Validation of a Multi-Foci Segmentation Method for Measuring Metabolic Tumor Volume in Hodgkin’s 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, Cármimo A. Souza 3,4, Juliano J. Cerci 7, Celso D. Ramos 1,2,3

Affiliations

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 (UNICAMP), Campinas, SP.
4. Division of Hematology and Hemotherapy, Department of Internal Medicine, University of Campinas, Campinas, Brazil.
5. Institute of Physics, University of Campinas (UNICAMP), Campinas, SP
6. Biomedical Engineering Center, University of Campinas (UNICAMP), Campinas, Brazil.
7. Quanta Diagnosis and Therapy Clinic, Curitiba, Brazil.

Author Contributions: M.R.C., N.T., J.J.C. and C.D.R.. made substantial contributions to the conception or design of the work. M.R.C., E.E., N.T., A.F.A.V., M.T.D., S.B.Q., M.E.S.T., C.D.R., C.A.S., I.G.H.L.M., J.J.C. and C.D.R. were responsible for acquisition and/or interpretation of data. M.R.C., M.E.S.T., S.B.Q., C.D.R. and I.G.H.L.M. were responsible for analysis of data. M.R.C., N.T. and C.D.R. wrote the initial draft of the paper, to which all authors contributed edits throughout. C.D.R. and E.E. have first approved the submitted version and were followed by all the other authors.

Corresponding Author
Celso Dario Ramos: Division of Nuclear Medicine, Department of Radiology, University of Campinas. Rua Roxo Moreira s/n, PO Box 6142, 13083-970, Campinas, Brazil.
Tel: +55(19)3521-7772. e-mail: cdramos@unicamp.br

First Author
Mariana Ramos Fernandes Camacho: MND Campinas
Rua Doutor Guilherme da Silva, 38, apto53, 13015028
Tel: +55(11)963692222. e-mail: marirfc@gmail.com
Word Count: 2356 words

Disclaimer: No potential conflicts of interest to disclose.

Short Running Title: Multi-Foci Segmentation for Measuring MTV
ABSTRACT

**Introduction:** Quantification of metabolic tumor volume (MTV) and total lesion glycolysis (TLG) can be time-consuming. We evaluated the performance of an automatic multi-foci segmentation method of quantification (MFS) in patients with different stages of Hodgkin’s Lymphoma, using the multiple VOI method (MV) as reference.

**Methods:** This prospective bicentric study included 50 patients with Hodgkin’s Lymphoma who underwent staging $^{18}$F-FGD PET/CT. The exams were centrally reviewed and processed with commercial MFS software in order to obtain MTV and TLG utilizing two fixed relative thresholds (40% and 20% of the maximum standardized uptake value) of each lesion. All PET/CTs were processed using the MV and MFS methods. Inter-class correlation coefficient and Bland & Altman plots were used for statistical analysis. Repeated calculations of MTV and TLG values by two observers with different degrees of PET/CT imaging experience were used to assess interobserver agreement of the MFS method.

**Results:** The mean and standard deviation values obtained for the MTV with MV and MFS were respectively 736mL ± 856mL and 660mL ± 699mL for the 20% threshold, and 313mL ± 359mL and 372mL ± 434mL for the 40% threshold. The time spent calculating the MTV was much shorter with the MFS than with the MV method (median time: 11.6 min. [1-30 min] and 64.4min. [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 two 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 two methods were related to difficulties in avoiding overlay of VOIs in the MV technique. MV and MFS perform equally well in in patients with small number of lesions.

**Key Words:** MTV, TLG, Hodgkin’s Lymphoma, $^{18}$F-FGD PET/CT, VOI, SUV.
INTRODUCTION

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

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 drawing VOIs, manually, 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 maximum SUV inside the master VOI drawn around the lesion.

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

Ideally, a software should determine automatically and simultaneously all areas containing metabolically active lesions in a few seconds. This is what the multi foci segmentation method (MFS) proposes: after determining one 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 MFS method for quantification of MTV and TLG in patients with different stages of Hodgkin’s Lymphoma (HL), using the MV method as the reference standard.
MATERIAL AND METHODS

This was a prospective bi-centric study. It was approved by the local ethics committees (CAAE: 07178612.0.1001.5405 and CAAE: 45797615.1.0000.5404) and 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 [U.C.] and 15 from PET-Center-2 [Q.D.T.C.]) had biopsy proven HL (28 females, 22 males; median 29 years; 3-84 years old). The histologic subtypes were: 35 (70.0 %) nodular sclerosis, 6 (12.0%) lymphocyte rich, 3 (6.0%) mixed cellularity, 1(2.0%) lymphocyte depleted and 5 (10.0%) unknown. Five patients (10%) were stage I, 15 (30%) stage II, 7 (14%) stage III and 23 patients (46%) were stage IV.

All PET/CTs were centrally reviewed in PET-Center-1. For MTV and TLG measurements, all images were first processed using the MV method by the same experienced observer (M.R.C.). To avoid bias of prior knowledge of patient image characteristics, images were processed using the MFS method by a different experienced observer. (A.F.A.V.). At least two months apart, 34 PET-Center-1 images were reprocessed using MFS by two different observers (M.R.C. and N.T.) to access interobserver agreement.

Image Acquisition

All patients fasted for at least 4-6 hours prior to the intravenous administration of 3.7 – 4.0 MBq/kg 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 images were performed in the craniocaudal direction from head to proximal thighs, in five to seven bed positions, 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, 50–80 mA.

Image Processing Using MV and MFS
MV and MFS were performed using the software Syngovia VB20 (Siemens Medical Solutions, Chicago, IL). A single experienced nuclear physician (M:R:C) calculated MTV and TLG for the 40% and 20% thresholds using the MV method.
The MV processing was executed by drawing elliptical VOIs surrounding each lesion and setting a threshold of 40% of lesion maximum SUV (SUVmax) for isocontour drawing. Total MTV and TLG were then automatically calculated by the software (Figure 1a). This same procedure was repeated using a threshold of 20% of lesion SUVmax for isocontour drawing.

The MFS was performed using the MFS tool of the Syngovia VB20 software by a different experienced observer (A.F.A.V.). A rectangular VOI was drawn around the entire body of the patient on the coronal axis. Afterwards, 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 with uptake higher than the mean SUV of the liver. All lesions were then automatically delineated with VOIs with thresholds of 40% or 20% of the SUVmax using isocontour drawings (Figure 1b). The image and VOIs were then reviewed to exclude physiologic areas incorrectly selected (brain, kidneys, bladder, ureters, etc.) and to include pathologic areas with relatively low uptake not selected by the software (e.g. small lymph nodes), using a single click of the mouse, and total MTV and TLG calculations were readily available.

Analysis of Interobserver Agreement of MFS Method

In order to access inter-observer agreement, at least two months apart, 34 of the 35 PET-Center-1 images were reprocessed using MFS by two different observers, one of them more experienced with FDG-PET/CT images (M.R.C.) than the other (N.T.). The two set of values obtained by the two observers were statistically compared. In one patient from PET-Center-1, the MFS method could not calculate MTV and TLG with both 20% and 40% thresholds.

Statistical Analysis

To evaluate the agreement between the MV and MFS methods, the intraclass correlation coefficient was used, and the Bland & Altman plots were constructed to compare the measurements obtained using the two techniques (16). The level of significance adopted was 5%. The statistical software SAS System for Windows (Statistical Analysis System, version 9.4. SAS Institute Inc, 2002-2008, Cary, NC, USA) was used.

The inter-observer agreement of the two sets of MTV and TLG values obtained by
the two different observers using the MFS method were compared using two-sample independent t-test.
RESULTS

The MV and MFS methods were initially performed for all 50 patients. In one 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 volume of interest the lesion and a physiological elimination area of the radiotracer near to it.

In three other patients, it was not possible to calculate the MTV and TLG with the 20% threshold. In one of these patients for the same reasons cited above and, in the other two cases, because the tool did not recognize the lesions due to their small dimensions or due to their low uptake (Figure 2).

The two methods of calculating MTV and TLG could be performed in all remaining PET images (49 patients using 40% threshold and 46 using 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 minutes, ranging from 1 minute in patients with small number of lesions and/or lesions distant from areas of physiological excretion of the radiopharmaceutical (kidneys, bladder, liver and heart, for example) to as much as 240 minutes in patients with disseminated lesions in multiple organs and/or confluent with areas of physiological excretion. Using the MFS tool the median time was 11.6 minutes ranging from 1 minute to 30 minutes. The time spent for the determination of MTV and TLG using the thresholds of 40% or 20% were similar.

The interclass correlation coefficients between the manual and the automatic values were high for all variables: MTV20%: 0.8 (0.73 – 0.91); TLG 20%: 0.96 (0.94 – 0.98); MTV40%: 0.93 (0.89 – 0.96); TLG40%: 0.94 (0.89 – 0.96). Although cross calibration was not performed in pediatric patients (3 - 15 years, 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 values of MTV and TLG of 20% and 40% were those that presented greater differences between the results obtained with MV and MFS processing. On the other hand, patients with low and intermediate values of MTV and TLG presented lower variation between the methods (Figure 3).

There were no statistically significant differences between the values obtained by the two different observers regarding the automatic method for calculation of MTV and TLG with both thresholds. The p-values for MTV and TLG using 20% threshold were, respectively 0.599 and 0.713, and using 40% threshold were respectively 0.309 and 0.415 (Table 2).
DISCUSSION

MTV and TLG are consolidated in the literature as important tools for tumor burden assessment of 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 is particularly important in HL a disease in which early modification of chemotherapy regimens and use of radiotherapy are directly related to the 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 two groups in the literature: threshold based and algorithm based (6). While algorithm-based methods are restricted to a few research centers, threshold-based methods are widespread worldwide. According to a recent study evaluating the pros and cons of each method of calculating MTV and TLG, although the fixed absolute method is one of the most used (30% of the published studies assessing the role of MTV in lung cancer patients), it overestimates lesions with 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 (32% of cases of lung cancer), and it seems to present good performance in metabolically homogeneous FDG avid neoplasms that form large bulky (6). As HL usually presents 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 reference, because the lesions are delimited one by one, which makes it very accurate. In patients with Hodgkin’s Lymphoma, a disease that may be widespread from the staging study, this task can be very complex and time consuming because the affected lymph node chains may be adjacent to organs with physiological uptake or excretion of the radiotracer. This is a common occurrence 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 MFS method utilizes VOIs of different shapes and sizes, the exclusion of areas with physiological uptake can be performed with just one click.

Several studies suggest the 40% (or 41%) threshold as the most accurate for delineating the margins of metabolically active lesions in both HL and non-Hodgkin's 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 values (40% and 20%). Although 20% threshold is not standard for MTV and TLG calculation, it has been used for some authors in lymphomas (12, 17). In addition, some authors have reported thresholds of 20% to 30% as the most adequate to delimit lesions with SUVmax between 20 and 30 (12, 17).

There are several software 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 with new equipment of this specific brand, it must be independently tested before routine use. Using software from the same manufacturer of the equipment itself is usually quite convenient for the user.

The software utilized in the present study cannot measure MTV according to the newest algorithm-based methods, such as gradient-based, fuzzy C means, artificial neural networks, fuzzy locally adaptive and multi-otsu method, all are very promising and apparently accurate (6). Unfortunately, these methods are not widely available and relatively few studies have been published using them. The software utilized 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. The differences observed between them 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. Therefore, those areas of intersection were counted twice and, thus, the values obtained for MTV and TLG using MV were higher than those obtained with the MFS, in 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 the reason for cases of MTV and TLG values being smaller for the MV method than those obtained with the MFS technique.

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

In 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 with similar accuracy as obtained with MV.
Acknowledgements: We thank the Nuclear and Energy Research Institute (IPEN-CNEN), São Paulo, Brazil, for kindly supplying the radiopharmaceuticals used in the present project (IPEN/UNICAMP agreement Nº01342000458/2017-15). We are grateful for financial support from FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo), grant numbers 2009/54065-0 and 2018/00654-4. CDR has a research grant from CNPq (National Council of Research) proc 311841/2018-0. ILM has a research grant from CNPq (National Council of Research) proc 305110/2018-7.

Disclosure: The authors declare that they have no conflict of interest.

Data availability: The datasets generated during and/or analyzed during the current study are not publicly available due to protect the identity of research subjects but are available from the corresponding author on reasonable request.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the local ethics committee (CAAE: 07178612.0.1001.5405 and CAAE: 45797615.1.0000.5404), University of Campinas Ethics Committee), and written informed consent was waived.
Figure 1: Comparison of the MV and the MFS methods of calculation of MTV and TLG in a 25-year-old female patient with widespread disease due to nodular sclerosis Hodgkin’s Lymphoma. (A) Using the MV method, elliptical VOIs surrounding each lesion or group of lesions were manually drawn avoiding areas of physiological excretion. Setting a threshold of 40% of SUVmax, total MTV and TLG were then automatically calculated by the software. (B) Using the MFS method a master VOI (black rectangle) was plotted surrounding the whole body and the threshold was set at 40%. The liver was set as the background reference and then all areas of disease were automatically drawn; areas of physiological uptake were subsequently deleted. Note that using the MV method, most bone marrow involvement of the iliac bones and proximal femurs was not included in the VOIs, because its SUVs were below the 40% threshold of the SUVmax of adjacent disease areas. For the same reason, also note the slight differences in the VOIs obtained with the two methods, especially in the armpits, upper chest and neck.
Figure 2: Limitation of the MFS method when there is similar metabolism intensity in lesions and adjacent areas with physiological uptake, verified in a 24-year-old female patient with nodular sclerosis Hodgkin’s Lymphoma. (A): MV method: multiple VOIs manually surrounding the lesions easily avoid the heart. (B): MFS method: a master VOI was placed around the whole body of the patient (a large rectangle whose side lines are seen in green). After setting a 40% threshold, multiple VOIs were automatically delineated, including areas of physiological uptake. (C): After trying to delete all areas of physiological uptake, the MFS tool remained including the heart and mediastinal lymph nodes in the same VOI, and the software was unable to separate those structures.
Figure 3: Bland & Altman plots. Note that patients with higher values of MTV and TLG were those that presented higher differences between the results obtained with MV and MFS methods.
Table 1: Mean, standard deviation, maximum and minimum values obtained for both manual and automatic MTV and TLG.

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTV 20%</td>
<td>MFS 46</td>
<td>660</td>
<td>699</td>
<td>10.6</td>
<td>2796.4</td>
</tr>
<tr>
<td></td>
<td>MV 46</td>
<td>736</td>
<td>856</td>
<td>10.9</td>
<td>3647.5</td>
</tr>
<tr>
<td>TLG 20%</td>
<td>MFS 46</td>
<td>3048</td>
<td>3149</td>
<td>34.7</td>
<td>11753.5</td>
</tr>
<tr>
<td></td>
<td>MV 46</td>
<td>3059</td>
<td>3211</td>
<td>35.4</td>
<td>12093.9</td>
</tr>
<tr>
<td>MTV 40%</td>
<td>MFS 49</td>
<td>371</td>
<td>434</td>
<td>3.8</td>
<td>1817.8</td>
</tr>
<tr>
<td></td>
<td>MV 49</td>
<td>313</td>
<td>359</td>
<td>3.7</td>
<td>1374.8</td>
</tr>
<tr>
<td>TLG 40%</td>
<td>MFS 49</td>
<td>2013</td>
<td>2253</td>
<td>18.7</td>
<td>8913.5</td>
</tr>
<tr>
<td></td>
<td>MV 49</td>
<td>1709</td>
<td>1855</td>
<td>18.4</td>
<td>7901.5</td>
</tr>
</tbody>
</table>

N = number of patients; SD = standard deviation; Min = minimum; Max = maximum; MTV unit: mL.
Table 2: Interobserver agreement between two observers using the MFS method. Mean, standard deviation and p values were calculated using two-sample independent t-test (n = 34).

<table>
<thead>
<tr>
<th>Observer</th>
<th>Mean</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTV 20%</td>
<td>A</td>
<td>721.1</td>
<td>729.9</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>635.7</td>
<td>595.3</td>
</tr>
<tr>
<td>TLG 20%</td>
<td>A</td>
<td>3261.7</td>
<td>3174.2</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2989.2</td>
<td>2922.8</td>
</tr>
<tr>
<td>MTV 40%</td>
<td>A</td>
<td>423.4</td>
<td>450.2</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>326.8</td>
<td>314.5</td>
</tr>
<tr>
<td>TLG 40%</td>
<td>A</td>
<td>2248.4</td>
<td>2268.2</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1843.2</td>
<td>1779.6</td>
</tr>
</tbody>
</table>

SD: standard deviation
REFERENCES


