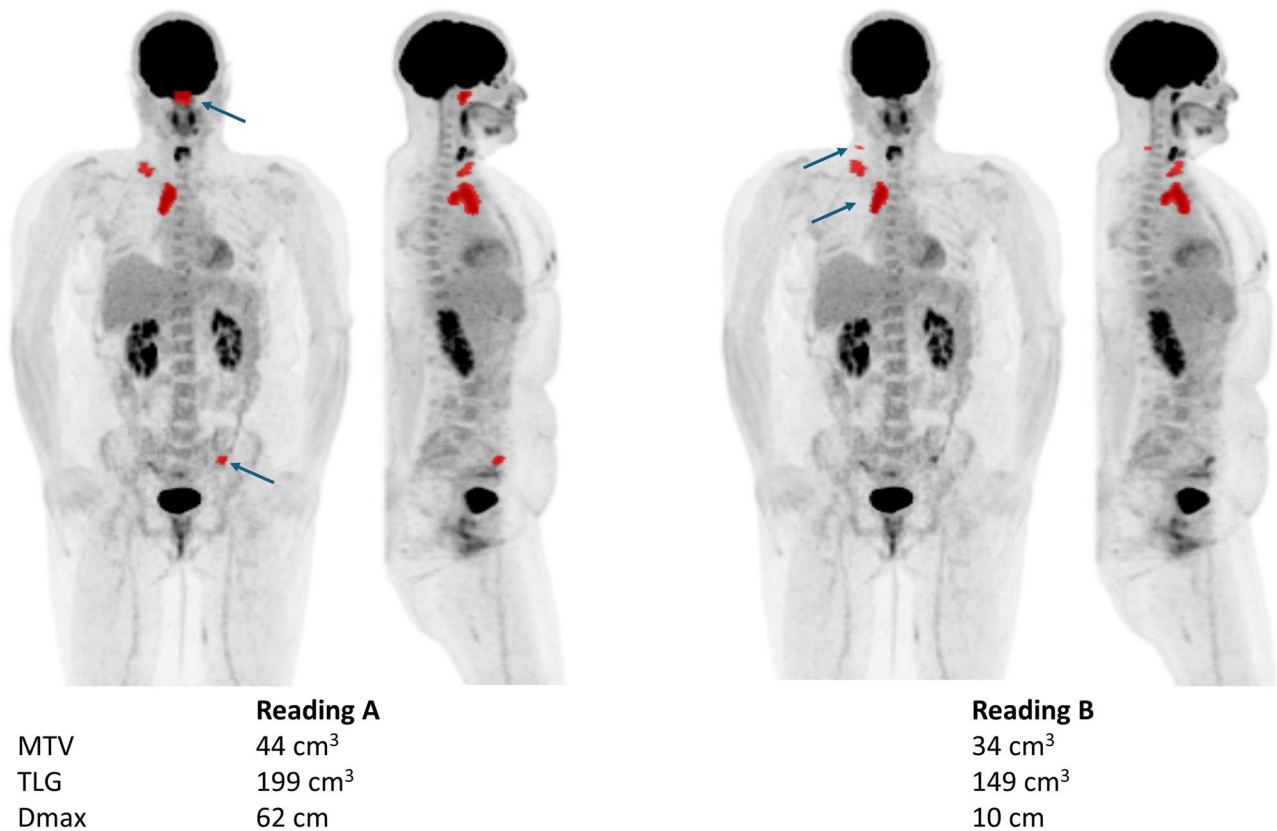
	Reading A	**Reading B**
MTV	44 cm^3	34 cm^3
TLG	199 cm^3	149 cm^3
Dmax	62 cm	10 cm

Fig. 2 Patient example. Maximum projection intensity PET images (frontal and sagittal views) from one patient with segmentations performed by two physicians (Reading **A** and **B**). The values of MTV, TLG and Dmax are shown below. The arrows point to the two segmentations used for the Dmax calculations

we investigated inter-reader agreement by dichotomizing MTV, TLG and Dmax based on the median values.

Quantitative PET/CT biomarkers are increasingly recognized for their potential to enhance risk stratification and guide treatment decisions in lymphoma. While the Deauville score remains the standard for response assessment, it is limited to post-treatment evaluations [2]. In contrast, MTV and TLG offer objective, volumetric assessments of tumour burden that can be applied at baseline and throughout treatment. The high reproducibility of MTV and TLG observed in this study supports their integration into clinical workflows and multicentre trials.

Dmax, which reflects spatial tumour dissemination, has shown prognostic value in previous studies [15–18]. However, its moderate inter-reader agreement in this study highlights the need for standardized segmentation guidelines if this metric is to be used reliably. This is particularly important given that small differences in lesion inclusion, especially for distant or borderline lesions, can disproportionately affect Dmax, without significantly altering MTV or TLG. Also, automated calculations of the centroids and distance between the centroids, as was done in this study, are crucial. If the Dmax had been measured manually, we believe the difference between the

readings would be even larger. As can be seen in Fig. 2, as well as in the results when excluding lesions smaller than 3 cm^3, inclusion or not of small spatially separated lesions have a large impact on Dmax.

Robustness across different imaging systems, reconstruction algorithms, and experience levels is essential for the clinical adoption of advanced PET/CT biomarkers. Various methods exist for delineation of MTV, including fixed SUV thresholds (e.g., SUV≥4.0) and relative thresholds (i.e. 41% of SUVmax in individual lesions). The use of different thresholds leads to different median, and consequently, to different optimal cutoffs to separate patients into high-and low-risk groups. The ideal threshold may vary depending on patient characteristics, lymphoma subtype, and treatment regimen, and should be tailored to specific clinical contexts [20]. In our study, readers were not instructed to use a specific segmentation method, yet MTV and TLG still showed strong agreement. This suggests that even greater consistency could be achieved with standardized or semi-automated approaches.

Manual segmentation is time-consuming, limiting its practicality in clinical workflows. Semi-automated or, preferably, fully automated methods are needed to improve efficiency, as well as to reduce observer dependency and

enhance reproducibility across institutions. Artificial intelligence (AI)-based methods can help with this task in the future. We have previously trained an AI-based tool to segment lymphoma lesions and compared the results with manual segmentations [22]. The AI-based tool could be used without major manual adjustments in 69% of HL patients. We are currently working on improving the method.

Previous studies have explored Inter-reader variability in the context of automated or semi-automated workflow. Burggraaff et al. [23] reported improved interobserver reliability for MTV and TLG using a fully automated pre-selection workflow compared to manual methods. Choi et al. [24] demonstrated high concordance in MTV and TLG measurements between expert readers using different methodologies, further supporting the value of semi-automated tools. Recently an international benchmark has been proposed for MTV measurement using SUV4 and a minimum volume of 3 cm^3 with high interobserver agreement with a publicly available benchmark dataset against which end-users can test their ability using local software to measure MTV within an acceptable range [25]. To our knowledge, no prior studies have specifically evaluated inter-reader variability for Dmax, making this study a novel contribution in that regard.

A major strength of this study is the inclusion of a relatively large and diverse patient cohort, along with multiple experienced nuclear medicine physicians from different institutions. This enhances the generalizability of the findings. The use of a cloud-based platform (RECOMIA) ensured consistent access to imaging tools and visualization settings, minimizing technical variability.

However, several limitations should be acknowledged. First, the segmentations were performed manually without standardized instructions beyond general recommendations, which may have introduced variability. Second, intra-reader variability was not assessed, which could provide additional insights into measurement consistency. Third, although the readers were instructed to segment all hypermetabolic lymphoma-related findings, the primary purpose of the segmentation was for MTV assessment in separate studies [19, 22]. As a result, small or distant lesions that minimally impact MTV or TLG but significantly affect Dmax may have been inconsistently included. Fourth, since the physicians were randomly assigned cases to segment, the segmentations of one physician may be in Reading A for some cases and in Reading B for some cases. Thus, assessment of "true" inter-reader variability was not possible – for example, the bias between Readings in this study close to 0, but would probably have been larger if one physician was consequently assigned as either Reading A or B. Also, the number of cases segmented by the physicians varied

between 12 and 81. In an ideal situation, the cases should have been equally distributed. Fifth, the median value was used as cutoff in this study, but in future prognostic models, other cutoff values are likely to be more relevant.

Conclusions

This study demonstrates that MTV and TLG are highly reproducible quantitative PET/CT biomarkers in lymphoma, supporting their reliability for use in both clinical practice and research settings. In contrast, the lower inter-reader agreement observed for Dmax highlights the need for standardized segmentation protocols to ensure its consistent application. These findings contribute to the growing body of evidence supporting the integration of quantitative imaging biomarkers into lymphoma management and highlight the importance of harmonization efforts to ensure reproducibility across institutions.

Abbreviations

[^{18}F]FDG	[^{18}F]fluorodeoxyglucose
AI	Artificial intelligence
DLBCL	Diffuse large B cell lymphoma
Dmax	Maximum tumour dissemination
HL	Hodgkin lymphoma
IQR	Inter-quartile range
MTV	Metabolic tumour volume
PET/CT	Positron emission tomography/computed tomography
SD	Standard deviation
TLG	Total lesion glycolysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01937-1.

Supplementary Material 1

Declarations

Ethics approval and consent to participate
This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Regional Ethics Committee in

Gothenburg (#2019 − 01274), and the need for written informed consent was waived.

Author details
[1]Department of Clinical Physiology and Nuclear Medicine, Skåne University Hospital, Inga Marie Nilssons g 47, Malmö 20502, Sweden
[2]Department of Translational Medicine, Wallenberg Center for Molecular Medicine, Lund University, Malmö, Sweden
[3]Department of Molecular and Clinical Medicine, Clinical Physiology, Sahlgrenska University Hospital, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[4]Eigenvision AB, Lund, Sweden
[5]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
[6]School of Biomedical Engineering and Imaging Sciences, King's College London and Guy's and St Thomas' PET Centre, Kings College, London, UK
[7]Department of Oncololgy, Skåne University Hospital, Lund, Sweden
[8]Department of Clinical Sciences, Lund University, Lund, Sweden

References
1. Barrington SF, Mikhaeel NG, Kostakoglu L, Meignan M, Hutchings M, Mueller SP, Schwartz LH, Zucca E, Fisher RI, Trotman J 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.
2. Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the first international workshop on Interim-PET-Scan in lymphoma. Leuk Lymphoma. 2009;50(8):1257–60.
3. Hutchings M, Loft A, Hansen M, Pedersen LM, Buhl T, Jurlander J, Buus S, Keiding S, D'Amore F, Boesen AM, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in hodgkin lymphoma. Blood. 2006;107(1):52–9.
4. Cerci JJ, Trindade E, Pracchia LF, Pitella FA, Linardi CC, Soares J Jr., Delbeke D, Topfer LA, Buccheri V, Meneghetti JC. Cost effectiveness of positron emission tomography in patients with hodgkin's lymphoma in unconfirmed complete remission or partial remission after first-line therapy. J Clin Oncol. 2010;28(8):1415–21.
5. Engert A, Haverkamp H, Kobe C, Markova J, Renner C, Ho A, Zijlstra J, Kral Z, Fuchs M, Hallek M, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. Lancet. 2012;379(9828):1791–9.
6. Dupuis J, Berriolo-Riedinger A, Julian A, Brice P, Tychyj-Pinel C, Tilly H, Mounier N, Gallamini A, Feugier P, Soubeyran P, et al. Impact of [(18)F]fluorodeoxyglucose positron emission tomography response evaluation in patients with high-tumor burden follicular lymphoma treated with immunochemotherapy: a prospective study from the groupe d'etudes des lymphomes de l'adulte and GOELAMS. J Clin Oncol. 2012;30(35):4317–22.
7. Trotman J, Fournier M, Lamy T, Seymour JF, Sonet A, Janikova A, Shpilberg O, Gyan E, Tilly H, Estell J, et al. Positron emission tomography-computed tomography (PET-CT) after induction therapy is highly predictive of patient outcome in follicular lymphoma: analysis of PET-CT in a subset of PRIMA trial participants. J Clin Oncol. 2011;29(23):3194–200.
8. Rigacci L, Vitolo U, Nassi L, Merli F, Gallamini A, Pregno P, Alvarez I, Salvi F, Sancetta R, Castagnoli A, et al. Positron emission tomography in the staging of patients with hodgkin's lymphoma. A prospective multicentric study by the intergruppo Italiano linfomi. Ann Hematol. 2007;86(12):897–903.
9. Le Dortz L, De Guibert S, Bayat S, Devillers A, Houot R, Rolland Y, Cuggia M, Le Jeune F, Bahri H, Barge ML, et al. Diagnostic and prognostic impact of 18F-FDG PET/CT in follicular lymphoma. Eur J Nucl Med Mol Imaging. 2010;37(12):2307–14.
10. van Heek L, Stuka C, Kaul H, Muller H, Mettler J, Hitz F, Baues C, Fuchs M, Borchmann P, Engert A, et al. Predictive value of baseline metabolic tumor volume in early-stage favorable hodgkin Lymphoma - Data from the prospective, multicenter phase III HD16 trial. BMC Cancer. 2022;22(1):672.
11. van Heek L, Weindler J, Gorniak C, Kaul H, Muller H, Mettler J, Baues C, Fuchs M, Borchmann P, Ferdinandus J, et al. Prognostic value of baseline metabolic tumor volume (MTV) for forecasting chemotherapy outcome in early-stage unfavorable hodgkin lymphoma: data from the phase III HD17 trial. Eur J Haematol. 2023;111(6):881–7.
12. Zucca E, Cascione L, Ruberto T, Facchinelli D, Schar S, Hayoz S, Dirnhofer S, Giovanella L, Bargetzi M, Mamot C, 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):715–25.
13. Manohar K, Mittal BR, Bhattacharya A, Malhotra P, Varma S. Prognostic value of quantitative parameters derived on initial staging 18F-fluorodeoxyglucose positron emission tomography/computed tomography in patients with high-grade non-Hodgkin's lymphoma. Nucl Med Commun. 2012;33(9):974–81.
14. Kim TM, Paeng JC, Chun IK, Keam B, Jeon YK, Lee SH, Kim DW, Lee DS, Kim CW, Chung JK, et al. Total lesion Glycolysis in positron emission tomography is a better predictor of outcome than the international prognostic index for patients with diffuse large B cell lymphoma. Cancer. 2013;119(6):1195–202.
15. Cui Y, Li Y, Hu W, Wu Z, Li S, Wang H. Evaluating deltamtv%, DeltaD(max)%, and %DeltaSUV(max) of (18)F-FDG PET/CT for mid-treatment efficacy and prognosis in diffuse large B-cell lymphoma. Discov Oncol. 2025;16(1):411.
16. Albano D, Bianchetti N, Talin A, Dondi F, Re A, Tucci A, Bertagna F. Prognostic role of pretreatment tumor burden and dissemination features from 2-[(18)F]FDG PET/CT in advanced mantle cell lymphoma. Hematol Oncol. 2025;43(1):e70009.
17. Wang F, Cui S, Lu L, Shao X, Yan F, Liu Y, He B, Wang J, Cao Y, Yue Y, et al. Dissemination feature based on PET/CT is a risk factor for diffuse large B cell lymphoma patients outcome. BMC Cancer. 2023;23(1):1165.
18. Dang J, Peng X, Wu P, Yao Y, Tan X, Ye Z, Jiang X, Jiang X, Liu Y, Chen S, et al. Predictive value of Dmax and %DeltaSUVmax of (18)F-FDG PET/CT for the prognosis of patients with diffuse large B-cell lymphoma. BMC Med Imaging. 2023;23(1):173.
19. Sadik M, López-Urdaneta J, Ulén J, Enqvist O, Andersson PO, Kumar R, Trägårdh E. Artificial intelligence increases the agreement among physicians classifying focal skeleton/bone marrow uptake in hodgkin's lymphoma patients staged with [(18)F]FDG PET/CT-a retrospective study. Nucl Med Mol Imaging. 2023;57(2):110–6.
20. 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):1096–102.
21. Tragardh E, Borrelli P, Kaboteh R, Gillberg T, Ulen J, Enqvist O, Edenbrandt L. RECOMIA-a cloud-based platform for artificial intelligence research in nuclear medicine and radiology. EJNMMI Phys. 2020;7(1):51.
22. Sadik M, Barrington SF, Tragardh E, Saboury B, Nielsen AL, Jakobsen AL, Gongora JLL, Urdaneta JL, Kumar R, Edenbrandt L. Metabolic tumour volume in hodgkin lymphoma-A comparison between manual and AI-based analysis. Clin Physiol Funct Imaging. 2024;44(3):220–7.
23. Burggraaff CN, Rahman F, Kassner I, Pieplenbosch S, Barrington SF, Jauw YWS, Zwezerijnen GJC, Muller S, Hoekstra OS, Zijlstra JM, et al. Optimizing workflows for fast and reliable metabolic tumor volume measurements in diffuse large B cell lymphoma. Mol Imaging Biol. 2020;22(4):1102–10.
24. Choi JW, Dean EA, Lu H, Thompson Z, Qi J, Krivenko G, Jain MD, Locke FL, Balagurunathan Y. Repeatability of metabolic tumor burden and lesion Glycolysis between clinical readers. Front Immunol. 2023;14:994520.

Trägårdh *et al. BMC Medical Imaging*   (2025) 25:368

Page 8 of 8

25. Boellaard R, Buvat I, Nioche C, Ceriani L, Cottereau AS, Guerra L, Hicks RJ, Kanoun S, Kobe C, Loft A, et al. International benchmark for total metabolic tumor volume measurement in baseline (18)F-FDG PET/CT of lymphoma patients: A milestone toward clinical implementation. J Nucl Med. 2024;65(9):1343–8.