

## Metabolic Tumor Volume: we still need a platinum-standard metric.

A Gallamini  
A Lacassagne Cancer Center – Nice  
L Kostakoglu  
Mount Sinai Medical Center – New York

In this issue of the Journal of Nuclear Medicine, Cottereau et al retrospectively evaluated data compiled from a cohort of 106 peripheral T-cell lymphoma (PTCL) patients, 50% of whom was previously enrolled in LYSA studies participated by five French and Belgian centers between 2006 and 2014. The objective was to determine the prognostic value of baseline total metabolic tumor volume (TMTV) calculated with the use of adaptive thresholding methods as compared to TMTV measured with a fixed threshold. The methods utilized to calculate TMTV included the Daisne method, based on the tumor to background ratios to segment tumor volumes and the Nestle method, which compares the intensity of the tumor to that of the background. Despite substantial differences in cutoff values across different TMTV computing methods ( $\pm 31\%$ ), the authors reported excellent intra-class correlation coefficients (from 0.97 to 0.98), for discriminating low vs. high TMTV.

See article by Cottereau

Tumor burden has long been considered an important prognostic marker in several lymphoma subtypes, hence its surrogates like bulky lesions, stage III-IV or extranodal disease have been incorporated in several prognostic models including IPI (International Prognostic Index), IPS (International Prognostic Score for Hodgkin Lymphoma), FLIPI 1 and 2 (Prognostic score for follicular lymphoma) MIPI (prognostic score for mantle cell lymphoma), PIT (Prognostic Index for Peripheral T-cell Lymphoma), AIP (Prognostic model

for Angioimmunoblastic Lymphoma). These prognostic models, however, are not considered sufficient to accurately stratify disease risk categories across patient populations. With the advent of advanced imaging techniques such as FDG-PET/CT and the availability of software offering sophisticated computer algorithms necessary for accurate tumor segmentation to calculate the sum of voxels of the tumor bulk it became possible to quantify the functional activity and the total tumor burden.

Earlier FDG-PET/CT studies found that a high MTV was independently associated with PFS and OS in HL patients treated with standard ABVD, with or without IFRT [1-2], suggesting that pre-therapy MTV as a measure of metabolically active whole body tumor bulk was a predictor of outcome when the conventionally described tumor burden did not reach a significance. However, there is also evidence for contrasting results, showing that baseline MTV could not predict survival when IPS did, while percent change ( $\Delta$ ) in both MTV and  $SUV_{max}$  at interim PET were associated with PFS and OS [3]. It is difficult to generalize these divergent results on the basis of retrospective analyses associated with no statistical design to determine sufficiently high number of patient cohorts, inherent risk of bias for population selection, treatment protocols, segmentation methodologies resulting in various MTV cut-offs. These limitations raise significant concerns for the internal validity of these published results.

In aggressive B-cell lymphoma (DLBCL), multiple retrospective studies investigated pre-treatment PET-derived volumetrics as a potential predictor of survival in patients undergoing R-CHOP therapy [4-8]. A systematic review of seven retrospective DLBCL studies (n=703) suggested that both  $SUV_{max}$  and MTV were significant prognostic factors for PFS ( $p=0.038$  and  $p=0.000$ , respectively) [4]. For OS, only high MTV was a strong predictor of poor prognosis ( $p=0.000$ ). When Ann Arbor staging did not predict survival, higher MTV was associated with significantly inferior EFS or PFS compared with the lower

MTV group, independent of the IPI [4,5]. Similar to HL data, the combination of baseline MTV and PET2 findings improved the predictive value [7]. There are, however, contradictory reports [6-8], one of which showed that the baseline  $SUV_{max}$  was a better predictor of EFS ( $p= 0.0002$ ) than MTV and TLG [6] and only the IPI score 3 was significantly associated with poor outcome. In the other study, the NCCN-IPI was the only significant predictor of PFS ( $P= 0.024$ ), while both NCCN-IPI and MTV were significant predictors of OS ( $P = 0.039$  and  $P = 0.043$ , respectively) [8]. More recently in a prospective cohort of 103 primary mediastinal (thymic) large B-cell lymphoma (PMBCL) patients enrolled in the International Extranodal Lymphoma Study Group trial IELSG-26, who received combination chemo-immunotherapy, Ceriani et al showed that only TLG retained statistical significance for both OS ( $P = .001$ ) and PFS ( $P < .001$ ) [10].

The above summarized studies used different segmentation techniques varying from fixed threshold methods based on absolute  $SUV_{max}$  [7] or percent thresholding using 25% of the  $SUV_{max}$  [10] 37% [8], 40% [9], 41% (Cottreau, present article ) or 42% [6]. This methodological variability resulted in widely disparate cutoff values ranging from 11 to 30 for  $SUV_{max}$ , from 130 ml to 550 ml for MTV and from 415 to 2955 for TLG in the prediction survival in various lymphoma subtypes. Moreover, these studies did not uniformly compare the prognostic values of  $SUV_{max}$  and conventional prognostic factors with MTV or TLG in a systematic fashion. Overall, the variability in methodology, the lack of demonstration of comparative superiority of any of these volumetric parameters over the conventional prognostic systems and the commonly used variable,  $SUV_{max}$ , or the comparison with varying prognostication systems, as well as the application of these methods in non-controlled heterogeneous populations, have generated skepticism about the internal and external validity of metabolic volumetric measures for the establishment of these parameters as an independent prognostic marker. In general, gradient-based methods that

factor in the background activity are considered a more accurate tumor volume segmentation method compared with fixed thresholding methods [11]. On the other hand, the importance of harmonization and cross-calibration across scanners in the multicenter studies should also be stressed. Another caveat is that all quantitative imaging surrogates should be prospectively validated before consideration for any preclinical or clinical application. Consequently, although there is suggestion towards a potential for PET-derived quantitative volumetrics to better prognosticate disease, one should realize that the previously published studies were not optimally designed to discriminate between risk groups to individualize DLBCL treatment. The shortcomings of the study by Cottreau et al largely follows the deficiencies of the prior studies published on this topic; retrospective design, no statistical sample analysis to attain the objective with a meaningful margin, patients being treated with different regimens, the use of different generation PET cameras and, non-cross calibration of the cameras across centers, all of which might have potentially affected the calculation of SUV and TMTV. Lastly, propagation of a systematic error cannot be ruled out with the results of this study, again bringing the focus on the internal validity of the results.

Briefly, some preliminary conclusions could be derived from the published literature: (1) In the entire series of reported cases, baseline TMTV or Total Lesion Glycolysis (TLG) were suggested to be relevant prognostic factors; (2) in a given lymphoma subtype such as HL or DLBCL, significantly varying cutoff values have been reported (3) the lack of technical information on methods to standardize the quality of the PET images as well as the absence of stringent methodological procedures to ensure the reproducibility of the results potentially undermines the robustness of the prognostic information. Essentially, a shared unit of optimal measure, a platinum-standard metric, proving accurate

and comparable results quantitative, still remains to be developed (4). The reported results for TMTV computing, the variability of MTV cutoff values to predict treatment outcome point towards a prognostic role of MTV as continuous instead of a dichotomous variable to identify patient outcome. The same conclusions could be drawn from the current data presented here by Cottureau et al: despite a variability of cutoff values, higher than 30%, all the models showed an equivalent predictive value with an exceptionally high concordance rate across different methods. However, although prognostically relevant, the results generated by this retrospective analysis remain to be of limited utility in a clinical setting. Hence, the predictive superiority of MTV or TLG over SUVmax for survival is yet to be proven in large-scale prospective, multicenter, well-designed studies. Only a validated cutoff value would be useful in clinical practice to modulate the intensity of treatment for patients showing significantly different risks of treatment failure.

1304 words

## REFERENCES

1. Song, M., Chung, J., Lee, J., Jeong, S., Lee, S., Hong, J. *et al.* Metabolic tumor volume by positron emission tomography/computed tomography as a clinical parameter to determine therapeutic modality for early stage Hodgkin's lymphoma. *Cancer Sci* 2013;104: 1656–1661
2. Kanoun, S, Rossi, C, Berriolo-Riedinger, *et al.* Baseline metabolic tumor volume is an independent prognostic factor in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging* 2014; 41: 1735–1743.
3. Tseng D, Rachakonda LP, Su Z, et al. Interim-treatment quantitative PET parameters predict progression and death among patients with Hodgkin's disease. *Radiat. Oncol.* 2012;7: 5.
4. Xie M, Wu K, Liu Y, Jiang Q, Xie Y. Predictive value of F-18 FDG PET/CT quantization parameters in diffuse large B cell lymphoma: a meta-analysis with 702 participants. *Med Oncol* 2015; 32: 446.
5. Kim J, Hong J, Kim SG, et al. Prognostic Value of Metabolic Tumor Volume Estimated by (18) F-FDG Positron Emission Tomography/Computed Tomography in Patients with Diffuse Large B-Cell Lymphoma of Stage II or III Disease. *Nucl Med Mol Imaging.* 2014;48:187-95.
6. Gallicchio R, Mansueto G, Simeon V, et al. F-18 FDG PET/CT quantization parameters as predictors of outcome in patients with diffuse large B-cell lymphoma. *Eur. J. Haematol.* 2014;92:382–389.
7. Mikhaeel NG, Smith D, Dunn JT, et al. Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL. *Eur J Nucl Med Mol Imaging.* 2016;43:1209-19.
8. Malek E, Sendilnathan A, Yellu M, Petersen A, Fernandez-Ulloa M, Driscoll JJ. Metabolic tumor volume on interim PET is a better predictor of outcome in diffuse large B-cell lymphoma than semiquantitative methods. *Blood Cancer J.* 2015; 5: e326.
9. Adams HJ, de Klerk JM, Fijnheer R, et al. Prognostic superiority of the National Comprehensive Cancer Network International Prognostic Index over pretreatment whole-body volumetric-metabolic FDG-PET/CT metrics in diffuse large B-cell lymphoma. *Eur J Haematol.* 2015; 94: 532-9.
10. Ceriani L, Martelli M, Zinzani PL et al. Utility of baseline 18FDG-PET/CT functional parameters in defining prognosis of primary mediastinal (thymic) large B-cell lymphoma. *Blood* 2015; 126(8): 950-956.
11. Cheebsumon P, Yaqub M, van Velden FH, et al. Impact of [(18)F]FDG PET imaging parameters on automatic tumor delineation: need for improved tumor delineation methodology. *Eur J Nucl Med Mol Imaging* 2011; 38: 2136–2144.