Journal of Nuclear Medicine, published on December 29, 2022 as doi:10.2967/jnumed.122.264740

Prognostic value of ^{18}F-FDG PET/CT in diffuse large B-cell lymphoma treated with a risk-adapted immunochemotherapy regimen

Laure Michaud*[1], Kurt Bantilan*[2], Audrey Mauguen[3], Craig H. Moskowitz[4], Andrew D. Zelenetz[2], Heiko Schöder[1]

*Contributed equally to the manuscript

[1]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[2]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[3]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[4]Department of Medicine, University of Miami Health System, Miami, FL, USA

Word count: Abstract 250; text 4,125

Running title: Prognostic value of FDG-PET in large cell lymphoma

Address for correspondence:

Heiko Schöder, MD
Molecular Imaging and Therapy Service
Memorial Sloan Kettering Cancer Center
1275 York Ave
New York, NY 10065
Email: schoderh@mskcc.org

Abstract

Early identification of patients with diffuse large B-cell lymphoma (DLBCL) who are likely to experience disease recurrence or refractory disease after rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) would be useful for improving risk-adapted treatment strategies. We aimed to assess the prognostic value of ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (FDG-PET/CT) parameters at baseline, interim, and end of treatment (EOT).

Methods: We analyzed the prognostic impact of FDG-PET/CT in 166 patients with DLBCL treated with a risk-adapted immunochemotherapy regimen. Scans were performed at baseline, after four cycles of R-CHOP or three cycles of RR-CHOP and one cycle of CHOP alone (interim) and six weeks after completing therapy (EOT). Progression-free survival (PFS) and overall survival (OS) were estimated using Kaplan-Meier and the impact of clinical/PET factors assessed with Cox models. We also assessed the predictive ability of the recently proposed International Metabolic Prognostic Index (IMPI).

Results: The median follow-up was 7.9 years. International Prognostic Index (IPI), baseline metabolic tumor volume (MTV), and change in maximum standardized uptake value (ΔSUVmax) at interim scans were statistically significant predictors for OS. Baseline MTV, interim ΔSUVmax, and EOT Deauville score were statistically significant predictors of PFS. Combining interim PET parameters demonstrated that patients with Deauville 4-5 and positive ΔSUVmax ≤ 70% at restaging (approximately 10% of the cohort) had extremely poor prognosis. The IMPI had limited discrimination and slightly overestimated the event rate in our cohort.

Conclusion: Baseline MTV and interim ΔSUVmax predicted both PFS and OS with this sequential immunochemotherapy program. Combining interim Deauville score with interim ΔSUVmax may identify an extremely high-risk DLBCL population.

Keywords: FDG-PET/CT, diffuse large B-cell lymphoma, metabolic tumor volume, delta SUVmax, Deauville score

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is a common and aggressive lymphoma subtype. The treatment regimen of rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) is considered the standard first-line DLBCL treatment, with a long-term remission rate of 60 to 70% (*1*). However, patients who do not respond to R-CHOP have a poor prognosis, and pre-treatment prognostic models such as the International Prognostic Index (IPI) that are used to predict survival (*2*) fail to identify these high-risk patients. Several studies have evaluated more aggressive first-line treatments using risk-adapted strategies for patients with good versus poor prognosis (*3,4*). Hence, early identification of patients who are likely to experience disease recurrence or refractory disease after R-CHOP is important for improving stratification to modified and innovative regimens.

^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (FDG-PET/CT) scans at baseline have proven to be highly sensitive in determining sites of disease for DLBCL (*5,6*). Furthermore, PET/CT scans at the end of treatment (EOT) have demonstrated high prognostic value for assessing long-term remission (*6*). However, there is still no consensus on the predictive value of interim PET/CT scans in the management of patients with DLBCL. Evidence that changing treatment strategy based on interim PET/CT scans improves outcome remains to be confirmed (*4,6,7*).

Imaging biomarkers have often been evaluated separately. Parameters calculated from PET/CT, such as metabolic tumor volume (MTV) at baseline and change in maximum standardized uptake value between baseline and interim scans (ΔSUVmax), were demonstrated to be prognostic in

DLBCL (*1,3,7-12*) and may prove useful for risk stratification. Recently, a simple prognostic model, the International Metabolic Prognostic Index (IMPI), which combines baseline MTV, age, and stage was shown to predict outcomes in DLBCL with higher accuracy than the IPI (*13*). Against this background, we aimed to assess the prognostic value of baseline, interim, and EOT FDG-PET/CT scans and validate IMPI in patients with DLBCL who were uniformly treated with a risk-adapted immunochemotherapy regimen.

METHODS

Study Population

Two risk-adapted studies treating patients with advanced-stage large cell lymphomas were approved by Memorial Sloan Kettering Cancer Center (MSK)'s Institutional Review Board. From March 2002 to November 2006, 98 patients were enrolled onto protocol 01-142 (NCT00039195) and from July 2008 to May 2013, 99 patients were enrolled onto 08-026 (NCT00712582). All patients provided written informed consent. From November 2006 through September 2010, 26 patients were treated at MSK with a non-cross-reactive chemotherapeutic program consistent with that of 01-142 but performed off-protocol since 01-142 was closed at the time.

Patients were treated with R-CHOP x4 or RR-CHOP x3 + CHOP x1 induction, and either three cycles of ifosfamide, carboplatin, and etoposide (ICE), ICE x2 + rituximab-ICE (R-ICE) x1, or augmented R-ICE x2 consolidation chemotherapy. Those with both an interim FDG-PET-positive result and confirmatory positive biopsy of the FDG-positive site went on to receive high-dose therapy and autologous stem cell rescue.

The 223 patients had similar pre-treatment characteristics and similar outcome after a median follow-up of 7.7 years (95% CI: 7.0-8.7), which justified combining the three cohorts. From the total cohort of 223 patients, 166 patients with baseline, interim, and/or EOT PET/CT scans available in MSK's Picture Archiving and Communication System were included in this analysis. A consort diagram of evaluable patients is shown in **Supplemental Figure 1**. No clinical (**Supplemental Table 1**) or follow-up (**Supplemental Figure 2**) differences were observed between the 166 patients in the PET/CT cohort analyzed in this paper and the 57 patients who were excluded. Only a sex difference was observed (**Supplemental Table 1**).

FDG PET/CT Imaging and Analysis

FDG PET/CT scans were obtained at baseline, after four cycles of R-CHOP (interim), and six weeks after completing immunochemotherapy (EOT). Patients fasted for six hours before injection of 444±44 MBq of ^{18}F-FDG. PET/CT scans from mid-skull to upper thighs were performed on Discovery scanners (GE Healthcare) after a standardized uptake time of approximately 60 minutes.

Baseline, interim, and EOT PET/CT scans were interpreted by an experienced nuclear medicine physician (LM) blinded to patient outcome. Mediastinal blood pool and normal liver were used as reference regions for background activity. Sites of abnormal FDG uptake, defined as intensity greater than surrounding local background, were recorded. The intensity of FDG uptake was measured using the maximum standardized uptake value (SUVmax), defined as the highest standardized uptake value (SUV) recorded among all lesions for each scan. Focal bone uptake corresponded to bone metastasis. Diffuse marrow uptake was defined visually and may represent lymphoma involvement or reactive hyperplasia. SUVmax of diffuse uptake was not recorded.

All measurable lesions were identified at baseline. Volumetric regions of interest were placed over all sites of abnormal uptake in lymph nodes, soft tissue organs, or focal bone lesions. Total MTV was obtained by summing the metabolic volumes of all measurable lesions and applying a 41% SUVmax threshold. The semi-automatic software Beth Israel plugin for FIJI was used (*14*) to record focal bone involvement and diffuse marrow uptake. The IMPI score, which represents the probability of being progression-free at 36 months, was calculated for each patient on the basis of age, stage, and baseline MTV as described by Mikhaeel et al. (*13*).

The visual Deauville/Lugano five-point scale was applied to the interim and EOT scans, with scores of 1-3 (indicating uptake $\leq$ that of the liver) considered negative and scores 4-5 (indicating uptake > the liver) considered positive. To measure metabolic change after induction therapy, ΔSUVmax was assessed using the most intense tumor in any region or organ at the interim scan — even if the location differed from the original tumor at baseline — calculated as follows: ΔSUVmax = (baseline SUVmax – interim SUVmax) / baseline SUVmax (*15*). Patients with ΔSUVmax $\leq$ 70% were considered positive and patients with ΔSUVmax > 70% were considered negative. The 70% threshold was chosen for this series based on the previously identified optimal cutoff to predict progression or death for ΔSUVmax after four cycles in the LNH2007-3B trial (16). As outlined by Meignan et al. based on the PETAL trial (NCT00554164), LNH2007-3B (NCT00498043), and International validation studies (17), patients with low baseline SUVmax (<10) and/or high interim SUVmax (>5) were deemed unsuitable for ΔSUVmax calculations. Visual assessment was used for these patients.

Statistical Analysis

Progression-free survival (PFS) and overall survival (OS) were used to evaluate the prognostic value of clinical and PET/CT parameters. PFS was defined as the time from the start of treatment to the date of disease progression/relapse or death from any cause. Patients without progression/relapse or death were censored at their last follow-up. OS was defined as the time from the start of treatment to the date of death from any cause. Surviving patients were censored at their last follow-up. To assess the prognostic value of parameters measured at interim or EOT, landmark analyses were used where PFS and OS were defined from the interim or EOT, respectively. Patients with the events of interest before the landmark time or without the corresponding PET/CT scans were excluded.

IPI, baseline PET/CT parameters (SUVmax, MTV, focal bone uptake, diffuse marrow uptake), interim PET/CT parameters (ΔSUVmax [positive vs. negative or continuous], Deauville scores [1-3 vs. 4-5]), and EOT PET/CT parameters (Deauville scores [1-3 vs. 4-5]) were evaluated as prognostic factors. We used 510 mL as the optimal cutoff for MTV as proposed by Meignan *et al*. (18), which we validated for PFS and OS in our cohort (**Supplemental Figure 3**). PFS and OS rates were estimated using a Kaplan-Meier estimator. The impact of candidate factors on survival were assessed using univariable and multivariable Cox proportional hazard models. The median follow-up was estimated using the reverse Kaplan-Meier method. The comparison between the patients included and excluded from the cohorts were done using the Wilcoxon rank sum test for continuous variables and the Fisher's exact test for categorical variables. A two-sided *P*-value < 0.05 was considered statistically significant.

To assess the predictive ability of IMPI (probability of being progression-free at 36 months), its complement, cIMPI (probability of a progression event by 36 months), was analyzed using three methods: measures of discrimination (Harrell's c-index), prediction error (Brier score), and calibration (calibration plot). Analyses were performed using R (version 4.1.0).

RESULTS

The median follow-up for the 166 patients included in this analysis was 7.9 years (95% CI: 6.7-8.8). Clinical characteristics and quantitative PET parameters are summarized in **Table 1**. Of the total, 48 patients experienced a progression event and 31 died (2 of these deaths were unrelated to cancer). The five-year PFS and OS rates were 76% and 85%, respectively. The ten-year rates were 69% and 80%, respectively.

All 166 patients underwent baseline PET/CT scans. The median SUVmax was 24.35 (range: 6.30-60.36). Median MTV was 297.82 mL (range: 6.45-5145.85) and average MTV was 522.32 mL. Fifty-five patients had FDG-positive focal bone lesions, and 20 patients had diffuse marrow uptake; among these, 5 patients had mixed focal bone lesions and diffuse uptake. Of the total, 157 patients underwent interim PET/CT after R-CHOP. For the remaining 9 patients, interim PET/CT was either not performed or not available (**Supplemental Figure 5**). One patient progressed before interim scanning and was excluded in PFS landmark analysis. By visual Deauville/Lugano classification, there were 39 interim PET/CT-positive patients (25%) and 118 interim PET/CT-negative patients (75%). The median ΔSUVmax was 0.90% (-0.33%-0.98%). Using ΔSUVmax criteria, 17 patients were classified as positive (11%), and 140 patients were classified as negative (89%) at interim. Among them, 23 had initial SUVmax <10 (6 patients) or interim SUVmax >5

(17 patients); Deauville scores were used to classify them as positive or negative. All but 15 patients, for whom imaging was not performed or not available, were analyzed for EOT PET/CT (**Supplemental Figure 5**). Three patients progressed before or on the day of EOT scan and were excluded in PFS landmark analysis. Visual Deauville/Lugano assessment was positive for 27 patients (17 of 27 also had a positive interim PET/CT result per Deauville/Lugano response criteria) and 124 were considered negative at EOT.

IPI, baseline MTV, and interim ΔSUVmax were statistically significant predictors of OS (**Table 2; Figure 1**). IPI (p=0.059) and baseline MTV (p=0.066) were independent prognostic factors of OS in a multivariable model with borderline significance. Baseline MTV, interim ΔSUVmax, and EOT Deauville score were statistically significant predictors of PFS (**Table 2; Figure 2**). Casasnovas et al. showed that combining visual (International Harmonization Project criteria) and quantitative (ΔSUVmax) PET assessments after four cycles of induction treatment identifies patients at extremely high risk of induction failure or early relapse (*16*). We performed a similar analysis looking at the prognostic relevance of interim PET parameters (Deauville score and ΔSUVmax) to outcome by combining these two interim response criteria. This Kaplan-Meier analysis demonstrated that patients with Deauville of 4-5 and positive ΔSUVmax at restaging (approximately 10% of the cohort) had extremely poor prognosis (**Figure 3**). Among these, nine patients also had high initial MTV.

The IMPI was calculated for all patients as a probability of being progression-free at 36 months. The predicted event rate was compared to the actual event rate (**Supplemental Figure 4**), and we found that the IMPI overestimated the event rate.

DISCUSSION

Early prediction of poor prognosis during the course of DLBCL therapy would be helpful for improving long-term outcome. While assessing early response to treatment using PET/CT scans has identified potential prognostic factors, there is currently no consensus on how to adapt treatment strategies based on molecular imaging parameters. For example, studies with large DLBCL cohorts have identified baseline MTV as a significant predictor for PFS and OS (*9,12*). Other studies showed ΔSUVmax on interim PET to be associated with both PFS and OS (*3,10*). Data reported by Casasnovas et al. also suggest that interim ΔSUVmax is more discriminant of outcome after four cycles of treatment than after two cycles (*3*). However, another large prospective trial reported interim PET/CT having limited prognostic relevance. Mamot et al. demonstrated that when interim PET/CT after two cycles was already positive, PET scans after four cycles of chemotherapy provided no additional predictive value compared to two cycles, and that only scans at EOT identified a significant difference in outcome (*7*).

To explore the prognostic value of PET/CT in DLBCL, we looked at the prognostic value of several PET/CT parameters in a group of 166 patients uniformly treated with a risk-adapted immunochemotherapy regimen. Our results showed that baseline MTV and interim ΔSUVmax were significantly predictive of PFS and OS. We also found that EOT Deauville score was prognostic for PFS. To note, EOT PET demonstrated less prognostic value in our study compared to what was reported by Mamot et al. *(7)*. This difference may be due to the risk-adapted treatment regimen as well as the longer follow-up in our series.

The recently proposed IMPI (*13*), which combines baseline MTV and age as continuous variables to predict patient outcome in DLBCL, is potentially useful for identifying patients with worse prognosis who might benefit from more aggressive or investigational treatment. We sought to validate this model in our cohort. In our series, the IMPI predictions overestimated the event rate. There are a number of potential explanations for the lower predictive accuracy in our population. Our patients were treated with R-CHOP followed by ICE/RICE, whereas Mikhaeel et al. used clinical data from patients treated with R-CHOP alone. Second, baseline MTV was calculated using different software. Finally, MTV was measured by including tumor with different SUV cutoffs (current analysis used the 41% SUVmax threshold method, whereas Mikhaeel et al. used SUVmax $\geq$ 4.0). Nevertheless, the median MTV in the current study was similar to theirs (298 mL vs. 308 mL).

In our series, combining interim PET parameters Deauville score and ΔSUVmax demonstrated that patients with Deauville scores of 4-5 and positive ΔSUVmax (10% of the cohort) had extremely poor prognosis. These results combining visual and quantitative assessments are similar to those previously reported in an independent cohort after four cycles of induction treatment (*16*). Thus, it appears that adding ΔSUVmax to visual analysis may be a robust and reproducible tool for identifying high-risk patients with DLBCL. Combining the two interim PET parameters identifies patients who have a poor outcome with standard chemoimmunotherapy and may help define a cohort of patients for evaluation of alternative therapeutic approaches, such as CAR T-cell therapy. ZUMA-12 attempted to identify patients with a poor prognosis for early intervention with axicabtagene ciloleucel (21); however, that trial has been criticized for the means of selecting the poor risk cohort. The interim PET evaluation described herein could potentially identify a more

uniform group of patients with a poor outcome. A prospective trial could randomize these high-risk patients to CAR T-cell vs second-line therapy followed by high dose therapy and autologous stem cell rescue, similar to the ZUMA-7 *(22)* and TRANSFORM *(23)* clinical trials. Other studies evaluating the role of PET/CT metrics for treatment guidance in DLBCL have reported other parameter combinations to be relevant. Cottereau et al. demonstrated that baseline MTV and Standardized Dmax (the largest distance between two lesions) complement each other in characterizing tumor burden and disease spread (*11*), whereas Vercellino et al. combined baseline MTV with the Eastern Cooperative Oncology Group performance status to identify a very high-risk DLBCL subgroup (*12*). Recently, Eertink et al., on behalf of PETRA investigators, demonstrated in 217 patients that MTV, $Dmax_{bulk}$, SUVpeak, WHO, and age identify patients at risk of relapse at baseline *(24)*.

To determine the optimal combination of PET/CT parameters and prognostic indices to improve the prediction of outcome in clinical practice, standardized methods of measurement are needed across all PET/CT centers internationally. Some examples include whether interim PET/CT scans should be done after two vs. four cycles, standardized definitions of ΔSUVmax, and methods for determination of MTV *(25)*. Once a robust set of parameters or score is determined, multiple large studies would need to validate the results for a consensus to be reached. Standardization is potentially complicated by different initial regimens. For the results to be applicable across studies, the parameters would ideally be independent of treatment. To move from being a prognostic tool to a predictive tool, well designed clinical trials need to evaluate new treatment strategies for the high-risk DLBCL patient and show improved outcome.

CONCLUSION

Our study confirmed the prognostic value of baseline MTV and interim ΔSUVmax in DLBCL. Combining interim Deauville score with interim ΔSUVmax could improve risk stratification for patients with extremely poor prognosis. These results warrant large multicenter studies to develop standardized practices and refine existing prognostic indices in DLBCL.

Disclosures: HS, NS, AM, CM, KB, and LM have no disclosures to report. AZ serves or has served as a consultant to Genentech/Roche, Gilead/Kite, BMS/Celgene/Juno, Janssen, Novartis, Adaptive Biotechnology, MorphoSys, Abbvie, AstraZeneca, MEI Pharma, and BeiGene; has collaborated on research with MEI Pharmaceuticals, Genentech/Roche, and Beigene; and has served as a DMC member for BMS/Celgene/Juno. No other potential conflicts of interest relevant to this article exist.

Acknowledgments: This work was supported in part by the NIH/NCI Cancer Center Support Grant P30 CA008748.

KEY POINTS

QUESTION: Do baseline MTV, alone or in combination with ΔSUV, and the recently proposed IMPI score predict outcome in patients with diffuse large cell lymphoma treated with the RCHOP-ICE drug regimen?

PERTINENT FINDINGS: Baseline MTV and ΔSUVmax at interim predict overall survival. Patients with Deauville scores of 4-5 and positive ΔSUVmax ≤ 70% at interim (approximately 10% of the cohort) had extremely poor prognosis. The new IMPI score had limited discrimination and slightly overestimated the event rate in our cohort.

IMPLICATIONS FOR PATIENT CARE: Combining interim Deauville scores with interim ΔSUVmax could improve risk stratification for DLBCL patients with extremely poor prognosis.

REFERENCES

1. Duhrsen U, Muller S, Hertenstein B, et al. Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL): A Multicenter, Randomized Phase III Trial. *J Clin Oncol.* 2018;36:2024-2034.

2. International Non-Hodgkin's Lymphoma Prognostic Factors P. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993;329:987-994.

3. Casasnovas RO, Ysebaert L, Thieblemont C, et al. FDG-PET-driven consolidation strategy in diffuse large B-cell lymphoma: final results of a randomized phase 2 study. *Blood.* 2017;130:1315-1326.

4. Moskowitz CH, Schoder 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.* 2010;28:1896-1903.

5. Juweid ME, Stroobants S, Hoekstra OS, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *J Clin Oncol.* 2007;25:571-578.

6. 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.

7. Mamot C, Klingbiel D, Hitz F, et al. Final Results of a Prospective Evaluation of the Predictive Value of Interim Positron Emission Tomography in Patients With Diffuse Large

B-Cell Lymphoma Treated With R-CHOP-14 (SAKK 38/07). *J Clin Oncol.* 2015;33:2523-2529.

8. Zhao P, Yu T, Pan Z. Prognostic value of the baseline 18F-FDG PET/CT metabolic tumour volume (MTV) and further stratification in low-intermediate (L-I) and high-intermediate (H-I) risk NCCNIPI subgroup by MTV in DLBCL MTV predict prognosis in DLBCL. *Ann Nucl Med.* 2021;35:24-30.

9. 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.

10. Schoder H, Polley MC, Knopp MV, et al. Prognostic value of interim FDG-PET in diffuse large cell lymphoma: results from the CALGB 50303 Clinical Trial. *Blood.* 2020;135:2224-2234.

11. 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(dagger). *Ann Oncol.* 2021;32:404-411.

12. Vercellino L, Cottereau AS, Casasnovas O, et al. High total metabolic tumor volume at baseline predicts survival independent of response to therapy. *Blood.* 2020;135:1396-1405.

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:JCO2102063.

14. 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.

15. Itti E, Lin C, Dupuis J, et al. Prognostic value of interim 18F-FDG PET in patients with diffuse large B-Cell lymphoma: SUV-based assessment at 4 cycles of chemotherapy. *J Nucl Med.* 2009;50:527-533.

16. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al. SUVmax reduction improves early prognosis value of interim positron emission tomography scans in diffuse large B-cell lymphoma. *Blood.* 2011;118:37-43.

17. Meignan M, Gallamini A, Itti E, Barrington S, Haioun C, Polliack A. Report on the Third International Workshop on Interim Positron Emission Tomography in Lymphoma held in Menton, France, 26-27 September 2011 and Menton 2011 consensus. *Leuk Lymphoma.* 2012;53:1876-1881.

18. 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;34:3618-3626.

19. Burggraaff CN, Eertink JJ, Lugtenburg PJ, Hoekstra OS, Arens AIJ, de Keizer B, Heymans MW, van der Holt B, Wiegers SE, Pieplenbosch S, Boellaard R, de Vet HCW, Zijlstra JM; HOVON Imaging Working Group and the HOVON Lymphoma Working Group. ^{18}F-FDG PET Improves Baseline Clinical Predictors of Response in Diffuse Large B-Cell Lymphoma: The HOVON-84 Study. J Nucl Med. 2022 ;63:1001-1007.

20. Kostakoglu L, Mattiello F, Martelli M, Sehn LH, Belada D, Ghiggi C, Chua N,González-Barca E, Hong X, Pinto A, Shi Y, Tatsumi Y, Bolen C, Knapp A, Sellam G, Nielsen T, Sahin D, Vitolo U, Trněný M. Total metabolic tumor volume as a survival predictor for patients with diffuse large B-cell lymphoma in the GOYA study. Haematologica. 2022;107:1633-1642.

21. Neelapu SS, Dickinson M, Munoz J, et al. Axicabtagene ciloleucel as first-line therapy in high-risk large B-cell lymphoma: the phase 2 ZUMA-12 trial. *Nat Med.* 2022;28:735-742.

22. Elsawy M, Chavez JC, Avivi I, et al. Patient-reported outcomes in ZUMA-7, a phase 3 study of axicabtagene ciloleucel in second-line large B-cell lymphoma. *Blood.* 2022.

23. Kamdar M, Solomon SR, Arnason J, et al. Lisocabtagene maraleucel versus standard of care with salvage chemotherapy followed by autologous stem cell transplantation as second-line treatment in patients with relapsed or refractory large B-cell lymphoma (TRANSFORM): results from an interim analysis of an open-label, randomised, phase 3 trial. *Lancet.* 2022;399:2294-2308.

24. Eertink JJ, van de Brug T, Wiegers SE, Zwezerijnen GJC, Pfaehler EAG,Lugtenburg PJ, van der Holt B, de Vet HCW, Hoekstra OS, Boellaard R, Zijlstra JM. ^{18}F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2022;49:932-942.

25. 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.

TABLES AND FIGURES

Table 1. Patient characteristics (n = 166)

Clinical Characteristics	**n = 166**
Median age (years)	50 [range: 20 – 71]
Ann Arbor Stage	
II	34 (20%)
III-IV	132 (80%)
Median LDH	332 [range: 130 – 1,925]
KPS	
≤ 70	49 (30%)
> 70	117 (70%)
Standard IPI Score	
0	33 (20%)
1	39 (23%)
2	53 (32%)
3	41 (25%)
Baseline PET	**n = 166**
Focal Bone Uptake	55 (33%)
Diffuse Marrow Uptake	20 (12%)
Median Liver SUVmax	2.42 [range: 0.81 – 7.20]
Unknown	2
Median SUVmax	24.35 [range: 6.30 – 60.36]
Median TMTV	297.82 [range: 6 – 5,145.85]
≤ 510 mL	117 (70%)
> 510 mL	49 (30%)
Interim PET	**n = 157**
ΔSUVmax	
Median	0.90 [range: -0.33 – 0.98]
Negative	140 (89%)
Positive	17 (11%)
Deauville score	
1-3	118 (75%)
4	36 (23%)
5	3 (2%)
EOT PET	**n = 151**
Deauville score	
1-3	124 (82%)
4	19 (13%)
5	8 (5%)

LDH = lactate dehydrogenase; KPS = Karnofsky Performance Scale; IPI = International Prognostic Index; PET = positron emission tomography; TMTV = total metabolic tumor volume; SUV = standardized uptake value; SUVmax = maximum SUV; ΔSUVmax = change in SUVmax; EOT = end of treatment

Table 2. Univariable Cox regression analyses

	Overall Survival			Progression-Free Survival		
	HR	95% CI	p-value	HR	95% CI	p-value
Clinical Characteristics						
Standard IPI score			0.015			0.13
0	—	—		—	—	
1	0.91	0.23, 3.65		0.71	0.26, 1.97	
2	1.12	0.33, 3.82		1.00	0.42, 2.39	
3	3.35	1.12, 10.0		1.81	0.79, 4.14	
Baseline PET						
Focal bone uptake	0.88	0.42, 1.88	0.75	0.73	0.39, 1.36	0.31
Diffuse marrow uptake	1.52	0.58, 3.98	0.41	1.12	0.47, 2.64	0.81
SUVmax (per 5 units)	1.07	0.88, 1.30	0.53	0.92	0.79, 1.09	0.34
TMTV (dichotomized)			0.011			0.004
≤ 510 mL	—	—		—	—	
> 510 mL	2.54	1.25, 5.13		2.33	1.32, 4.12	
Interim PET (landmark)						
ΔSUVmax (continuous)	0.03	0.01, 0.14	<0.001	0.08	0.02, 0.32	0.007
ΔSUVmax (dichotomized)			0.007			0.015
Negative	—	—		—	—	
Positive	3.75	1.60, 8.80		2.91	1.35, 6.29	
Deauville score			0.15			0.21
1-3	—	—		—	—	
4-5	1.79	0.83, 3.84		1.54	0.80, 2.95	
EOT PET (landmark)						
Deauville score			0.092			0.010
1-3	—	—		—	—	
4-5	2.24	0.93, 5.41		2.72	1.34, 5.51	

HR = Hazard ratio; CI = Confidence interval; IPI = International Prognostic Index; PET = positron emission tomography; SUV = standardized uptake value; SUVmax = maximum SUV; TMTV = total metabolic tumor volume; ΔSUVmax = change in SUVmax; EOT = end of treatment

Figure 1. Overall survival stratified by baseline TMTV (A), interim Deauville score (B), interim ΔSUVmax (C), and end-of-treatment (EOT) Deauville score (D)

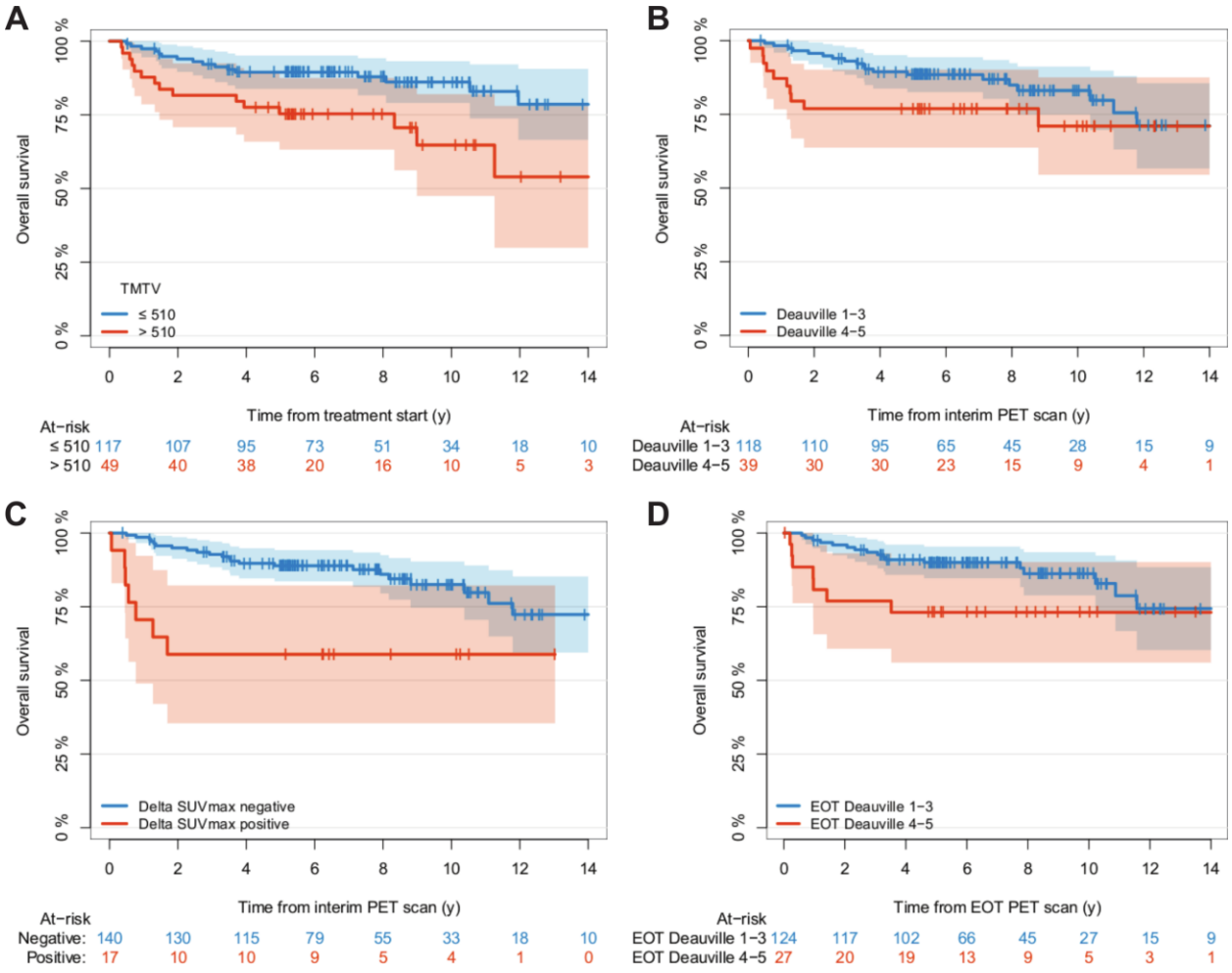
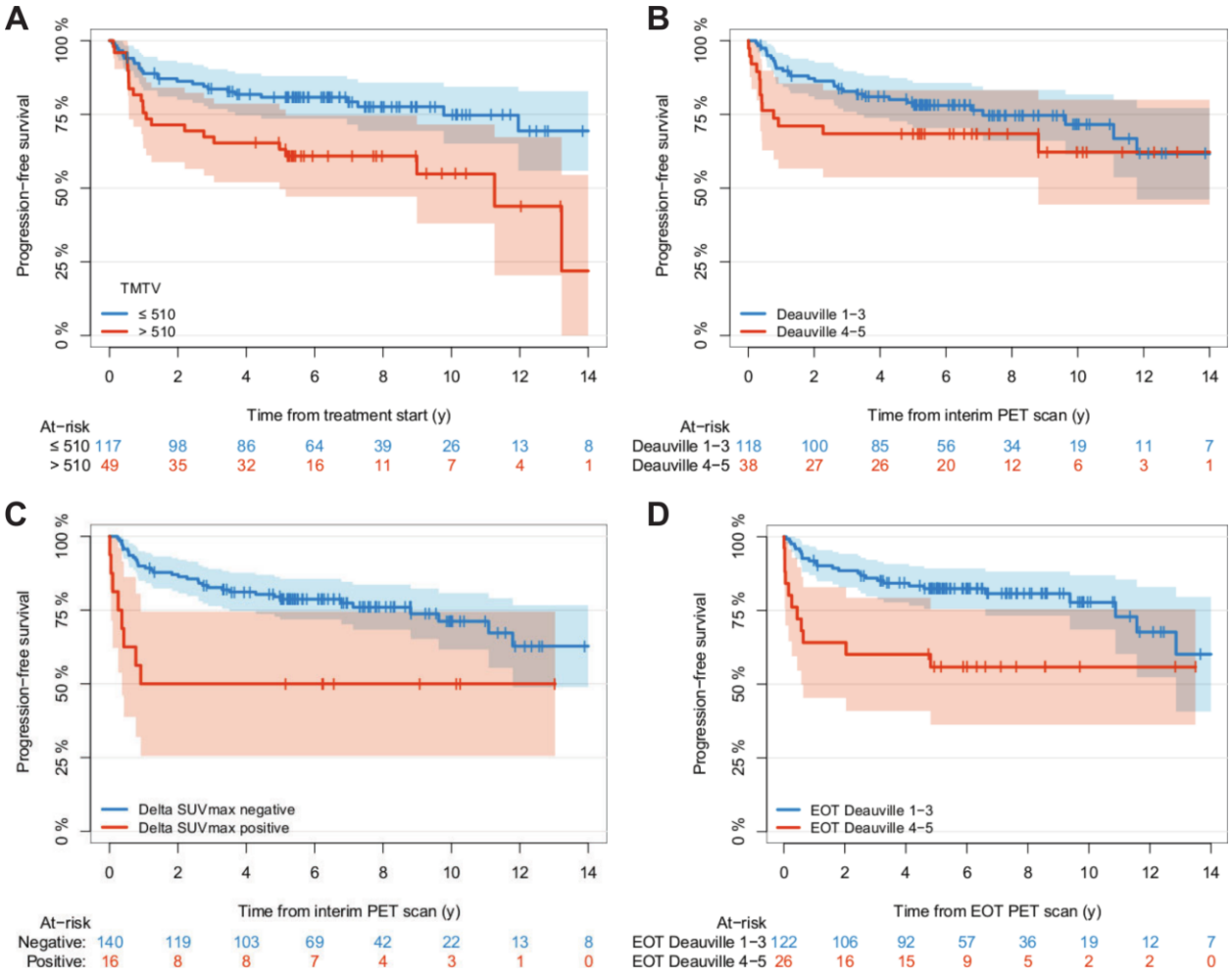
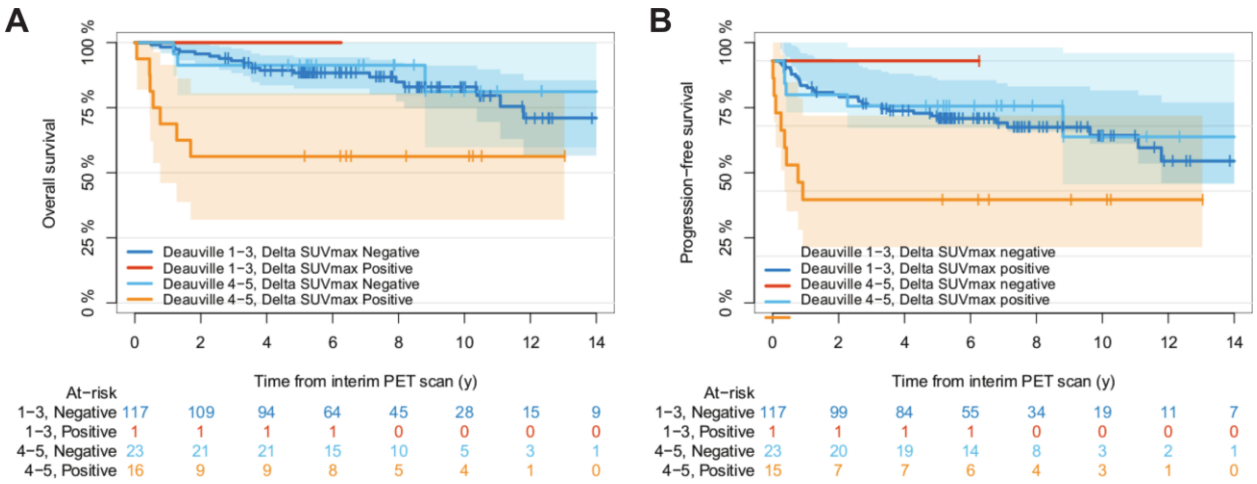


Figure 2. Progression-free survival stratified by baseline TMTV (A), interim Deauville score (B)*, interim ΔSUVmax (C)*, and end-of-treatment (EOT) Deauville score (D)**



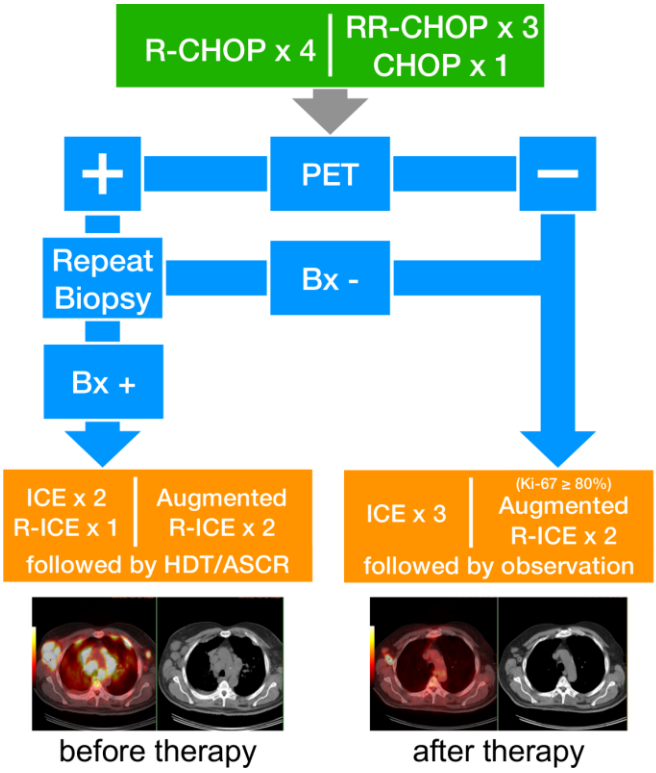
*1 patient progressed before interim scan and excluded from landmark analysis
**3 patients progressed before or on the day of EOT scan and excluded from landmark analysis

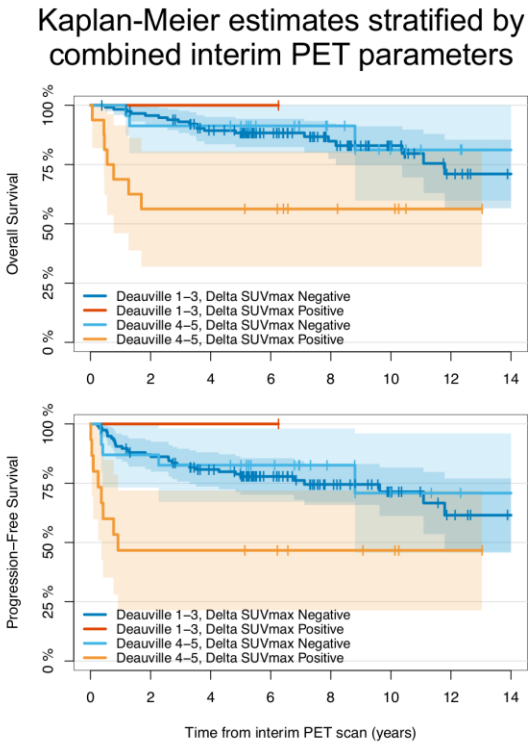
Figure 3. Overall survival (A) and progression-free survival (B)* stratified by a combination of interim Deauville score and interim ΔSUVmax



*1 patient progressed before interim scan and excluded from landmark analysis

Graphical Abstract

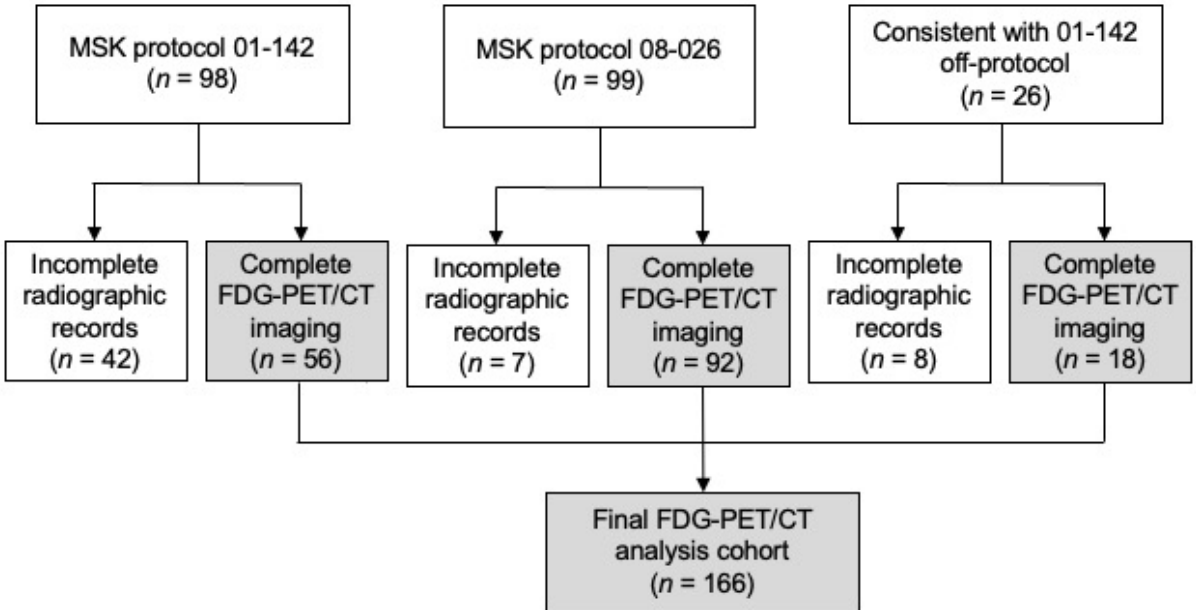




Supplemental Table 1. Clinical and biological differences between FDG PET cohort and entire cohort. Continuous variables were compared using the Wilcoxon rank-sum test and categorical variables were compared using the Fisher's exact test. Only a difference in sex was observed.

Characteristics	PET Cohort (n=166)	Excluded Patients (n=57)	p-value
Age at diagnosis (years)			
Median (range)	50 (20 – 71)	46 (18 – 65)	0.34
Sex			
Male	79 of 166 (48%)	36 of 57 (63%)	0.05
Female	87 of 166 (52%)	21 of 57 (3748%)	
LDH			
> ULN	134 of 166 (81%)	50 of 57 (88%)	0.31
≤ ULN	32 of 166 (19%)	7 of 57 (12%)	
KPS			
≤ 70%	49 of 166 (30%)	18 of 57 (32%)	0.88
> 70%	117 of 166 (70%)	39 of 57 (68%)	
Ann Arbor Stage			
IV	108 of 166 (65%)	36 of 57 (63%)	0.45
III	24 of 166 (14%)	12 of 57 (21%)	
II	34 of 166 (21%)	9 of 57 (16%)	
Extranodal sites > 1			
> 1	81 of 166 (49%)	26 of 57 (46%)	0.76
≤ 1	85 of 166 (51%)	31 of 57 (54%)	
Bone marrow involvement			
Yes	34 of 166 (23%)	8 of 57 (16%)	0.42
No	114 of 166 (77%)	41 of 57 (84%)	
Standard IPI			
High	41 of 166 (25%)	9 of 57 (16%)	0.26
High-intermediate	53 of 166 (32%)	25 of 57 (44%)	
Low-intermediate	39 of 166 (23%)	15 of 57 (26%)	
Low	33 of 166 (20%)	8 of 57 (14%)	
Age-adjusted IPI			
High	35 of 166 (21%)	13 of 57 (23%)	0.17
High-intermediate	81 of 166 (49%)	34 of 57 (60%)	
Low-intermediate	50 of 166 (30%)	10 of 57 (18%)	

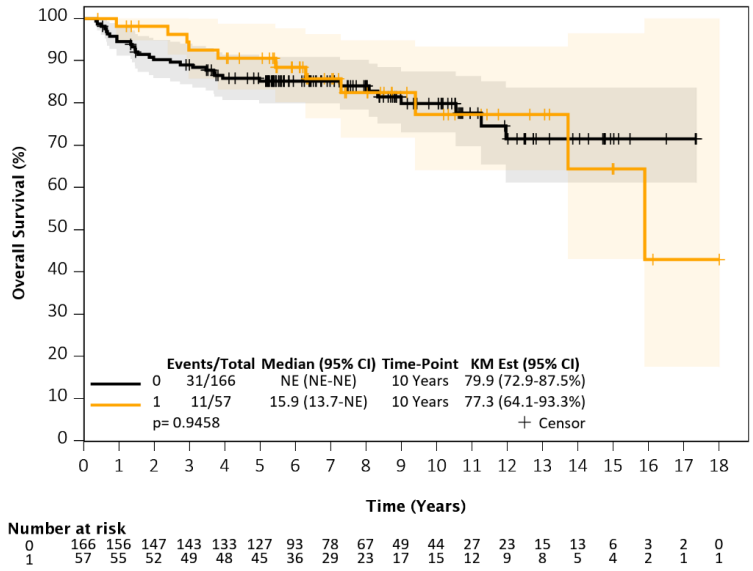
PET=positron emission tomography. LDH=lactate dehydrogenase. ULN=upper limit of normal. KPS=Karnofsky performance status. aaIPI=age-adjusted International Prognostic Index.



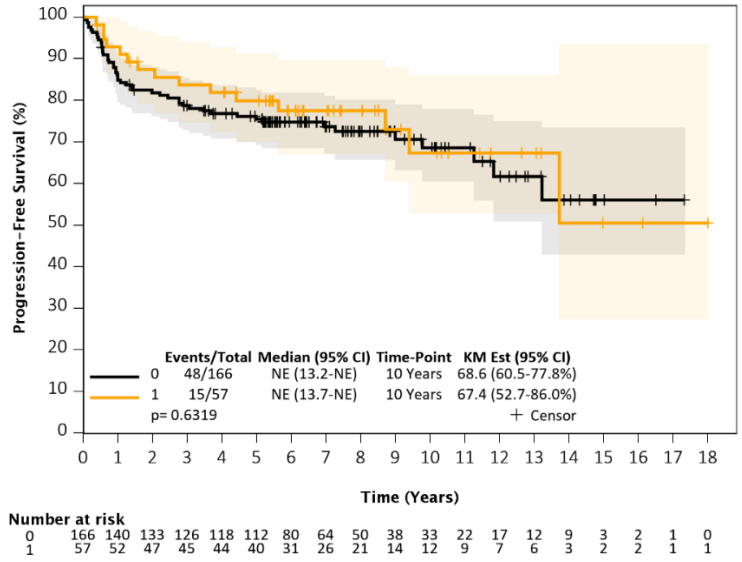
Supplemental Figure 1. Consort diagram of evaluable patients at MSK treated on IRB-approved protocols 01-142 and 08-026, and patients receiving treatment consistent with 01-142 off-protocol.

MSK = Memorial Sloan Kettering; IRB = Institutional Review Board. FDG-PET/CT = fluorodeoxyglucose positron emission tomography and computed tomography.

A.

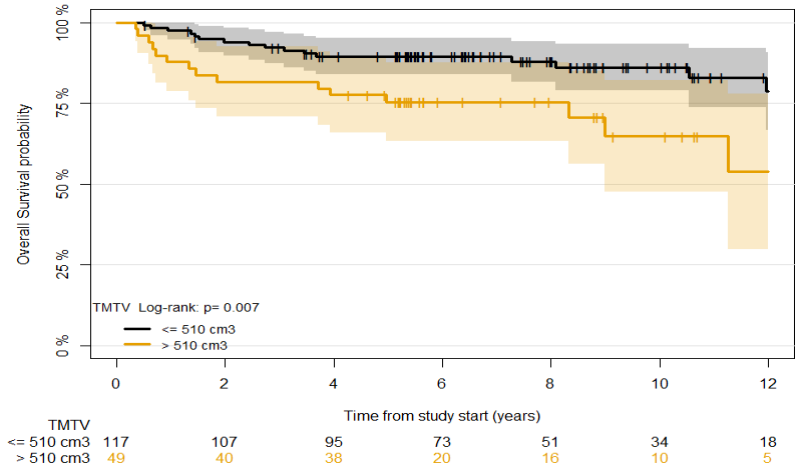


B.



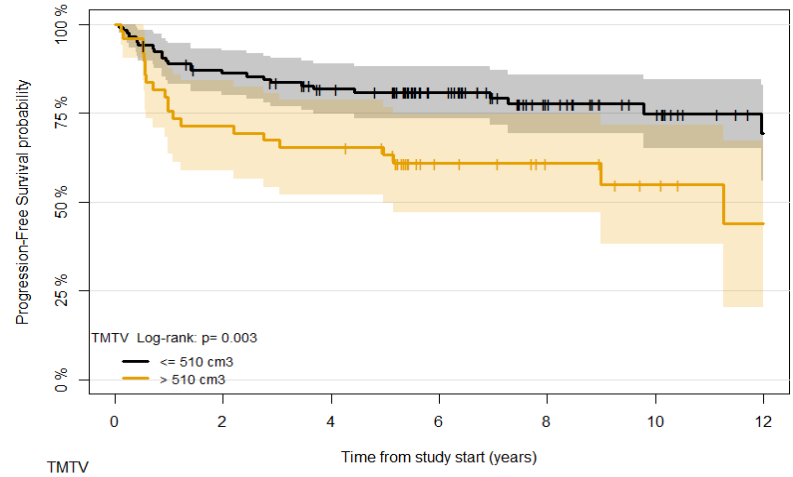
Supplemental Figure 2. Overall survival (A) and progression-free survival (B) comparison between positron emission tomography (PET) cohort (denoted by "0"; n=166) and excluded patients (denoted by "1"; n=57). PET cohort median follow-up: 7.9 years (95% CI: 6.7-8.8). Excluded patient's median follow-up: 7.4 years (95% CI: 6.3-10.2). There were no follow-up differences between the PET cohort and excluded patients.

A.



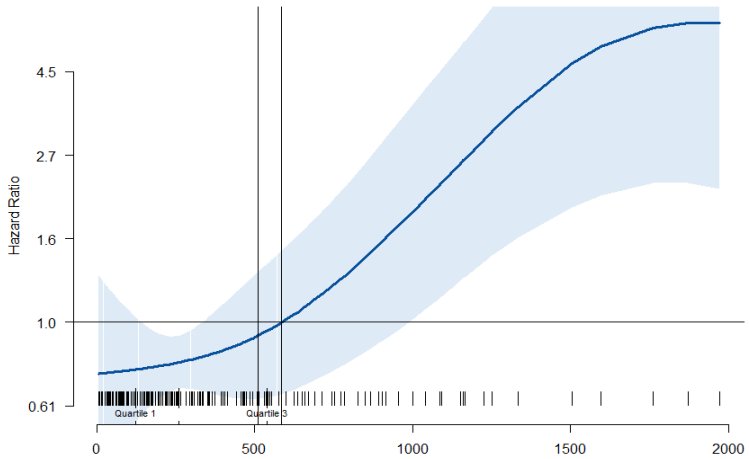
TMTV							
<= 510 cm3	117	107	95	73	51	34	18
> 510 cm3	49	40	38	20	16	10	5

B.

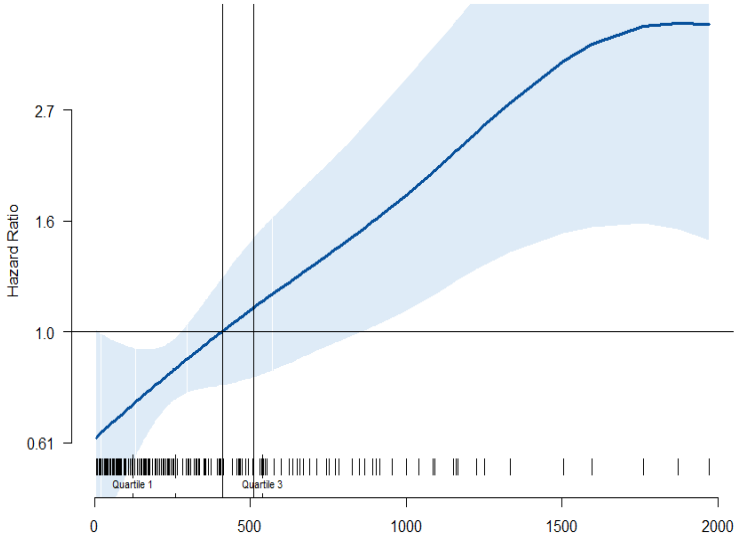


TMTV							
<= 510 cm3	117	98	86	64	39	26	13
> 510 cm3	49	35	32	16	11	7	4

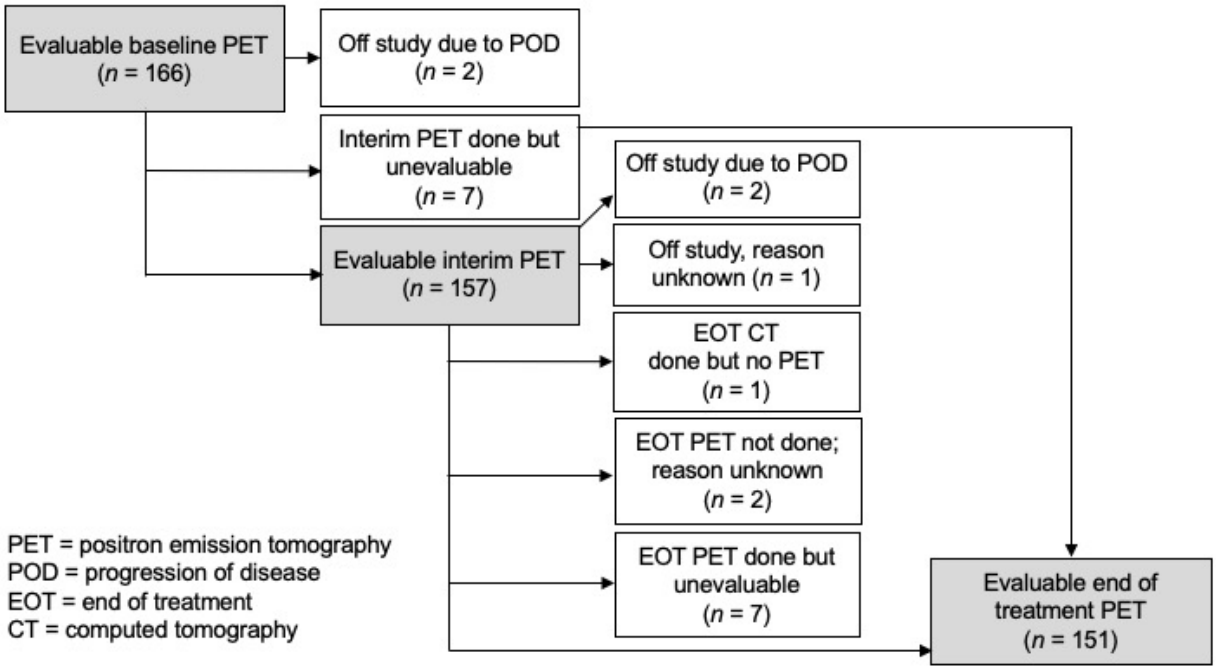
C.



D.

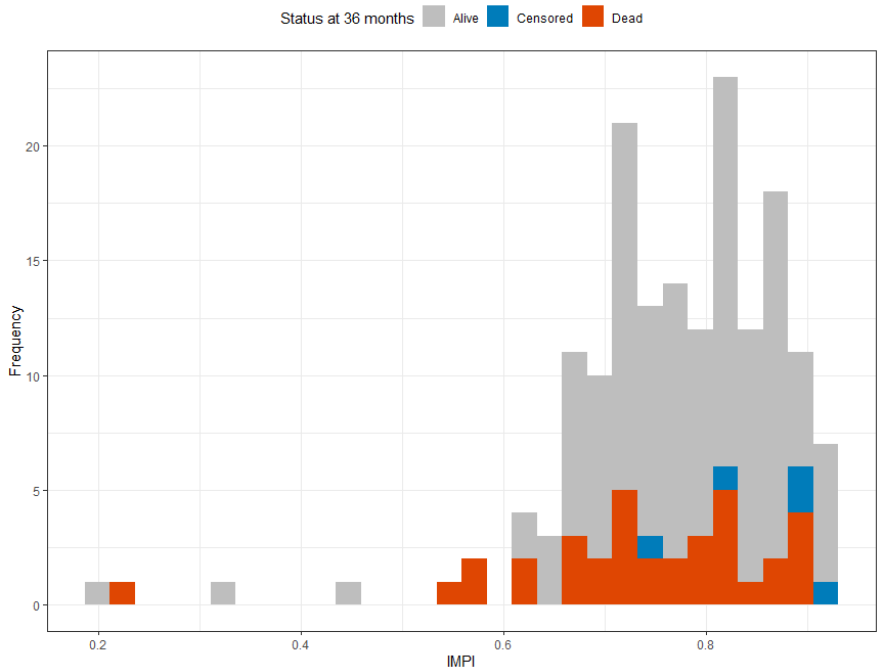


Supplemental Figure 3. An optimal cut-off of 510 mL for total metabolic tumor volume (TMTV) was described by Meignan et al. in a follicular lymphoma population (Meignan et al. JCO 2016). We wanted to validate this threshold for progression-free survival (PFS) and overall survival (OS) in our cohort of patients with diffuse large B-cell lymphoma. At baseline, a total of 117 (70%) patients had TMTV ≤ 510 mL and 49 (30%) had TMTV > 510 mL. Both OS (A) and PFS (B) were significantly different when stratified by this TMTV threshold. Survival curves were compared using a log rank test. To explore a new threshold, we estimated the relationship between TMTV and OS through splines to allow for non-linear relations. The graph represents the Hazard Ratio (HR) according to the TMTV (HR > 1 means an increased risk of death). HR of OS by TMTV using splines (C); the horizontal line represents an HR of 1, while the two vertical lines represent the previously proposed threshold (510) and the new threshold (585). HR of PFS by TMTV using splines (D); the horizontal line represents an HR of 1, while the two vertical lines represent the previously proposed threshold (510) and the new threshold (410). The new thresholds were not very different from the previously proposed one. Since the OS analysis suggested a higher threshold (585) and the PFS analysis suggested a lower one (410), the previously proposed threshold of 510 was suitable to be applied to both.

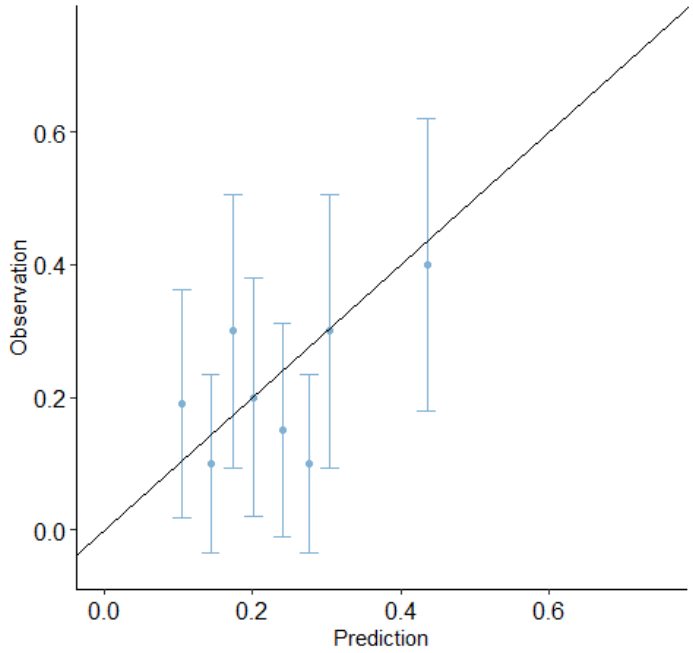


Supplemental Figure 4. Consort diagram showing patients that were removed from initial study population.

A.



B.



Supplemental Figure 5. The International Metabolic Prognostic Index (IMPI) was calculated for all patients as a probability of being progression-free at 36 months. The median IMPI was 0.79 (range 0.21 - 0.93). The distribution of IMPI was represented by people who died or were known to be alive at 36 months (A). To assess the predictive ability of IMPI, its complement, cIMPI (the

probability of a progression event by 36 months), was analyzed. The Harrell's c-index for the cIMPI prediction was 0.59 (standard deviation=0.05). This represents the chances that, when taking 2 random subjects, the one with a higher IMPI will have a longer PFS (concordance between the ranking of the predictions and observations). The mean square error (Brier score), calculated excluding 5 patients censored before 36 months, was 0.17. This represents the square distance between the predicted value and the actual status (0 for alive without event and 1 for event). A calibration plot to assess the IMPI's prediction of progression status at 36 months was performed (censored patients were excluded) (B). The population was separated into groups with similar predicted event rates (8 groups of about 20 patients), and the predicted event rate (x-axis) was plotted against the actual average event rate in those patients (y-axis). If the IMPI prediction was well calibrated, the value would be close to the 45-degree line (perfect agreement between prediction and observation). In all but one of the eight groups, the confidence intervals for the observation included the predicted value. The predictions slightly overestimated the event rate as most of the points were below the 45-degree line, but seems especially well calibrated for the high-risk patients (two groups with the highest predicted risk).