ORIGINAL RESEARCH
Open Access

A comparative study of standardized quantitative and visual assessment for predicting tumor volume and outcome in newly diagnosed diffuse large B-cell lymphoma staged with 18F-FDG PET/CT

Lars C. Gormsen[1*], Mikkel H. Vendelbo[1,7], Mette Abildgaard Pedersen[1], Ate Haraldsen[1], Karin Hjorthaug[1], Trond Velde Bogsrud[1,6], Lars J. Petersen[3], Karen Juul Jensen[4], Rasmus Brøndum[2] and Tarec C. El-Galaly[2,5]

Abstract

Background: Semi-automated quantitative measurement of metabolic tumor volume (MTV) for prognosis in diffuse large B-Cell lymphoma (DLBCL) has gained considerable interest lately. However, simple tumor volume measures may be inadequate for assessment of prognosis in DLBCL as other characteristics such as growth pattern and metabolic heterogeneity may be just as important. In addition, MTV measurements require delineation of tumor lesions by semi-automated software, which can be time-consuming. We hypothesized that a simple visual assessment of tumor volume performs as well as standardized MTV measurements in DLBCL prognostication.

Materials and methods: Quantitative and visual analyses of pre-therapy 18F-FDG PET/CT scans in 118 patients with newly diagnosed DLBCL were conducted. Quantitative analyses were performed using Hermes TumourFinder® to obtain $MTV^{2.5}$ (SUV 2.5 cut-off) and MTV^{41} (41% SUVmax isocontour cut-off). Visual assessments included a binary prediction (good/poor prognosis) as well as tumor burden based on a visual analog scale (MTV^{VAS}) and an estimated volume (eMTV). Three experienced nuclear medicine physicians who were blinded to clinical outcome performed visual evaluations. Progression-free survival was evaluated by Kaplan-Meier curves and log-rank test. Inter-observer variability was evaluated by Fleiss' kappa for multiple observers.

Results: In the quantitative analysis, a ROC-determined $MTV^{2.5}$ cut-off (log-rank $p = 0.11$) seemed to outperform MTV^{41} (log-rank $p = 0.76$) for PFS prediction. TLG2.5 (log-rank $p = 0.14$) and TLG41 (log-rank $p = 0.34$) were not associated with outcomes. By visual analysis, all three reviewers were able to stratify patients into good/poor prognosis (reviewer A log-rank $p = 0.002$, reviewer B log-rank $p = 0.016$, and reviewer C log-rank $p = 0.012$) with fair inter-observer agreement (Fleiss' kappa 0.47). MTV^{VAS} and eMTV were not consistently correlated with the outcome.

Conclusion: Predictions of outcome after first-line treatment for DLBCL were surprisingly good when left to the unsupervised, subjective judgment of experienced readers of lymphoma 18F-FDG-PET/CT. The study highlights the importance of non-standardized clinical judgments and shows potential loss of valuable prognostic information when relying solely on semi-automated MTV measurements.

Keywords: DLBCL, MTV, 18F-FDG PET/CT, Prognosis

---

* Correspondence: lars.christian.gormsen@clin.au.dk
[1]Department of Nuclear Medicine & PET Centre, Aarhus University Hospital, DK 8000 Aarhus C, Aarhus, Denmark
Full list of author information is available at the end of the article

Background

Diffuse large B-cell lymphoma (DLBCL) is the most common type of non-Hodgkin lymphoma in the Western world. The majority of patients with DLBCL respond to standard immunochemotherapy with a combination of the drugs rituximab, cyclophosphamide, doxorubicin hydrochloride (hydroxydaunorubicin), vincristine sulfate, and prednisone (R-CHOP), but 30–40% of the patients are refractory or relapse following initial response [1]. These patients have dismal outcomes and only a minority can be cured by salvage high-dose therapy and autologous stem cell transplantation [2, 3]. No major progress has been made in the treatment of DLBCL since the introduction of rituximab. Attempts to improve outcomes by intensifying chemotherapy and rituximab-dosing schedules have failed [4–10]. A treatment escalation strategy based on early PET/CT response did not improve outcomes in poor responders over and above what was achieved by continuing standard therapy [11]. Also, the predictive value of interim PET/CT in DLBCL, unlike in Hodgkin lymphoma [12], remains controversial [11]. The key to improving outcomes in DLBCL may reside in better risk stratification of DLBCL using an integrative approach that combines baseline clinical risk factors, genomic features, and early treatment response [13, 14]. Combinations of tumor volume measurements on baseline PET/CT and early treatment response have been proposed to identify high-risk DLBCL patients likely to fail standard therapy [14]. Providing metabolic tumor volume measurements (quantitative PET/CT), however, is time-consuming and requires highly standardized algorithms for image acquisition and reconstruction. Furthermore, relying increasingly on "automatization" for prognostics minimizes the influence of clinical judgments by readers with years of experience. Numerous automated or machine learning-based algorithms for outcome prediction, particularly in radiology have been proposed in the past decade; however, only few have been found to be on par or outperform human-level performance [15]. To test the superiority of quantitative PET/CT over simple and unsupervised assessment of prognosis by experienced nuclear medicine specialists, we compared the prognostic value of both methods in a consecutive cohort of 118 patients.

Methods

One hundred and eighteen patients with newly diagnosed DLBCL treated with R-CHOP/R-CHOP-like therapy and undergoing diagnostic work-up including baseline 18F-FDG PET/CT in the period 2007–2012 at Aalborg University Hospital were included in this retrospective study. Clinical information including outcome parameters was collected from medical records. The Danish Data Protection agency (ref 1-16-02-88-15) and the

Danish Ministry of Health approved the study (ref 3-3013-860/1/).

PET/CT acquisition

Staging 18F-FDG PET/CT scans were carried out after a 6-h fast on a GE Discovery VCT 710 integrated PET/CT system in accordance with local protocols and manufacturer guidelines. Images from the base of the skull to the proximal thigh were acquired a median of 71 (range 57–166) minutes after injection of 4 MBq/kg 18F-FDG and were reconstructed using iterative reconstruction.

Quantitative image analysis (Table 1)

18F-FDG PET/CT images were analyzed by author LCG using Hybrid Viewer (Hermes Medical Solutions, Sweden) with the plug-in TumourFinder. Lymphoma lesions were segmented using both a SUV > 2.5 threshold ($SUV^{2.5}$) as well as > 41% of the SUVmax (SUV^{41}) thresholding. Quantitation parameters metabolic tumor volume ($MTV^{2.5}$ and MTV^{41}) and total lesion glycolysis ($TLG^{2.5}$ and TLG^{41}) were recorded for both segmentation methods.

Visual image analysis (Table 1)

Three experienced (> 10 years) nuclear medicine physicians (A.H., T.V.B., and K.H.) blinded to clinical information except patient age performed visual analysis of images. Observers were asked to rate the following tentatively predictive parameters: (1) prognosis (poor/favorable), (2) tumor volume estimate (eMTV) (in milliliters), and (3) tumor volume (MTV^{vas}) (visual analog score; VAS, 1–9). Observers were asked to restrict the time spent evaluating these parameters to a maximum of 30 s per evaluation in order to reflect the intuitive interpretation of the images. Figure 1 depicts the VAS scale used by our reviewers as well as examples of patients with poor and favorable prognosis. To visually assess prognosis, reviewers were asked to take into account tumor heterogeneity, tumor volume, infiltration into adjoining tissue, extra-nodal spread, and diffuse involvement of, e.g., the pleura or peritoneum.

In case of disagreement between reviewers, prognosis was analyzed as a trichotomous variable (favorable/intermediate/poor), whereas eMTV and MTV^{vas} were averaged for all reviewers.

Statistics

Fleiss' kappa was used to evaluate the inter-observer variability when reporting the subjective assessment of prognosis, while the variability for MTV^{vas} and eMTV was evaluated using the intra-class correlation (ICC). For MTV^{VAS}, we used the ICC described in [16] while the ICC for absolute agreement between a random sample of readers was used for eMTV (ICC2 from the

Table 1 Semi-automated methods and visual assessment of tumor volume and prognosis

Method	Semi-automated measurements		Visual assessment		
	MTV$^{2.5}$	MTV^{41}	eMTV	MTV^{VAS}	Prognosis
Description	Metabolic tumor volume delineation > SUV 2.5	Metabolic tumor volume delineation > 41% of SUVmax	Metabolic tumor volume assessed visually	Degree of metabolic tumor volume assessed on VAS scale	Prognosis assessed visually based on volume, heterogeneity, involvement of extra-nodal organs
Parameter	Continous (ml)	Continous (ml)	Continous (ml)	Continous (1–9)	Dichotomous (poor/favorable)

R-package *psych*). The Pearson correlation coefficient was used to measure consistency between MTV values obtained by the automated software (MTV$^{2.5}$ and MTV^{41}) and visual estimates (eMTV).

Survival analyses were performed according to prognosis subjectively assessed (good/bad), eMTV, MTV^{vas}, MTV$^{2.5}$, MTV^{41}, TLG$^{2.5}$, and TLG^{41}. Progression-free survival (PFS) was defined as the time from diagnosis to relapse/progression or death from any cause. Overall survival (OS) was defined as the time from diagnosis to death from any cause. Optimal PFS cut-offs for continuous variables were determined with receiver operating
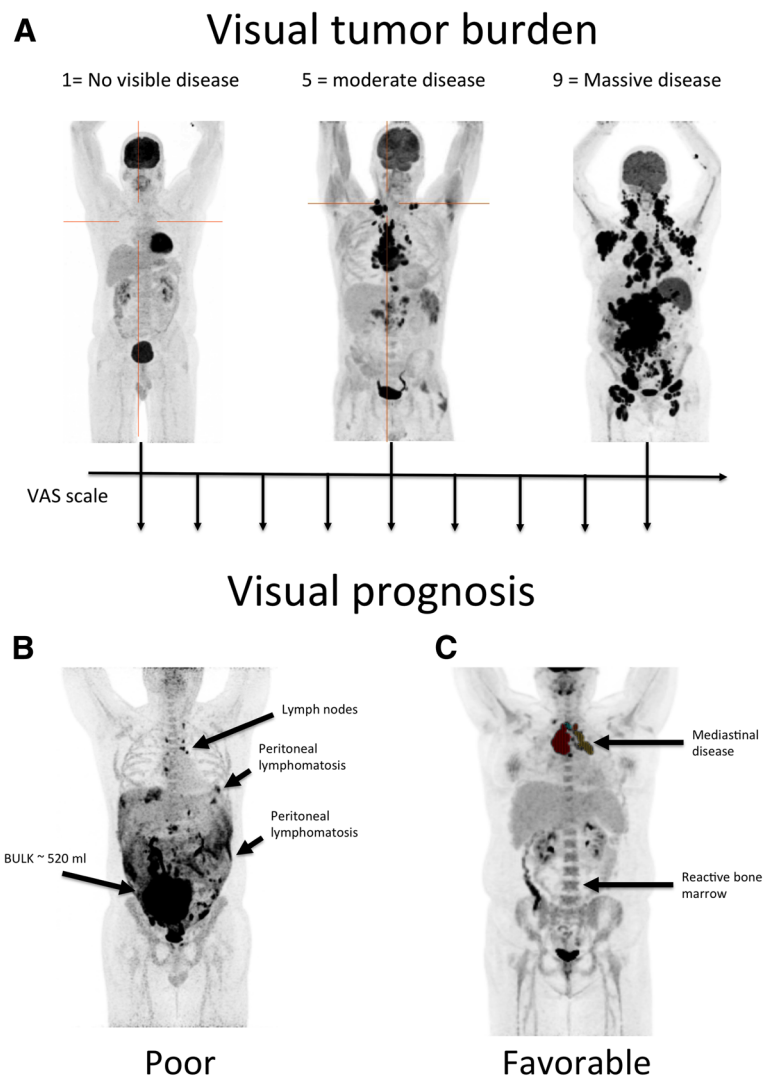

Fig. 1 Metabolic tumor burden assessed on a VAS scale with representative examples of patients graded as 1, 5, and 9 (**a**). All reviewers used these examples to grade individual cases in the study. **b** An example of a patient with a poor prognosis based on the widespread peritoneal lymphomatosis, bulky disease, and high metabolic tumor volume. Pointedly, the peritoneal lymphomatosis may be difficult to accurately quantify using semi-automatic tumor delineation tools whereas the severity of disease is readily perceived by visual inspection using a maximum intensity projection FDG PET/CT. **c** A patient with limited mediastinal disease used as an example of a patient with favorable prognosis

characteristics (ROC) curves for three-year survival and three-year progression-free survival as the point with the shortest distance from (0, 1). Survival curves were computed using the Kaplan-Meier method and differences were tested using the log-rank test. Statistics were calculated using R version (3.4.3).

Results

A total of 118 patients (76 males and 42 females) with median age 66 (range 18–88) were included (Table 2). According to the international prognostic index (IPI), 60 patients (50.8%) were low/low-intermediate risk, 29 (24.6%) were high-intermediate, and 23 (19.5%) were high-risk. Information on IPI was not available in 5 patients. Median follow-up was 67 months (reverse Kaplan-Meier method) and the overall 5-year PFS and overall survival (OS) estimates were 65.7% and 68.9%, respectively (Fig. 2 for IPI-specific survival curves). Treatment regimens were R-CHOP ($n = 104$, 88.1%), R-CHOEP ($n = 12$, 10.2%), and R-CEOP ($n = 2$, 1.7%). Two patients turned out to be CD20 negative and thus had rituximab removed from their treatment regimens.

The inter-observer agreement evaluated by Fleiss' kappa was moderate at 0.47 when PET/CT readers assessed prognosis (poor/favorable), whereas the agreement for MTV^{vas} and eMTV were good to excellent with values of 0.81 and 0.66, respectively. Finally, moderate to high correlations between automated MTV measurements (MTV$^{2.5}$, MTV^{41}, TLG$^{2.5}$, and TLG41) and visual eMTV estimates were found, as shown in Table 1. There was, however, no unanimous highest correlation between the automated and visually assessed MTV across the three reviewers.

Prediction of outcome using automated software

The areas under the ROC curves for 3-year PFS predictions were 0.59, 0.54, 0.58, and 0.56 for MTV$^{2.5}$, MTV^{41}, TLG$^{2.5}$, and TLG^{41}, respectively. The corresponding values for 3-year OS predictions were 0.58, 0.53, 0.57, and 0.54 for MTV$^{2.5}$, MTV^{41}, TLG$^{2.5}$, and TLG^{41}, respectively. The best dichotomous cut-off values for PFS

Table 2 IPI classification for patients. Pre-therapy IPI was unavailable in five patients

Patient characteristics	
Age	66 (16–88)
Ann Arbor I–II	40 (34%)
Ann Arbor III–IV	76 (64%)
Nodal > 1	31 (26%)
ECOG > 1	18 (15%)
LDL > UNL	62 (53%)
IPI > 2	52 (44%)

based on tumor burden were 542 ml, 147 ml, 3237, and 1485 for MTV$^{2.5}$, MTV^{41}, TLG$^{2.5}$, and TLG^{41}, respectively. Corresponding values for OS were 585 ml, 105 ml, 3560 ml, and 2557 ml for MTV$^{2.5}$, MTV^{41}, TLG$^{2.5}$, and TLG^{41}, respectively. Using these thresholds, only MTV$^{2.5}$ was borderline associated with OS (log-rank $p = 0.065$) and PFS (log-rank $p = 0.11$) (Fig. 2a and e), whereas MTV^{41} was not associated with OS (log-rank $p = 0.76$) or PFS (log-rank $p = 0.67$) (Fig. 2b and f). TLG2.5 (log-rank $p = 0.14$) or TLG41 (log-rank $p = 0.34$) was not prognostic in PFS or OS analyses.

Visual estimation of tumor volume

For MTV^{vas} and eMTV, the areas under ROC curves for the three reviewers for PFS and OS were 0.62, 0.63, and 0.58 and 0.61, 0.61, and 0.56 for MTV^{vas} and 0.60, 0.62, and 0.58 and 0.60, 0.60, and 0.56 for eMTV. Best cut-offs for MTV^{vas} were 4, 4, and 6 for both PFS and OS. Best eMTV cut-off for high and low tumor burden was 288, 450, and 425 ml for PFS and 288, 450, and 1250 ml for OS. Subsequent Kaplan-Meier analyses showed that only the eMTV threshold for one of the reviewers (450 ml) could separate patients in high or low risk (Kaplan-Meier log-rank $p = 0.034$ for PFS). However, the areas under the ROC curves were comparable or better for eMTV than for MTV. Table 3 shows the correlation between eMTV of individual reviewers and the semi-automated measurements, which was surprisingly good for two reviewers.
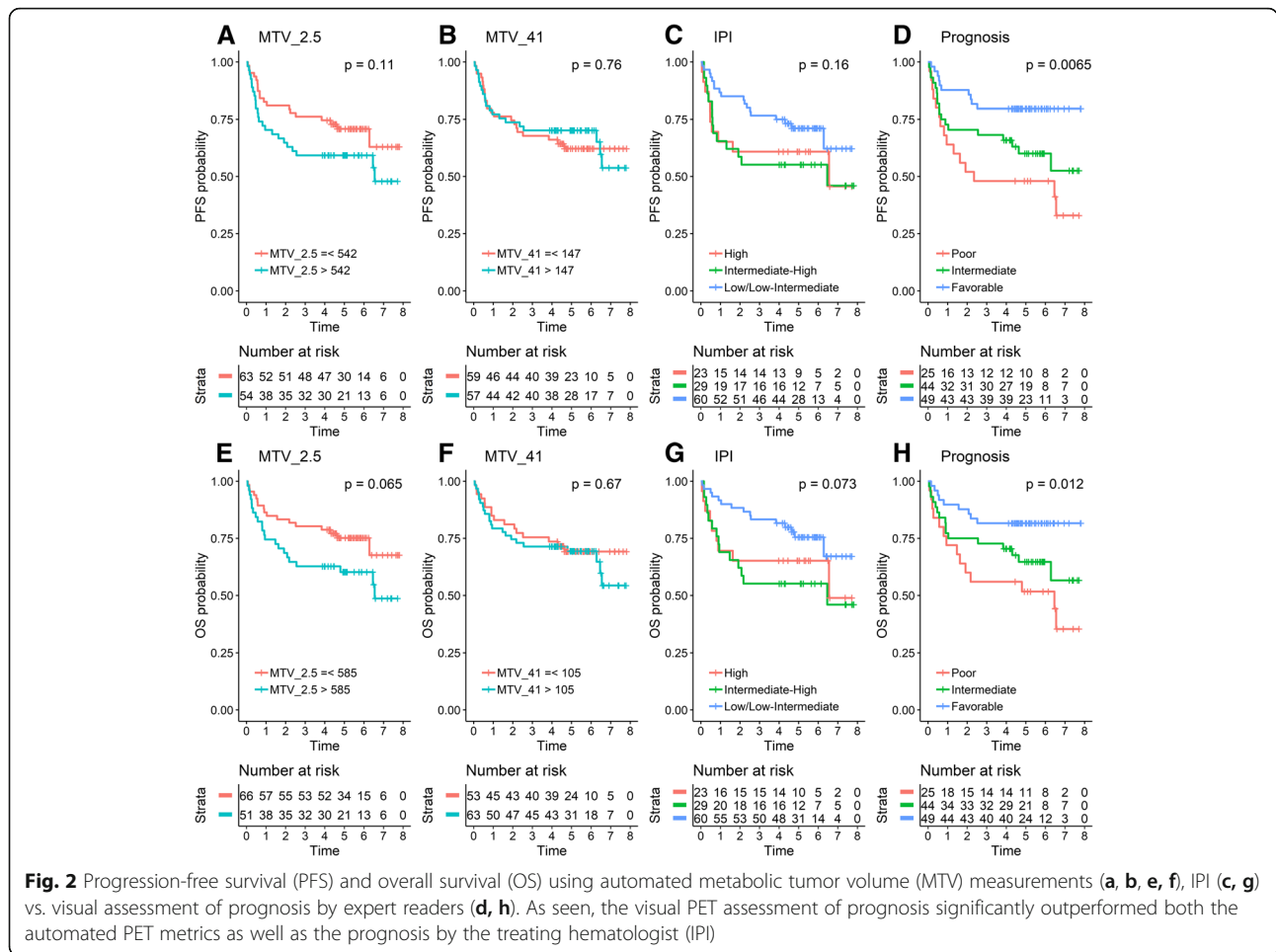
Visual prognosis

The subjective visual assessments of prognosis (poor/favorable) were highly associated with outcomes both in PFS and OS analyses for all three reviewers, and the consensus prognosis for the three reviewers (all poor/mixed/all favorable) also showed strong association ($p < 0.01$ for PFS and $p = 0.01$ for OS (Fig. 2d and g). For the 44 patients in the mixed group, the majority vote was "poor" for 14 and "favorable" for the remaining 30. Ranges of MTV2.5 were 287–8462, 8–2672, and 14–2260 for "all poor," "all favorable," and "mixed" consensus prognosis patients, respectively. The overlap in tumor volume between patients classified by visual prognosis underlines the importance of other factors than tumor burden, when reviewers assessed prognosis.

Using the majority vote, the log-rank test also showed significant differences between the two groups, differences in 2-year survival were, however, larger when including a mixed group, with differences between favorable and poor of respectively 28% and 36% for OS and PFS using the consensus diagnosis and differences of 23% and 26% for OS and PFS using the majority vote (Fig. 3).

Gormsen *et al. EJNMMI Research* (2019) 9:36

Page 5 of 8

Fig. 2 Progression-free survival (PFS) and overall survival (OS) using automated metabolic tumor volume (MTV) measurements (**a**, **b**, **e**, **f**), IPI (**c**, **g**) vs. visual assessment of prognosis by expert readers (**d**, **h**). As seen, the visual PET assessment of prognosis significantly outperformed both the automated PET metrics as well as the prognosis by the treating hematologist (IPI)

Discussion

Understanding of DLBCL and the outcome diversity has evolved substantially over the last decade, in particular with the recognition of distinct and prognostic molecular genetic subtypes of DLBCL [13, 17], although none of these classifications have well-established influence on treatment decisions [18–21]. In parallel, the use of staging 18F-FDG PET/CT has enabled a more precise estimate of total tumor volume, one of the historically most important prognostic factors in lymphoma usually determined indirectly by the Ann Arbor stage and/or LDH level. Not surprisingly, most [14, 22–24] but not all [25] studies have demonstrated tumor burden as an adverse prognostic factor regardless of how the lymphoma lesions are delineated. This has led to increased interest in using automated measurements of pre-therapy MTV to

Table 3 Pearson correlation between visually assessed eMTV and semi-automated measurements

eMTV—Pearson correlation

	Reviewer 1	Reviewer 2	Reviewer 3	MTV$^{2.5}$	MTV^{41}	TLG$^{2.5}$	TLG^{41}
Reviewer 1	1.00	0.66	0.66	0.90	0.82	0.82	0.72
Reviewer 2	0.66	1.00	0.69	0.80	0.74	0.89	0.84
Reviewer 3	0.66	0.69	1.00	0.68	0.68	0.77	0.76
MTV$^{2.5}$	0.90	0.80	0.68	1.00	0.92	0.89	0.79
MTV^{41}	0.82	0.74	0.68	0.92	1.00	0.80	0.81
TLG$^{2.5}$	0.82	0.89	0.77	0.89	0.80	1.00	0.95
TLG^{41}	0.72	0.84	0.76	0.79	0.81	0.95	1.00

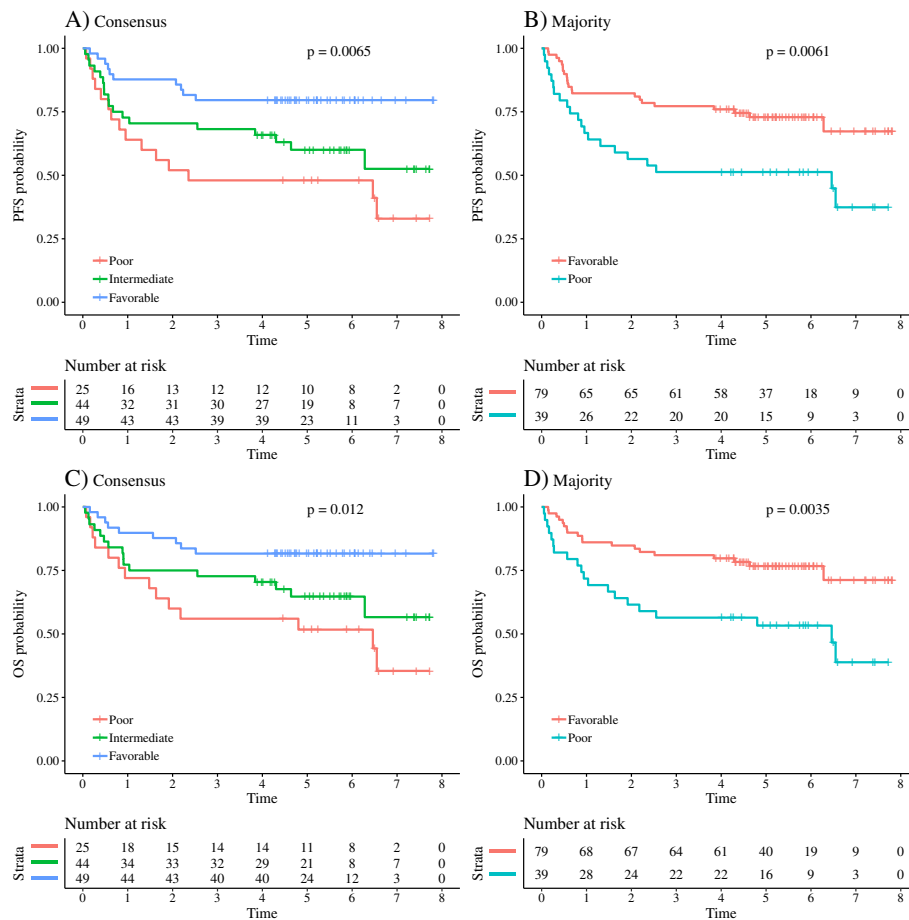Gormsen et al. EJNMMI Research (2019) 9:36

Page 6 of 8

Fig. 3 Progression-free survival (PFS) and overall survival (OS) using the visual assessment of prognosis by expert readers. **a**, **c** The consensus diagnosis of the three readers with an intermediate group reflecting disagreement between reviewers. **b**, **d** Prognosis based on majority decision. As seen, both consensus and majority decisions were able to significantly identify high- and low-risk patients

guide treatment decisions, occasionally in combination with already established response assessment criteria [14].

There are several examples of successful automated scoring systems to report disease severity in the field of pathology and radiology, e.g., Ki-67 index [26], brain atrophy by MRI [27], breast density by mammography [28]. In general nuclear medicine, automatic scoring of stress-induced myocardial ischemia by myocardial perfusion imaging has been shown to perform as well as visual interpretation [29, 30]. However, in order to replace expert analysis in routine practice, an automated technique should either outperform expert analysis or perform as well but with less human effort to avoid reader fatigue. Currently, full measurement of lymphoma MTV has some drawbacks that limit clinical implementation: First, MTV measurements are time-consuming using available software solutions. Second, consensus regarding the best method for lymphoma lesion delineation has not been reached. Third, it remains to be shown that MTV measurements in DLBCL add prognostic information

over and above what can be determined from simple visual image assessment and commonly used clinical risk scores.

We used Hermes TumorFinder to evaluate the two most commonly used lesion delineation algorithms: one based on a fixed threshold of SUV2.5 ($MTV^{2.5}$) [14] and one using a 41% of SUVmax isocontour (MTV^{41}) [31]. Both methods require manual adjustments of segmentations provided by the software (e.g., cropping out of the kidneys, ureters, bladder, liver, heart, and brain), which took 10–15 min on average per patient, similar to what was reported in a previous study [32]. As detailed in Fig. 2, $MTV^{2.5}$ performed significantly better than MTV^{41} in predicting PFS and OS, although we were unable to separate high- and low-risk groups as well as previously reported [14, 22–24]. This discrepancy may well be explained by the ROC-based $MTV^{2.5}$ cut-off for high tumor burden, which in our cohort was approximately 550 ml. Others have reported markedly lower cut-off values of ~ 400 ml [32] and 220 ml [24]. Importantly, our cohort was relatively small and weak

Gormsen et al. EJNMMI Research   (2019) 9:36

Page 7 of 8

prognostic factors may have been missed due to the risk of type II errors.

Our expert readers outperformed the automated prediction of outcome when using all available information from the PET/CT scans including visually estimated MTV, heterogenous lesion pattern, and extra-nodal involvement in a binary score (good/poor prognosis). Moreover, they did so in less than a minute per patient and therefore with far less effort and without the need for additional software packages than a standard PET/CT viewer. Interestingly, the three-decade old IPI classification based on simple clinical risk factors was as prognostic as the automated MTV, although not on par with the binary assessment of prognosis provided by the expert readers (Fig. 2).

The present study also points out the flaws of visual image interpretation, in particular inter-observer variability. When asked to grade metabolic tumor volume on a VAS, agreement among reviewers was poor (Fleiss' kappa 0.28) and the consensus MTV^{vas} could not satisfyingly stratify patients into high- or low-risk groups. Our reviewers were also unable to estimate absolute MTV in milliliters (eMTV) to a level that reflected true measured levels. In addition, although individual reviewers were able to significantly prognosticate outcome on a dichotomous scale (poor/favorable prognosis), agreement among reviewers was only moderate (Fleiss kappa 0.47). In other words, our reviewers did not uniformly identify the same patients as having a good or poor prognosis.

Conclusions

In summary, automated MTV measurements in DLBCL are time-consuming and may not add prognostic information over and above what is delivered in an unsupervised expert analysis, which on the other hand is prone to inter-reader variability and limited generalizability. Importantly, this study is not intended to challenge, devaluate, or oppose the important progress made in the use of quantitative PET/CT in lymphoma. However, it highlights the fact that any prognostic information based on automated measures should be held up against the information obtained in an unsupervised clinical judgment based on physician experience. In fact, a general reminder that algorithms do not always outperform good clinical judgment and that their value should be tested against the unmeasurable knowledge of trained clinical experts.

Abbreviations
DLBCL: Diffuse large B-cell lymphoma; ECOG: Eastern Cooperative Oncology Group; eMTV: Visually estimated MTV (ml); ICC: Intra-class correlation; IPI: International prognostic index; MRI: Magnetic resonance imaging; MTV: Metabolic tumor volume; MTV$^{2.5}$: Semi-automated MTV based on SUV 2.5 threshold; MTV^{41}: Semi-automated MTV based on 41% of SUV max threshold; MTV^{vas}: Visually assessed tumor volume (0–9); OS: Overall Survival; PFS: Progression-free survival; R-CHOP: Rituximab, cyclophosphamide, doxorubicin hydrochloride (hydroxydaunorubicin), vincristine sulfate, and prednisone; ROC: Receiver operating characteristics; TLG: Tumor lesion glycolysis; TLG$^{2.5}$: Semi-automated TLG based on SUV 2.5 threshold; TLG^{41}: Semi-automated TLG based on 41% of SUV max threshold

Availability of data and materials
Please contact the authors for data request.

Authors' contributions
LCG contributed to the design of the study, data acquisition, data analysis and interpretation, and drafted the manuscript. MHV contributed to the data acquisition, interpretation, and critically revised the manuscript. MAP contributed to the data acquisition, interpretation, and critically revised the manuscript. AH contributed to the data acquisition, interpretation, and critically revised the manuscript. KH contributed to the data acquisition, interpretation, and critically revised the manuscript. TVB contributed to the data acquisition, interpretation, and critically revised the manuscript. LJP contributed to the data acquisition, interpretation, and critically revised the manuscript. KJJ contributed to the design of the study and critically revised the manuscript. RB contributed to the design of the study, data analysis and interpretation, and critically revised the manuscript. TCG contributed to the design of the study, data analysis, interpretation, and finalized the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Due to the retrospective nature of this study, the need for individual informed consent was waived. The Danish Ministry of Health approved the study (ref 3-3013-860/1/). The Danish Data Protection agency (ref 1-16-02-88-15) approved data collection.

Consent for publication
Not applicable.

Competing interests
L.J.P has received grants from The Obel Family Foundation, he has received speaker's fee from Bayer, Janssen, Astellas, Sanofi-Genzyme, and Janssen Pharma, and he has received consultancy fees from KLIFO Drug Development Council, Copenhagen, and ljpmedical, Aalborg, Denmark. The other authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Nuclear Medicine & PET Centre, Aarhus University Hospital, DK 8000 Aarhus C, Aarhus, Denmark. [2]Department of Hematology, Aalborg University Hospital, Aalborg, Denmark. [3]Department of Nuclear Medicine, Clinical Cancer Research Center, Aalborg University Hospital, Aalborg, Denmark. [4]Department of Hematology, Odense University Hospital, Odense, Denmark. [5]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark. [6]PET Centre, University Hospital of North Norway, Tromso, Norway. [7]Department of Biomedicine, Aarhus University, Aarhus, Denmark.

References
1. Ziepert M, Hasenclever D, Kuhnt E, Glass B, Schmitz N, Pfreundschuh M, et al. Standard international prognostic index remains a valid predictor of

Gormsen *et al. EJNMMI Research*   (2019) 9:36

Page 8 of 8

outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. J Clin Oncol. 2010;28:2373–80. https://doi.org/10.1200/JCO.2009.26.2493.

2. Gisselbrecht C, Glass B, Mounier N, Singh Gill D, Linch DC, Trneny M, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. J Clin Oncol. 2010;28:4184–90. https://doi.org/10.1200/JCO.2010.28.1618.

3. Crump M, Neelapu SS, Farooq U, Van Den Neste E, Kuruvilla J, Westin J, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. Blood. 2017;130:1800–8. https://doi.org/10.1182/blood-2017-03-769620.

4. Cortelazzo S, Tarella C, Gianni AM, Ladetto M, Barbui AM, Rossi A, et al. Randomized trial comparing R-CHOP versus high-dose sequential chemotherapy in high-risk patients with diffuse large B-cell lymphomas. J Clin Oncol. 2016;34:4015–22. https://doi.org/10.1200/JCO.2016.67.2980.

5. Chiappella A, Martelli M, Angelucci E, Brusamolino E, Evangelista A, Carella AM, et al. Rituximab-dose-dense chemotherapy with or without high-dose chemotherapy plus autologous stem-cell transplantation in high-risk diffuse large B-cell lymphoma (DLCL04): final results of a multicentre, open-label, randomised, controlled, phase 3 study. Lancet Oncol. 2017;18:1076–88. https://doi.org/10.1016/S1470-2045(17)30444-8.

6. Jaeger U, Trneny M, Melzer H, Praxmarer M, Nawarawong W, Ben Yehuda D, et al. Rituximab maintenance for patients with aggressive B-cell lymphoma in first remission: results of the randomized NHL13 trial. Haematologica. 2015;100:955–63. https://doi.org/10.3324/haematol.2015.125344.

7. Murawski N, Pfreundschuh M, Zeynalova S, Poeschel V, Hanel M, Held G, et al. Optimization of rituximab for the treatment of DLBCL (I): dose-dense rituximab in the DENSE-R-CHOP-14 trial of the DSHNHL. Ann Oncol. 2014;25:1800–6. https://doi.org/10.1093/annonc/mdu208.

8. Wilson WH, Sin-Ho J, Pitcher BN, Hsi ED, Friedberg J, Cheson B, et al. Phase III randomized study of R-CHOP versus DA-EPOCH-R and molecular analysis of untreated diffuse large B-cell lymphoma: CALGB/Alliance 50303. Blood. 2016;128.

9. Fisher RI, Gaynor ER, Dahlberg S, Oken MM, Grogan TM, Mize EM, et al. Comparison of a standard regimen (CHOP) with three intensive chemotherapy regimens for advanced non-Hodgkin's lymphoma. N Engl J Med. 1993;328:1002–6. https://doi.org/10.1056/NEJM199304083281404.

10. Lugtenburg PJ, Brown PDN, van der Holt B, D'Amore F, Koene HR, Berenschot HW, et al. Randomized phase III study on the effect of early intensification of rituximab in combination with 2-weekly CHOP chemotherapy followed by rituximab or no maintenance in patients with diffuse large B-cell lymphoma: results from a HOVON-Nordic lymphoma group study. J Clin Oncol. 2016;34. https://doi.org/10.1200/JCO.2016.34.15_suppl.7504.

11. Duhrsen U, Muller S, Hertenstein B, Thomssen H, Kotzerke J, Mesters R, et al. Positron emission tomography-guided therapy of aggressive non-hodgkin lymphomas (PETAL): a multicenter, randomized phase III trial. J Clin Oncol. 2018. https://doi.org/10.1200/JCO.2017.76.8093.

12. Andre MPE, Girinsky T, Federico M, Reman O, Fortpied C, Gotti M, et al. Early positron emission tomography response-adapted treatment in stage I and II Hodgkin lymphoma: final results of the randomized EORTC/LYSA/FIL H10 trial. J Clin Oncol. 2017;35:1786–94. https://doi.org/10.1200/JCO.2016.68.6394.

13. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. N Engl J Med. 2018;378:1396–407. https://doi.org/10.1056/NEJMoa1801445.

14. Mikhaeel NG, Smith D, Dunn JT, Phillips M, Moller H, Fields PA, 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;43:1209–19. https://doi.org/10.1007/s00259-016-3315-7.

15. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am College Radiol. 2018;15:504–8. https://doi.org/10.1016/j.jacr.2017.12.026.

16. Berry KJ, Johnston JE, Mielke PW Jr. Weighted kappa for multiple raters. Percept Mot Skills. 2008;107:837–48. https://doi.org/10.2466/pms.107.3.837-848.

17. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med. 2018;24:679–90. https://doi.org/10.1038/s41591-018-0016-8.

18. Staiger AM, Ziepert M, Horn H, Scott DW, Barth TFE, Bernd HW, et al. Clinical impact of the cell-of-origin classification and the MYC/ BCL2 dual expresser status in diffuse large B-cell lymphoma treated within prospective clinical trials of the German high-grade non-Hodgkin's lymphoma study group. J Clin Oncol. 2017;35:2515–26. https://doi.org/10.1200/JCO.2016.70.3660.

19. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127:2375–90. https://doi.org/10.1182/blood-2016-01-643569.

20. Ninan MJ, Wadhwa PD, Gupta P. Prognostication of diffuse large B-cell lymphoma in the rituximab era. Leuk Lymphoma. 2011;52:360–73. https://doi.org/10.3109/10428194.2010.543716.

21. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002;346:1937–47. https://doi.org/10.1056/NEJMoa012914.

22. Sasanelli M, Meignan M, Haioun C, Berriolo-Riedinger A, Casasnovas RO, Biggi A, 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;41:2017–22. https://doi.org/10.1007/s00259-014-2822-7.

23. Chang CC, Cho SF, Chuang YW, Lin CY, Chang SM, Hsu WL, et al. Prognostic significance of total metabolic tumor volume on (18)F-fluorodeoxyglucose positron emission tomography/ computed tomography in patients with diffuse large B-cell lymphoma receiving rituximab-containing chemotherapy. Oncotarget. 2017;8:99587–600. https://doi.org/10.18632/oncotarget.20447.

24. Song MK, Chung JS, Shin HJ, Lee SM, Lee SE, Lee HS, et al. Clinical significance of metabolic tumor volume by PET/CT in stages II and III of diffuse large B cell lymphoma without extranodal site involvement. Ann Hematol. 2012;91:697–703. https://doi.org/10.1007/s00277-011-1357-2.

25. Adams HJ, de Klerk JM, Fijnheer R, Heggelman BG, Dubois SV, Nievelstein RA, et al. Prognostic superiority of the National Comprehensive Cancer Network International Prognostic Index over pretreatment whole-body volumetric-metabolic FDG-PET/CT metrics in diffuse large B-cell lymphoma. Eur J Haematol. 2015;94:532–9. https://doi.org/10.1111/ejh.12467.

26. Mohammed ZM, McMillan DC, Elsberger B, Going JJ, Orange C, Mallon E, et al. Comparison of visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. Br J Cancer. 2012;106:383–8. https://doi.org/10.1038/bjc.2011.569.

27. Ross DE, Ochs AL, Seabaugh JM, Shrader CR. Alzheimer's disease neuroimaging I. man versus machine: comparison of radiologists' interpretations and NeuroQuant(R) volumetric analyses of brain MRIs in patients with traumatic brain injury. J Neuropsychiatry Clin Neurosci. 2013;25:32–9. https://doi.org/10.1176/appi.neuropsych.11120377.

28. Brandt KR, Scott CG, Ma L, Mahmoudzadeh AP, Jensen MR, Whaley DH, et al. Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. Radiol. 2016;279:710–9. https://doi.org/10.1148/radiol.2015151261.

29. Arsanjani R, Xu Y, Hayes SW, Fish M, Lemley M Jr, Gerlach J, et al. Comparison of fully automated computer analysis and visual scoring for detection of coronary artery disease from myocardial perfusion SPECT in a large population. J Nuclear Med. 2013;54:221–8. https://doi.org/10.2967/jnumed.112.108969.

30. Driessen RS, Raijmakers PG, Danad I, Stuijfzand WJ, Schumacher SP, Leipsic JA, et al. Automated SPECT analysis compared with expert visual scoring for the detection of FFR-defined coronary artery disease. Eur J Nucl Med Mol Imaging. 2018;45:1091–100. https://doi.org/10.1007/s00259-018-3951-1.

31. Meignan M, Sasanelli M, Casasnovas RO, Luminari S, Fioroni F, Coriani C, et al. Metabolic tumour volumes measured at staging in lymphoma: methodological evaluation on phantom experiments and patients. Eur J Nucl Med Mol Imaging. 2014;41:1113–22. https://doi.org/10.1007/s00259-014-2705-y.

32. Ilyas H, Mikhaeel NG, Dunn JT, Rahman F, Moller H, Smith D, 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. https://doi.org/10.1007/s00259-018-3953-z.