Prognostic value of ^{18}F-FDG-PET in patients with mantle cell lymphoma: results from the LyMa-PET Project

Recently, Bailly *et al*. reported results from the LyMa-PET project regarding the prognostic value of image-derived ^{18}F-fluoro-deoxyglucose positron emission-tomography (^{18}F-FDG-PET) quantitative indices in patients with mantle cell lymphoma (MCL) having autologous stem-cell transplantation (ASCT).[1] The authors concluded that the maximal standard uptake value (SUV_{max}) corresponding to the hottest voxel of the lesion with the highest uptake at diagnosis, has a strong prognostic value for both progression-free survival and overall survival. A new scoring system combining the mantle cell lymphoma international prognostic index (MIPI) score and SUV_{max} was proposed to improve the patient outcome prediction. In particular, the authors argued that, since the prognostic value of SUV_{max} and SUV_{peak} (corresponding to the average SUV obtained from a 1-mL sphere centered over the most active region of the highest-uptake lesion) was similar, the former metric was preferred over the latter one owing to its wide use.

We suggest that, for such a choice, the comparison between the measurement uncertainty (MU) of SUV_{max} and SUV_{peak} should be taken into account and discussed in details. Indeed, SUV_{peak}, or any averaged SUV from several voxels such as the $SUV_{max\text{-}N}$ that pool several N hottest voxels regardless of their location within one or different ^{18}F-FDG-positive lesion, have a lower MU than the former, and, hence, are more reliable for a clinical decision to be taken.[2,3] In a previously published lung cancer series, the relative measurement error (MEr) (i.e., the relative difference between a single SUV estimate and its average true value) of SUV_{peak} and $SUV_{max\text{-}40}$ was found to be significantly lower than that of the SUV_{max}: 9.4 and 8.8 *versus* 13.9 % (with 95 % reliability), respectively.[3] These results may be applied to MCL patients because positron emission tomography (PET) imaging does not allow identifying the disease type underlying an ^{18}F-FDG uptake. Noteworthy, they were obtained with SUV_{max} values ranging between 6.6 and 23.2 g/mL, and it should be stressed that the MEr percentage does increase when applied to SUV values lower than 6.6 g/mL, as clearly demonstrated by de Langen (in terms of repeatability percentage): the lower the SUV value, the greater its MU.[4] In the study by Bailly *et al*. *Online Supplementary Table S2* reports a minimal value at baseline of 1.8 and zero g/mL for SUV_{max} and SUV_{peak}, respectively, whose MEr is substantial, even incalculable.[1] As a consequence, the inclusion of such low SUV outcomes may very likely explain the huge range of ΔSUV_{max} and ΔSUV_{peak} outcomes between baseline and before transplantation that are reported in *Online Supplementary Table S4*: [-100,+271] and [-95,+278] (expressed in %), respectively.[1] Furthermore, the new scoring system combining the MIPI score and SUV_{max} that was proposed by the authors, used a SUV_{max} cutoff value of 10.3 g/mL. Since MEr is the relative difference between a single estimate and its average true value, we suggest that the cutoff value of 10.3 g/mL be completed by lower and upper limits of 8.8 and 11.7 g/mL obtained from the above-reported ± 13.9 % MEr for SUV_{max} (95 % reliability). The use of such limits, which may be refined by taking into account the MU of the cutoff outcome and the PET system employed, could enable physicians to adjust their decision according to the clinical trial design, that is, to avoid a false-negative/positive PET scan leading to patient's under-treatment/therapy-escalation, respectively. We believe that this point emphasizes why a metric with a reduced MU should be preferred. Finally, it is worth noting that the MU argument also applies to the therapeutic evaluation since the repeatability, *i.e.*, the minimal relative change between two SUVs assessed from two successive examinations that is required to consider a significant difference, can be computed as $2^{1/2}\times$MEr.[5]

To conclude, the relevant study by Bailly *et al*. clarifies the prognostic value of quantitative ^{18}F-FDG-PET imaging in MCL patients. With respect to establishing the prognosis or to assessing the response to treatment, we suggest that the criteria for the choice of a metric should involve the magnitude of its MU and, therefore, averaged quantitative indices should be a priori preferred. How acceptable the MU magnitude of a metric should be, that might cause difficulties in clinical decision making and, hence, that would rule it out, is a question of judgment and consensus.

Eric Laffon[1,2,3] and Roger Marthan[1,2,3]

[1]CHU de Bordeaux, Bordeaux; [2]Univ. Bordeaux, Centre de Recherche Cardio-Thoracique de Bordeaux and [3]INSERM U-1045, Centre de Recherche Cardio-Thoracique de Bordeaux, Bordeaux, France

Correspondence: ERIC LAFFON
elaffon@u-bordeaux.fr

doi:10.3324/haematol.2019.236869

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Bailly C, Carlier T, Berriolo-Riedinger A, et al. Prognostic value of FDG-PET in patients with mantle cell lymphoma: results from the LyMa-PET Project. Haematologica. 2019 Aug 1. [Epub ahead of print].
2. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumours. J Nucl Med. 2009;50 Suppl 1:122S-150S.
3. Laffon E, Burger IA, Lamare F, de Clermont H, Marthan R. SUV_{peak} performance in lung cancer: comparison to average SUV from the 40 hottest voxels. J Nucl Med. 2016;57(1):85-88.
4. de Langen AJ, Vincent A, Velasquez LM, et al. Repeatability of 18F-FDG uptake measurements in tumors: a meta-analysis. J Nucl Med. 2012;53(5):701-708.
5. JCGM 2008. Evaluation of measurement data – Guide to the expression of uncertainty in measurement. www.bipm.org, September 2008.