Journal of Nuclear Medicine, published on November 27, 2020 as doi:10.2967/jnumed.120.255034

Interim positron emission tomography in diffuse large B-cell lymphoma

Lars Kurch,[1] Andreas Hüttmann,[2] Thomas W. Georgi,[1] Jan Rekowski,[3] Osama Sabri,[1] Christine Schmitz,[2] Regine Kluge,[1] Ulrich Dührsen,[2*] and Dirk Hasenclever[4*]

*These authors contributed equally

[1]Klinik und Poliklinik für Nuklearmedizin, Universitätsklinikum Leipzig; [2]Klinik für Hämatologie, Universitätsklinikum Essen; [3]Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universität Duisburg-Essen; [4]Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig; all in Germany

Acknowledgment of research support:
This work is based on a study funded by Deutsche Krebshilfe (grant nos. 107592 and 110515), Amgen Germany, and Roche Pharma.

Contact information for the corresponding author:
Dr. Lars Kurch, MD
Klinik und Poliklinik für Nuklearmedizin
Universitätsklinikum Leipzig
Liebigstraße 18
04103 Leipzig, Germany
Telephone: 0049 341 9718008
Telefax: 0049 341 9718009
Email: lars.kurch@medizin.uni-leipzig.de

Short running title:

Interim PET in DLBCL

Word count:

Abstract	289
Text	2,783
Total	4,991

Figures	7
Tables	4
References	31
Data Supplement	1 supplemental Table, 1 supplemental Figure

ABSTRACT

Rationale

In diffuse large B-cell lymphoma, early assessment of treatment response by 18-fluorodeoxyglucose positron emission tomography (PET) may trigger treatment modification. Reliable identification of good and poor responders is important. We compared three competing methods of interim PET evaluation.

Methods

Images from 449 patients participating in the 'Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas' trial were re-analyzed by applying the visual Deauville score and the standardized uptake value (SUV)-based qPET and ΔSUV_{max} scales to interim PET scans performed after two cycles of chemotherapy. qPET relates residual lymphoma 18-fluorodeoxyglucose uptake to physiological liver uptake, converting the ordinal Deauville scale into a continuous scale and permitting a direct comparison with the continuous ΔSUV_{max} scale, which is based on SUV_{max} changes between baseline and interim scans. Positive and negative predictive values were calculated for progression-free survival.

Results

Using established thresholds to distinguish between good and poor responders (visual Deauville score 1-3 vs. 4-5; ΔSUV_{max} >66% vs. ΔSUV_{max} ≤66%), the positive predictive value was significantly lower with Deauville than ΔSUV_{max} (38.4% versus 56.6%; p=0.03). qPET and ΔSUV_{max} were strongly correlated on the log scale (Pearson's r=0.75). When plotted along corresponding percentiles, the positive predictive value curves for qPET and ΔSUV_{max} were superimposable, with low values up to the 85th percentile and a steep rise thereafter. The recommended threshold of 66% SUV_{max} reduction for the identification of poor responders was equivalent to qPET=2.26 corresponding to score 5 on the visual Deauville scale. The negative predictive value curves were also superimposable, but remained flat between 80% and 70%.

Conclusions

Continuous scales are better suited for interim PET-based outcome prediction than the ordinal Deauville scale. qPET and ΔSUV_{max} essentially carry the same information. The proportion of poor risk patients identified is less than 15%.

KEYWORDS

Diffuse large B-cell lymphoma, positron emission tomography, interim evaluation, Deauville scale, ΔSUV_{max}

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the most frequent cancer of the immune system (*1*). It is cured in about two thirds of patients (*2*). Treatment response is among the most important factors determining outcome. Remissions are more durable in rapid than slow responders, which was first demonstrated by computed tomography (CT) (*3*), and later by positron emission tomography (PET) using the tracer 18-fluorodeoxyglucose (*4,5*). To adapt treatment to treatment response, reliable identification of good and poor responders is of utmost importance.

Current guidelines recommend the Deauville scale for PET-based evaluation of early treatment response (*6*). It is based on a visual comparison of residual lymphoma-related uptake with areas of physiologically increased activity, such as mediastinal blood pool or liver (*7*). At the present time, any residual uptake exceeding that of the liver is considered a poor metabolic response.

A drawback of the Deauville scale is its ordinal nature with no more than five response categories. An alternative way of evaluating interim scans is a quantitative comparison of the maximum standardized uptake value (SUV) before and during treatment. The ratio between the two values (ΔSUV_{max}) results in a continuous scale, which can be dichotomized to distinguish between good and poor responders (*8*). In DLBCL, thresholds of 66% SUV_{max} reduction after two and 73% after four cycles of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) lead to maximum separation of the survival curves of good and poor responders (*8-11*).

The SUV_{max} method requires both baseline and interim PET scans. To restrict interim analysis to a single scan, while maintaining the advantage of a continuous scale, we developed the qPET method (q, quotient) where the mean SUV of the four most intense connected voxels of residual lymphoma-related uptake are put into relation with the mean SUV of a large volume in the liver. Pioneered in Hodgkin's lymphoma, this approach converted the ordinal Deauville scale into a well-defined quantitative scale (*12-14*).

The goal of the present study was to apply the qPET approach to DLBCL and compare it to the visual Deauville scale and the ΔSUV_{max} method. To this end, we re-analyzed the data of the 'Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas' (PETAL) trial which set out – and failed – to improve treatment by adapting it to the response to the first two cycles of R-CHOP (*15*). Because none of the PET-driven treatment changes had an impact on outcome compared to standard R-CHOP, all treatment arms were combined for this analysis.

PATIENTS AND METHODS

Study Design

The PETAL trial (ClinicalTrials.gov NCT00554164; EudraCT 2006-001641-33) was a multicenter study for newly diagnosed aggressive non-Hodgkin lymphomas (*15*). The study was approved by the Federal Institute for Drugs and Medical Devices and the ethics committees of all participating sites. All patients gave written informed consent including permission to use their data for post-hoc scientific analyses.

Patients were treated with bi-weekly R-CHOP, with a 3-week interval between cycles 2 and 3 to prevent false-positive results at interim staging uniformly performed after cycle 2. Patients with favorable interim PET response received four more cycles of R-CHOP or the same treatment plus two extra doses of rituximab. Patients with unfavorable response were randomly assigned to receive six additional cycles of R-CHOP or six blocks of a more intensive Burkitt's lymphoma protocol (*15*).

PET/CT Imaging and Evaluation

Imaging conditions have been described previously (*15*). Baseline and interim scans (median chemotherapy-free interval, 20 days) were evaluated by local investigators using the ΔSUV_{max} method. Scans were then pseudonymized and transferred to a central server for re-analysis including verification of the ΔSUV_{max} findings and evaluation according to the Deauville criteria by experienced nuclear medicine physicians (*15*).

In the present analysis, archived images were re-analyzed by a single physician (LK; >10 years of working experience, >5,000 evaluated PET scans from lymphoma patients) employing three different methods of interim scan evaluation. To ensure that visual Deauville scoring was not affected by quantitative measurements, assessment by purely visual criteria was performed before qPET and ΔSUV_{max}. The results were compared to the reports of the initially involved physicians (*15*), and any inconsistencies

were resolved by further image evaluation taking account of clinical data available at the time of first analysis. Discrepant results will be the subject of a future report.

The Deauville scale comprises five categories which are defined as score 1, no residual uptake; 2, residual uptake not exceeding mediastinal uptake; 3, residual uptake above mediastinal, but not exceeding liver uptake; 4, residual uptake above liver uptake; and 5, residual uptake markedly above liver uptake and/or new lesions (*7*). qPET was calculated by dividing the mean SUV of the four hottest connected voxels (SUV_{peak}) of the hottest residual lesion by the SUV_{mean} of a 30 ml volume of interest in the right lobe of the liver (*12*). ΔSUV_{max} was determined by dividing the SUV_{max} of the hottest residual lesion on the interim scan by the SUV_{max} of the hottest lesion on the baseline scan (*8*).

Statistical Analysis

The endpoint of the current analysis was progression-free survival defined as time from interim PET scanning to disease progression, relapse, or death from any cause, subsequently referred to as treatment failure. Progression and relapse were defined by clinical and imaging criteria and confirmed by biopsy in the majority of cases. For simplicity, we treated progression-free survival as a binary variable (events within 60 months). This appeared justified, because, in DLBCL, the majority of events occur within the first two years (*16*), and, with a median follow-up of 52 months, the data was mature (*15*).

We plotted empirical cumulative distribution functions of qPET by visual Deauville score and used Receiver Operating Characteristic (ROC) and Youden index to derive plausible thresholds between individual scores of the visual scale. These cut-off values were compared to the thresholds found in a study with nearly 900 pediatric Hodgkin's lymphoma patients (*12*).

With regard to ΔSUV_{max}, we used 1-ΔSUV_{max}, i.e. the remaining proportion of maximum 18-fluorodeoxyglucose uptake. This allowed us to use the log scale (no negative values), and assured the correlation with qPET to be positive. The area-under-

the-ROC-curve was used to quantify the prognostic value of interim scanning. All analyses were carried out using R, version 3.5.1 (R Core Team, Vienna, Austria).

RESULTS

Patient Characteristics

Of 862 patients treated in the PETAL trial, 609 had DLBCL (*15*). Baseline and interim PET scans for post-hoc analyses were available from 449 patients. In 65 cases, the scans were not transferred to the central server, in 75, the transferred data was incomplete, and in 20, quantitative evaluation was not possible for technical reasons.

Baseline features and treatment results of the subgroup studied here were similar to the subgroup excluded from the analysis and the entire DLBCL population of the PETAL trial (Table 1, Supplemental Table 1 and Supplemental Fig. 1; no statistically significant differences).

Response Assessment

Among the 449 patients included in this study, 117 (26.0%) were assigned to visual Deauville score 1, 42 (9.4%) to score 2, 113 (25.2%) to score 3, 120 (26.7%) to score 4, and 57 (12.7%) to score 5.

Patients with a visual score of 1 (n=117) had a qPET value of zero, because the interim scan showed no measurable lymphoma-related activity. The qPET distribution of the 332 scans with measurable activity was symmetrical on the log scale, with some outliers (representing poorly responding patients) on the right, and a mode (most frequent value) near 1.3 (Fig. 1A).

qPET measurements within a single visual score category were well separated (Fig. 2). Some overlap was observed between neighboring categories, particularly between scores 3 and 4, which were located near the mode of the distribution. Optimal qPET thresholds between individual visual categories were 0.87 for the distinction between scores 2 and 3, 1.31 for scores 3 and 4, and 2.02 for scores 4 and 5, confirming the thresholds previously determined in pediatric Hodgkin's lymphoma (0.95, 1.3, 2.0) (*12*). These thresholds allow translation of qPET measurements into quantitative Deauville scores (qDS 1, qPET not measurable; qDS 2, qPET measurable,

but <0.95; qDS 3, qPET 0.95 to <1.3; qDS 4, qPET 1.3 to <2.0; qDS 5, qPET ≥2). The concordance between visual and quantitative scores was 82.4% (Table 2).

As for ΔSUV_{max}, 100% SUV_{max} reduction, was found in 117 patients. In the remainder, the distribution of measurable 1-ΔSUV_{max} values resembled the qPET distribution, with symmetry on the log scale, outliers on the right, and a mode at 0.156 corresponding to 84.4% SUV_{max} reduction (Fig. 1B).

Outcome Prediction

Employing the recommended thresholds for the visual Deauville scale (scores 1-3 versus 4-5) and the quantitative ΔSUV_{max} scale (>66% versus ≤66% SUV_{max} reduction) to distinguish between good and poor responders, the positive predictive values were 38.4% and 56.6% (p=0.03), and the negative predictive values were 75.4% and 73.5%, respectively. Raising the threshold to score 5 (versus 1-4) of the Deauville scale improved the positive predictive value, with a concomitant decrease in the proportion of high-risk patients (Table 3).

Treatment responses measured by qPET and ΔSUV_{max} were highly correlated on the log scale (Pearson's r=0.75, 95% confidence interval (CI) 0.70-0.80), suggesting that the methods provided similar information (Fig. 3). Neither of them reliably distinguished between patients in continued remission and patients who progressed, relapsed or died, except at very high qPET and 1-ΔSUV_{max} values where patients experiencing treatment failure were enriched. The similarities between qPET and ΔSUV_{max} were confirmed by ROC analysis yielding superimposable curves (Fig. 4). The area-under-the-ROC-curve was low for both methods (0.623 and 0.612, respectively), consistent with limited overall prognostic value.

To compare the positive and negative predictive values of qPET and ΔSUV_{max} at comparable thresholds, the values were plotted along their respective percentiles. Again, the curves were superimposable (Fig. 5). The positive predictive value was low up to the 85^{th} percentile, followed by a steep increase (Fig. 5A). The negative predictive value curves remained flat between 80% and 70% (Fig. 5B).

Corresponding percentiles were used to translate between the ordinal Deauville scale and the quantitative qPET and ΔSUV_{max} scales. Table 4 displays the clinically relevant thresholds on the visual 5-point scale and the 66% SUV_{max} reduction threshold on the ΔSUV_{max} scale which identifies patients with a high risk of treatment failure.

Outcome According to Prognostic Group

Figure 6 shows progression-free survival of patients categorized according to the Deauville scales (five categories) or the ΔSUV_{max} scale which was divided into three categories (SUV_{max} reduction by 100%, <100 to >66 %, ≤66%). The agreement between visual and quantitative Deauville scale was good. Progression-free survival did not differ between scores 2, 3, and 4, while score 5 was associated with significantly worse outcome (visual scale, hazard ratio (HR) 2.56, CI 1.68-3.90, p<0.0001; quantitative scale, HR 2.52, CI 1.71-3.73, p<0.0001), similar to SUV_{max} reduction ≤66% on the ΔSUV_{max} scale (HR 3.27, CI 2.16-4.96, p<0.0001).

Patients with complete normalization of the interim scan (Deauville score 1, 100% SUV_{max} reduction) tended to have better outcome than patients with good response, but residual activity. This observation was of borderline statistical significance (visual Deauville scale, HR 1.63, CI 1.03-2.60, p=0.036; quantitative Deauville scale, HR 1.57, CI 0.98-2.51, p=0.054; ΔSUV_{max} scale, HR 1.57, CI 0.99-2.51, p=0.051).

The 66% ΔSUV_{max} threshold identified a high-risk group comprising 13% of the total PETAL trial population (Fig. 5A) (*15*). The equivalent qPET value of 2.26 detected the same percentage, but not the same individuals. Of 75 high-risk patients, 31 were tested positive by both methods, while 22 each were tested positive only by qPET or ΔSUV_{max}, respectively. Single-positive patients had better outcome than double-positive patients (p=0.0011). In our interpretation, qPET and ΔSUV_{max} are different methods measuring the same response concept, i.e. replicates with independent measurement errors. Taking the average of the percentiles to reduce measurement errors, the mean percentile of double-negative patients was 0.47, of single-positive patients 0.83, and of double-positive patients 0.95 (p<0.0001). Thus, the observed outcomes (Fig. 7)

corresponded to what was expected from the superimposable curves shown in Figure 5A.

DISCUSSION

The main result of our study is that, in DLBCL, qPET carries the same prognostic information as ΔSUV_{max}. The ROC and predictive value curves as a function of percentiles were superimposable. Thus, the methods can be used interchangeably. Combining them may help in individual cases, but only to reduce measurement errors by averaging.

The results obtained with both quantitative methods imply that the currently recommended threshold to identify high-risk patients by virtue of the visual Deauville scale (score 1-3 versus 4-5) is of limited value. This conclusion complements our previous finding that ΔSUV_{max} is superior to Deauville for interim PET-based outcome prediction when recommended thresholds are employed (*17*). A cut-off between scores 4 and 5 of the visual scale may be more appropriate, in particular when interim PET is used to select patients for more aggressive therapies. The current definition of Deauville score 5, however, is imprecise (*7*). Therefore, a better alternative is utilizing a quantitative scale.

The visual Deauville scale is easy to use, but standardization remains difficult because of physiological limitations of the human eye. The perception of light intensity depends on the surrounding background. In addition, any visual comparison is compounded by the distance between the areas of interest. Therefore, the reproducibility of visual assessments remains limited (*18-20*). The qPET method circumvents these problems because it relies on objective measurements rather than subjective impressions. Perhaps more importantly, it converts the Deauville categories into a continuous scale, allowing the definition of risk groups independent of the somewhat arbitrary thresholds of the visual scale. qPET is similar to rPET (r, ratio) which compares the SUV_{max}, i.e. the single most intense voxels, in residual lymphoma and liver. In two small studies evaluating rPET after two cycles of R-CHOP, the best threshold for prognostic dichotomization was determined to be 1.4- or 1.6-times the SUV_{max} of the liver (*21,22*). After one or four treatment cycles, the most appropriate thresholds were 3.1- or 1.4-times the liver SUV_{max}, respectively (*20,23*). Advanced image reconstructions, however, may over-estimate SUV_{max} compared to SUV_{peak} and

SUV_{mean} (*24*). We therefore chose the SUV_{mean} of a large volume within the liver as the reference standard and a very small SUV_{peak} volume, comprising a low number of connected voxels, to represent the residual lymphoma lesion (*12*). Because of the systemic nature of DLBCL, its often rapid response to therapy, and the small size of post-treatment remnants, larger SUV_{peak} volumes commonly used in solid tumors appeared less suitable.

ΔSUV_{max} is more firmly established for quantitative interim scan evaluation than qPET. In contrast to ΔSUV_{max}, interim PET interpretation by qPET is based on a single scan, which minimizes the influence of factors known to impair SUV_{max} measurements. Prominent examples are blood glucose levels, adipose tissue, plasma clearance, paravenous injection, calibration and correction errors, and reconstruction algorithms (*25*).

While the Deauville scale distinguishes five response categories, dichotomization of the ΔSUV_{max} scale results in only two groups. In our study, patients with Deauville scores 2-4 did not differ in outcome. Interestingly, with all three methods, patients with complete interim PET normalization comprising >25% of the total population, tended to fare better than patients with good response, but remaining uptake. Our conclusion that interim PET may identify three rather than two prognostic groups needs to be confirmed in an independent data set.

Both qPET and ΔSUV_{max} convey the same information, but the proportion of patients identified to be at high-risk of treatment failure is less than 15%. Most patients eventually failing therapy remain undetected. Treatment response is only one of several factors determining outcome. Others include lymphoma burden and distribution which can readily be assessed at baseline PET scanning (*26,27*), gene expression (*28*), and genetic abnormalities (*29*). Combining one or several of these factors with early response assessment is likely to improve outcome prediction (*10,30*). Future studies will show whether radiation exposure can be eliminated by substituting PET/CT by serial measurement of circulating tumor DNA (*31*).

Strenghts of our study include rigorously defined conditions for PET performance and treatment delivery, a large sample size encompassing the entire spectrum of

DLBCL, and re-evaluation of all scans by a single specialist whose interpretation was reconciled with previous assessments of the same scans. In a comparative study, overall inter-observer agreement was found to be 'almost perfect' for the ΔSUV_{max} approach, but no more than 'substantial' for the Deauville scale (*19*). As for qPET, data on inter-observer concordance are not yet available. In rPET relying on similar principles as qPET, inter-observer agreement was found to be 'almost perfect' (*20*).

CONCLUSION

The currently recommended method for the identification of high-risk patients at interim PET scanning appears of limited value in diffuse large B-cell lymphoma. The visual Deauville scale should be replaced by one of the quantitative methods, such as qPET or ΔSUV_{max}, that minimize the confounding factors of visual assessment and permit outcome prediction on a continuous scale.

ACKNOWLEDGEMENTS

We thank the patients, investigators, and PET centers for their participation, and Deutsche Krebshilfe (grant nos. 107592 and 110515), Amgen Germany, and Roche Pharma for financial support.

DISCLOSURES

UD received institutional research grants and honoraria from Amgen Germany and Roche Pharma. No other potential conflicts of interest relevant to this article exist.

KEY POINTS

QUESTION:

Should the visual Deauville scale be replaced by a quantitative method of interim FDG-PET evaluation in DLBCL?

PERTINENT FINDINGS:

In a post-hoc analysis of the PETAL trial, the positive predictive value of the categorical Deauville scale was lower than that of the continuous ΔSUV_{max} and qPET scales. The continuous scales conveyed similar information.

IMPLICATIONS FOR PATIENT CARE:

ΔSUV_{max} and qPET are better suited for the identification of high-risk DLBCL patients than the visual Deauville scale.

REFERENCES

1. Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, Linet MS. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood*. 2006;107:265-276.

2. Cunningham D, Hawkes EA, Jack A, et al. Rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisolone in patients with newly diagnosed diffuse large B-cell non-Hodgkin lymphoma: A phase 3 comparison of dose intensification with 14-day versus 21-day cycles. *Lancet*. 2013;381:1817-1826.

3. Armitage JO, Weisenburger DD, Hutchins M, et al. Chemotherapy for diffuse large-cell lymphoma – rapidly responding patients have more durable remissions. *J Clin Oncol*. 1986;4:160-164.

4. Mikhaeel NG, Hutchings M, Fields PA, O'Doherty MJ, Timothy AR. FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol*. 2005;16:1514-1523.

5. Haioun C, Itti E, Rahmouni A, et al. [18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood*. 2005;106:1376-1381.

6. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol*. 2014;32:3059-3068.

7. 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.

8. Lin C, Itti E, Haioun C, et al. Early 18F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med*. 2007;48:1626-1632.

9. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al. Groupe d'étude des lymphomes de l'adulte (GELA). SUVmax reduction improves early prognosis value

of interim positron emission tomography scans in diffuse large B-cell lymphoma. *Blood*. 2011;118:37-43.

10. Schmitz C, Hüttmann A, Müller SP, et al. Dynamic risk assessment based on positron emission tomography scanning in diffuse large B-cell lymphoma: Post-hoc analysis from the PETAL trial. *Eur J Cancer*. 2020;124:25-36.

11. 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.

12. Hasenclever D, Kurch L, Mauz-Körholz C, et al. qPET – a quantitative extension of the Deauville scale to assess response in interim FDG-PET scans in lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41:1301-1308.

13. Georgi T, Kurch L, Hasenclever D, et al. Quantitative assessment of interim PET in Hodgkin lymphoma: An evaluation of the qPET method in adult patients in the RAPID trial. *PLoS One*. 2020;15:e0231027.

14. Kurch L, Hasenclever D, Kluge R, et al. Only strongly enhanced residual FDG uptake in early response PET (Deauville 5 or qPET ≥ 2) is prognostic in pediatric Hodgkin lymphoma: Results of the GPOH-HD2002 trial. *Pediatr Blood Cancer*. 2019;66: e27539.

15. Dührsen U, Müller 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.

16. Maurer MJ, Ghesquières H, Jais JP, et al. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *J Clin Oncol*. 2014;32:1066-1073.

17. Rekowski J, Hüttmann A, Schmitz C, et al. Interim PET evaluation in diffuse large B-cell lymphoma employing published recommendations: comparison of the Deauville 5-point scale and the ΔSUV_{max} method. *J Nucl Med*. 2020 (in press).

18. Horning SJ, Juweid ME, Schöder H, et al. Interim positron emission tomography scans in diffuse large B-cell lymphoma: an independent expert nuclear medicine evaluation of the Eastern Cooperative Oncology Group E3404 study. *Blood*. 2010;115:775-777.

19. Itti E, Meignan M, Berriolo-Riedinger A, et al. An international confirmatory study of the prognostic value of early PET/CT in diffuse large B-cell lymphoma: comparison between Deauville criteria and ΔSUVmax. *Eur J Nucl Med Mol Imaging*. 2013;40:1312-1320.

20. Toledano MN, Vera P, Tilly H, Jardin F, Becker S. Comparison of therapeutic evaluation criteria in FDG-PET/CT in patients with diffuse large-cell B-cell lymphoma: Prognostic impact of tumor/liver ratio. *PLoS One*. 2019;14:e0211649.

21. Itti E, Juweid ME, Haioun C, et al. Improvement of early 18F-FDG PET interpretation in diffuse large B-cell cymphoma: Importance of the reference background. *J Nucl Med*. 2010;51:1857-1862.

22. Fan Y, Zhang Y, Yang Z, et al. Evaluating early interim fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography with the SUVmax-liver-based interpretation for predicting the outcome in diffuse large B-cell lymphoma. *Leuk Lymphoma*. 2017;58:1-9.

23. Mylam KJ, Kostakoglu L, Hutchings M, et al. 18F-fluorodeoxyglucose-positron emission tomography/computed tomography after one cycle of chemotherapy in patients with diffuse large B-cell lymphoma: results of a Nordic/US intergroup study. *Leuk Lymphoma*. 2015;56: 2005-2012.

24. Kaalep A, Sera T, Rijnsdorp S, et al. Feasibility of state of the art PET/CT systems performance harmonisation. *Eur J Nucl Med Mol Imaging*. 2018;45:1344-1361.

25. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging*. 2010;37:181-200.

26. 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. 2014;41:2017-2022.

27. Cottereau AS, Nioche C, Dirand AS, et al. 18F-FDG PET dissemination features in diffuse large B-cell lymphoma are predictive of outcome. *J Nucl Med*. 2020;61:40-45.

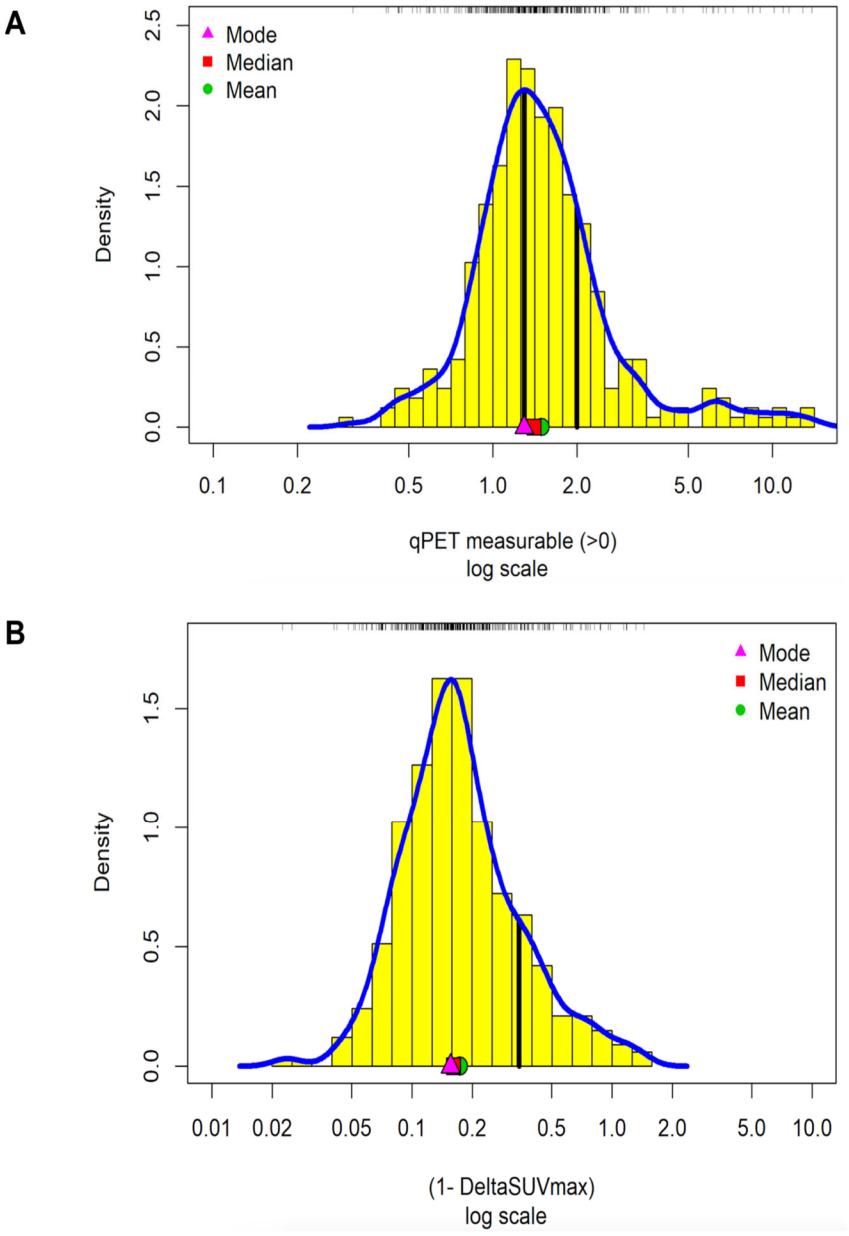
28. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403:503-511.

29. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24:679-690.

30. Kim J, Lee JO, Paik JH, Lee WW, Kim SE, Song YS. Different predictive values of interim 18F-FDG PET/CT in germinal center like and non-germinal center like diffuse large B-cell lymphoma. *Ann Nucl Med*. 2017;31:1-11.

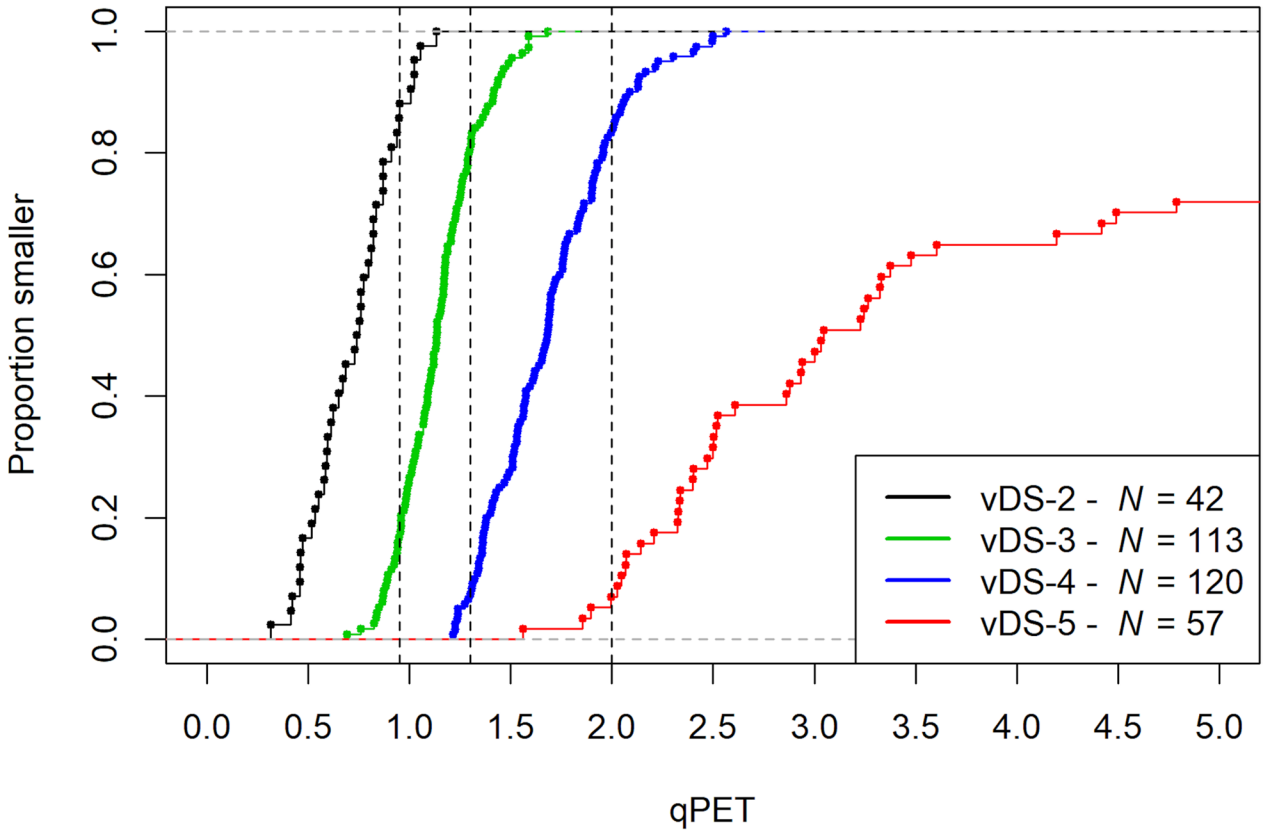
31. Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. *J Clin Oncol*. 2018;36:2845-2853.

Figure 1



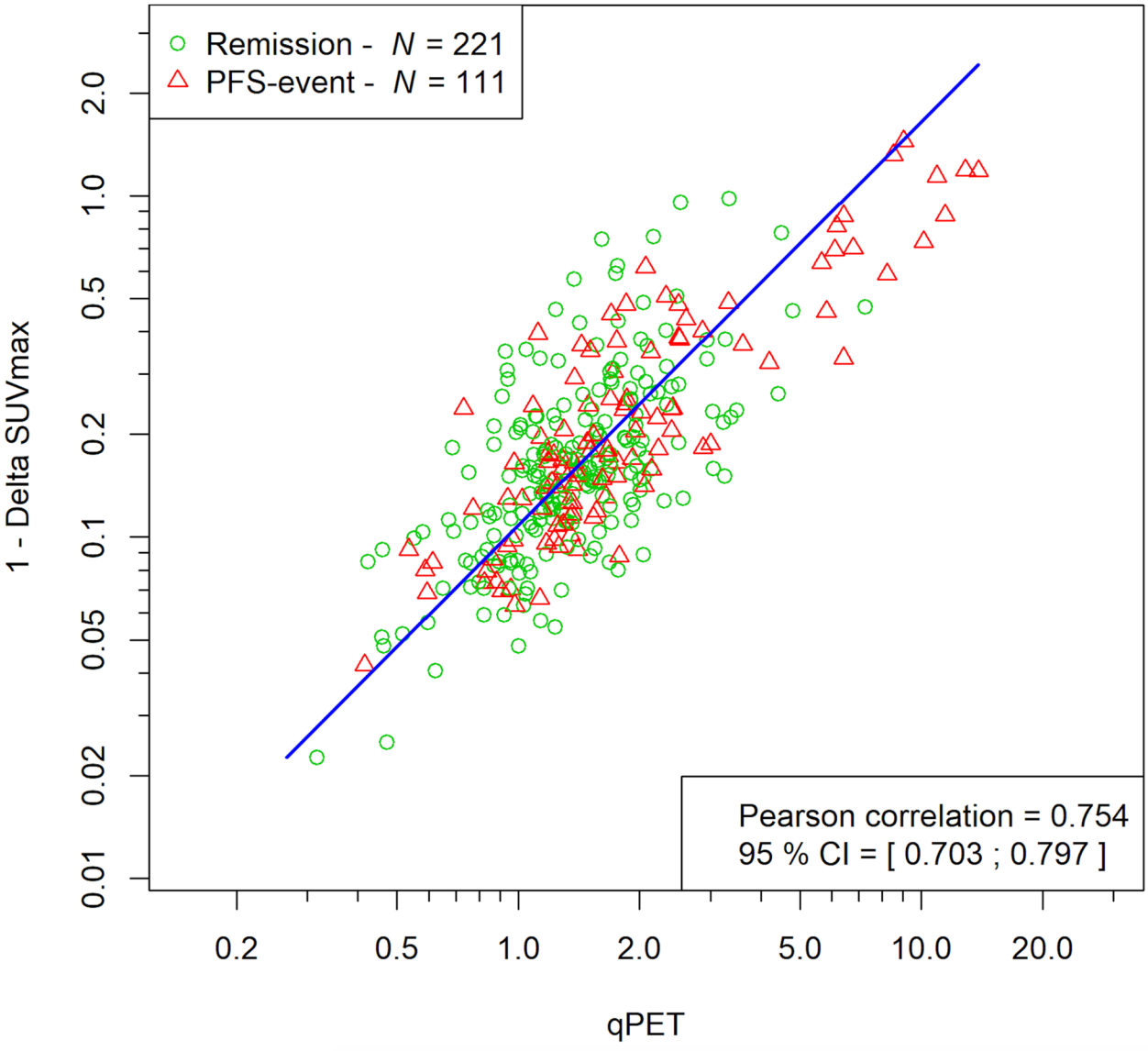
Density histograms for patients with measurable residual uptake at interim PET scanning (n=332) evaluated by qPET (A) or ΔSUV_{max} (B) on a log scale. The vertical lines in panel A indicate the published thresholds between the visual Deauville scores 3 and 4 (1.3) and 4 and 5 (2.0), respectively. ΔSUV_{max} values in panel B are expressed as 1-ΔSUV_{max}. The vertical line indicates the published threshold of 0.66, here 1-0.66 = 0.34.

Figure 2



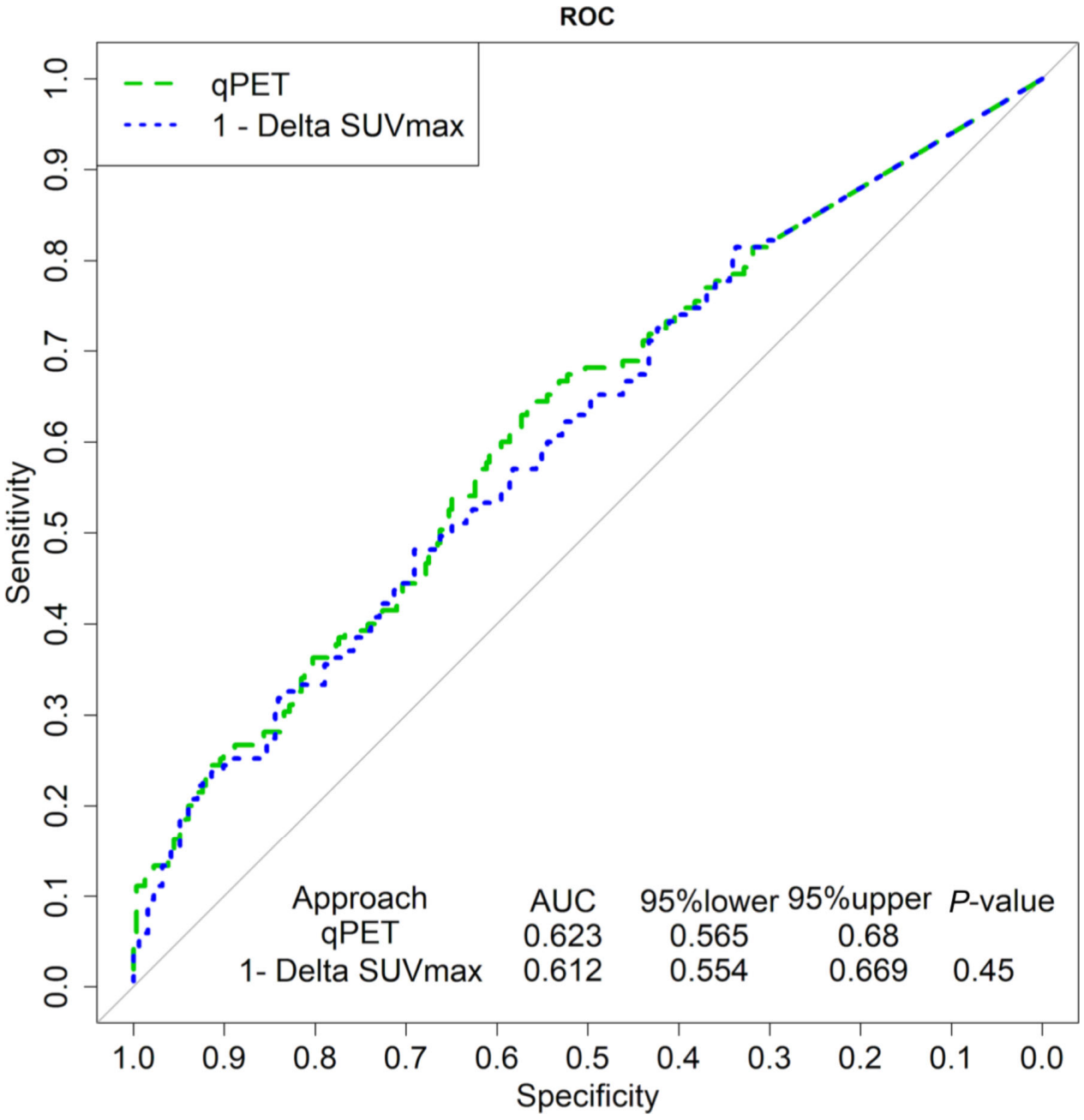
Empirical cumulative distribution functions of qPET measurements by visual Deauville categories. Vertical lines indicate the published thresholds to map qPET values to individual categories. vDS, visual Deauville score.

Figure 3



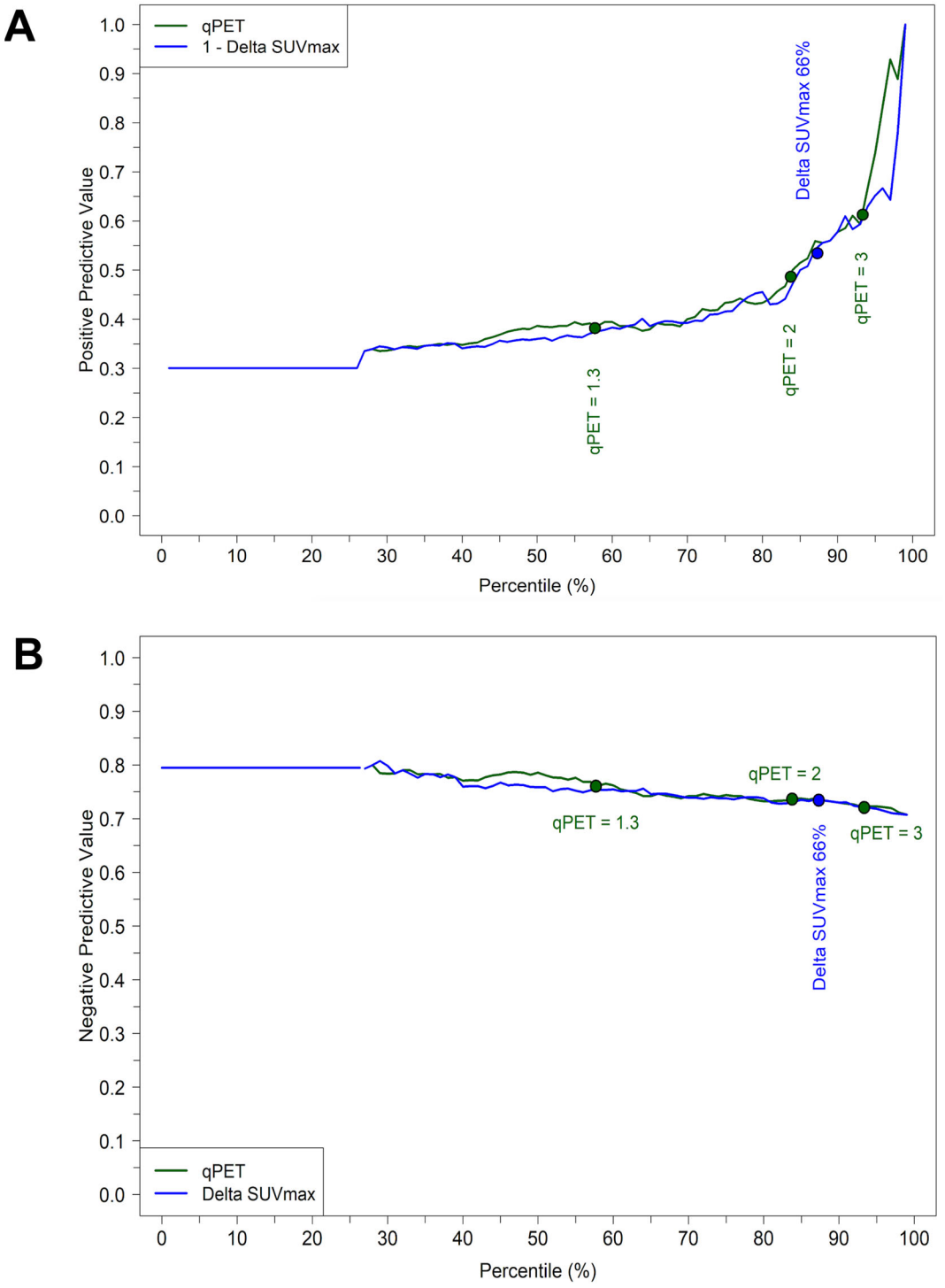
Scatterplot of qPET and ΔSUV_{max} values. The red triangular symbols refer to patients experiencing treatment failure whereas the green symbols refer to patients who remained in remission. The blue line is the principal axis illustrating the correlation. CI, confidence interval; PFS, progression-free survival.

Figure 4



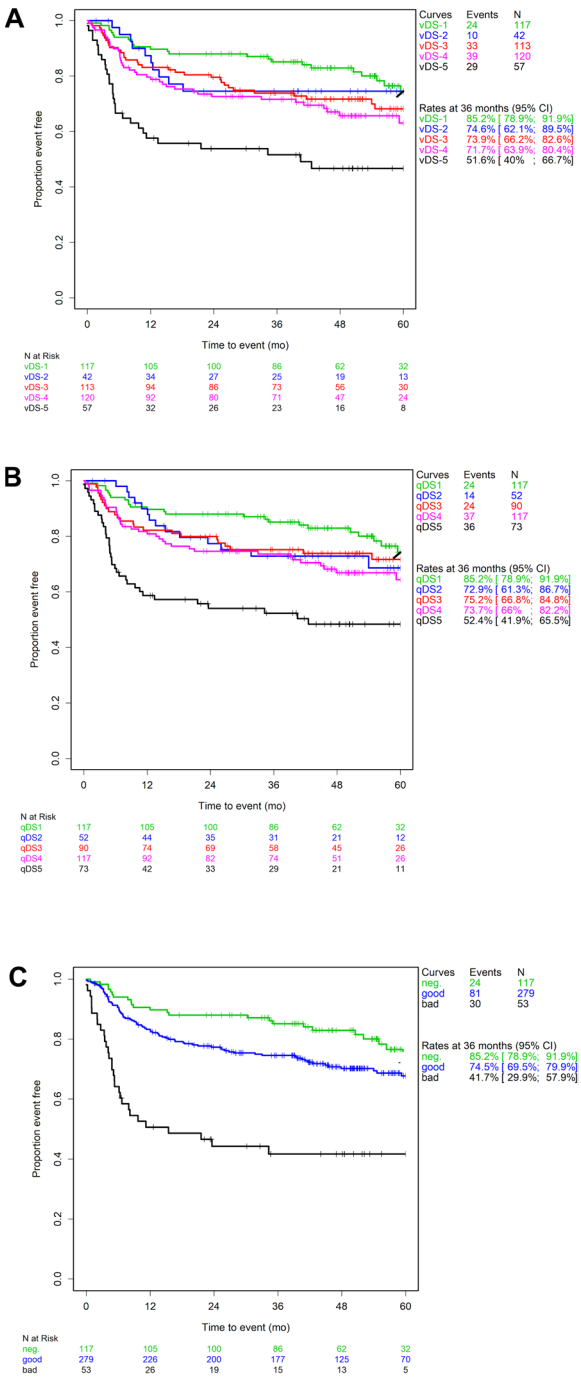
Receiver Operating Characteristic (ROC) curves of qPET and 1-ΔSUV_{max} for progression-free survival. AUC, area-under-the-curve.

Figure 5



Positive predictive value (A) and negative predictive value (B) of corresponding percentiles of qPET and ΔSUV_{max} measurements. The constant part of the curves at low percentiles is due to the inclusion of non-measurable values set at zero (n=117).

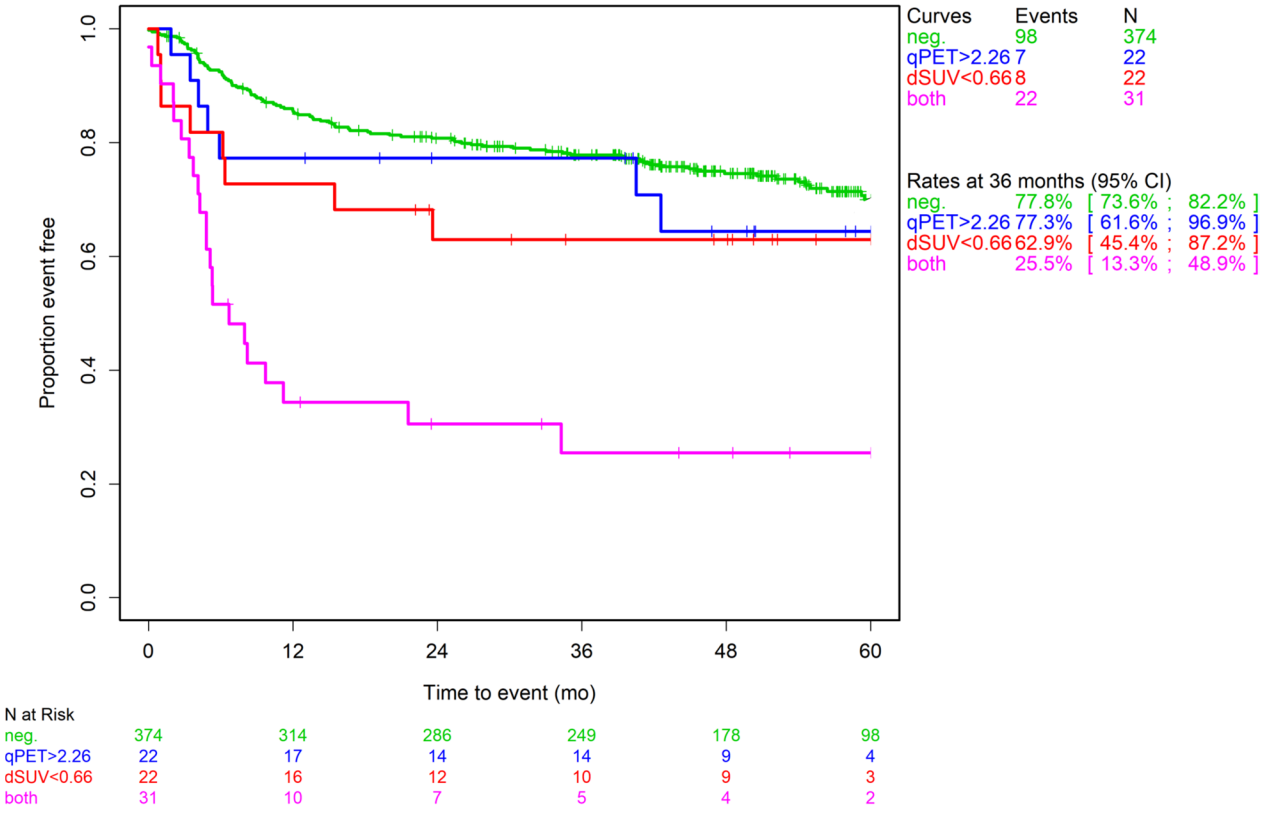
Figure 6



Progression-free survival in prognostic subgroups derived from the visual Deauville scale (A), the quantitative Deauville scale (B), or the ΔSUV_{max} scale (C) (Kaplan-Meier

analysis). CI, confidence interval; vDS, visual Deauville score, qDS, quantitative Deauville score.

Figure 7



Progression-free survival in patients with a good interim PET response according to both qPET and ΔSUV_{max}, only qPET or only ΔSUV_{max}, or a poor interim PET response according to both methods (Kaplan-Meier analysis).

Table 1

Baseline characteristics of patients included in the present analysis in comparison to excluded patients and all diffuse large B-cell lymphoma patients participating in the PETAL trial

Characteristic	Pts. included	Pts. excluded	All patients
No. of patients	449	160	609
Age – median (range), years	62 (18 – 80)	59.5 (18 – 79)	62 (18 – 80)
Age >60 years	236 (52.6%)	78 (48.8%)	314 (51.6%)
Male sex	249 (55.5%)	93 (58.1%)	342 (56.2%)
ECOG performance status ≥2	48 (10.7%)	11 (6.9%)	59 (9.7%)
Ann Arbor stage III or IV	258 (57.5%)	100 (62.5)	358 (58.8%)
Extranodal sites >1	148 (33.0%)	50 (31.2%)	198 (32.6%)
Lactate dehydrogenase >ULN	257 (57.4%)	78 (48.8%)	335 (55.1%)
International Prognostic Index			
Low risk	160 (35.7%)	64 (40.0%)	224 (36.8%)
Low-intermediate risk	111 (24.8%)	47 (29.4%)	158 (26.0%)
High-intermediate risk	102 (22.8%)	25 (15.6%)	127 (20.9%)
High risk	75 (16.7%)	24 (15.0%)	99 (16.3%)

Data are given as number of patients affected (% of total number of patients with documented data), unless otherwise noted. ECOG, Eastern Cooperative Oncology Group; ULN, upper limit of normal

Table 2

Comparison of visual (vDS) and quantitative Deauville scores (qDS)

	qDS1	qDS2	qDS3	qDS4	qDS5	Sum
vDS1	**117**	0	0	0	0	117
vDS2	0	**35**	7	0	0	42
vDS3	0	17	**74**	22	0	113
vDS4	0	0	9	**91**	20	120
vDS5	0	0	0	4	**53**	57
Sum	117	52	90	117	73	449

Table 3

Positive and negative predictive values and proportion of high-risk patients identified by interim positron emission tomography – comparison of methods and thresholds

Definition of high-risk patients	PPV	NPV	Proportion of high-risk pts.
Visual Deauville score 4 or 5	38.4%	75.4%	39.4%
Visual Deauville score 5	50.9%	73.0%	12.7%
Quantitative Deauville score 4 or 5	38.4%	76.1%	42.3%
Quantitative Deauville score 5	49.3%	73.7%	16.3%
qPET ≥2.26	54.7%	73.2%	11.8%
ΔSUV_{max}, ≤66% SUV_{max} reduction	56.6%	73.5%	11.8%

NPV, negative predictive value; PET, positron emission tomography; PPV, positive predictive value; pts., patients; qPET, quotient lymphoma/liver; SUV_{max}, maximum standardized uptake value

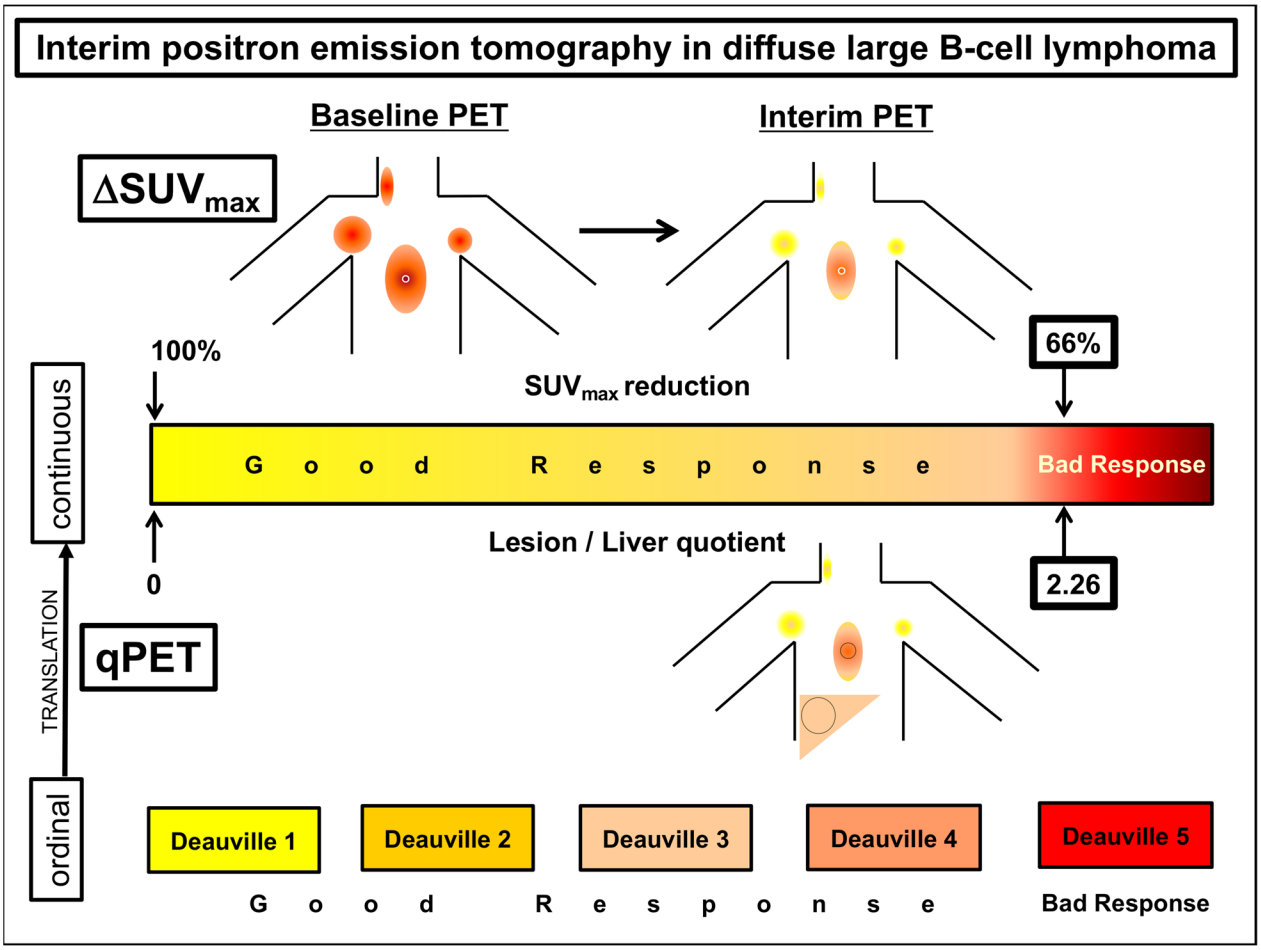
Table 4

Selected corresponding thresholds in categorical and continuous interim PET response scales

vDS	qPET	ΔSUV_{max}
2/3	0.95	91%
3/4	1.30	85%
4/5	2	73%
5	2.26	66%

PET, positron emission tomography; qPET, quotient lymphoma/liver; SUV_{max}, maximum standardized uptake value; vDS, visual Deauville scale

Graphical Abstract



Supplemental Table 1

Treatment results of patients included in the present analysis in comparison to excluded patients and all diffuse large B-cell lymphoma patients participating in the PETAL trial

	Pts. included	Pts. excluded	All patients
No. of patients	449	160	609
Treatment allocation[1]			
6 x R-CHOP	225 (50.1%)	69 (43.1%)	294 (48.3%)
6 x R-CHOP + 2 x R	182 (40.5%)	70 (43.8%)	252 (41.4%)
8 x R-CHOP	22 (4.9%)	10 (6.2%)	32 (5.3%)
2 x R-CHOP + 6 x Burkitt protocol	20 (4.5%)	11 (6.9%)	31 (5.1%)
Outcome			
Overall response[2]	380 (84.6%)	129 (80.6%)	509 (83.6%)
Complete remission[2]	280 (67.8%)	97 (68.8%)	377 (68.1%)
3-year progression-free survival[3]	73.5% (69.5 – 77.8)	73.5% (66.7 – 80.9)	73.5% (70.0 – 77.2)
3-year overall survival[3]	82.4% (78.9 – 86.1)	82.6% (76.7 – 88.9)	82.4% (79.3 – 85.6)

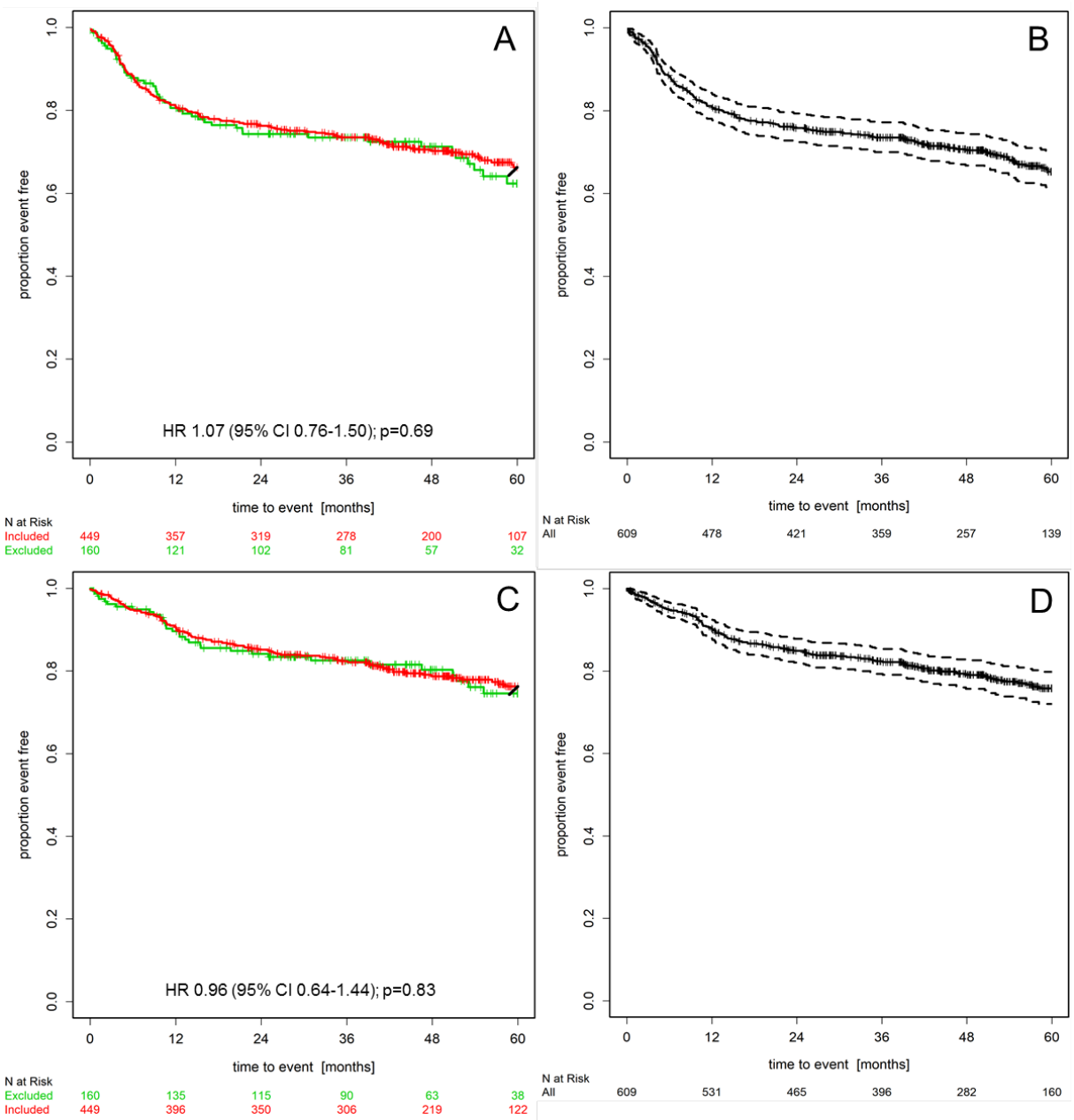
CHOP, cyclophosphamide, doxorubicin, vincristine, prednisone; pts., patients; R, rituximab

[1] Data are given as number of patients affected (% of total number of patients)

[2] No. of patients responding / total no. of patients reaching the end-of-treatment evaluation (%)

[3] Kaplan-Meier estimate of percentage of patients surviving after 3 years (95% confidence interval)

Supplemental Figure 1



Progression-free survival (A, B) and overall survival (C, D) of patients included in the present analysis in comparison to excluded patients (A, C) and all diffuse large B-cell lymphoma patients (B, D) participating in the PETAL trial.