Published Ahead of Print on September 6, 2016 as 10.1200/JCO.2016.69.3747
The latest version is at http://jco.ascopubs.org/cgi/doi/10.1200/JCO.2016.69.3747

JOURNAL OF CLINICAL ONCOLOGY

EDITORIAL

Metabolic Tumor Volume in Lymphoma: Hype or Hope?

Heiko Schöder and Craig Moskowitz, *Memorial Sloan Kettering Cancer Center, New York, NY*

See accompanying article doi:10.1200/JCO.2016.66.9440

Follicular lymphoma (FL) is a heterogeneous disease with a spectrum of genetic, biologic, laboratory, and clinical features that determine the need for, and the variable response to, first-line therapy and, ultimately, patient outcome.[1-4] As in other lymphomas, [^{18}F]fluorodeoxyglucose (FDG) –positron emission tomography (PET) and computed tomography (CT) imaging are widely used for staging and response assessment in FL. The intensity of FDG uptake—usually measured by standardized uptake values (SUVs)—varies widely among patients with FL, from barely detectable (corresponding to SUV numbers in the low single digits) to very intense (corresponding to high SUV numbers greater than 10). At the extreme points (very low uptake or very high uptake), the SUV may provide clinically relevant information. Low uptake is generally associated with low proliferation and less aggressive disease, while the opposite pertains to very high uptake, which may also herald transformation from indolent to aggressive lymphoma.[5]

Beyond SUV, other metrics for analysis of tumor FDG uptake include the metabolic tumor volume (MTV)[6] and an index called total lesion glycolysis (TLG).[7] MTV measures the volume of FDG-avid disease, for which three-dimensional regions of interest are drawn (typically by autosegmentation using computer software, with some manual adjustment by the reader) around individual lesions. Volumes for all lesions are then added to derive the total body metabolic tumor volume. TLG additionally considers the intensity of FDG uptake in each disease site (TLG = MTV × mean SUV within the lesion).

Historically, the interest in quantifying volumes of metabolically active disease was driven by three major intentions: to derive data that could be used for lesional dosimetry, to estimate patient prognosis, and to quantify the response to therapy. For many years, efforts to derive these quantitative indices remained an academic exercise that relied on homegrown software packages and manual contouring of each individual tumor site.[8-10] This only changed when computer algorithms and user-friendly, commercially available software packages became available for clinical research. Accordingly, the number of publications investigating the technical features and prognostic value of MTV has grown exponentially since the turn of the century, from fewer than 20 papers per year to more than 200 papers per year in 2015. Whereas most initial studies focused on solid tumors with regionally confined disease, the improvements in algorithms and automation in recent years have enabled the evaluation of total-body metabolic tumor volumes (TMTVs), even in patients with widespread systemic disease, including patients with lymphoma.[11-18]

In the article accompanying this editorial, Meignan et al[19] report on the prognostic value of TMTV in FL. The authors retrospectively compiled data from three clinical trials[20-22] and analyzed data from 185 patients. Using a threshold of 41% of maximum, the median TMTV in this population was 297 cm^3. The authors used three different approaches to define the optimal cutoff for TMTV as a predictor of survival: X-tile analysis,[23] receiver operating curve (ROC) analysis, and restricted cubic spline. Of these, X-tile is the primary reliable source of cut-point definition. This method uses a training set and a validation data set, improving the robustness of the analysis. Splines are useful for modeling the relationship between TMTV as a continuous variable and survival time, but their contribution to optimal cut-point definition is minimal. The authors also seem to have used ROC analysis with survival as a binary end point, ignoring the follow-up time. This would be inappropriate.[24] By X-tile analysis, the authors derived the TMTV (510 cm^3) that provided the best combined sensitivity and specificity for predicting progression-free survival (PFS). The 2-year PFS was 58% in patients with TMTV $>$ 510 cm^3, and 87% in patients with TMTV $<$ 510 cm^3. Interestingly, high-baseline TMTV and a persistently positive FDG-PET scan after induction therapy were both equally independent prognostic factors. Of course, baseline TMTV has the advantage of providing this prognostic information at the outset of treatment.

Shortcomings of the study by Meignan et al[19] largely relate to the retrospective nature of the project. For instance, the vast majority of patients (1,634 of 1,819) from the three clinical trials were not eligible for analysis, presumably because baseline staging PETs had not been performed or were not available for electronic volumetric analysis. Patients were treated with three different drug regimens and 16 patients remained on 2-year maintenance therapy with rituximab. PET scans were obtained on scanners of different generations from three different vendors, which may potentially affect the calculation of SUV and, therefore, also the calculation of TMTV. Uptake times for FDG were not standardized; although only patients with scans obtained less than 90 minutes after FDG injection were included, the range or even median of uptake times was not provided. Because tumor FDG uptake, in general, increases with time after injection, lack of

standardization of uptake times can lead to variable FDG SUV that is not due to differences in tumor biology but simply technical reasons.

It seems intuitive that patients with high tumor burden should be at higher risk for treatment failure and shorter survival than those with low tumor burden. Current staging systems are built on this paradigm, and prognostic scoring systems use clinical and biochemical surrogates (eg, lactate dehydrogenase levels) to account for overall tumor burden. Both the term "tumor burden" and the idea that this is associated with prognosis in patients with lymphoma date back at least 30 years.[25] Several prior studies investigated the prognostic utility of TMTV or TLG in lymphoma (Table 1). Potential shortcomings in most of these studies relate to a lack of standardization with regard to PET scanning and treatment regimens. In nearly all of these studies, so-called optimal cutoff values were derived from retrospective ROC analysis, the limitations of which were previously discussed. More appropriate and sophisticated statistical tests do exist[23,27,28] but are rarely used, thus limiting the value of many publications.

To place the study by Meignan et al[19] further into context, it may be helpful to shed some light on the technical aspects and potential pitfalls in calculating TMTV. The authors used a threshold of 41% of maximum for PET volume autosegmentation, which originally dates back to phantom studies conducted about two decades ago.[29] With this approach, the calculation of MTV is based on finding the area of highest FDG uptake in a particular disease site (eg, with SUV 12), and then calculating a three-dimensional volume that encompasses all volume elements (voxels) up to a certain threshold. That is, with a threshold of 40% of maximum, all voxels with SUV $>$ 4.8 would be included to calculate the MTV for this particular lesion. The outer boundaries of the volume can be confined to anatomic boundaries, as seen on the corresponding CT of the PET/CT scans. However, appropriate thresholds may well depend on primary tumor SUV and on the anatomic location of a lesion. The latter determines the contrast between FDG uptake in that lesion and regional background activity in the surrounding normal tissue. This is particularly important for lesions with relatively low FDG

Table 1. Studies on the Prognostic Value of MTV in Lymphoma

Study	Type of Lymphoma	Patients (No.)	Tumor Volume Parameters				Predictors of PFS	Determination of MTV Cutoff
			Median SUV	Threshold (%)	Median MTV (cm^3)	Range (cm^3)*		
Kanoun et al[15]	HL	59	NR	41	117	4-1,611	MTV 225 cm^3 yields 4-year PFS 85% *v* 42%	ROC analysis†, no validation sample
Sasanelli et al[17]	DLBCL	114	NR	41	313	4-2,650	MTV 550 cm^3 yields 3-year PFS 77% *v* 60%	ROC analysis, no validation sample
Adams et al[11]	DLBCL	73	22.0	40	272	6-2,454	Neither MTV nor TLG predicted outcome	N/A
Mikhaeel et al[16]	DLBCL	147	27.2	41	595	2-7,360	MTV 396 cm^3 yields 5-year PFS 92% *v* 42% Best predictive model combines MTV with i-PET Deauville score	ROC analysis, no validation sample
Cottereau et al[26]	DLBCL	81	18	41	320	IQR: 106-668	MTV 300 cm^3 yields 5-year PFS 75% *v* 42%	ROC analysis, no validation sample
Schöder et al[18]	DLBCL	65	23.4	Various‡	226	9-3,453	MTV did not predict outcome	N/A
Ceriani et al[12]	PMBL	103	18.8	25	406	NR	MTV 703 cm^3 yields 5-year PFS 97% *v* 60% TLG 5,814 yields 5-year PFS 99% *v* 64%	ROC analysis, no validation sample
Cottereau et al[13]	PTCL	108	14	41	224	3-3,824	MTV 230 cm^3 yields 2-year PFS 71% *v* 26%	ROC analysis, no validation sample
Meignan et al[19]	FL 1-3a	185	10.0	41	297	IQR: 135-567	MTV 510 cm^3 yields 2-year PFS 87% *v* 58%	X-tile analysis

Abbreviations: DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; HL, Hodgkin lymphoma; i-PET, interim positron emission tomography; IQR, interquartile range; MTV, metabolic tumor volume; N/A, not applicable; NR, not reported; PFS, progression-free survival; PMBL, primary mediastinal B-cell lymphoma; PTCL, peripheral T-cell lymphoma; ROC, receiver operator curve; SUV, standardized uptake value; TLG, total lesion glycolysis.
*Unless given as IQR.
†ROC analysis is performed to derive best combined sensitivity and specificity along the ROC; study population is then dichotomized by so-called optimal threshold derived from ROC analysis.
‡Tested various proposed thresholds, including 41%.

uptake. True SUV is underestimated in smaller lesions because of partial volume effects (the exact size up to which this underestimation occurs depends on the specific PET scanner and its recovery coefficient). Even for two equally sized lesions that differ only in their maximum SUV, the metabolic volume will differ considerably when the same percent-of-maximum SUV threshold is applied. This was recognized in a recent study in primary mediastinal B-cell lymphoma in which investigators chose a threshold of 25% of maximum for measuring MTV, on the basis of their observation that this provided the best agreement between metabolic and anatomic boundaries in large mediastinal masses with high SUV.[12] In other words, choosing a higher percent-of-maximum threshold would have led to calculating smaller metabolic volumes, potentially excluding large portions of the anatomic mass or areas of central necrosis. Finally, SUV measurements, and, hence, MTV calculations on the basis of fixed percent-of-maximum SUV thresholds, also depend on many other biologic and technical factors[30] that need to be standardized to assure reproducibility of measurements.

Several other methods for autosegmentation of PET volumes exist (eg, threshold-based, gradient-based, statistical, and texture-based methods). All have specific advantages and disadvantages, and none may be completely accurate in measuring tumor volumes in all organs and settings. However, one might argue that accuracy (ie, our ability to derive an MTV that reflects the true viable tumor tissue volume in cm^3) is perhaps less important than reproducibility (ie, our ability to arrive at the same volume regardless of the equipment and software used for image acquisition, reconstruction, and analysis) as long as the outcome of our measurements yields prognostic information. For instance, Meignan et al[19] used two different software programs for calculating MTV (PET-VCAR [GE Healthcare, Little Chalfont, UK] and Imagys [Keosys, Saint-Herblain, France]). To their credit, the investigators established the reproducibility of measurements in a random subset of 20% of scans with reasonable results.

Meignan et al[19] pooled data from three clinical trials conducted across many countries and continents. Given the retrospective nature of the study, we must assume that the PET scanners were not cross-calibrated to ensure that the same SUV was measured when the same lesion was assessed on different scanners. Thus, while the authors can assure us that the TMTV calculations were reasonably reproducible when performed by different observers using two distinct software packages, it is unknowable whether the same TMTV would have been calculated if the same patient, under the same biologic conditions, would have undergone PET imaging on two different scanners. Finally, modern PET scanners using the time-of-flight technique (with or without point-spread function reconstruction) generally provide better detection of smaller FDG-avid lesions and often yield higher SUV numbers than scanners of older generations, in particular for smaller lesions.[31,32] Although application of appropriate image reconstruction and filtering techniques can help to standardize SUV measurements across participating sites in multicenter trials,[33] this obviously could not be done in the study by Meignan et al in view of its retrospective nature.

Is it too early to declare victory and call TMTV a new prognostic biomarker in lymphoma? Probably yes. As a single parameter, TMTV remains far from perfect for prognostication. In the Meignan et al study,[19] added prognostic information was derived when TMTV was used in combination with the Follicular Lymphoma International Prognostic Index score. In other studies, a combination of TMTV with treatment response assessed on interim PET imaging led to better segregation of prognostic groups.[16] Quantitatively, the situation also seems less clear; all proposed so-called optimal TMTV cutoffs (Table 1) were derived retrospectively, varied from study to study, and will require prospective validation. Importantly, cutoffs are only indicators of probability (in this case, the probability of poor PFS) and always reflect a trade-off between sensitivity and specificity. Cutoffs may also vary depending on the specific patient population, the range of tumor volumes in this population, and, possibly, the drug regimen used. The latter seems particularly important because biomarkers may potentially lose their prognostic power when standard treatment is changed to a more aggressive regimen,[34] or when novel targeted therapies are administered. Going forward, we should better understand the biologic reasons for why TMTV is associated with patient outcome and ask how these data could be put to good clinical use. Does a large disseminated tumor volume make a patient less likely to respond to a certain drug regimen and dose? Is this related to suboptimal drug concentration at each tumor cell? Is it not surprising that, currently, patients with a TMTV of 300 cm^3 often receive exactly the same treatment (regimen and dose) as do patients with a TMTV of 3,000 cm^3? Do patients with larger TMTVs simply need more of the same treatment, or would it be reasonable to offer these patients more aggressive therapies while sparing patients with lesser tumor volumes (and better prognosis) the adverse effects associated with such treatments? On the contrary, should patients with smaller TMTVs receive less chemotherapy?

On other fronts, it is becoming increasingly clear how the tumor microenvironment contributes to the prognosis and drug response in patients with FL,[35] and a "clinicogenetic" prognostic score incorporating a seven-gene signature has been proposed for better risk stratification.[4] In the end, the most meaningful and actionable information will probably come from a combination of clinical, imaging, and biologic factors rather than from any single parameter alone. It will take some time until all of these novel imaging and biologic features are confirmed, tests become standardized and widely available, and results can be integrated into the routine armamentarium of the hematologic oncologist. In the meantime, we need hypothesis-testing studies that apply imaging features and biologic signatures for better risk stratification and treatment selection in patients with lymphoma. Despite their shortcomings, the data provided by Meignan et al[19] move us one step along in this direction.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at www.jco.org.

AUTHOR CONTRIBUTIONS

Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Bouska A, Zhang W, Gong Q, et al: Combined copy number and mutation analysis identifies oncogenic pathways associated with transformation of follicular lymphoma. Leukemia doi: 10.1038/leu [e-pub ahead of print on July 8, 2016]

2. Dave SS, Wright G, Tan B, et al: Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. N Engl J Med 351: 2159-2169, 2004

3. Nooka AK, Nabhan C, Zhou X, et al: Examination of the follicular lymphoma international prognostic index (FLIPI) in the National LymphoCare study (NLCS): A prospective US patient cohort treated predominantly in community practices. Ann Oncol 24:441-448, 2013

4. Pastore A, Jurinovic V, Kridel R, et al: Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: A retrospective analysis of a prospective clinical trial and validation in a population-based registry. Lancet Oncol 16:1111-1122, 2015

5. Schöder H, Noy A, Gönen M, et al: Intensity of 18fluorodeoxyglucose uptake in positron emission tomography distinguishes between indolent and aggressive non-Hodgkin's lymphoma. J Clin Oncol 23:4643-4651, 2005

6. Zasadny KR, Kison PV, Francis IR, et al: FDG-PET determination of metabolically active tumor volume and comparison with CT. Clin Positron Imaging 1: 123-129, 1998

7. Larson SM, Erdi Y, Akhurst T, et al: Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging. The visual response score and the change in total lesion glycolysis. Clin Positron Imaging 2:159-171, 1999

8. Kolbert KS, Sgouros G, Scott AM, et al: Implementation and evaluation of patient-specific three-dimensional internal dosimetry. J Nucl Med 38:301-308, 1997

9. Sgouros G, Chiu S, Pentlow KS, et al: Three-dimensional dosimetry for radioimmunotherapy treatment planning. J Nucl Med 34:1595-1601, 1993

10. Wang W, Larson SM, Fazzari M, et al: Prognostic value of [^{18}F]fluorodeoxyglucose positron emission tomographic scanning in patients with thyroid cancer. J Clin Endocrinol Metab 85:1107-1113, 2000

11. 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 94:532-539, 2015

12. 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 126:950-956, 2015

13. 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 27:719-724, 2016

14. 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 22:3801-3809, 2016

15. 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 41:1735-1743, 2014

16. 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 43:1209-1219, 2016

17. Sasanelli M, Meignan M, Haioun C, et al: Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging 41:2017-2022, 2014

18. Schöder H, Zelenetz AD, Hamlin P, et al: Prospective study of 3′-deoxy-3′-^{18}F-fluorothymidine PET for early interim response assessment in advanced-stage B-cell lymphoma. J Nucl Med 57:728-734, 2016

19. Meignan M, Cottereau AS, Vesari 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 doi: 10.1200/JCO.2016.66.9440

20. Dupuis J, Berriolo-Riedinger A, Julian A, et al: Impact of [(18)F]fluorodeoxyglucose positron emission tomography response evaluation in patients with high-tumor burden follicular lymphoma treated with immunochemotherapy: A prospective study from the Groupe d'Etudes des Lymphomes de l'Adulte and GOELAMS. J Clin Oncol 30:4317-4322, 2012

21. Federico M, Luminari S, Dondi A, et al: R-CVP versus R-CHOP versus R-FM for the initial treatment of patients with advanced-stage follicular lymphoma: Results of the FOLL05 trial conducted by the Fondazione Italiana Linfomi [Erratum: J Clin Oncol 32:1095, 2014]. J Clin Oncol 31:1506-1513, 2013

22. Salles G, Seymour JF, Offner F, et al: Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): A phase 3, randomised controlled trial. Lancet 377:42-51, 2011

23. Camp RL, Dolled-Filhart M, Rimm DL: X-tile: A new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 10:7252-7259, 2004

24. Altman DG, Lausen B, Sauerbrei W, et al: Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. J Natl Cancer Inst 86:829-835, 1994

25. Jagannath S, Velasquez WS, Tucker SL, et al: Tumor burden assessment and its implication for a prognostic model in advanced diffuse large-cell lymphoma. J Clin Oncol 4:859-865, 1986

26. Cottereau AS, Lanic H, Mareschal S, et al: Molecular profile and FDG-PET metabolic volume at staging in DLBCL-response. Clin Cancer Res 22:3414-3415, 2016

27. Heagerty PJ, Lumley T, Pepe MS: Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 56:337-344, 2000

28. Zheng Y, Cai T, Feng Z: Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. Biometrics 62:279-287, 2006

29. Erdi YE, Mawlawi O, Larson SM, et al: Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer 80(12 suppl):2505-2509, 1997

30. Huang SC: Anatomy of SUV: Standardized uptake value. Nucl Med Biol 27: 643-646, 2000

31. Akamatsu G, Mitsumoto K, Taniguchi T, et al: Influences of point-spread function and time-of-flight reconstructions on standardized uptake value of lymph node metastases in FDG-PET. Eur J Radiol 83:226-230, 2014

32. Knäusl B, Rausch IF, Bergmann H, et al: Influence of PET reconstruction parameters on the TrueX algorithm. A combined phantom and patient study. Nucl Med (Stuttg) 52:28-35, 2013

33. 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 42:2072-2082, 2015

34. Moskowitz CH, Schöder H, Teruya-Feldstein J, et al: Risk-adapted dose-dense immunochemotherapy determined by interim FDG-PET in advanced-stage diffuse large B-cell lymphoma. J Clin Oncol 28:1896-1903, 2010

35. Kahl BS, Yang DT: Follicular lymphoma: Evolving therapeutic strategies. Blood 127:2055-2063, 2016

DOI: 10.1200/JCO.2016.69.3747; published online ahead of print at www.jco.org on September 6, 2016.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Metabolic Tumor Volume in Lymphoma: Hype or Hope?

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or jco.ascopubs.org/site/ifc.

Heiko Schöder
No relationship to disclose

Craig H. Moskowitz
No relationship to disclose