Journal of Nuclear Medicine, published on October 30, 2020 as doi:10.2967/jnumed.120.257030

Reply to Laffon and Marthan: Automated segmentation of Ben MTV in DLBCL patients: what about method measurement uncertainty?

Sally F Barrington[1], Ben GJC Zwezerijnen[2], Henrica C W de Vet[3], Martijn W Heymans[3], Ronald Boellaard[2].

Corresponding author: Sally Barrington, St Thomas Hospital, London SE1 7EH UK

00 44 207 188 8364 (phone) 00 44 207 620 0790 (fax) ORCID ID 0000-0002-2516-5288

sally.barrington@kcl.ac.uk

[1] School of Biomedical Engineering and Imaging Sciences, Kings College London, UK, [2] Department of Radiology and Nuclear Medicine and [3] Department of Epidemiology & Data Science at Amsterdam UMC, Vrije Universiteit Amsterdam, Netherlands.

Running title : Automated segmentation of MTV

We thank Laffon and Marthan for their interest in our work *(1)* and for acknowledging that bias in 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 the agreement between observers for assessment of MTV measurements using the same software was 91% for the SUV41% method and > 95% for all other methods and considered this as good agreement *(1)*. The success rate of MTV measurement was unaffected by scanning conditions (EARL compliant or not) and the presence or absence of subsequent disease progression. The uptake time influenced the success rate of measurements for the SUV41% and majority vote 3 (MV3) methods, which were less successful with longer uptake times.

Laffon and Marthan propose that MTV cut-offs derived from PET data to guide discrimination of prognosis should be accompanied by upper and lower confidence limits based on measurement uncertainty. The main purpose of our work was not to derive cut-offs to discriminate prognosis, but as a first step to answer a methodological question, which was to determine the optimal automatic segmentation method/s for MTV to apply in a larger cohort. The criteria in our study focussed on two aspects. Firstly, did the MTV measurement methods generate plausible total tumor burden segmentations? This was prioritized over precision, as good repeatability do not necessarily provide meaningful results. Thereby, if such (known) precision should subsequently be used to define a threshold uncertainty or grey zone is a matter of effect size in the studied population and the intended use of the biomarker. Secondly, to apply a method clinically or in trials, the segmentation and workflow should be fast, easy to use with minimal observer interaction. By applying these criteria, we identified two candidate methods (SUV4.0 and MV2) that can be considered for further MTV biomarker validation. For individual patient assessment to guide prognosis and where the ultimate goal is to offer personalised treatment, MTV should ideally be assessed as a continuous variable. Then cut-points and measurement error/misclassification become less relevant.

We presented data on discriminatory power to confirm similarity for the different segmentation methods as shown previously *(2)*, to support the argument that choice of method can be based on ease of use and success rates giving plausible volumes under various conditions. For the current study, we used a case-control design to test parameters that might influence the best segmentation method meaning that the patient population and any derived cut-offs would not be representative of usual clinical practice. We are progressing with MTV measurement in a large warehouse of clinical and scan data in patients with non-Hodgkin lymphoma; https://petralymphoma.org/. Sufficient data are required to derive robust optimal MTV cut-off/s for training, validation and test datasets. In these studies measurement error, confidence limits and uncertainty will be taken into account.

Finally, MTV is a robust predictor of prognosis in DLBCL but will likely need to be factored into an algorithm with baseline clinical factors, including the international prognostic index *(3)*, and potentially with emerging biomarkers that reflect tumor dissemination and molecular heterogeneity *(4, 5)* and dynamic response markers *(3, 4)*.

DISCLOSURES

SFB acknowledges support from the National Institute for Health Research (NIHR) [RP-2-16-07-001]. King's College London and UCL Comprehensive Cancer Imaging Center is funded by the CRUK and EPSRC in association with the MRC and Department of Health and Social Care (England).The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The PETRA project is supported by the Alpe d' HuZes/KWF fund, provided by the Dutch Cancer Society (# VU 2012-5848).

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.

2. Ilyas H, Mikhaeel NG, Dunn JT, et al. Defining the optimal method for measuring baseline metabolic tumor volume in diffuse large B cell lymphoma. Eur J Nucl Med Mol Imaging. 2018;45:1142-1154.

3. Mikhaeel NG, Smith D, Dunn JT, et al. Combination of baseline metabolic tumor volume and early response on PET/CT improves progression-free survival prediction in DLBCL. Eur J Nucl Med Mol Imaging. 2016;43:1209-1219.

4. Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. J Clin Oncol. 2018;36:2845-2853.

5. Cottereau AS, Nioche C, Dirand AS, et al. (18)F-FDG PET dissemination features in diffuse large B-cell lymphoma are predictive of outcome. J Nucl Med. 2020;61:40-45.