Baseline Total Metabolic Tumor Volume measured with fixed or different adaptive thresholding methods equally predicts outcome in Peripheral T cell lymphoma.

Authors

A-S. Cottereau¹, S. Hapdey^{2,3}, L.Chartier⁴, R.Modzelewski^{2,3}, O. Casasnovas⁵, E. Itti¹, H. Tilly⁶, P. Vera^{2,3}, M. Meignan¹, S. Becker^{2,3}.

Affiliations

1 Nuclear Medicine Department, Hôpital Henri Mondor, University Paris-Est Créteil, France

2 Nuclear Medicine Department, Henri Becquerel Cancer Center and Rouen University Hospital, Rouen, France.

3 QuantIF-LITIS (EA[Equipe d'Accueil] 4108), Faculty of Medicine, University of Rouen, Rouen, France.

4 Department of Biostatistics (LYSARC), Centre Hospitalier Lyon Sud, Pierre Bénite, France.

5 Hematology Department, Hopital Le Bocage - CHU Dijon, France.

6 Hematology Department, UMR918, Henri Becquerel Cancer Center and Rouen University Hospital, Rouen, France.

Corresponding author

Anne-Ségolène Cottereau, MD

Address : Nuclear Medicine Department,

AP-HP, Groupe Henri Mondor Albert Chenevier, CHU Henri Mondor, 51 Avenue du Marechal de Lattre de Tassigny, 94010 Creteil, France

E-mail : annesegolene.cottereau@aphp.fr

Phone : +33 1 49 81 21 11

Conflict of interest

All authors have reviewed and approved the manuscript. The authors have no conflicts of interest to disclose.

This study has been presented during the annual meeting of the Society of Nuclear Medicine and Molecular Imaging (San Diego 11-15 june 2016).

Words count Abstract: 346 Total: 4270

Keywords:

Metabolic tumor volume-lymphoma-PTCL-adaptive thresholds

Running Title Methods of TMTV measurement in lymphoma

ABSTRACT

To compare in a large series of Peripheral T cell lymphoma (PTCL), as a model of diffuse disease, the prognostic value of baseline Total Metabolic Tumor Volume (TMTV) measured on FDG-PET/CT with adaptive thresholding methods to TMTV measured with a fixed 41% SUVmax threshold method. Methods: 106 patients with PTCL, staged with a PET/CT were enrolled from 5 LYSA centers. In this series TMTV computed with the 41% SUVmax threshold is a strong predictor of outcome (Ann Oncol, 2016). On a dedicated workstation, we measured the TMTV with four adaptive thresholding methods based on characteristic image parameters: Daisne (Da) modified based on signal/background ratio, Nestle (Ns) on tumor and background intensities, Fit including a 3D geometric model based on spatial resolution (Fit) and Black (Bl) based on mean SUVmax. The TMTV values obtained with each adaptive method were compared to those obtained with 41% SUVmax method. Their respective prognostic impacts on outcome prediction were compared using ROC analysis and Kaplan Meier survival curves. Results: The median value of TMTV_{41%}, TMTV_{Da}, TMTV_{Ns}, TMTV_{Fit}, TMTV_{Bl} were respectively 231 cm³ (range 5-3824), 175 cm³ (8-3510), 198 cm³ (3-3934), 175 cm³ (8-3512), and 333 cm³ (3-5113). The intra-class correlation coefficients were excellent from 0.972 to 0.988 for TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns}, less good for TMTV_{Bl} (0.856). The mean differences obtained from the Bland Altman plots were 48.5, 47.2, 19.5 and -253.3 cm³ respectively. Except for Black there was no significant difference within the methods between the ROC curves (p>0.4) for Progression Free Survival (PFS) and for Overall Survival (OS). Survival curves with the ROC optimal cutoff for each method separated the same groups of low risk (volume secutoff) from high risk patients (volume>cutoff) with similar 2v-PFS (range 66-72% vs 26-29%; HR 3.7-4.1) and 2y-OS (79-83% vs 50-53%, HR 3.0-3.5). Conclusion: The prognostic value of TMTV remained quite similar whatever the methods, adaptive or 41% SUVmax. This supports its use as a strong prognosticator in lymphoma. However for implementation of TMTV in clinical trials one single method easily applicable in a multicentric PET review must be selected and kept all along the trial.

INTRODUCTION

18 FDG-PET/CT has been recognized as the more valuable imaging tool in FDG-avid lymphoma for staging and response assessment. The last ICML recommendations (1) encourage investigating the quantitative analysis of FDG-PET/CT at staging. In this regard the measurement of the Total metabolic tumor volume (TMTV), which gives an estimation of the total tumor burden, has gained special interest. Indeed several series have shown that TMTV was predictive of outcome in different lymphoma subtypes, Diffuse Large B cell lymphoma (DLBCL) (2,3), Hodgkin lymphoma (HL) (4), Peripheral T cell lymphoma (PTCL) (5) and Follicular lymphoma (FL) (6). In these studies, different methods of TMTV measurement were used; all were based on a fixed thresholding principle to determine the metabolic volume of local tumors. The threshold can be absolute: a SUV value of 2.5 was generally chosen (2). It can be relative, using a percentage of the maximum uptake. A threshold of 41% of the SUVmax within the lesion, recommended by the European Association of Nuclear Medicine (EANM) for solid tumor (7) has been used in patients with HL (4), DLBCL (3) and PTCL (5) with a good inter observer reproducibility. However, since lymphomas are heterogeneous disease with several tumor sites with a wide range of volumes, SUV and tumor background ratios, the adaptive segmentation methods might be of interest for TMTV measurement and could be proposed as an alternative to fixed thresholding methods. The principle of these adaptive methods developed for radiotherapy planning of solid tumors is to adapt the threshold following a fitting model according to one or two characteristic image parameters, such as the SUV or the contrast. In a previous study we have demonstrated in a retrospective group of PTCL patients that TMTV measured with the 41%SUVmax threshold method was a good predictor of outcome (5). The aim of the present study was to compare in the same series, taken as a model of diffuse lymphoma, different adaptive thresholding methods to this fixed 41% method and to evaluate if they were better predictors of outcome than a fixed relative threshold.

MATERIALS AND METHODS

Patients

A group of consecutive patients with PTCL newly diagnosed during the time period 2006 to 2014, from 5 LYmphoma Study Association (LYSA) centers (Creteil, Dijon, Marseille, Rouen, and Liege) were included in this study. All were part of a previous study already published on the prognostic value of baseline TMTV (5). Patients had a baseline PET/CT, with central data available for review. The median follow up of this population was 23 months. The 2-year PFS and OS of the population were 49% and 67% respectively. The characteristics of the patients were: median age of 58 year old, 91 % of advanced stage, half of them with International prognostic index (IPI) > 2 and 45% with Prognostic index for PTCL (PIT) > 1, 80% treated with cyclophosphamide, doxorubicin, vincristine, prednisone (CHOP) /CHOP like therapy, 20% with cyclophosphamide, doxorubicin, vindesine, bleomycin (ACVBP). The study was conducted in accordance with the precepts of the Helsinki declaration and received approval by the Ethical Committee with a waiver of informed consent due to its retrospective nature (5).

PET acquisition

All the centers adhered to EANM guidelines for patient preparation and PET/CT acquisition. All patients were instructed to fast for at least 6 hours before the injection of 4-5 MBq per kilogram of 18F FDG, to ensure that serum glucose and endogenous serum insulin levels were low. Non contrast enhanced CT images were acquired before PET data acquisition. Whole body PET was

acquired sequentially using a dedicated PET/CT system. For the PET imaging, the emission data were acquired from the base of the skull to the proximal thigh with 3 to 3.5 min of acquisition per bed position. Biograph Sensation 16 Hi-Rez (Siemens Medical Solution, Knoxville, TN, USA) or Gemini GXL or Gemini TOF (Philips, Da Best, The Netherlands) were used by the 5 centers. All the devices used in this study followed a QC program insuring that the data were quantitatively correct (quarterly SUV verification). Four centers had obtained EARL accreditation at the time of the study and one was accredited according to RTEP procedure (8). The similarity of performances of the different equipment was confirmed with the analysis of the recovery curves obtained from their NEMA phantoms in terms of volume and contrast.

TMTV measurement

The baseline FDG PET/CT was processed with a Planet Onco workstation (Planet Onco v2.0, DOSISoft, Cachan, France) localized in Henri Becquerel Center, Rouen.

TMTV was computed following these steps:

First, the volumetric regions of interest (VOI) were placed around each lesion, avoiding physiological uptake (urinary elimination, heart). The reproducibility of the ROI setting has been evaluated, as previously published (5). Then the tumor volume was delineated with 5 thresholding methods: one fixed, 41% SUVmax considered thereafter as the reference (9) and 4 adaptive based on mathematical algorithms: Daisne modified by Vauclin *et al* (TMTV_{Da}) which iteratively adapt the threshold according to the local signal to background ratio (*10*), Fitting (TMTV_{Fit}) which fit the sphere image using a 3D geometric model based on the spatial resolution in the reconstructed images and on a tumor shape derived from activity thresholding (*11,12*), Nestle (TMTV_{Ns}) according to tumor and background intensities (*13*), and Black (TMTV_{BI}) according to the SUVmean (*14*). The tumor SUVmax and the liver SUVmax were also reported.

Statistical analysis:

Quantitative variables were expressed as median and ranges. Comparison of TMTV between methods was analyzed with the Friedman non-parametric test. When a significant difference was detected, Wilcoxon signed-rank test was performed, with Bonferroni correction for multiple testing. Correlation between TMTV values from the 41% SUVmax method and those from adaptive methods (Daisne modified, Fitting, Nestle and Black) was tested using Spearman coefficients. Agreement between methods were represented on the Bland Altman plots and quantified with the intra-class correlation coefficient (ICC) based on Shrout-Fleiss formulae.

For each method Receiver Operating Curve (ROC) were obtained to define the optimal TMTV cutoff for survival prediction. The prognostic relevance of each method to predict PFS and OS was pairwise comparison of these ROC curves. Survival functions were calculated using Kaplan-Meier (KM) estimates for each method using their optimal TMTV cutoff. Comparison between categories was made using the log-rank test and Cox proportional-hazards models. The agreement between the dichotomization of patients in low and high TMTV group obtained with the 41% SUVmax method and each adaptive method was tested with the Cohen's kappa coefficient. Survival functions were also calculated for each method by using the threshold of 230 cm³ determined with the 41% SUVmax method for PFS and OS (*5*).

Since PET/CT were performed in several institution and adaptive methods were optimized on Dosisoft workstation for the Rouen equipment (Biograph Sensation 16 Hi-Rez), we verified that the Spearman coefficients as well as the ICC based on Shrout-Fleiss formulae existing between methods were similar between Rouen subset and the whole population. Statistical analyses used SAS 9.2 and Med Calc.

RESULTS

One-hundred and six patients with PTCL newly diagnosed and for whom PET/CT could be retrieved for analysis on Dosisoft software were included in the present study. Characteristics of this group was similar to the initial population

TMTV measurements

The median value of TMTV_{41%}, TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns} and TMTV_{BI} were 231 cm³ (range 5-3824), 175 (8-3510), 175 (8-3512), 198 (3-3934), 333 (3-5113) respectively (Fig. 1). We observed no significant difference between Nestle and 41%SUVmax methods (p=0.7) but significant differences existed between the others adaptive methods and 41%SUVmax (p<0.001): from 24% for Fit and Daisne to 44% for Black. Median tumor SUVmax was 14, with a wide range of values (3.4 to 39.0). Tumor SUVmax/liver SUVmax ratio, taken as an index of tumor/background ratio, ranged from 1 to 15. Reproducibility of volume of interest setting was excellent with a Lin concordance correlation coefficient of p=0.995 (IC 95= 0.992 to 0.997) (5).

Comparison of TMTV values between the 41% SUVmax method and adaptive methods

The correlation between TMTV values from the 41% SUVmax method and those from adaptive methods was excellent with a Spearman coefficient of 0.99, 0.99, 0.97 and 0.96 for $TMTV_{Da}$, $TMTV_{Fit}$, $TMTV_{Ns}$, $TMTV_{Bl}$ respectively;

The Bland Altman plots are presented in Fig. 2. The mean of difference between TMTV_{41%} and TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns}, TMTV_{Bl} were 48.5 \pm 97.3 SD (95% CI -147;243), 47.2 \pm 96.7 (-

146;241), -19.5 \pm 170.6 (-360;322), and -253.3 \pm 393.8 (-1041;534) respectively. A major overestimation was observed with TMTV_{B1} compared to TMTV_{41%}. The differences between methods increased for the high TMTV, with a proportional error for TMTV values greater than 500cm³. Looking at TMTV under 500 cm³, the mean differences between 41%SUVmax and adaptive methods were really reduced: 21.8 \pm 36.3 for TMTV_{Da}, 21 \pm 34.8 for TMTV_{Fit}, 0.4 \pm 54.5 for TMTV_{Ns}, and -105.8 \pm 124.5 for TMTV_{B1}.

The intra-class correlation coefficient (ICC) varied from 0.972 and 0.988 for $TMTV_{Da}$, $TMTV_{Fit}$, $TMTV_{Ns}$ and was 0.856 for $TMTV_{Bl}$. The coefficient of variation between $TMTV_{41\%}$ and $TMTV_{Da}$, $TMTV_{Fit}$, $TMTV_{Ns}$, $TMTV_{Bl}$ were respectively 14%, 14%, 21% and 48%.

Prognostic value of the different methods

The respective optimal cutoff found with ROC analysis for TMTV_{41%}, TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns} and TMTV_{Bl} were 230 cm³, 132 cm³, 147 cm³, 277 cm³ and 345 cm³ for PFS and 260 cm³, 132 cm³, 147 cm³, 191 cm³ and 345 cm³ for OS (table 1). The respective AUC of TMTV_{41%}, TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns} and TMTV_{Bl} varied from 0.68 to 0.71 for PFS and from 0.60 to 0.62 for OS. For PFS the AUC obtained with TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns} were not significantly different (p>0.4 for each pairwise comparison) from the AUC of TMTV_{41%} (Fig. 3). A significant difference was observed for TMTV_{Bl} on PFS (p=0.02). No significant difference was observed for OS.

TMTV was significantly associated with inferior PFS (p<0.001) and OS (p<0.001) whatever the method of computation, with no significant difference between them and similar hazard ratio (table 1). The hazard ratio ranged from 3.7 (TMTV_{B1}) to 4.1 (TMTV_{41%}) on PFS and from 3.0

(TMTV_{BI}) to 3.5 (TMTV_{Da}) on OS. The 2y-PFS ranged from 66% to 72% for the low TMTV groups vs 26-29% for the high TMTV groups and 2y-OS from 79% to 83% vs 50 to 53% (Fig. 4). Based on the 230cm³ cutoff (threshold determined with the 41% SUVmax method), the agreement between the 41% SUVmax method and adaptive methods to dichotomize the population was almost perfect with a kappa of 0.87, 0.87, 0.87 for TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns} respectively. As expected due to the major overestimation, no agreement was observed with Black method. Applying to the various methods this single cutoff of 230 cm³ comparable PFS prediction was found (p<0.0001 HR=3.4 for Daisne modified and Fit, p<0.0001 HR=3.3 for Nestle, and p=0.0056 HR=2.4 for Black). This TMTV cutoff remained significant to predict overall survival except for Black (p=0.058). For the other adaptative methods, this TMTV cutoff was slightly less significant than when using specific optimal cutoff for each method (p=0.0055 HR=2.4 for Daisne modified and Fit, and p=0.0037 HR=2.6 for Nestle).

Comparison between Rouen and the others centers

Among the 106 patients, 28 were included in Rouen and 78 in the others centers. The coefficient of variation observed in the Rouen center between TMTV_{41%} and TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns}, TMTV_{B1} were respectively 11,3%, 11,2%, 20,7% and 40,1%, similar to those observed in the whole population. Correlations between methods were similar in the Rouen subgroup and in the other centers with a Spearman coefficient from 0.96 to 0.99 for both subgroups. ICC was also similar for TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns} between both subgroups, from 0.974 to 0.993 in the Rouen center and from 0.972 to 0.985 in the other centers. Regarding the Black method, ICC

observed in the other centers were slightly lower than in the Rouen center with an ICC of 0.836 compare to 0.9.

DISCUSSION

The major result of this study is to show that the prognostic value of baseline TMTV computed with several adaptive methods was similar to TMTV computed with 41% SUVmax threshold method in a large series of PTCL patients taken as an example of diffuse tumor disease.

Retrospective studies have demonstrated that TMTV was a powerful predictor of outcome in different lymphoma subtypes. TMTV measurement at baseline is important since it could help stratifying patient in different risk categories and has been suggested as a possible tool for early guiding therapy. However until now in lymphoma different TMTV methodologies have been used: an absolute cutoff of SUV>2.5 or a relative SUVmax thresholding of the tumor sites. The absolute threshold using SUV>2.5 is limited by the variability of SUV values, due to PET/CT devices, PET acquisition protocol and reconstruction methods. In addition due to partial volume effect non tumor regions located between small distant nodes with high uptake could be included (*15,16*). Therefore relative thresholds have been used in several lymphoma studies: a 41% SUVmax cutoff as recommended by EANM guidelines for solid tumors was applied in DLBCL, HL, FL, and PTCL and a 25% SUVmax recently evaluated in PBMCL (*17*).

Lymphoma characteristics, i.e a disseminated disease with different size of lesion, different sites with nodal or extra nodal lesions and heterogeneous FDG uptake (tumor/liver ratios varying from one to 15 in our series), might limit the efficiency of both fixed threshold methods. Conversely, adaptive methods may be more accurate and even easier to use in routine but they had not been yet tested on lymphoma.

In a previous study we have shown in PTCL that baseline TMTV with a 230 cm³ threshold was a good tool for outcome prediction and predicted progression free and overall survival much better than the currently used clinical index.

In the same series of patients we observed that the intra-class correlation coefficient found between TMTV values obtained with the three adaptive methods (Daisne modified, Fit and Nestle) and those from the 41% SUVmax method were excellent. The optimal threshold dichotomizing the population in low and high volumes groups for each adaptive method were different but, despite these differences, all these methods predicted PFS and OS with similar p and HR values for small and large volumes. The only slight incremental prognostic value compared to 41% threshold was observed for Daisne modified for OS prediction. Moreover when the same threshold of TMTV obtained with the 41%SUVmax method (230cm³) was used for Daisne modified, Fit and Nestle, the Hazard ratio obtained for PFS and OS prediction were comparable, which supported the fact that the values obtained with each methods were really closed. Only Blake was out of range with a major overestimation. This is probably explained because Black is based on the SUVmean whereas the others methods are based on the SUVmax.

Comparison between different methods of metabolic volume measurements has already been done in Hodgkin lymphoma. Kanoun et al (18) have compared in a monocentric study 41% SUVmax threshold to a per-patient adapted threshold based on SUVmax of the liver (>125% and >140% SUVmax of the liver background). They found no significant difference between ROC curves and similar prediction of PFS and OS according to high volume. It emphasized the strong prognostic value of metabolic imaging since TMTV remain prognostic throughout the different methods used. The current study is the first comparing adaptive methods to a fixed threshold method in a large series of patients with a diffuse subtype of lymphoma, PTCL. Our results

further confirm the strong prognostic value of baseline TMTV and demonstrate that these different methods of TMTV measurement equally predict outcome.

In this study we analyzed retrospective data acquired with different PET system on a workstation already calibrated to one of them for adaptive methods. However standardized phantom experiments, confirmed the similarity of image characteristics between centers and patient's data comparison between adaptive techniques and the 41% threshold was similar within centers. Indeed TMTV were all over 5 cm³ with a median of 231 cm³ an order of magnitude of volume where PET systems have similar detection capability.

Several quantitative measurements including TMTV and TLG have been done in ancillary studies of prospective trials based on quality controlled PET (2,6,17). However, to our knowledge, no ongoing trials have been launched using the TMTV to guide therapy. Even if the prerequisite for this type of trial is quality control, as done using various existing control systems (7,19,20), it is anyway required for good PET clinical practice. The main problem is which TMTV technique measurement should be chosen as there is no established consensus. We think that relative methods (SUVmax thresholding or adaptive) have the advantage to minimize the errors linked to the use of different devices and the participation of different centers.

In addition our results suggest that it is possible to conduct a prospective trial based on TMTV measurement provided one single relative method of TMTV measurement is used by all participating centers. The 41% TMTV threshold method is currently available in all commercial software and can be used in the majority of the PET/CT system. It has demonstrated a good reproducibility among trained observers but requires accurate manual drawing of the VOI around each lesion. Adaptive methods might be an option if available in all the centers.

14

CONCLUSION

Our results demonstrate that adaptive methods can be used with the same efficacy as 41% TMTV method in PTCL and would open the way to automatic procedures of volume computation. These conclusions should be confirmed for other types of diffuse aggressive lymphoma and new generation devices.

ACKNOWLEDGMENTS

Radiophysicists of all centers participating to the study: C Bernard, J Darreon, S Hapdey, H. Masset and JM Vrigneaud.

REFERENCES

1. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *J Clin Oncol.* 2014;32:3048-3058.

2. 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;43:1209-1219.

3. Cottereau AS, Lanic H, Mareschal S, et al. Molecular Profile and FDG-PET/CT Total Metabolic Tumor Volume Improve Risk Classification at Diagnosis for Patients with Diffuse Large B-Cell Lymphoma. *Clin Cancer Res.* 2016;22:3801-3809.

4. Kanoun S, Rossi C, Berriolo-Riedinger A, et al. Baseline metabolic tumour volume is an independent prognostic factor in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41:1735-1743.

5. Cottereau AS, Becker S, Broussais F, et al. Prognostic value of baseline total metabolic tumor volume (TMTV0) measured on FDG-PET/CT in patients with peripheral T-cell lymphoma (PTCL). *Ann Oncol.* 2016;27:719-724.

6. Meignan M, Cottereau AS, Versari A et al. Baseline Metabolic Tumor Volume predicts outcome in high tumor burden follicular lymphoma: a pooled analysis of three multicenter studies. *J Clin Oncol.* 2016;on line.

7. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328-354.

8. Vera P, Mezzani-Saillard S, Edet-Sanson A, et al. FDG PET during radiochemotherapy is predictive of outcome at 1 year in non-small-cell lung cancer patients: a prospective multicentre study (RTEP2). *Eur J Nucl Med Mol Imaging*. 2014;41:1057-1065.

9. Meignan M, Sasanelli M, Casasnovas RO, et al. Metabolic tumour volumes measured at staging in lymphoma: methodological evaluation on phantom experiments and patients. *Eur J Nucl Med Mol Imaging*. 2014;41:1113-1122.

10. Vauclin S, Doyeux K, Hapdey S, Edet-Sanson A, Vera P, Gardin I. Development of a generic thresholding algorithm for the delineation of 18FDG-PET-positive tissue: application to the comparison of three thresholding models. *Phys Med Biol.* 2009;54:6901-6916.

11. Tylski P GN, Giraud P, Rosenwald J, Buvat I. Experimental comparison of three methods for estimating tumor volume in FDG PET. *J Nucl Med.* 2007;48:43P.

12. Tylski P, Stute S, Grotus N, et al. Comparative assessment of methods for estimating tumor volume and standardized uptake value in (18)F-FDG PET. *J Nucl Med.* 2010;51:268-276.

13. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med.* 2005;46:1342-1348.

14. Black QC, Grills IS, Kestin LL, et al. Defining a radiotherapy target with positron emission tomography. *Int J Radiat Oncol Biol Phys.* 2004;60:1272-1282.

15. Quak E, Le Roux PY, Hofman MS, et al. Harmonizing FDG PET quantification while maintaining optimal lesion detection: prospective multicentre validation in 517 oncology patients. *Eur J Nucl Med Mol Imaging.* 2015;42:2072-2082.

16. Quak E, Hovhannisyan N, Lasnon C, et al. The importance of harmonizing interim positron emission tomography in non-Hodgkin lymphoma: focus on the Deauville criteria. *Haematologica*. 2014;99:e84-85.

17. 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;126:950-956.

18. Kanoun S, Tal I, Berriolo-Riedinger A, et al. Influence of Software Tool and Methodological Aspects of Total Metabolic Tumor Volume Calculation on Baseline [18F]FDG PET to Predict Survival in Hodgkin Lymphoma. *PLoS One.* 2015;10:e0140830.

19. Chauvie S, Bergesio F, Fioroni F, et al. The (68)Ge phantom-based FDG-PET site qualification program for clinical trials adopted by FIL (Italian Foundation on Lymphoma). *Phys Med.* 2016;32:651-656.

20. Palie O, Michel P, Menard JF, et al. The predictive value of treatment response using FDG PET performed on day 21 of chemoradiotherapy in patients with oesophageal squamous cell carcinoma. A prospective, multicentre study (RTEP3). *Eur J Nucl Med Mol Imaging*. 2013;40:1345-1355.



Figure 1: TMTV distribution according to each methodology, with median and interquartile range (box), mean (diamond) and outliers (circle).



Figure 2: Bland Altman analysis comparing TMTV values of $TMTV_{Da}$ (A), $TMTV_{Fit}$ (B), $TMTV_{Ns}$ (C), $TMTV_{Bl}$ (D) to $TMTV_{41\%}$. Mean bias and limits of agreements are represented by solids lines.



Figure 3: Progression-free survival (PFS) ROC curves comparison according to each methodology.



Figure 4 : Kaplan Meier estimates of progression-free survival (PFS) according to $TMTV_{Da}$ (A), $TMTV_{Fit}$ (B), $TMTV_{Ns}$ (C) and $TMTV_{B1}$ (D).

	PFS			OS		
	Specific threshold	Р	HR (CI 95%)	Specific threshold	Р	HR (CI 95%)
TMTV _{41%}	230 cm ³	<0.0001	4.1 (2.3-7.3)	260 cm ³	0.0005	3.1 (1.6-6.0)
TMTV _{Da}	132 cm ³	< 0.0001	4.0 (2.3-7.0)	132 cm^3	0.0007	3.5 (1.8-6.6)
TMTV _{Fit}	147 cm ³	< 0.0001	4.0 (2.3-6.9)	147 cm ³	0.0009	3.3 (1.7-6.2)
TMTV _{Ns}	277 cm ³	< 0.0001	3.9 (2.2-7.0)	191 cm ³	0.0007	3.2 (1.7-6.2)
TMTV _{Bl}	345 cm ³	< 0.0001	3.7 (2.1-6.6)	345 cm^3	0.0009	3.0 (1.6-5.8)

Table 1: ROC optimal TMTV cutoff, PFS and OS survival analyzing according to TMTV_{41%}, TMTV_{Da}, TMTV_{Fit}, TMTV_{Ns} and TMTV_{Bl}