Automated Segmentation of TMTV in DLBCL Patients: What About Method Measurement Uncertainty?

TO THE EDITOR: In baseline ^{18}F-FDG PET imaging of patients with diffuse large B-cell lymphoma (DLBCL), Barrington et al. recently confirmed that different outlining methods providing total metabolic tumor volume (TMTV) can be used to predict prognosis (*1*). An automated tool was applied 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 2 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 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 would like to stress that the study did not provide any quantitative information about the reproducibility percentage for each method—a quantification of the closeness of the agreement between TMTV outcomes obtained under changed conditions of measurement (*2*). These changed conditions may consist of different observers, as in Barrington's study, but also, in clinical practice, interscan time, scanning, and patient's conditions (including uptake time). Going further with the suggestion of Barrington et al., we believe that an outlining method providing a biased TMTV estimate—in other words, a surrogate—but accompanied by a significantly lower measurement uncertainty (here, for single scan) than that of SUV4.0 should be preferred for DLBCL prognosis (*2*). As a supporting example, although the ^{18}F-FDG SUV is only a surrogate for the metabolic rate of glucose consumption, its use no longer needs to be justified, because measurement uncertainty and availability are reasonable (*3*). It is noteworthy that such a reduced measurement uncertainty might compensate for the substantial measurement uncertainty expected for the TMTV cutoff from Figure 4 by Barrington, showing poor ($<$0.65) areas under the receiver-operating-characteristic curves (*1,4*). To summarize, the issue of a quick and easy method is indeed relevant in clinical practice, but we believe that it should not dominate the crucial measurement-uncertainty issue, even if too many clicks may affect inter- and intraobserver reproducibility. A 3- to 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 cutoff for DLBCL staging should involve measurement uncertainty and, hence, be accompanied by asymmetric confidence limits of $100 \times \{\exp[\pm 1.96 \times \mathrm{SD}(d)/\mathrm{sqrt}(2)] - 1\}\%$, where SD($d$) is the SD of the differences in the test–retest TMTV-value logarithms (95% confidence) (*3,4*). Unlike a strict cutoff, these measurement uncertainty–derived upper and lower limits may reduce the number of false-positive and -negative scans for avoiding therapy escalation or undertreatment, respectively. This rationale offers the same flexibility as the use of liver or mediastinum SUV for assessing complete metabolic response in lymphoma patients according to treatment strategy. Strategy may also help to arbitrarily decide whether an outcome is false-positive or -negative when the outcome is close to a limit. The limits 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 the optimal TMTV cutoff to separate patients according to good or poor prognosis, are important issues for making treatment decisions. However, without any quantitative information about the measurement uncertainty of each method, we believe that recommendations are of limited scope. Repeated comments about the prognostic use of a strict cutoff for 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.* July 17, 2020 [Epub ahead of print].
2. Evaluation of measurement data—guide to the expression of uncertainty in measurement. Bureau International des Poids et Mesures website.https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Published September 2008. Accessed December 10. 2020.
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.

Eric Laffon*
Roger Marthan
**Hôpital Haut-Lévèque*
Avenue de Magellan
33604 Pessac, France
E-mail: elaffon@u-bordeaux.fr

Published online Oct. 9, 2020.
DOI: 10.2967/jnumed.120.256214

REPLY: We thank Laffon and Marthan for their interest in our work (*1*) and for acknowledging that bias in metabolic tumor volume (MTV) outcome is less clinically relevant than good reproducibility. We agree that estimation of the reproducibility of MTV measurement methods is important to determine measurement uncertainty. We reported that agreement between observers for assessment of