Is ^{18}F-FDG Metabolic Tumor Volume in Lymphoma Really Happening?

Sally F. Barrington[1], Anne-Ségolène Cottereau[2], and Josée M. Zijlstra[3]

[1]King's College London and Guy's and St. Thomas' PET Centre, School of Biomedical Engineering and Imaging Sciences, King's College London, United Kingdom; [2]Department of Nuclear Medicine, Cochin Hospital, APHP, Paris Cité University, Paris, France; and [3]Department of Hematology and Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

Tumor burden influences prognosis in lymphoma, with unidimensional bulk used for risk assessment and decisions about radiotherapy consolidation (1). Total metabolic tumor volume (TMTV) using ^{18}F-FDG PET emerged 10 y ago as a promising biomarker (2) that was superior to bulk (3). However, until now, metabolic tumor volume (MTV) has not been used in clinical practice or trial design. We attribute this to lack of common methodology, the perception that measurement is difficult, and the unavailability of software tools. Consensus is also required about which tumor areas to include (4). Furthermore, MTV has been evaluated in datasets providing binary cutoffs to divide patients into prognostic groups, which are data-driven and population-dependent (5).

WHAT IS CHANGING?

A New Benchmark Method Has Been Established

Because prognostication and interobserver agreement are equally good irrespective of the measurement method (6), choice should reflect ease of use. One method has emerged as simple, quick to perform using academic and commercial software, and closely matching the visual perception of nuclear medicine reads from 6 published methods in diffuse large B-cell lymphoma (DLBCL) (7) and Hodgkin lymphoma (8). The delineation method uses an SUV of at least 4.0 and a minimum individual lesion volume of 3 cm^3. The SUV of 4.0 limits physiologic uptake that requires editing compared with lower thresholds and reduces underestimation of heterogeneous lesions compared with percentage SUV thresholds. The 41% SUV$_{max}$ threshold, although frequently studied, has increased variability across software depending on whether SUV$_{max}$ is defined as the maximum in the TMTV or as lesional. If lesional, the clustering algorithm can influence how lesions are outlined. The minimum volume reduces measurement complexity without significantly influencing TMTV. The method using an SUV of 4.0 is insensitive to uptake time and to the presence or absence of later progression in patients (7) and is the least sensitive to reconstruction method, including ultra-high-sensitivity reconstructions (9) used in advanced technologies such as total-body PET/CT. In approximately 80% of DLBCL cases, minimal reader interaction was required—for example, removing physiologic uptake with single clicks, achievable in 2–3 min (7) with additional manual editing in 20% of cases.

The vision for standardization of MTV measurement was outlined in this journal (4) with a proposal for a benchmark dataset using a common method with consensus MTV values and segmentations as outputs. Data from the MTV road map have been presented involving 12 readers from 9 countries using 3 academic software programs. Readers analyzed 60 cases from 3 lymphoma subtypes (10). TMTV measurement was unaffected by the software used, with close reader agreement in 52 of 60 cases. Disagreement was mainly due to interpretation of diffuse splenic uptake with smaller, less clinically relevant differences due to manual editing of physiologic uptake.

The benchmark will soon be made publicly available so readers can check the reliability of MTV measurements using local software and their clinical interpretation. New measurement methods, including artificial intelligence approaches, can be evaluated against the benchmark, with both being tested in the same dataset provided patient outcomes are known, to determine whether newer methods improve prognostication, reduce reader time, or improve agreement. The benchmark can also be used to explore questions such as the prognostic relevance of the spleen.

A concern about the transition to one method could be whether research using other methods might be wasted and that the method using an SUV of 4.0 has not been widely tested in indolent, albeit less common, subtypes. A statistical method for combining batches (ComBat) of data using different methods (11) has been successfully applied in retrospective trial datasets (12). Nonetheless, having a standardized approach for prospective study in aggressive lymphoma is a critical step for universal adoption of MTV as a biomarker.

New Prognostic Indices Have Been Incorporated

Other major developments are testing of TMTV in large datasets, expression as a continuous variable, and incorporation with established risk factors.

The method using an SUV of 4.0 was explored in 1,214 patients with newly diagnosed DLBCL (13) by the Positron Emission Tomography Reanalysis (PETRA) consortium in 5 international trials. First, the best statistical relationship was derived to associate MTV with progression-free and overall survival. The relationship was a linear spline with 2 coefficients, such that the same incremental change had different impacts on survival above and below the median. MTV performed better than the International Prognostic Index (comprising binary cutoffs for age, lactate dehydrogenase, stage, performance status, and more than one extranodal site) (14). Most International Prognostic Index factors proved redundant when combined with MTV. The optimal "International Metabolic Prognostic Index" included 3 factors: MTV and age (as continuous variables) and stage (I–IV). Its continuous nature means progression-free

survival can be predicted for individual patients by entering MTV, age, and stage in a simple Excel (Microsoft) spreadsheet (https://petralymphoma.org/impi). The International Metabolic Prognostic Index allows for intelligent trial design, selecting a progression-free survival cutoff at which the benefit of a novel treatment will likely outweigh the risk of standard treatment in high-risk patients. The integration of MTV with patient factors has also been explored in 2,174 patients from trial and real-world datasets, with performance status being identified as an independent risk factor and MTV plus performance status outperforming the International Prognostic Index (12). Optimal selection of high-risk patients is relevant because new treatments such as CAR T-cell therapy and bispecific monoclonal antibodies are being tested in phase III trials in first- and second-line DLBCL.

HOW SHOULD WE BUILD ON THE SUCCESS OF MTV?

The success of MTV has generated interest in other radiomic features, which can be measured once TMTV is delineated. The independent prognostic value of disease dissemination—for example, the maximum distance between lesions—was first reported by Cottereau et al. (15). Biologic explanations for this phenomenon were recently explored in Hodgkin lymphoma (16). The PETRA consortium suggested the potential to replace "stage" in the International Metabolic Prognostic Index by "Dmaxbulk," the maximum distance between the bulkiest and the furthest lesion, with a small incremental benefit for a radiomics score that also included performance status and SUV_{peak} (17). Preliminary reports that integrate PET with emerging molecular markers in circulating tumor DNA (18) may further improve baseline and dynamic risk. Methods to establish reliable dynamic MTV measurement are also being explored (19).

Confidence in TMTV is growing with agreement about standardization. A similar pragmatic approach of a simple, albeit not perfect, method led to widespread adoption of the Deauville score for lymphoma (20).

Now MTV needs to feature in trial design, either alone or within prognostic indices such as the International Metabolic Prognostic Index for risk stratification. To develop clinical decision tools, MTV (with or without other radiomic features) should be prospectively evaluated at baseline and interim with liquid biomarkers for minimal residual disease.

The first trial using MTV, Deauville score, and circulating tumor DNA to risk-adapt treatment is already under way in Hodgkin lymphoma (www.clinicaltrials.gov/study/NCT04866654).

In conclusion, MTV for risk stratification in DLBCL is feasible now in the clinic and being evaluated in a clinical trial on Hodgkin lymphoma. A benchmark dataset will be available soon for standardization of measurement by PET centers, software developers, and vendors. "The time to prepare for risk adaptation in lymphoma by standardizing measurement of metabolic tumor burden is over" (4): it is time to get on board.

REFERENCES

1. Specht L, Nordentoft AM, Cold S, Clausen NT, Nissen NI. Tumor burden as the most important prognostic factor in early stage Hodgkin's disease. Relations to other prognostic factors and implications for choice of treatment. *Cancer*. 1988;61:1719–1727.
2. 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.
3. 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.
4. Barrington SF, Meignan M. Time to prepare for risk adaptation in lymphoma by standardizing measurement of metabolic tumor burden. *J Nucl Med*. 2019;60:1096–1102.
5. El-Galaly TC, Villa D, Cheah CY, Gormsen LC. Pre-treatment total metabolic tumour volumes in lymphoma: does quantity matter? *Br J Haematol*. 2022;197:139–155.
6. Ilyas H, Mikhaeel NG, Dunn JT, et al. Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2018;45:1142–1154.
7. Barrington SF, Zwezerijnen BGJC, de Vet HCW, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method is most successful? A study on behalf of the PETRA consortium; 32680929. *J Nucl Med*. 2021;62:332–337.
8. Driessen J, Zwezerijnen GJC, Schöder H, et al. The impact of semiautomatic segmentation methods on metabolic tumor volume, intensity, and dissemination radiomics in ^{18}F-FDG PET scans of patients with classical Hodgkin lymphoma. *J Nucl Med*. 2022;63:1424–1430.
9. Ferrández MC, Eertink JJ, Golla SSV, et al. Combatting the effect of image reconstruction settings on lymphoma [^{18}F]FDG PET metabolic tumor volume assessment using various segmentation methods. *EJNMMI Res*. 2022;12:44–49.
10. Boellaard R, Buvat I, Ceriani L, et al. International benchmark for total metabolic tumor volume assessment in baseline FDG PET/CT of lymphoma patients [abstract]. *J Nucl Med*. 2023;64(suppl 1):83.
11. Orlhac F, Eertink JJ, Cottereau A, et al. A guide to ComBat harmonization of imaging biomarkers in multicenter studies. *J Nucl Med*. 2022;63:172–179.
12. Thieblemont C, Chartier L, Dührsen U, et al. A tumor volume and performance status model to predict outcome before treatment in diffuse large B-cell lymphoma. *Blood Adv*. 2022;6:5995–6004.
13. Mikhaeel NG, Heymans MW, Eertink JJ, et al. Proposed new dynamic prognostic index for diffuse large B-cell lymphoma: international metabolic prognostic index. *J Clin Oncol*. 2022;40:2352–2360.
14. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med*. 1993;329:987–994.
15. Cottereau AS, Meignan M, Nioche C, et al. Risk stratification in diffuse large B-cell lymphoma using lesion dissemination and metabolic tumor burden calculated from baseline PET/CT. *Ann Oncol*. 2021;32:404–411.
16. Durmo R, Donati B, Rebaud L, et al. Prognostic value of lesion dissemination in doxorubicin, bleomycin, vinblastine, and dacarbazine-treated, interimPET-negative classical Hodgkin lymphoma patients: a radio-genomic study. *Hematol Oncol*. 2022;40:645–657.
17. Eertink JJ, Zwezerijnen GJC, Heymans MW, et al. Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma. *Blood*. 2023;141:3055–3064.
18. Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. *J Clin Oncol*. 2018;36:2845–2853.
19. Zwezerijnen GJC, Eertink JJ, Burggraaff CN, et al. Interobserver agreement on automated metabolic tumor volume measurements of Deauville score 4 and 5 lesions at interim ^{18}F-FDG PET in diffuse large B-cell lymphoma. *J Nucl Med*. 2021;62:1531–1536.
20. 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.