Validation of a Multifocal Segmentation Method for Measuring Metabolic Tumor Volume in Hodgkin Lymphoma

Mariana R. Camacho[1], Elba Etchebehere[1,2], Natalia Tardelli[3], Marcia T. Delamain[4], Aline F.A. Vercosa[2], Maria E.S. Takahashi[5], Sergio Q. Brunetto[6], Irene G.H.L. Metze[3,4], Cármino A. Souza[3,4], Juliano J. Cerci[7], and Celso D. Ramos[1–3]

[1]MND Campinas, Campinas, Brazil; [2]Division of Nuclear Medicine, Department of Radiology, University of Campinas, Campinas, Brazil; [3]School of Medical Sciences, University of Campinas, Campinas, Brazil; [4]Division of Hematology and Hemotherapy, Department of Internal Medicine, University of Campinas, Campinas, Brazil; [5]Institute of Physics, University of Campinas, Campinas, Brazil; [6]Biomedical Engineering Center, University of Campinas, Campinas, Brazil; and [7]Quanta Diagnosis and Therapy Clinic, Curitiba, Brazil

Quantification of metabolic tumor volume (MTV) and total lesion glycolysis (TLG) can be time-consuming. We evaluated the performance of an automatic multifocal segmentation (MFS) method of quantification in patients with different stages of Hodgkin lymphoma, using the multiple VOI (MV) method as reference. **Methods:** This prospective bicentric study included 50 patients with Hodgkin lymphoma who underwent staging ^{18}F-FGD PET/CT. The examinations were centrally reviewed and processed with commercial MFS software to obtain MTV and TLG using 2 fixed relative thresholds (40% and 20% of SUV_{max}) for each lesion. All PET/CT scans were processed using the MV and MFS methods. Interclass correlation coefficients and Bland–Altman plots were used for statistical analysis. Repeated calculations of MTV and TLG values by 2 observers with different degrees of PET/CT imaging experience were used to ascertain interobserver agreement on the MFS method. **Results:** The means and SDs obtained for the MTV with MV and MFS were, respectively, 736 ± 856 mL and 660 ± 699 mL for the 20% threshold and 313 ± 359 mL and 372 ± 434 mL for the 40% threshold. The time spent calculating the MTV was much shorter with the MFS method than with the MV method (median time, 11.6 min [range, 1–30 min] and 64.4 min [range, 1–240 min], respectively), especially in patients with advanced disease. Time spent was similar in patients with localized disease. There were no statistical differences between the MFS values obtained by the 2 different observers. **Conclusion:** MTV and TLG calculations using MFS are reproducible, generate similar results to those obtained with MV, and are much less timing-consuming. Main differences between the 2 methods were related to difficulties in avoiding overlay of VOIs in the MV technique. MV and MFS perform equally well in patients with a small number of lesions.

Key Words: MTV; TLG; Hodgkin lymphoma; ^{18}F-FDG PET/CT; VOI; SUV

The role of metabolic tumor volume (MTV) and total lesion glycolysis (TLG), both obtained from PET/CT using ^{18}F-FDG, has been extensively debated in the literature in solid tumors, especially in lung neoplasms. Most of the studies show a correlation between those variables and patient prognosis (*1–5*). However, those metrics have not been adopted in clinical practice, mainly because of difficulties in the standardization of tumor segmentation (*6–9*). Another limitation is the difficulty in segmenting all lesions in patients with disseminated disease, such as advanced lymphomas. This process can be very time-consuming because multiple volumes of interest (VOIs) have to be drawn to include all sites of disease (*8–11*).

One of the most widely used methods for obtaining MTV and TLG is a fixed relative threshold method with multiple VOIs (MV method), which consists of manually drawing VOIs surrounding each metabolically active lesion (*6,7,9*). After determining the VOI, the software automatically defines the lesions' boundaries according to the selected threshold. For example, if the chosen threshold is 40%, the lesion limits are determined by selecting all voxels above 40% of the SUV_{max} inside the master VOI drawn around the lesion.

However, the MV method is considerably time-consuming, especially when performed on patients with disseminated diseases. Another difficulty of the MV method is that, when MVs are placed over the metabolically active lesions, overlap with areas of physiologic uptake of the radiotracer may occur.

Ideally, a software program should determine automatically and simultaneously, in a few seconds, all areas containing metabolically active lesions. This is what the multifocal segmentation method (MFS) proposes: after determining 1 VOI over the liver or mediastinum and drawing a master VOI around the entire body of the patient, all lesions are automatically drawn at the same time (*6,12–15*).

The main objective of this study was to evaluate the performance of the MFS method for quantification of MTV and TLG in patients with different stages of Hodgkin lymphoma, using the MV method as the reference standard.

MATERIALS AND METHODS

This was a prospective, bicentric study. It was approved by the local ethics committees (CAAE 07178612.0.1001.5405 and CAAE 45797615.1.0000.5404), and the requirement for written informed consent was waived.

Patients who underwent a staging ^{18}F-FDG PET/CT scan were studied. All 50 patients (35 from PET Center 1 [University of Campinas] and 15 from PET Center 2 [Quanta Diagnosis and Therapy Clinic]) had biopsy-proven Hodgkin lymphoma (28 female and 22 male patients; median age, 29 y; range, 3–84 y). The histologic subtypes were 35 cases of nodular sclerosis (70.0%), 6 cases of the lymphocyte-rich subtype (12.0%), 3 cases of the mixed-cellularity subtype (6.0%), 1 case of the lymphocyte-depleted subtype (2.0%), and 5 cases of unknown subtype (10.0%). Five patients (10%) were stage I, 15 (30%) were stage II, 7 (14%) were stage III, and 23 (46%) were stage IV.

All PET/CT scans were centrally reviewed at PET Center 1. For MTV and TLG measurements, all images were first processed using the MV method by the same experienced observer. To avoid bias from prior knowledge of patient image characteristics, images were processed using the MFS method by a different experienced observer. At least 2 mo apart, 34 PET Center 1 images were reprocessed using MFS by 2 different observers to ascertain interobserver agreement.

Image Acquisition

All patients fasted for at least 4–6 h before intravenous administration of a 3.7–4.0 MBq/kg dose of ^{18}F-FDG. Acquisition of whole-body ^{18}F-FDG PET/CT images followed standard protocols regarding uptake time (60–90 min). Calibration of scanners and scaling of images for reading were performed according to the local protocols in each institution.

PET imaging was performed in the craniocaudal direction from head to proximal thighs at 5–7 bed positions and at a rate of 1.5–2.0 min/bed position. The CT portion of the PET/CT study was performed as a low-dose acquisition with 130 kV and 50–80 mA.

Image Processing Using MV and MFS

MV and MFS were performed using syngo.via VB20 software (Siemens Medical Solutions). A single experienced nuclear physician calculated MTV and TLG for the 40% and 20% thresholds using the MV method.

The MV processing was executed by drawing elliptic VOIs surrounding each lesion and setting a threshold of 40% of lesion SUV_{max} for isocontour drawing. Total MTV and TLG were then automatically calculated by the software (Fig. 1A). This same procedure was repeated using a threshold of 20% of lesion SUV_{max} for isocontour drawing.

The MFS was performed using the MFS tool of the syngo.via VB20 software by a different experienced observer. A rectangular VOI was drawn around the entire body of the patient on the coronal axis. Afterward, if necessary, the VOI was adjusted on the axial and sagittal axes. The liver was set as the background reference and then the areas of interest were automatically determined around each lesion that had uptake higher than the SUV_{mean} of the liver. All lesions were then automatically delineated with VOIs with thresholds of 40% or 20% of the SUV_{max} using isocontour drawings (Fig. 1B). The image and VOIs were then reviewed to—using a single click of the mouse—exclude physiologic areas incorrectly selected by the software (e.g., brain, kidneys, bladder, or ureters) and include pathologic areas with relatively low uptake not selected by the software (e.g., small lymph nodes). Total MTV and TLG calculations were readily available.

Analysis of Interobserver Agreement of MFS Method

To ascertain interobserver agreement, at least 2 mo apart, 34 of the 35 PET Center 1 images were reprocessed using MFS by 2 different observers, one of them more experienced with FDG PET/CT images than the other. The 2 sets of values obtained by the 2 observers were statistically compared. In 1 patient from PET Center 1, the MFS method could not calculate MTV and TLG with both 20% and 40% thresholds.

Statistical Analysis

To evaluate agreement between the MV and MFS methods, the intraclass correlation coefficient was used, and Bland–Altman plots were constructed to compare the measurements obtained using the 2 techniques (*16*). The level of significance adopted was 5%. Statistical Analysis System software (version 9.4; SAS Institute Inc.) for Windows (Microsoft) was used.

Interobserver agreement on the 2 sets of MTV and TLG values obtained by the 2 different observers using the MFS method were compared using a 2-sample independent *t* test.

Data Availability

The datasets generated during or analyzed during the current study are not publicly available, to protect the identity of research subjects, but are available from the corresponding author on reasonable request.

RESULTS

The MV and MFS methods were initially performed for all 50 patients. In 1 patient, the MFS was not able to calculate the MTV and TLG automatically with both the 20% and the 40% thresholds. In this case, the tool included in the same VOI the lesion and a nearby area of physiologic elimination of radiotracer.

In 3 other patients, it was not possible to calculate the MTV and TLG with the 20% threshold: in one of these patients, the reasons were the same as cited above; in the other 2 patients, the tool did not recognize the lesions because of their small dimensions or low uptake (Fig. 2).

The 2 methods of calculating MTV and TLG could be performed for all remaining PET images (49 patients using a 40% threshold and 46 using a 20% threshold). The MTV and TLG values obtained using the MV and MFS quantifications are described in Table 1.

The median time required to calculate the MTV and TLG by the MV method was 64.4 min, ranging from 1 min in patients with lesions few in number or distant from areas of physiologic excretion of radiopharmaceutical (kidneys, bladder, liver, and heart, for example) to as much as 240 min in patients with lesions disseminated to multiple organs or confluent with areas of physiologic excretion.
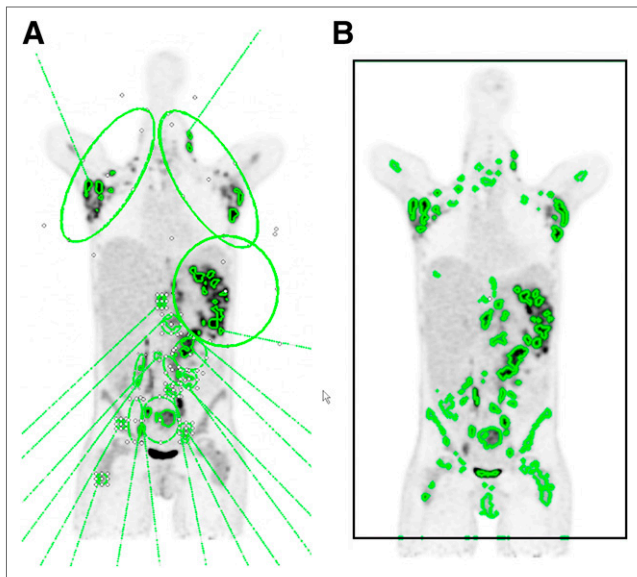
FIGURE 1. Comparison of MV and MFS methods of calculation of MTV and TLG in 25-y-old woman with widespread disease due to nodular sclerosis Hodgkin lymphoma. (A) Using MV method, elliptic VOIs surrounding each lesion or group of lesions were manually drawn avoiding areas of physiologic excretion. With threshold of 40% of SUV$_{max}$, total MTV and TLG were then automatically calculated by software. (B) With MFS method, master VOI (black rectangle) was plotted surrounding whole body, and threshold was set at 40%. Liver was set as background reference, and then all areas of disease were automatically drawn; areas of physiologic uptake were subsequently deleted. With MV method, most bone marrow involvement in iliac bones and proximal femurs was not included in VOIs, because SUVs there were below 40% threshold of SUV$_{max}$ of adjacent disease areas. For same reason, slight differences can be seen in VOIs obtained with the 2 methods, especially in armpits, upper chest, and neck.

With the MFS tool, the median time was 11.6 min, with a range of 1–30 min. The time to determine MTV and TLG using the 40% and 20% thresholds was similar.

The interclass correlation coefficients between the manual and the automatic values were high for all variables (MTV 20%, 0.8 (confidence interval, 0.73–0.91); TLG 20%, 0.96 (confidence interval, 0.94–0.98); MTV 40%, 0.93 (confidence interval, 0.89–0.96); and TLG 40%, 0.94 (range, 0.89–0.96)). Although cross calibration was not performed in pediatric patients (3–15 y old, 9 patients), there were no significant variations between the results obtained between MTV and TLG calculated by MV and MFS in this population.

Bland–Altman plots showed that the patients with higher MTV and TLG values of 20% and 40% were those who presented greater differences between the results obtained with MV and MFS processing. On the other hand, patients with low and intermediate MTV and TLG values presented lower variation between the methods (Fig. 3).

There were no statistically significant differences between values obtained by the 2 observers regarding the automatic method for calculation of MTV and TLG with either threshold. The P values for MTV and TLG were, respectively, 0.599 and 0.713 using a 20% threshold and 0.309 and 0.415 using a 40% threshold (Table 2).

DISCUSSION

MTV and TLG are consolidated in the literature as important tools for tumor burden assessment in cancer patients. Many studies report these variables as important for clinical decision making, contributing to prognostic assessment and to personalizing therapeutic strategies (1–5,8–13). This ability is particularly important in Hodgkin lymphoma, a disease in which early modification of chemotherapy regimens and use of radiotherapy are directly related to morbidity and prognosis (8,11–13). However, standardization of the method for calculating these variables is still lacking in the literature (6–8,12,14).

MTV and TLG can be calculated using several methodologies that are subdivided into 2 groups in the literature:
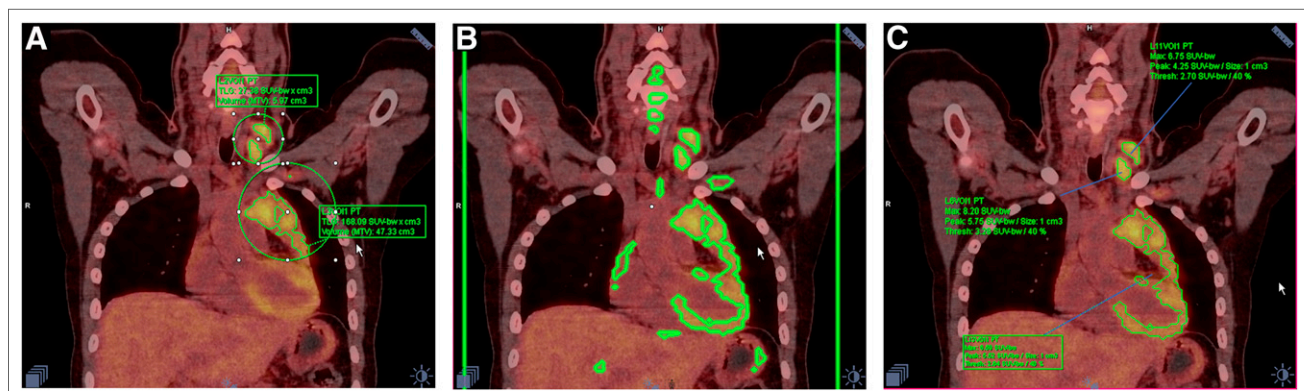

FIGURE 2. Limitation of MFS method when there is similar metabolism intensity in lesions and adjacent areas with physiologic uptake, verified in 24-y-old woman with nodular sclerosis Hodgkin lymphoma. (A) MV method: MVs manually surrounding lesions easily avoid heart. (B) MFS method: master VOI was placed around whole body of patient (large rectangle, whose side lines are seen in green). After 40% threshold was set, MVs were automatically delineated, including areas of physiologic uptake. (C) After trying to delete all areas of physiologic uptake, MFS tool kept heart and mediastinal lymph nodes included in same VOI, and software was unable to separate those structures.

TABLE 1
Manual and Automatic MTV and TLG Data

Method	n	Mean	SD	Minimum	Maximum
MTV 20% (mL)					
MFS	46	660	699	10.6	2,796.4
MV	46	736	856	10.9	3,647.5
TLG 20%					
MFS	46	3,048	3,149	34.7	11,753.5
MV	46	3,059	3,211	35.4	12,093.9
MTV 40% (mL)					
MFS	49	371	434	3.8	1,817.8
MV	49	313	359	3.7	1,374.8
TLG 40%					
MFS	49	2,013	2,253	18.7	8,913.5
MV	49	1,709	1,855	18.4	7,901.5

According to a recent study evaluating the pros and cons of each method, although the fixed absolute method is one of the most used (having been applied in 30% of the published studies assessing the role of MTV in lung cancer patients), it overestimates lesions with an SUV greater than 15 (6). The fixed relative method, used in the present study, is also one of the most reported methods in the literature (having been applied in 32% of cases of lung cancer) and seems to present good performance in metabolically homogeneous ^{18}F-FDG–avid neoplasms that are large and bulky (6). Because Hodgkin lymphoma usually has homogeneous metabolic activity, we chose to keep the same threshold for all lesions using the fixed relative method. The MV, using the fixed relative threshold-based method, was chosen as a reference because the lesions are delimited one by one, making the method accurate. In patients with Hodgkin lymphoma, a disease that may already be spread in the body in the staging study, this task can be complex and time-consuming because the affected lymph node chains
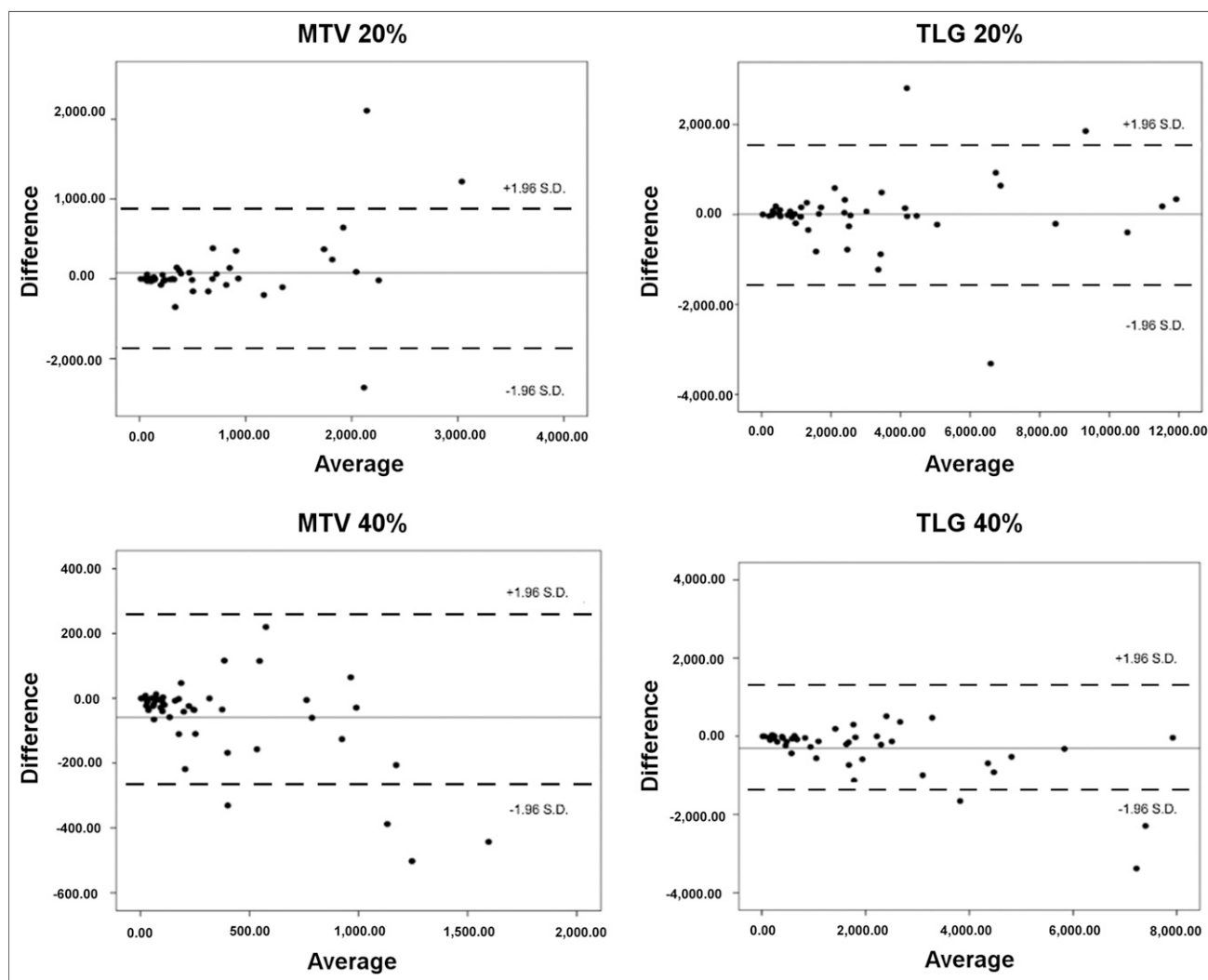
threshold-based and algorithm-based (6). Although algorithm-based methods are restricted to a few research centers, threshold-based methods are widespread worldwide.

FIGURE 3. Bland–Altman plots. Patients with higher values of MTV and TLG were those who presented higher differences between results obtained with MV and MFS methods.

TABLE 2
Agreement Between 2 Observers Using MFS Method

Method	Observer	Mean	SD	P
MTV 20%	A	721.1	729.9	0.599
	B	635.7	595.3	
TLG 20%	A	3,261.7	3174.2	0.713
	B	2,989.2	2922.8	
MTV 40%	A	423.4	450.2	0.309
	B	326.8	314.5	
TLG 40%	A	2,248.4	2268.2	0.415
	B	1,843.2	1779.6	

Mean, SD, and P values were calculated using 2-sample independent t test (n = 34).

may be adjacent to organs with physiologic uptake or excretion of the radiotracer. This occurrence is common in the mediastinum, where lymph node conglomerates may be contiguous to physiologic cardiac uptake, or in the retroperitoneum, where the lymphatic chains follow the ureters. Since the MFS method uses VOIs of different shapes and sizes, areas with physiologic uptake can be excluded with just 1 click.

Several studies suggest a 40% (or 41%) threshold as most accurate for delineating the margins of metabolically active lesions in both Hodgkin lymphoma and non-Hodgkin lymphoma and other neoplasms, such as lung cancer (2,6–9,12,14,16,17). In the present study, since we were evaluating new software, we sought to test its performance in different thresholds (40% and 20%). Although a 20% threshold is not standard for MTV and TLG calculation, this threshold has been used for lymphomas by some authors (12,17). In addition, some authors have reported thresholds of 20%–30% as the most adequate to delimit lesions with an SUV$_{max}$ of 20–30 (12,17).

There are several software programs for MTV and TLG calculation, many of them available for free download from the Internet. However, since the commercial software used in this paper usually comes included with new equipment of this specific brand, it must be independently tested before routine use. Choosing software from the same manufacturer as the equipment itself is usually quite convenient for the user.

The software used in the present study cannot measure MTV according to the newest algorithm-based methods, such as gradient-based, fuzzy C means, artificial neural network, fuzzy locally adaptive, and multi-Otsu methods— all of which are promising and apparently accurate (6). Unfortunately, these methods are not widely available, and relatively few studies using them have been published. The software used here provides only threshold-based methods to obtain MTV and TLG: fixed absolute, fixed relative, background-based, and adaptive.

Our results obtained with both MV and MFS processing methods are similar. Any differences were not significant and occurred mostly because the MV method provides only ellipses or spheres to delineate the lesions. For this reason, when the operator attempted to include all areas containing metabolically active lesions, there was an overlap between some of the VOIs. Areas of intersection were therefore counted twice, and the values for MTV and TLG using MV were thus higher than those for MFS, for which there was no superposition of VOIs.

On the other hand, when the operator decided to avoid VOI overlap during MV processing, some small parts of the lesion could not be included in the VOIs. This is probably why MTV and TLG values were sometimes smaller for the MV method than for the MFS method.

Finally, we found that MFS is more practical and faster than MV for MTV and TLG calculations in patients with a moderate to high tumor burden and can replace MV in those cases, with the same accuracy. We also verified that, using MFS, observers with different ^{18}F-FDG PET/CT imaging skills can calculate MTV and TLG values with similar accuracy. MV and MFS perform equally well in patients with lesions that have a low MTV, lesions of similar intensity to blood pool, and lesions near areas of physiologic excretion of the radiopharmaceutical.

CONCLUSION

In clinical practice, the use of MFS can render the calculation of MTV and TLG reproducible, fast, and practical in patients with disseminated diseases and has an accuracy similar to that of MV.

DISCLOSURE

REFERENCES

1. Salavati A, Duan F, Snyder BS, et al. Optimal FDG PET/CT volumetric parameters for risk stratification in patients with locally advanced non-small cell lung cancer: results from the ACRIN 6668/RTOG 0235 trial. *Eur J Nucl Med Mol Imaging.* 2017;44:1969–1983.
2. Wang XY, Zhao YF, Liu Y, Yang YK, Wu N. Prognostic value of metabolic variables of [18F]FDG PET/CT in surgically resected stage I lung adenocarcinoma. *Medicine (Baltimore).* 2017;96:e7941.
3. Tamandl D, Fueger B, Haug A, et al. A diagnostic algorithm that combines quantitative 18F-FDG PET parameters and contrast-enhanced CT improves posttherapeutic locoregional restaging and prognostication of survival in patients with esophageal cancer. *Clin Nucl Med.* 2019;44:e13–e21.

4. Choi WR, Oh JS, Roh JL, et al. Metabolic tumor volume and total lesion glycolysis predict tumor progression and survival after salvage surgery for recurrent oral cavity squamous cell carcinoma. *Head Neck.* 2019;41:1846–1853.

5. Wang L, Bai J, Duan P. Prognostic value of ^{18}F-FDG PET/CT functional parameters in patients with head and neck cancer: a meta-analysis. *Nucl Med Commun.* 2019;40:361–369.

6. Im HJ, Bradshaw T, Solaiyappan M, Cho SY. Current methods to define metabolic tumor volume in positron emission tomography: which one is better? *Nucl Med Mol Imaging.* 2018;52:5–15.

7. Im HJ, Solaiyappan M, Lee I, et al. Multi-level Otsu method to define metabolic tumor volume in positron emission tomography. *Am J Nucl Med Mol Imaging.* 2018;8:373–386.

8. Meignan M, Itti E, Gallamini A, Younes A. FDG PET/CT imaging as a biomarker in lymphoma. *Eur J Nucl Med Mol Imaging.* 2015;42:623–633.

9. Moskowitz AJ, Schöder H, Gavane S, et al. Prognostic significance of baseline metabolic tumor volume in relapsed and refractory Hodgkin lymphoma. *Blood.* 2017;130:2196–2203.

10. Akhtari M, Milgrom SA, Pinnix CC, et al. Re-classifying patients with early-stage Hodgkin lymphoma based on functional radiographic markers at presentation. *Blood.* 2018;131:84–94.

11. Song MK, Chung JS, Lee JJ, et al. Metabolic tumor volume by positron emission tomography/computed tomography as a clinical parameter to determine therapeutic modality for early stage Hodgkin's lymphoma. *Cancer Sci.* 2013;104:1656–1661.

12. 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;32:410–416.

13. Xie M, Zhai W, Cheng S, Zhang H, Xie Y, He W. Predictive value of F-18 FDG PET/CT quantization parameters for progression-free survival in patients with diffuse large B-cell lymphoma. *Hematology.* 2016;21:99–105.

14. Mhlanga JC, Chirindel A, Lodge MA, Wahl RL, Subramaniam RM. Quantitative PET/CT in clinical practice: assessing the agreement of PET tumor indices using different clinical reading platforms. *Nucl Med Commun.* 2018;39:154–160.

15. Brito AE, Mourato F, Santos A, Mosci C, Ramos C, Etchebehere E. Validation of the semiautomatic quantification of ^{18}F-fluoride PET/CT whole-body skeletal tumor burden. *J Nucl Med Technol.* 2018;46:378–383.

16. Schöder H, Moskowitz C. Metabolic tumor volume in lymphoma: hype or hope? *J Clin Oncol.* 2016;34:3591–3594.

17. Ceriani L, Martelli M, Zinzani PL, et al. Utility of baseline [18]FDG-PET/CT functional parameters in defining prognosis of primary mediastinal (thymic) large B-cell lymphoma. *Blood.* 2015;126:950–956.