Journal of Nuclear Medicine, published on October 9, 2020 as doi:10.2967/jnumed.120.256214

Automated segmentation of TMTV in DLBCL patients: what about method measurement uncertainty?

Eric Laffon[1,2,3*], Roger Marthan[1,2,3].

[1]CHU de Bordeaux, F-33000 Bordeaux, France.

[2]Univ. Bordeaux, Centre de Recherche Cardio-Thoracique de Bordeaux, F-33000 Bordeaux, France.

[3]INSERM U-1045, Centre de Recherche Cardio-Thoracique de Bordeaux, F-33000 Bordeaux, France.

*Correspondence: Eric Laffon, Service de Médecine Nucléaire, Hôpital Haut-Lévèque, avenue de Magellan, 33604 Pessac, France. elaffon@u-bordeaux.fr; ORCID: **0000-0001-9561-8521**.

TO THE EDITOR:

In baseline ^{18}F-FDG PET imaging of diffuse large B-cell lymphoma (DLBCL) patients, Barrington et al. recently confirmed that different outlining methods providing total metabolic tumor volume (TMTV) can be used for predicting prognosis [1]. An automated tool was used for segmentation, focusing on the need in clinical practice for a fast, easy and robust method. From the success/failure ratings of the visible-tumor delineation by two independent observers, involving minimal user interaction, the method based on a fixed-SUV threshold of 4.0 g/mL (i.e., SUV4.0) was recommended for further evaluation, as well as MV2, i.e., a majority-vote method usually combining SUV4.0 and SUV2.5 (i.e, 2.5-g/mL-fixed-SUV threshold). Although different methods may provide significantly different TMTV outcomes, the authors suggested that bias in TMTV outcome is clinically less relevant than good reproducibility.

We fully agree with this suggestion but we would like to stress that the current study did not provide any quantitative information about reproducibility percentage of each method, which quantifies the closeness of the agreement between TMTV outcomes obtained under changed conditions of measurement [2]. These changed conditions may be different observers as in Barrington's study, but also in clinical practice, inter-scan time, scanning and patient's conditions (including uptake time) … Going further with Barrington et al.'s suggestion, we believe that an outlining method providing a biased TMTV estimate, in other words, a

surrogate, but accompanied by a significantly lower measurement uncertainty (MU; here, for single scan) than that of the SUV4.0, should be preferred for DLBCL prognosis [2]. As a supporting example, the ^{18}F-FDG SUV is only a surrogate for the metabolic rate of glucose consumption, however, its use does not longer need to be justified owing to reasonable MU and availability [3]. Noteworthy, such a reduced MU might compensate for the substantial MU expected for TMTV cut-off from Figure 4 by Barrington, showing poor (<0.65) areas under the ROC curves [1,4]. To summarize, the issue of a quick-easy method is indeed relevant in clinical practice, but we believe that it should not dominate the crucial MU one, even if too many clicks may affect inter- and intra-observer reproducibility. A 3–6-min TMTV measurement for most scans, depending on the method, seems to us a reasonable price to pay for patient management [1].

Furthermore, since the Quantitative-Imaging-Biomarkers-Alliance profile for ^{18}F-FDG as an imaging biomarker for treatment-response assessment did not address the prognosis issue from a single scan, we take the opportunity to suggest that a TMTV-cut-off value for DLBCL staging should involve MU and, hence, be accompanied by asymmetric confidence limits of $100\times\{\exp[\pm 1.96\times SD(d)/\mathrm{sqrt}(2)] - 1\}\%$ (SD(d): standard deviation of the differences of the test-retest TMTV-value logarithms; 95%-confidence)[3,4]. Unlike a strict cut-off value, these MU-derived upper/lower limits may reduce the number of false positive/negative scans for avoiding patient therapy-escalation/under-treatment, respectively. This rationale offers the same flexibility as the use of liver/mediastinum SUV for assessing complete metabolic response in lymphoma patients according to treatment strategy. Strategy may also help to arbitrarily decide a false positive/negative when an outcome is close to a limit. Noteworthy, the limit values may be relevantly adjusted by expert consensus (e.g., changing 1.96 to 1 for 68%-confidence).

To conclude, evaluating the best outlining method in clinical practice for assessing TMTV in DLBCL at baseline, along with determining optimal TMTV cut-off to separate patients with good/poor prognosis, are important issues for treatment-decision making. However, without any quantitative information about MU of each method, we believe that recommendations are of limited scope. Repeated comments about the prognostic use of a strict cut-off value of a continuous parameter, as well as a proposal for avoiding TMTV computing, might be taken into consideration [4,5].

REFERENCES

1. Barrington SF, Zwezerijnen BG, de Vet HC, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method is most successful? *J Nucl Med*. 2020 Jul 17:jnumed.119.238923.
2. JCGM 2008. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. www.bipm.org, September 2008.
3. Kinahan PE, Perlman ES, Sunderland JJ, et al. The QIBA Profile for FDG PET/CT as an imaging biomarker measuring response to cancer therapy. *Radiology*. 2020;294:647–657.
4. Laffon E, Marthan R. On the cutoff of baseline total metabolic tumor volume in high-tumor-burden follicular lymphoma. *J Clin Oncol*. 2017;35:919–920.
5. Laffon E, Marthan R. Could we avoid computing TMTV of DLBCL patients in routine practice? *Eur J Nucl Med Mol Imaging*. 2018;45:2235–2237.