Dear Sir

Thank you for the opportunity to reply to the letter by Laffon and Marthan.

We disagree with their statement that the prognostic value of baseline metabolic tumour volume (MTV) depends on the outlining method. One of our study aims was to compare accuracy of various segmentation methods to predict survival [1]. We found in our cohort of 147 consecutive patients with diffuse large B cell lymphoma (DLBCL) treated at a single institution that all methods (SUV ≥ 2.5, SUV ≥ 41% of max SUV and SUV ≥ mean liver uptake) predicted progression free- and overall survival with similar accuracy. Instead, it was the optimal cut-offs for predicting survival that were dependent on the method, which is an important distinction.

This indicates that baseline MTV is a robust predictor of outcomes in DLBCL, but as stated in the discussion in our manuscript and reiterated by Laffon and Marthan, standardisation of the methodology used to measure MTV will be required. We believe this will need collaborative efforts to validate measurement of tumour burden in international patient datasets, in a manner previously adopted to standardise methods for response assessment in lymphoma, using the Deauville criteria [2-4].

Standardisation of the methodology should include the method used to segment the tumour volume and PET acquisition [5], including uptake time and reconstruction methods with which we concur with Laffon and Marthan.

Regarding the statistical analyses, none of the comparative combinations described by Laffon and Marthan revealed a normal distribution, so we remain convinced that a non-parametric approach was
more appropriate. In addition, the aim was to describe quantitative differences between the MTV methods rather than explore the relationships between them. With our study it would not be possible to explore in detail how agreement varies with MTV value, however, the large percentage variances calculated by Laffon and Marthan are based on median values, while the larger individual discrepancies contributing to the high upper limits of agreement in figures 1 and 3a in [1] appear well above the median (indicating a lower relative difference for those measurements). While we agree that the discrimination revealed by the ROC curves is fair but not strong, the survival analysis revealed highly significant differences between the between the high risk and low risk groups according to MTV for the cut-off values found for all methods.

We thank Laffon and Marthan for concluding that our work has addressed an important issue. We stand by the findings: that in our dataset, acknowledging the limitations discussed that the 2.5 method was the easiest to use and gave the best interobserver agreement, using 2 different softwares, although all methods were prognostic. The 400ml cut off using the SUV ≥2.5 method was in line with cut-offs reported by an independent group using SUV ≥ 2.5 in different populations of patients with DLBCL [6, 7]. It is a given that there will be more uncertainty about how MTV affects prognosis, for patients with data points that lie close to the 400ml cut off. This applies to any measurement that lies close to a binary threshold and refinement with larger datasets will reduce the degree of uncertainty.

Nonetheless measurement of disease burden remains an important parameter for study[8] and will no doubt become integrated into preclinical risk assessment of patients in the not too distant future.
Yours faithfully

Hajira Ilyas

N George Mikhaeel

Joel T Dunn

Fareen Rahman

Henrik Møller

Daniel Smith

Sally F Barrington

References


