18F-FDG PET as biomarker in aggressive lymphoma; technical and clinical validation

Burggraaff, Corry Neli

2022

document version
Publisher's PDF, also known as Version of record

^{18}F-FDG PET as biomarker in aggressive lymphoma; technical and clinical validation

Coreline Burggraaff

^{18}F-FDG PET as biomarker in aggressive lymphoma; technical and clinical validation

Coreline Burggraaff

2022

^{18}F-FDG PET as biomarker in aggressive lymphoma; technical and clinical validation

ISBN/EAN: 978-90-832556-5-1

Cover design and lay-out: Ferdinand van Nispen tot Pannerden
Citroenvlinder DTP & Vormgeving, *my*-thesis.nl

Printed by: Proefschriftenprinten.nl, Ede, The Netherlands

Financial support for printing of this thesis was kindly provided and supported by Cancer Center Amsterdam.

VRIJE UNIVERSITEIT

^{18}F-FDG PET as biomarker in aggressive lymphoma; technical and clinical validation

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op vrijdag 9 september 2022 om 9.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door
Corry Neli Burggraaff
geboren te Woerden

promotoren:	prof.dr. J.M. Zijlstra-Baalbergen prof.dr.ir. H.C.W. de Vet
copromotoren:	prof.dr. R. Boellaard prof.dr. O.S. Hoekstra
promotiecommissie:	prof.dr. S. Stroobants prof.dr. S. Zweegman prof.dr. Y.M. Smulders prof.dr. L. Hooft prof.dr. C. Kobe prof.dr. N.G. Mikhaeel

Contents

CHAPTER 1

General introduction

Introduction lymphoma

Lymphoma is a malignant proliferative disease of lymphoid tissue. Historically, lymphoma has been divided into 2 types; Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL). HL was first described in a report by Thomas Hodgkin in 1832. The disease is characterized by the presence of Reed-Sternberg (RS) cells in pathological specimens. The RS cells originate from B lymphocytes and were later named after the pathologists Dorothy Reed Mendenhall and Carl Sternberg who discovered and described these multinuclear cells in their publications of 1902 and 1898, respectively. Young adults (<40 years) are most often affected with HL and the chance of curing HL patients is high (survival above 90%). However, survival rates for older patients (>60 years) are generally lower (median survival between 57% and 70%) [1].

NHL consists of a diverse group of lymphoma subtypes derived from (progenitors of) B-cells, T-cells or NK-cells [2]. The entity has a wide clinical spectrum from (very) indolent subtypes until highly aggressive subtypes. Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of NHL, consisting of about 40% of all NHL, and has an aggressive clinical behavior [3]. DLBCL is the most prevalent hematologic malignancy, with about 1500 new diagnoses per year in the Netherlands [4]. The median age at diagnosis is 70 years at population level [3], but in recent randomized clinical trials the average age ranges between 60 and 65 years due to the study eligibility criteria [5-7]. Patients with DLBCL have a heterogeneous clinical presentation and prognosis.

Most DLBCL patients are treated with a chemotherapy combination of cyclophosphamide, doxorubicin, vincristine and prednisone (CHOP) [8] as first-line therapy. This chemotherapy scheme was introduced in the seventies and given in 3-week intervals. The therapeutic scheme changed after the introduction of the monoclonal anti-CD20 antibody rituximab [9]. This antibody targets the CD20 cell surface protein of mature B-cells (present in most B-cell lymphomas) and leads to apoptosis induction and cell death by different mechanisms (direct signalling, complement-mediated cytotoxicity and antibody-dependent cellular cytotoxicity). Survival of DLBCL has improved clearly after addition of rituximab to the CHOP scheme (R-CHOP) [10,11]. Current 5-year overall survival is 78% for patients <65, 64% for patients between 65 and 75 and 46% for patients above

75 years old in the Netherlands in population based data [3]. Patients with a relapse or progression after first-line therapy with R-CHOP often have a poor response to second-line treatment [12-14]. Theoretically, it might be relevant to early identify the non-responders to maximize their chances of a successful second-line treatment and minimize side-effects of less effective first-line therapy.

Positron emission tomography

Positron emission tomography (PET) combined with computed tomography (CT) scan (together PET/CT) was announced as medical invention of the year 2000 by Time Magazine. PET/CT is a non-invasive imaging technique that provides visual and quantitative information on physiological and pathological processes in the body. A radioactive tracer is needed to visualize and quantify a specific process. ^{18}F-fluoro-2-deoxy-D-glucose ([^{18}F]FDG) is a radiolabelled glucose analogue and this tracer is nowadays widely used in the clinic for several malignancies known to show high glucose metabolism, resulting in high [^{18}F]FDG uptake [15]. These areas with high [^{18}F]FDG uptake can be either physiological tissues using glucose (e.g. brain), tissues involved in the elimination process of [^{18}F]FDG (e.g. kidney and bladder) and tissues with pathologically increased use of glucose (e.g. malignant tissues such as lymphoma) [16]. The combination with a CT scan is needed for attenuation and scatter correction and adds information about the anatomical location of the increased [^{18}F]FDG uptake. Nowadays [^{18}F]FDG PET/CT is used for both staging (i.e. to assess the extent of a certain malignancy) before treatment and response assessment after treatment.

The role of PET in lymphoma

Staging

Staging of lymphoma was originally based on symptoms, physical examination, radiological studies, laboratory tests of urine and blood and initial biopsy results. Acknowledging the importance of reproducible results from one center to another, the Ann Arbor classification system, consisting of 4 clinical stages, was developed in 1971 for HL [17], but later also adopted for NHL. During the Cotswolds meeting in 1988 the staging system was modified with the addition

of CT for evaluation of intrathoracic and infradiafragmatic lymph nodes [18]. The use of baseline [^{18}F]FDG PET/CT led to a higher sensitivity especially for extranodal disease compared to CT (e.g. bone marrow, liver/spleen involvement) [19]. The international harmonization project (IHP) criteria strongly encouraged the use of baseline [^{18}F]FDG PET when [^{18}F]FDG PET is used for response assessment [20]. Since the Lugano classification guidelines in 2014, baseline [^{18}F]FDG PET has a firm role for the staging of all [^{18}F]FDG avid lymphomas such as HL and DLBCL according to a modified Ann Arbor staging system [16,21,22]. The Lugano classification guidelines contain recommendations for the initial evaluation, staging and response evaluation for both HL and NHL.

End-of-treatment response assessment

The use of [^{18}F]FDG PET for response assessment after treatment in lymphoma led to the revision of the International Working Group criteria [23] by the IHP imaging committee [20]. The IHP criteria were the first criteria for lymphoma patients based on the visual interpretation of [^{18}F]FDG PET scans that were recommended in clinical guidelines [20]. These criteria dichotomized end-of-treatment [^{18}F]FDG PET results into positive and negative based on an assessment of [^{18}F]FDG uptake in the tumor compared to the mediastinal blood pool activity or the surrounding normal background in lesions smaller than 2 cm [20]. Nowadays, [^{18}F]FDG PET/CT is the standard for the end-of-treatment response assessment in DLBCL and HL and should be assessed according to the Lugano classification guidelines [21,22].

Interim response assessment by [^{18}F]FDG PET

In the search for an early predictor of outcome to distinguish responders from non-responders, interim [^{18}F]FDG PET response assessment has been identified as a very promising tool. The hypothesis is that response to R-CHOP treatment can be predicted using interim [^{18}F]FDG PET, i.e. PET during treatment. Conceptually, it provides an excellent basis for 'personalized medicine', where DLBCL patients without sufficient response (i.e. positive interim [^{18}F]FDG PET scan) may be shifted early during R-CHOP treatment to another potentially curative therapy. In this way effectiveness of therapy is maximized, while unnecessary delays, toxicity and costs are minimized. For HL a PET guided approach is nowadays common practice in the Netherlands and recommended internationally (if PET is available), with an interim PET after 2 cycles of treatment [24,25].

The value of [^{18}F]FDG PET as a tool for interim response assessment in DLBCL is still highly debated. Observational studies have indicated that interim [^{18}F]FDG PET may be predictive, but the results reveal inconsistencies and clinical heterogeneity [26]. It is unclear to which extent these inconsistencies are due to differences in the timing of PET during therapy and/or different PET positivity criteria or clinical heterogeneity resulting from studies including patients with different prognostic characteristics. So far these inconsistencies and heterogeneity preclude the standard use of interim [^{18}F]FDG PET for DLBCL in daily clinical practice.

PET response criteria for interim and end-of-treatment [^{18}F]FDG PET

In literature 2 types of [^{18}F]FDG PET response assessment are currently used in lymphoma:

1. Visual (qualitative) methods.
2. SUV-based (standardized uptake value, semi-quantitative) methods.

Visual methods

In clinical practice [^{18}F]FDG PET scans are typically interpreted visually. The question arose whether the IHP criteria for end-of-treatment response assessment [20] could also be used for interim [^{18}F]FDG PET response assessment. It was hypothesized that a more liberal [27] or higher cut-off (liver instead of mediastinal blood pool) is needed for this earlier response assessment [28]. A systematic review from 2009 reported that various definitions for positive and negative diagnostic criteria were used for interim [^{18}F]FDG PET [26]. The need for uniform and flexible criteria led to the introduction of a 5-point scale. This system was originally developed in London [29,30], but later called Deauville criteria [31,32], because of the adoption of these criteria during the first lymphoma consensus workshop in Deauville. The Deauville criteria (**Table 1**) use the mediastinal blood pool and liver as reference for tumor uptake, and are the recommended response evaluation criteria for both interim- and end-of-treatment [^{18}F]FDG PET in the Lugano classification guidelines [22,23].

Table 1. Deauville 5-point scale (based on [^{18}F]FDG avidity during and after treatment)

Deauville score	Interpretation*
1	No uptake
2	Uptake ≤ mediastinum
3	Uptake > mediastinum but ≤ liver
4	Uptake moderately^ higher than liver
5	Uptake markedly^ higher than liver and/or new lesions
X	New areas of uptake unlikely to be related to lymphoma

*The Deauville 5-point scale scores the most intense uptake in a site of initial disease
^The consensus guideline suggests that Deauville score 4 should be applied to uptake > the maximum SUV in a large region of normal liver and Deauville score 5 to uptake 2x to 3x > the maximum SUV in the liver [22].

Semi-quantitative methods

Standardized uptake value (SUV) quantifies the level of [^{18}F]FDG uptake in a lesion normalized for the injected [^{18}F]FDG activity and volume of distribution (e.g. body weight). With an [^{18}F]FDG PET scan being intrinsically a quantitative imaging method, it seems logical to assess therapy response by determination of the SUV and change of SUV (ΔSUV). To date most studies apply the SUVmax metrics (i.e. reflecting the [^{18}F]FDG uptake in the voxel with the highest [^{18}F]FDG uptake in a lesion). For lymphomas a rapid drop in SUVmax is common, reported cut-offs for a clinically relevant interim [^{18}F]FDG PET response assessment in DLBCL ranged from 66% to 73% [33-35]. In a retrospective validation study in 114 DLBCL patients it was concluded that both Deauville and ΔSUVmax criteria for interim [^{18}F]FDG PET assessment after 2 cycles of chemotherapy were valid for progression-free survival outcome prediction [36]. There was a better performance and interobserver reproducibility for ΔSUVmax, however the requirement of a baseline [^{18}F]FDG PET is obvious to allow a ΔSUVmax analysis [36]. Residual questions are which criteria best predict the 2-years progression-free survival and whether these criteria can be validated in other DLBCL cohorts.

Timing of interim [^{18}F]FDG PET

At any observation time during first-line chemotherapy, [^{18}F]FDG uptake reflects a dynamic metabolic state, that is a balance between tumor growth, death of chemosensitive tumor components and later regrowth of chemoresistant

components. When interim [^{18}F]FDG PET is performed after 1 or 2 treatment cycles, it evaluates the response of the cells with the highest mitotic index, thereby providing an early evaluation of chemosensitivity. On the other hand, after 3 or 4 cycles of therapy, the [^{18}F]FDG uptake of interim [^{18}F]FDG PET is more dependent upon the tumor regrowth and inflammatory response (e.g. by macrophages). At this time point, interim [^{18}F]FDG PET is also able to identify DLBCL patients with slow responding disease to R-CHOP treatment [37]. In studies that investigated the predictive value of interim [^{18}F]FDG PET, timing of interim [^{18}F]FDG PET varies considerably, both between and within studies, ranging from interim [^{18}F]FDG PET being performed after 1 to 4 treatment cycles [26]. The timing of interim [^{18}F]FDG PET may affect the visual and ΔSUVmax cut-off that should be used. Studies were consistent with a 66% threshold for interim [^{18}F]FDG PET after 2 cycles [33,35,36], whereas after 4 cycles the optimal cut-off is found at higher ΔSUVmax values, with reported thresholds ranging from 70% to 73% [35,38]. These findings illustrate that results of interim [^{18}F]FDG PET scans performed after different treatment cycles cannot be merged a priori, and that timing is an important factor to consider in the search for the most optimal interim [^{18}F]FDG PET assessment methodology and criteria.

Combination with established/new prognostic factors

Clinical prognostic indices

Estimation of prognosis of DLBCL is currently still based on the international prognostic index (IPI) or age-adjusted IPI (aaIPI), introduced in the early nineties [39]. This clinical risk score has several adaptations since then (e.g. R-IPI, NCCN-IPI) [40,41] and it was proven that is has still prognostic value after the introduction of rituximab [42]. However, almost half of the patients fall into the intermediate risk group [40,41], none of these criteria identify a poor prognosis group with survival clearly below 50% and the current criteria do not result into different treatment decisions outside clinical trials [43]. Therefore, the incremental predictive value of baseline and/or interim [^{18}F]FDG PET in addition to clinical prognostic baseline indices is of high interest.

Baseline metabolic tumor volume

More recently several publications report that, apart from the level of maximum [^{18}F]FDG uptake, the so-called metabolic tumor volume (MTV) may also harbor prognostic value in DLBCL [44,45]. In lung cancer it was demonstrated that test-retest variability of metabolic volume was high compared to SUV, thus careful optimization of imaging and delineation method parameters is needed when using metabolic volume as a prognostic parameter [46]. A more recent study in gastro-intestinal malignancies showed high feasibility of 96% and repeatability for MTV measured with a 50% threshold of mean SUV of a sphere of 12mm with highest local intensity and recommended this method for multicenter [^{18}F] FDG PET studies [47]. However, for lymphoma the optimal method in terms of interobserver agreement (fast, robust and reliable) is not yet fully investigated and determined. Therefore, technical standardization and validation is needed before implementation of MTV as prognostic factor in trials and possible future clinical use.

PETRA Consortium

It is clear that no single study is able to collect enough evidence to determine the optimal timing, response criteria and treatment effect of interim [^{18}F]FDG PET. Therefore, collaboration with national and international experts in lymphoma (hematologists) and PET imaging (nuclear medicine specialists) was needed. This idea was pitched by prof.dr. Josée Zijlstra (hematologist), prof.dr. Otto Hoekstra (nuclear medicine specialist), prof.dr.ir. H.C.W. de Vet (epidemiologist) and prof.dr. R. Boellaard (medical physicist) in the international lymphoma conference in Menton in 2011. Experts from HOVON, London and from a large German randomized trial group (PETAL) already intended to join before a formal collaboration was started. Other researchers involved in interim [^{18}F] FDG PET studies in DLBCL were identified by a systematic literature searches and were invited to join the consortium. Our research group successfully set up a collaboration with principal investigators from these major international interim [^{18}F]FDG PET studies in DLBCL by organizing meetings, development of policy documents and protocols, and solving several legal and practical issues. Finally, a shared database was established with imaging and clinical data of these studies, called PETRA [48]. PETRA is an abbreviation of PET Re-Analysis. For the

development and maintenance of the shared PETRA database our research group collaborates with Lygature's TraIT (Translational Research IT) project, currently transitioning into HEALTH-RI infrastructure [49]. This PETRA database consisting of individual patient data enabled us to investigate the above mentioned research questions about the optimal timing and response criteria for interim [^{18}F] FDG PET and to perform necessary steps to validate interim [^{18}F]FDG PET as a biomarker of response in first-line DLBCL treatment using a meta-analysis of individual patient data. A KWF/Alpe d'Huzes grant was obtained for performing this research (VU 2012-5848).

HOVON-84 study

HOVON-84 was the first international randomized phase III clinical trial where an observational interim [^{18}F]FDG PET scan was made after 4 cycles of R-CHOP14. This study included 574 newly diagnosed CD20 positive DLBCL patients (18-80 years) with Ann Arbor stage II-IV. Interim [^{18}F]FDG PET and CT were performed after 4 cycles of standard R-CHOP14 or rituximab intensified R-CHOP14, without treatment modification based on the interim scan. Rituximab intensification did not lead to improvement of outcome in patients with untreated DLBCL [50]. Thus, the HOVON-84 study provides excellent data to determine which interim [^{18}F]FDG PET response criteria perform best in this clinical setting. For this purpose, we included the HOVON-84 study in the PETRA database.

Central PET review

For the HOVON-84 study a central [^{18}F]FDG PET review of all interim and end-of-treatment [^{18}F]FDG PET scans was performed. Nuclear medicine physicians who are members from the HOVON imaging group performed this review [51]. An imaging platform for this study was setup in Keosys (Imagys platform) to perform the review in a similar digital environment [52]. This central PET review procedure allowed our research group to investigate the interobserver agreement of interim and end-of-treatment [^{18}F]FDG PET assessment with the Deauville 5-point scale. With HOVON-84 being the first HOVON study that

has been evaluated with central imaging review, this review process was optimized (regarding ease of use and speed of the viewer, improvements of instruction manuals and clinical record forms) for future HOVON studies.

This thesis is a result of the research questions and PETRA interim [^{18}F]FDG PET project outlined above. **Part I**: what is known about the predictive value of interim [^{18}F]FDG PET in HL and DLBCL patients? **Part II**: what is the interobserver agreement of using the Deauville criteria in interim [^{18}F]FDG PET? **Part III**: is the predictive value of interim [^{18}F]FDG PET also valid in other DLBCL patient cohorts and what interim [^{18}F]FDG PET criteria predict best for progression-free survival?

Part I of this thesis consists of overviews of the current evidence on interim [^{18}F]FDG PET and its role in clinical practice in DLBCL. **Chapter 2** summarizes the evidence why interim [^{18}F]FDG PET adapted therapy is (becoming) a clinical reality for HL. The evidence for DLBCL shows a high negative predictive value of interim [^{18}F]FDG PET but a less favorable positive predictive value. Besides that, switching therapy to currently available salvage therapies may not overcome early treatment resistance in DLBCL. In addition to this narrative review, a systematic review and meta-analysis was performed on the predictive value of visual response criteria used for interim [^{18}F]FDG PET in DLBCL (**Chapter 3**).

Part II is a more methodological part and focusses on important prerequisites for the technical validation of interim [^{18}F]FDG PET. **Chapter 4** describes the interobserver agreement of interim and end-of-treatment [^{18}F]FDG PET assessment with the Deauville 5-point scale by nuclear medicine physicians in the HOVON-84 study. **Chapter 5** is a pilot study on baseline [^{18}F]FDG PET scans focussing on workflow optimization strategies for a fast, robust and reliable (in terms of interobserver agreement) assessment of metabolic tumor volume.

Part III of this thesis comprises the clinical validation of interim [^{18}F]FDG PET in 2 HOVON studies and in an individual patient data meta-analysis of studies included in the PETRA database. **Chapter 6** is divided into 2 parts. Part A details the assessment of interim [^{18}F]FDG PET in the HOVON-84 study and compares the Deauville score and the semi-quantitative ΔSUVmax criteria in terms of predictive value and shows the added value of interim [^{18}F]FDG PET

to baseline clinical characteristics (aaIPI). Part B is the publication of the original HOVON-84 study in which the primary endpoint is complete metabolic response at the end of induction treatment PET. For this main outcome an extensive central review procedure was performed and the study forms the basis for the other chapters in this thesis. **Chapter 7** describes a phase II study of aggressive B-cell lymphoma with MYC rearrangement treated with a combination of R-CHOP and lenalidomide using [^{18}F]FDG PET for interim- and end-of-treatment response assessment with complete metabolic response as the primary endpoint. In this study we investigated the positive- and negative predictive value of interim [^{18}F] FDG PET for the prediction of end-of-treatment result as a secondary endpoint. The overall aim of our research project was to validate interim [^{18}F]FDG PET as a predictive biomarker of response to first-line therapy in DLBCL patients using a meta-analysis consisting of individual patient data from 1692 patients originally included in 8 international studies (**Chapter 8**).

In **Chapter 9** the results of the studies included in this thesis are summarized and discussed. Finally, the clinical implications are emphasized and future directions are suggested.

References

1. NKR cijfers Hodgkin Lymfoom. *Utrecht: Integraal Kankercentrum Nederland*; 2021. Available from: https://iknl.nl/kankersoorten/hemato-oncologie/nkr-cijfers/hl.
2. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016;127(20):2375-2390.
3. Durmaz M et al. Het diffuus grootcellig B-cellymfoom in Nederland, 2014-2016. Landelijk rapport van het hemato-oncologieregeister van de Nederlandse Kankerregistratie. *Utrecht: Integraal Kankercentrum Nederland*; 2019. Available from: https://www.iknl.nl/getmedia/acf796a4-47f0-4715-bc86-70be4ad5c0da/Hematologische_kankersoorten_landelijk_rapport_iknl_dlbcl_2014_2016.pdf".
4. NKR cijfers Diffuus grootcellig B-cellymfoom en varianten. *Utrecht: Integraal Kankercentrum Nederland*; 2021. Available from: https://iknl.nl/nkr-cijfers?fs%7Cepidemiologie_id=526&fs%7Ctumor_id=400&fs%7Cregio_id=550&fs%7Cperiode_id=564%2C565%2C566%2C567%2C568%2C569%2C570%2C571%2C572%2C573%2C574%2C575%2C576%2C577%2C578%2C579%2C580%2C581%2C582%2C583%2C584%2C585%2C586%2C587%-2C588%2C589%2C590%2C591%2C592%2C593%2C563%2C562%2C561&fs%7Cgeslacht_id=644&fs%7Cleeftijdsgroep_id=677&fs%7Cjaren_na_diagnose_id=687&fs%7Ceenheid_id=703&cs%7Ctype=line&cs%7CxAxis=periode_id&cs%7Cseries=epidemiologie_id&ts%7CrowDimensions=periode_id&ts%7CcolumnDimensions=&lang%7Clanguage=nl
5. 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(9880):1817-1826.
6. Vitolo U, Trněný M, Belada D, et al. Obinutuzumab or Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Previously Untreated Diffuse Large B-Cell Lymphoma. *J Clin Oncol.* 2017;35(31):3529-3537.
7. Dührsen U, Müller S, Hertenstein B, et al. PETAL Trial Investigators. Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL): A Multicenter, Randomized Phase III Trial. *J Clin Oncol.* 2018;36(20):2024-2034.
8. Gottlieb JA, Gutterman JU, McCredie KB, Rodriguez V, Frei E 3rd. Chemotherapy of malignant lymphoma with adriamycin. *Cancer Res.* 1973;33(11):3024-3028.
9. Leget GA, Czuczman MS. Use of rituximab, the new FDA-approved antibody. *Curr Opin Oncol.* 1998;10(6):548-551.
10. Coiffier B, Lepage E, Briere J, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med.* 2002;346(4):235-242.
11. Pfreundschuh M, Trümper L, Osterborg A, et al. MabThera International Trial Group. CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group. *Lancet Oncol.* 2006;7(5):379-391.
12. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28(27):4184-4190.
13. Crump M, Kuruvilla J, Couban S, et al. Randomized comparison of gemcitabine, dexamethasone, and cisplatin versus dexamethasone, cytarabine, and cisplatin chemotherapy before autologous stem-cell transplantation for relapsed and refractory aggressive lymphomas: NCIC-CTG LY.12. *J Clin Oncol.* 2014;32(31):3490-3496.
14. van Imhoff GW, McMillan A, Matasar MJ, et al. Ofatumumab Versus Rituximab Salvage Chemoimmunotherapy in Relapsed or Refractory Diffuse Large B-Cell Lymphoma: The ORCHARRD Study. *J Clin Oncol.* 2017;35(5):544-551.
15. Petersen H, Holdgaard PC, Madsen PH, et al. PET/CT Task Force of the Region of Southern Denmark. FDG PET/CT in cancer: comparison of actual use with literature-based recommendations. *Eur J Nucl Med Mol Imaging.* 2016;43(4):695-706.
16. Weiler-Sagie M, Bushelev O, Epelbaum R, et al. (18)F-FDG avidity in lymphoma readdressed: a study of 766 patients. *J Nucl Med.* 2010;51(1):25-30.

17. Carbone PP, Kaplan HS, Musshoff K, Smithers DW, Tubiana M. Report of the Committee on Hodgkin's Disease Staging Classification. *Cancer Res.* 1971;31(11):1860-1861.
18. Lister TA, Crowther D, Sutcliffe SB, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *J Clin Oncol.* 1989;7(11):1630-1636. Erratum in: *J Clin Oncol.* 1990 Sep;8(9):1602.
19. Schaefer NG, Hany TF, Taverna C, et al. Non-Hodgkin lymphoma and Hodgkin disease: coregistered FDG PET and CT at staging and restaging--do we need contrast-enhanced CT? *Radiology.* 2004;232(3):823-829.
20. Juweid ME, Stroobants S, Hoekstra OS, et al. Imaging Subcommittee of International Harmonization Project in Lymphoma. 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(5):571-578.
21. 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(27):3059-3068.
22. 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(27):3048-3058.
23. Cheson BD, Horning SJ, Coiffier B, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol.* 1999;17(4):1244-1253. Review. Erratum in: *J Clin Oncol.* 2000;18(11):2351.
24. Lymfoomwerkgroep HOVON. Richtlijn Hodgkin Lymfoom bij Volwassenen. Amsterdam: Stichting Hemato-Oncologie voor Volwassenen Nederland (HOVON); 2019. Available from: https://hematologienederland.nl/wp-content/uploads/2020/01/Richtlijn-Hodgkin-lymfoom-HOVON-2019-v-2019.11.13.pdf.
25. Eichenauer DA, Aleman BMP, André M, et al. ESMO Guidelines Committee. Hodgkin lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2018;29(Suppl 4):iv19-iv29.
26. Terasawa T, Lau J, Bardet S, et al. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. *J Clin Oncol.* 2009;27(11):1906-1914.
27. Kostakoglu L. Early prediction of response to therapy: the clinical implications in Hodgkin's and non-Hodgkin's lymphoma. *Eur J Nucl Med Mol Imaging.* 2008;35(8):1413-1420.
28. Itti E, Juweid ME, Haioun C, et al. Improvement of early 18F-FDG PET interpretation in diffuse large B-cell lymphoma: importance of the reference background. *J Nucl Med.* 2010;51(12):1857-1862.
29. 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(4):775-777; Erratum in: *Blood.* 2012;119(22):5340.
30. Barrington SF, Qian W, Somer EJ, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging.* 2010;37(10):1824-1833.
31. Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. *Leuk Lymphoma.* 2009;50(8):1257-1260.
32. Meignan M, Gallamini A, Haioun C, Polliack A. Report on the Second International Workshop on interim positron emission tomography in lymphoma held in Menton, France, 8-9 April 2010. *Leuk Lymphoma.* 2010;51(12):2171-2180.
33. 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(10):1626-1632.
34. 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(4):527-533.

35. 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(1):37-43.
36. 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(9):1312-1320.
37. Meignan M, Itti E, Gallamini A, Haioun C. Interim 18F-fluorodeoxyglucose positron emission tomography in diffuse large B-cell lymphoma: qualitative or quantitative interpretation--where do we stand? *Leuk Lymphoma.* 2009;50(11):1753-1756.
38. 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(4):527-533.
39. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993 Sep;329(14):987-994.
40. Sehn LH, Berry B, Chhanabhai M, et al. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood.* 2007;109(5):1857-1861.
41. Zhou Z, Sehn LH, Rademaker AW, et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood.* 2014;123(6):837-842.
42. Ziepert M, Hasenclever D, Kuhnt E, et al. Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28(14):2373-2380. Erratum in: *J Clin Oncol.* 2011;29(6):779.
43. Ruppert AS, Dixon JG, Salles GA, et al. International prognostic indices in diffuse large B-cell lymphoma (DLBCL): a comparison of IPI, R-IPI and NCCN-IPI. *Blood.* 2020;135(23):2041-2048.
44. Song MK, Chung JS, Shin HJ, 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(5):697-703.
45. 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(11):2017-2022.
46. Cheebsumon P, van Velden FH, Yaqub M, et al. Effects of image characteristics on performance of tumor delineation methods: a test-retest assessment. *J Nucl Med.* 2011;52(10):1550-1558.
47. Frings V, van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. *Radiology*. 2014;273(2):539-548.
48. PETRA consortium. Available from: https://petralymphoma.org/.
49. Lygature, Translation Research IT (TraIT) project. Available from: https://www.lygature.org/data-infrastructure.
50. Lugtenburg PJ, de Nully Brown P, van der Holt B, et al. Rituximab-CHOP with early rituximab intensification for diffuse large B-cell lymphoma: a randomized phase 3 trial of the HOVON and the Nordic Lymphoma Group (HOVON-84). *J Clin Oncol.* 2020;38(29):3377-3387.
51. HOVON imaging werkgroep. Available from: http://www.hovon.nl/werkgroepen/technische-commissies/imaging-werkgroep.html.
52. Keosys medical imaging. Available from: https://www.keosys.com/.

PART I

PET as biomarker of response in lymphoma

CHAPTER 2

FDG-PET as a biomarker for early response in diffuse large B-cell lymphoma as well as in Hodgkin lymphoma? Ready for implementation in clinical practice?

Josée M. Zijlstra
Coreline N. Burggraaff
Marie José Kersten
Sally F. Barrington
On behalf of the EHA Scientific Working Group on Lymphoma

Haematologica. 2016; 101: 1279-1283.

A short history

Major changes have taken place in the staging and response assessment of malignant lymphoma in the last two decades. With the introduction of fluorodeoxyglucose-positron emission tomography (FDG-PET) and positron emission tomography-computed tomography (PET-CT), the criteria for staging and monitoring response have changed dramatically. In the revised Cheson criteria published in 2007 [1], staging with FDG-PET was still optional, and end-of treatment assessment using FDG-PET and CT was obligatory for Hodgkin lymphoma (HL) and diffuse large B-cell lymphoma (DLBCL). In the Lugano criteria published in 2014 [2], PET-CT is recommended for staging as well as response assessment following therapy, as it is the most accurate imaging modality. However, one of the characteristics of (molecular) metabolic imaging is to be able to assess metabolic changes early in treatment. The question arises whether 'interim' FDG-PET-CT (iPET) can be used as a biomarker to differentiate good and poor responders during treatment, in order to modify therapy and to improve outcome. Recent clinical trials have addressed these questions, and we discuss the results and the implications for clinical practice.

Assessment of interim-PET scans

International guidelines recommend the use of a 5-point scale [also called the Deauville score (DS)] for grading FDG-uptake in lymphoma, compared to physiological uptake in the mediastinum and liver, for response assessment in daily practice and clinical trials [2-4]. No FDG uptake is graded as DS 1; uptake less than or equal in intensity to the mediastinum as DS 2; lesions with FDG uptake between mediastinum and liver are assessed as DS 3; uptake more intense than liver is scored as DS 4; and markedly increased uptake or new lymphoma-related lesions as DS 5 (Figure 1). This categorization has a high interobserver agreement in HL and DLBCL [5,6].

However, FDG-PET is also a quantitative imaging technique, allowing semi-quantitative imaging interpretation, using standardized uptake values (SUV). Reporting change of FDG uptake (usually expressed as a relative change) can also be used for interim response assessment. The reliability of the results depends on having comparable procedures for patient preparation and injection, and scanning

and image reconstruction protocols, as well as comparable data analysis. Quality control and quality assurance procedures are also required to maintain the accuracy and precision of quantification.
Recently, the European Association of Nuclear Medicine (EANM) guidelines for FDG-PET in tumor imaging for trials and clinical practice have been up-dated [7], and an accreditation system is available (EARL; *http://earl.eanm.org*). Within clinical studies, these changes in SUV are being compared with visual assessment. Besides SUV, metabolically active tumor volume defined with FDG-PET is being investigated.

Interim-PET in Hodgkin lymphoma

Hodgkin lymphoma is a lymphoma entity with cure rates of up to 90%. iPET predicts response early during treatment and PET-guided therapy is a new strategy in development for HL. The goal of current and recently completed

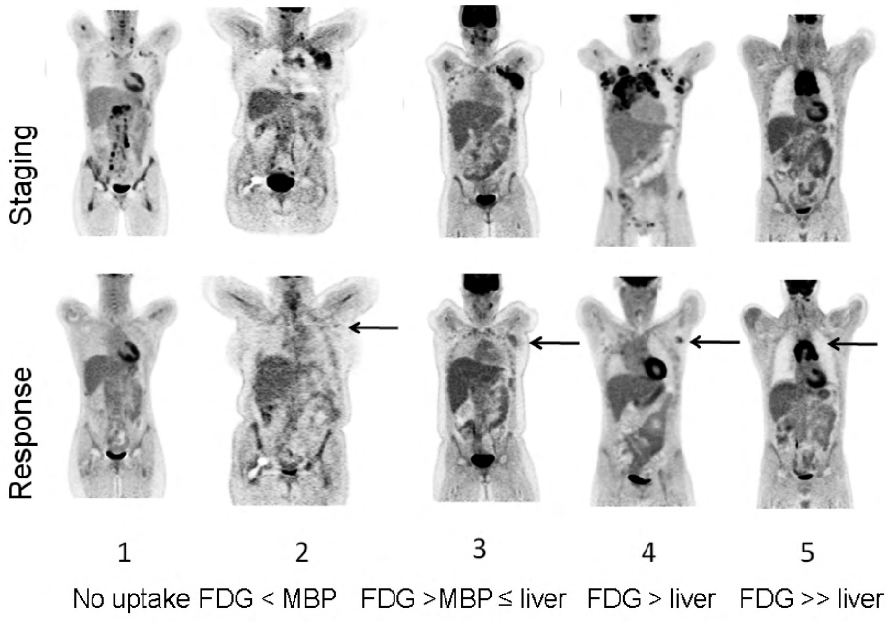


Figure 1. Coronal slices from 5 patients are shown at baseline and response.
The level of uptake at residual sites, where present (arrowed) is graded according to the 5-point Deauville score.

clinical trials is to achieve optimal efficacy in terms of progression-free survival (PFS) and overall survival (OS), and to reduce long-term adverse effects.

The first reports using iPET to de-escalate therapy in responding individuals with early-stage disease have been published. The UK RAPID study [8] and the EORTC H10 study [9] have randomized patients with complete metabolic response (CMR) on iPET after 2-4 cycles of doxorubicin, bleomycin, vinblastine, dacarbazine (ABVD) treatment to receive radiotherapy (RT) or no further treatment (NFT). Both were non-inferiority studies, with a slightly different design. Involved field was used in RAPID and involved node RT in H10. RAPID investigators accepted that by abandoning RT some loss of disease control was inevitable, whereas H10 investigators designed their trial to demonstrate that patients could be spared RT without any compromise in disease control. Both studies demonstrated a modest PFS advantage for patients receiving RT (Table 1).

In the RAPID trial, the 3-year PFS was 97.1% using RT versus 90.8% for NFT in a per-protocol analysis (HR 2.36; 1.13, 4.95). There was no significant difference in 3-year OS: 97.1% (RT) versus 99.0% (NFT). In the H10 study, 1-year PFS was 100% (favorable disease) and 97.3% (unfavorable disease) using RT versus 94.9% (favorable) and 94.7% (unfavorable) for NFT. The H10 study was halted early for patients with CMR as it was felt unlikely to demonstrate non-inferiority for the NFT option with a 10% decrease in 5-year PFS where the threshold for non-inferiority was set at a hazard ratio of respectively 3.2 and 2.1 for the favorable and unfavorable subgroups. Nonetheless, patients had excellent outcomes in both trials whether or not they received RT. However, follow up in both trials is still short, and (late) adverse effects of radiotherapy may become apparent over time [10]. Results from the HD16 and HD17 trials of the German Hodgkin Study Group are currently awaited. Both trials are comparing standard combined modality treatment with a PET-directed regimen, omitting radiotherapy for patients with complete metabolic response after chemotherapy (*www.ghsg.org*).

So de-escalation has become a real option in clinical practice, but requires detailed discussions between patients, hematologists and radiation oncologists. Balancing the risks and benefits of chemotherapy alone versus combined modality treatment

depends on patient age, fitness, disease distribution and, most importantly, the individual assessment of that risk in the decision-making process.

The recently published US Intergroup Trial of response adapted therapy for stage III-IV Hodgkin lymphoma used early interim PET after 2 cycles of ABVD to escalate therapy for patients with Deauville score 4 or 5 to BEACOPP escalated. The authors concluded that response-adapted therapy based on iPET imaging seemed promising with a 2-year PFS of 64% for PET2-positive patients compared to historical series with 2-year PFS of 15%-30% for PET-positive patients treated with ABVD [11].
Unpublished data presented in early and advanced disease from the EORTC H10 and the recently published UK Response Adapted Therapy in Advanced Hodgkin Lymphoma (RATHL) studies [12] also suggest that escalation from ABVD to BEACOPP may be beneficial in patients with an inadequate response on iPET after 2 cycles. In RATHL, patients randomized to receive AVD rather than ABVD on the basis of CMR on iPET had less pulmonary toxicity but no significant difference in 3-year PFS/OS. Published data are awaited for the EORTC H10 trial but in the meantime, at least in centers that participated in RATHL, this strategy is being offered to patients in clinical practice.

The H10 and RAPID trials used the mediastinal blood pool (equivalent to DS 2) as the reference region for CMR; the RATHL study used the liver (DS 3). To avoid under-treatment, it may be desirable to use the mediastinal blood pool in trials testing de-escalation. The RATHL study, which tested both treatment escalation and de-escalation, used DS 3 as a cutoff for CMR. The liver is a more reliable threshold for reporting iPET with respect to inter-reporter agreement and there was good agreement amongst reporters in local PET centers with expert central reviewers in RATHL [4]. This supports the use of DS 3 for assessment of CMR in patients undergoing standard treatment but, in the authors' opinion, in early stage disease for de-escalation it is still prudent to use DS2. It is imperative that those reporting PET results and clinicians understand how the DS should be used for response-adaptation in clinical practice. Nowadays, many imaging specialists are educated in using DS not only for clinical trials, but also for clinical practice.

Table 1. Studies with I-PET adapted therapy in Hodgkin lymphoma and diffuse large B-cell lymphoma.

Author/ Study	Year	Design	Type +stage	Number	i-PET after	Pos criteria	i-PET negative therapy	i-PET positive therapy	Median FUP	Outcome i-PET -/+
HL										
Radford/ RAPID[8]	2015	RCT	st IA/IIA non-bulky HL	571	3x ABVD	DS 3/4/5	IF RT or NFT	1x ABVD + RT	60 mo	3-yr PFS: IF RT: i-t-t: 94.6% p-p a: 97.1% vs NFT: 90.8%. 3-yr OS: IF RT: 97.1% vs NFT: 99.0%
Raemaekers/ EORTC H10[9]	2014	RCT	st I/II supra-diaphragmatic HL	1137	2x ABVD	IHP	Favorable: 2x ABVD or 1x ABVD+INRT Unfavorable: 4x ABVD or 2xABVD+INRT	Favorable: 2x BEACOPP-esc+ INRT Unfavorable: 2x BEACOPP-esc + INRT	1.1 yr	1-yr PFS fav. IN-RT: 100% NFT 94.9%. 1-yr PFS unfav. IN-RT: 97.3% NFT 94.7%
Press/US intergroup S0816[11]	2016	phase II	st III/IV HL	336	2x ABVD	DS 4/5	4x ABVD	6x BEACOPP-esc	39.7 mo	2-yr PFS: 82%/64% sign 2-yr OS: 98%
Johnson/ RATHL[12]	2015	RCT	st II-IV HL	1137	2x ABVD	DS 4/5	4x ABVD or 4x AVD	BEACOPP-14 or BEACOPP-esc	32 mo	3-yr PFS: ABVD: 85.5%; AVD: 84,5% / i-PET pos:68% 3-yr OS: ABVD:97.0%; AVD: 97,5% / i-PET pos: 86%
Straus/ CALGB Alliance 50604[20]	2015	phase II	non-bulky st I/II HL	164	2x ABVD	DS 4/5	2x ABVD	2x BEACOPP-esc + IF RT	2 yr	3-yr PFS: 92%/66% sign
Ganesan[21]	2015	phase II	st IIB/III/IV HL	50	2x ABVD	DS 4/5	2x ABVD	4x BEACOPP-esc	24.7 mo	2-yr EFS: 82%/50% sign
DLBCL										
Hertzberg[22]	2015	phase II	poor risk DLBCL	151	4x R-CHOP14	IHP	2x R-CHOP +2R	3x R-ICE + Z-BEAM ASCT	35 mo	2-yr PFS: 74% /67% NS 2-yr OS: 88%/78% NS
Swinnen/ E3404[23]	2015	phase II	DLBCL st II(bulky)/III/ IV	80	3x R-CHOP	'ECOG criteria'	2x R-CHOP	4th R-CHOP +4x R-ICE	4.6 yr	2-yr PFS: 76% /42% NS 3-yr OS: 93%/69% NS
Stewart [24]	2014	phase II	adv st DLBCL	70	2x R-CHOP21	>Liver at >1 site	4x R-CHOP	1x R-DICEP + R-BEAM ASCT	41 mo	3-yr PFS: 65.2%/52.7% NS 3-yr OS: 68.4%/70.5% NS
Pardal[25]	2014	phase II	DLBCL/ gr 3B FL	71	3x R-Mega CHOP	IHP	3x R-MegaCHOP	2x R-IFE + BEAM ASCT	42.8 mo	3-yr PFS: 81%/57% sign 3-yr OS: 89%/73% NS

Table 1. Continued

Author/ Study	Year	Design	Type +stage	Number	i-PET after	Pos criteria	i-PET negative therapy	i-PET positive therapy	Median FUP	Outcome i-PET -/+
DLBCL										
Dührsen/ PETAL[19]	2014	RCT	aggressive NHL (~80% DLBCL)	853	2x R-CHOP	<66% ΔSUV reduction	4x R-CHOP or 4x R-CHOP+2R	6x R-CHOP or 6x 'Burkitt protocol'	33 mo	2-yr TTTF: 79% i-PET+/47% i-PET- sign.
Sehn[26]	2014	phase II	adv stage DLBCL/ PMBCL	150	4x R-CHOP21	IHP	2x R-CHOP21	4x R-ICE (+RT if end of treatment PET pos)	45 mo	4-yr PFS: 91%/59% sign 4-yr OS: 96%/73% sign
Casasnovas[16]	2011	phase II	DLBCL/ PMBCL	102	2x R-CHOP14 or 2x R-ACVBP	IHP	R-CHOP14 or MTX+ R-ifos-VP-16 +AraC	MTXiv + Z-BEAM ASCT	19 mo	PET 2: 2-yr PFS 73%/77% NS 2-yr OS 93%/ 84% NS PET 4: 2-yr PFS 81%/73% NS 2-yr OS 94%/83% NS
Moskowitz[14]	2010	Pros-pective	adv stage DLBCL	98	4x R-CHOP14	>local bg	3x ICE	biopsy neg: 3x ICE; biopsy pos:2x ICE+ 1x R-ICE+ASCT	44 mo	PFS NS OS NS
Kasamon[27]	2009	phase II	aggressive B-cell lymphoma	59	2 or 3X (R-)CHOP	> bg	(R-)CHOP14 or 21	2x (R-)ESHAP or 2x R-ICE	33.6 mo	2-yr EFS 89%/75% 3-yr EFS: 82%/65%

HL: Hodgkin lymphoma; DLBCL: diffuse large B-cell lymphoma; PMBCL: primary mediastinal large B-cell lymphoma; NHL: non-Hodgkin lymphoma; iPET; interim positron emission tomography; RCT: randomized clinical trial; phase II: prospective phase II study; st: stage; adv: advanced; gr: grade; ABVD: doxorubicin, bleomycin, vinblastine, dacarbazine; (R-)CHOP: (rituximab,) cyclophosphamide, doxorubicin, vincristine, prednisone; R-ACVBP: rituximab, doxorubicin, vindesine, bleomycin, prednisone; DS: Deauville score; IHP: international harmonization project; SUV: standardized uptake value; bg: background; rand: randomization; IF RT: involved field radiotherapy; NFT: no further treatment; INRT: involved node radiotherapy; AVD: doxorubicin, vinblastine and dacarbazine; 2R: 2 cycles rituximab; MTX: methotrexate; R-ifos-VP-16: rituximab, ifosfamide, vindesine; AraC: cytosine arabinoside; RT: radiotherapy; (R-)ICE: (rituximab,) ifosfamide, carboplatin, etoposide; BEACOPP: bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, prednisolone; esc: escalated; BEAM: carmustine, etoposide, cytarabine, melphalan; ASCT: autologous stem cell transplantation; R-DICEP: rituximab, dose intensive cyclophosphamide, etoposide, cisplatin; R-IFE: rituximab, ifosfamide, etoposide; MTXiv: intravenous methotrexate; Z-BEAM: ibritumomab tiuxetan, carmustine, etoposide, cytarabine, melphalan; R-ESHAP: rituximab, etoposide, cisplatin, high-dose cytarabine, methylprednisone; FUP: follow-up; mo: months; yr: years; PFS: progression free survival; i-t-t: intention-to-treat; p-p A: per-protocol analysis; fav.=favorable; unfav=unfavorable; OS: overall survival; EFS: event-free survival; TTTF: time to treatment failure; NS: not significant; Sign: statistically significant; ECOG; Eastern Cooperative Oncology Group.

Interim-PET in diffuse large B-cell lymphoma

R-CHOP is the standard therapy in DLBCL and will cure approximately 60% of patients. Standard treatment for the significant proportion of patients up to the age of 70 years with relapsed or refractory disease is platinum-based immunochemotherapy followed by high-dose chemotherapy and autologous stem cell transplantation (ASCT). However, the results of second-line immunochemotherapy are disappointing, especially for patients who relapse within one year of completing R-CHOP treatment.
Early identification of non-responders is of the utmost importance to maximize the chances of successful second-line therapy and to decrease side-effects associated with ineffective first-line therapy.
To distinguish responders from non-responders, observational studies have indicated that iPET may be an effective predictive biomarker of outcome in DLBCL, but there are inconsistencies [13,14]. It is unclear to what extent these are due to differences in the timing of PET during therapy, the choice of therapy and/or different PET reporting criteria. The current recommendation is to use DS, but earlier studies used International Harmonization Project criteria which separated PET into 'positive' and 'negative' by comparing FDG uptake with the intensity of the blood pool or nearby normal structures, if less than 2 cm, to offset partial volume effects [15].

Standardized uptake value based methods have also been used to assess response in DLBCL. To date, most studies have applied the change in FDG uptake in the pixel with the highest uptake (SUVmax) before and during/after treatment (ΔSUV) [6]. Casasnovas et al. advocate ΔSUV as the most accurate criterion for response assessment. For lymphomas, in which cure is feasible and a rapid drop in SUV is common, cutoffs for a clinically relevant interim assessment of response have been reported to range from 66% to 91% [16]. Finally, metabolic tumor volume at baseline, perhaps combined with iPET response, has recently been reported as demonstrating predictive value [17]. Currently, an international consortium called PETRA (PET-Re-Analyses) is pooling clinical studies in DLBCL to perform an individual patient data meta-analysis and compare different methods in assessing interim-PET [18]. Hopefully, this will reveal the optimal time point and best visual or semi-quantitative PET-metrics to use for interim assessment.

Another important issue is whether early identification of patients who are likely to be refractory to R-CHOP will result in better outcomes if these patients can be salvaged with high-dose chemotherapy or novel non-chemotherapeutic agents. Progress in targeted therapies in DLBCL might shift treatment paradigms from broad-spectrum poly-chemotherapy towards more targeted therapies based on genetic heterogeneity and complexity. These new drugs are currently being tested within phase I-II trials and results are awaited. Predicting response or resistance to a specific therapy will not only expedite the introduction of the most effective therapy to the patient but will also most likely be necessary to reduce the overall costs.

Nowadays, international guidelines do not recommend changing standard treatment on iPET unless there is clear evidence of progression. Nonetheless, if mid-treatment imaging is performed, PET is better than CT at predicting prognosis and can be useful to exclude the possibility of progression. Preliminary published data and data presented only in abstract form suggest that, for patients with inadequate response on iPET, current chemotherapy-based escalation strategies may not overcome treatment resistance [19,23-24] (Table 1). For these patients, a more effective initial therapy regimen is needed.

Conclusions

FDG-PET is a reliable biomarker for assessing early response in HL. The high negative predictive value of CMR after 2-3 cycles of ABVD has been the basis for recent trials exploring de-escalation of therapy in early-stage disease. The high positive predictive value in advanced disease has also been the focus of clinical trials, with promising data presented for patients escalated from ABVD to BEACOPP if they do not achieve a CMR after 2 cycles. In HL, PET-adapted therapy based on early response is rapidly becoming a clinical reality.
In DLBCL, the ability to escalate treatment early for patients unlikely to respond to first-line immunochemotherapy is highly desirable, as these patients do not have good salvage options. Obtaining a CMR on interim PET has a high negative predictive value, but partial metabolic response is also often associated with good outcomes. Modifying treatment for patients who do not achieve an early CMR in DLBCL is likely to lead to overtreatment of a significant proportion of patients,

with associated costs and patient anxiety [28]. Early data suggest that patients with early failure also show treatment resistance with currently available salvage therapies, and novel, more targeted treatment strategies are clearly needed.

Acknowledgement

The authors wish to acknowledge prof.dr. GJ Ossenkoppele and prof.dr. OS Hoekstra for critically reviewing the manuscript.

References

1. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007;25(5):579-586.
2. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and NonHodgkin Lymphoma: The Lugano Classification. *J Clin Oncol.* 2014;32(27):3059-3068.
3. 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(27):3048-3058.
4. Barrington SF, Kirkwood AA, Franceschetto A, et al. PET-CT for staging & early response: results from 'Response Adapted Therapy in Advanced Hodgkin Lymphoma' (RATHL) (CRUK/07/033). *Blood.* 2016;127(12):1531-1538.
5. Biggi A, Gallamini A, Chauvi S, et al. International validation study for interim PET in ABVD-treated, advanced-stage Hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. *J Nucl Med.* 2013;54(5):683-690.
6. 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(9):1312-1320.
7. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42(2):328-354.
8. Radford J, Illidge T, Counsell N, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. *N Engl J Med.* 2015;372(17):1598-1607.
9. Raemaekers JM, André MP, Federico M, et al. Omitting radiotherapy in early positron emission tomography-negative stage I/II Hodgkin lymphoma is associated with an increased risk of early relapse: Clinical results of the preplanned interim analysis of the randomized EORTC/LYSA/FIL H10 trial. *J Clin Oncol.* 2014;32(12):1188-1194.
10. Meyer RM, Gospodarowicz MK, Connors JM, et al. ABVD alone versus radiation-based therapy in limited-stage Hodgkin's lymphoma. *N Engl J Med.* 2012;366(5):399-408.
11. Press OW, Li H, Schöder H, et al. US Intergroup Trial of Response-Adapted Therapy for Stage III to IV Hodgkin Lymphoma Using Early Interim Fluorodeoxyglucose-Positron Emission Tomography Imaging: Southwest Oncology Group S0816. *J Clin Oncol.* 2016;34(17):2020-2027.
12. Johnson PW, Federico M, Kirkwood A, Fossa A, et al. Adapted treatment guided by interim PET-CT scan in advanced Hodgkin's lymphoma. *N Engl J Med.* 2016;374(25):2419-2429.
13. Terasawa T, Lau J, Bardet S, et al. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. *J Clin Oncol.* 2009;27(11):1906-1914.
14. Moskowitz CH, Schöder H, Teruya-Feldstein J, et al. Risk-adapted Dose-Dense Immunochemotherapy Determined by Interim FDG-PET in Advanced-Stage Diffuse Large B-Cell Lymphoma. *J Clin Oncol.* 2010;28(11):1896-1903.
15. 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(5):571-578.
16. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al. SUVmax reduction improves early prognosis value of interim positron emission tomography scan in diffuse large B-cell lymphoma. *Blood.* 2011;118(1):37-43.
17. 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(7):1209-1219.
18. Zijlstra JM, Hoekstra OS, de Vet HCW. Validation of interim PET as a biomarker of response in NHL- a study on PET timing, therapies, response criteria, type of NHL and cost-effectiveness. Menton 2014. Available from: http://www.lymphomapet.com/files/Poster% 20Session%202014.pdf.

19. Dührsen U, Hüttmann A, Müller S, et al. Positron Emission Tomography (PET) Guided Therapy of Aggressive Lymphomas – a Randomized Controlled Trial Comparing Different Treatment Approaches Based on Interim PET Results (PETAL Trial). *Blood.* 2014;124(21):(Abstract 391).
20. Straus DJ, Pitcher B, Kostakoglu L , et al. Initial Results of US Intergroup Trial of Response-Adapted Chemotherapy or Chemotherapy/Radiation Therapy Based on PET for Non-Bulky Stage I and II Hodgkin Lymphoma (HL) (CALGB/Alliance 50604). *Blood.* 2015;126(23):(Abstract 578).
21. Ganesan P, Rajendranath R, Kannan K, et al. Phase II study of interim PET-CT-guided response-adapted therapy in advanced Hodgkin's lymphoma. *Ann Oncol.* 2015;26(6):1170-1174.
22. Hertzberg MS, Gandhi MK, Butcher B, et al. Early Treatment Intensification with R-ICE Chemotherapy Followed By Autologous Stem Cell Transplantation (ASCT) Using Zevalin-BEAM for Patients with Poor Risk Diffuse Large B-Cell Lymphoma (DLBCL) As Identified By Interim PET/CT Scan Performed after Four Cycles of R-CHOP-14: A Multicenter Phase II Study of the Australasian Leukaemia Lymphoma Study Group (ALLG.) *Blood.* 2015;126(23):(Abstract 815).
23. Swinnen LJ, Li H, Quon A, et al. Response-adapted therapy for aggressive non-Hodgkin's lymphomas based on early [18F] FDG-PET scanning: ECOG-ACRIN Cancer Research Group study (E3404). *Br J Haematol.* 2015;170(1):56-65.
24. Stewart DA, Kloiber R, Owen C, et al. Results of a prospective phase II trial evaluating interim positron emission tomography-guided high dose therapy for poor prognosis diffuse large B-cell lymphoma. *Leuk Lymph.* 2014;55(9):2064-2070.
25. Pardal E, Coronado M, Martín A, et al. Intensification treatment based on early FDG-PET in patients with high-risk diffuse large B-cell lymphoma: a phase II GELTAMO trial. *Br J Haematol.* 2014;167(3):327-336.
26. Sehn LH, Hardy ELG, Gill KK, et al. Phase 2 Trial of Interim PET Scan-Tailored Therapy in Patients with Advanced Stage Diffuse Large B-Cell Lymphoma (DLBCL) in British Columbia (BC). *Blood.* 2014;124:392.
27. Kasamon YL, Wahl RL, Ziessmann HA, et al. Phase II study of risk adapted therapy of newly diagnosed, aggressive Non-Hodgkin Lymphoma based on midtreatment FDG-PET scanning. *Biol Blood Marrow Transplant.* 2009;15(2):242-248.
28. Barrington SF, Mikhaeel NG. PET-scans for staging and restaging in diffuse large B-cell and follicular lymphoma. *Curr Hematol Rep.* 2016;11(3):185-195.

CHAPTER 3

Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis

Coreline N. Burggraaff
Otto S. Hoekstra
Rutger A.J. Nievelstein
Martijn W. Heymans
Josée M. Zijlstra
Antoinette de Jong
Nikie J. Hoetjes
Elise P. Jansma
Henrica C.W. de Vet

European Journal of Nuclear Medicine and Molecular Imaging. 2019; 46: 65-79.

Abstract

Purpose

Diffuse large B-cell lymphoma (DLBCL) represents the most common subtype of non-Hodgkin lymphoma. Most relapses occur in the first 2 years after diagnosis. Early response assessment with ^{18}F-fluoro-2-deoxy-D-glucose (^{18}F-FDG) positron emission tomography (PET) may facilitate early change of treatment, thereby preventing ineffective treatment and unnecessary side effects. We aimed to assess the predictive value of visually-assessed interim ^{18}F-FDG PET on progression-free survival (PFS) or event-free survival (EFS) in DLBCL patients treated with first-line immunochemotherapy regimens.

Methods

For this systematic review and meta-analysis Pubmed, Embase, and the Cochrane Library were searched until July 11, 2017. Prospective and retrospective studies investigating qualitative interim PET response assessment without treatment adaptation based on the interim PET result were eligible. The primary outcome was two-year PFS or EFS. Prognostic and diagnostic measures were extracted and analysed with pooled hazard ratios and Hierarchical Summary Receiver Operator Characteristic Curves, respectively. Meta-regression was used to study covariate effects.

Results

The pooled hazard ratio for 18 studies comprising 2,255 patients was 3.13 (95%CI 2.52–3.89) with a 95% prediction interval of 1.68–5.83. In 19 studies with 2,366 patients, the negative predictive value for progression generally exceeded 80% (64–95), but sensitivity (33–87), specificity (49–94), and positive predictive values (20–74) ranged widely.

Conclusions

These findings showed that interim ^{18}F-FDG PET has predictive value in DLBCL patients. However, (subgroup) analyses were limited by lack of information and small sample sizes. Some diagnostic test characteristics were not satisfactory, especially the positive predictive value should be improved, before a successful risk stratified treatment approach can be implemented in clinical practice.

Keywords
Aggressive non-Hodgkin's lymphoma. Diffuse large B-cell lymphoma. Positron-emission tomography. Systematic review. Meta-analysis.

Introduction

Diffuse large B-cell lymphoma (DLBCL) represents the most common subtype of adult non-Hodgkin lymphoma (NHL) cases, and is associated with an aggressive clinical course. There are several potentially effective first-line chemotherapy regimens of which most consist of cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP). The addition of the monoclonal antibody rituximab (R) to this regimen (R-CHOP) has significantly improved the outcome of DLBCL patients [1,2]. However, treatment failure is still an important problem as the 3-year progression-free survival (PFS) of DLBCL patients is approximately 60–70% [3].

Commonly used prognostic indices are the International Prognostic Index (IPI) [4,5], or the more powerful Revised-IPI (R-IPI) [6], and National Comprehensive Cancer Network IPI (NCCN-IPI) [7]. These indices can be used for risk-stratification to predict a poor outcome after R-CHOP. It is important to identify a poor outcome as soon as possible because these patients could benefit from a switch to a second-line treatment or high-dose chemotherapy (HDCT) with autologous stem cell transplantation (ASCT) as an upfront treatment [8]. 18F-fluoro-2-deoxy-D-glucose (^{18}F-FDG) positron emission tomography (PET) after a few cycles of therapy, also known as interim ^{18}F-FDG PET, is of increasing interest, as it may facilitate early change of treatment and prevent unnecessary side effects [9]. In recent decades several visual criteria for interpretation of ^{18}F-FDG PET have been developed, for example, the EORTC, PERCIST, and International Harmonization Project (IHP) criteria as well as the Deauville scoring system [9–13]. Nowadays the latter is widely adopted for interpretation of response evaluation with ^{18}F-FDG PET in DLBCL [9,13].

Interim ^{18}F-FDG PET has shown high predictive value in Hodgkin lymphoma [14]; however, according to previous reviews, the role of interim ^{18}F-FDG PET in DLBCL is still unknown [15–18]. From these studies it can be concluded that

heterogeneity in patient populations, therapy regimens, PET scanners, timing of the interim ^{18}F-FDG PET scans, and/or differences in the visual criteria used for interpretation of the interim ^{18}F-FDG PET scans made it hard to clarify the accuracy of interim ^{18}F-FDG PET to predict clinical outcome in DLBCL.

Therefore, we performed a new systematic review and meta-analysis, focusing on DLBCL patients only, assessing both the hazard ratio (HR) and diagnostic parameters (sensitivity, specificity, and predictive values) of interim ^{18}F-FDG PET on PFS or event-free survival (EFS) in patients with DLBCL treated with first-line immuno-chemotherapy regimens. The primary outcome measure was PFS (preferably) or EFS at 2 years, since DLBCL patients who are event-free after 24 months have demonstrated an overall survival (OS) comparable to an age- and sex-matched general population [19]. In order to reduce the previously described heterogeneity we performed several subgroup analyses, for example, by the type of ^{18}F-FDG PET scanner and the type of visual criteria used for interpretation of the interim ^{18}F-FDG PET scans. In this meticulously performed review we contacted the authors for additional information if necessary.

Materials and methods

Search strategy

For this systematic review and meta-analysis we searched in collaboration with a medical librarian Pubmed/MEDLINE, Embase, and the Cochrane Library databases from onset until July 11, 2017 with a language restriction to English, French, Dutch, or German. Our search strategy contained a combination of various indexed terms and free text words for "positron emission tomography" and "non-Hodgkin lymphoma" (full search strategy Supplemental Table 1). We included full-text publications of original prospective and retrospective studies. Excluded were conference abstracts, letters, comments, editorials, review articles, animal studies, and case reports. Reference lists of included articles were checked to identify additional eligible studies.

Study selection: Eligibility criteria

Patients

Adult patients treated with first-line immuno-chemotherapy regimens for stage I-IV DLBCL were considered as our target population. We excluded studies that investigated HIV-related lymphoma, central nervous system (CNS) lymphoma involvement, or post-transplant lymphoproliferative disease (PTLD). Studies containing less than 80% of DLBCL subtype were excluded, unless subgroup data for DLBCL were presented or if the remaining 20% had PMBCL or FL grade 3B [20]. Studies including ten patients or less were classified as case series and therefore also excluded.

Treatment procedures

Studies in which a change of treatment was based on the interim ^{18}F-FDG PET result and prospective PET-adapted trials were not included. However, we allowed a change of therapy in patients with clinical evidence of progressive disease during first-line treatment [9]. We included all R-CHOP-like treatments as first-line treatment strategies [1,2,21–23], but we excluded studies if ≤50% of patients received rituximab. Therapies using other (new generation) monoclonal antibodies were excluded. Studies with autologous stem cell transplantation (ASCT) were eligible if this strategy was part of the preplanned first-line treatment. Radiotherapy was accepted if the decision to give radiotherapy was preplanned or used for consolidation of PET positive sites at the end of first-line treatment, but not affected by interim ^{18}F-FDG PET results. If studies did not report on the use of ASCT or radiotherapy, we assumed that no ASCT or radiotherapy was given based on interim ^{18}F-FDG PET result.

Interim ^{18}F-FDG PET procedures

An interim ^{18}F-FDG PET scan should have been performed after the first, second, third, or fourth treatment cycle. PET only as well as PET/CT systems were considered eligible. Use of other radiopharmaceuticals than ^{18}F-FDG were not accepted. We focused on visual interpretation criteria only, as nowadays, semi-quantitative PET strategies are used for research purposes only and are not standard in the current guidelines yet [13]. PET response criteria were grouped into three categories: Deauville score (DS) on a 5-point scale [9,13], International Harmonization Project (IHP) [12], and custom visual criteria (i.e. not based on consensus guidelines).

Outcome measures

The primary outcome measure was defined as PFS (preferably) or EFS at 2 years. We included studies with a minimum median follow-up period of 24 months in surviving patients (or for the entire study population), because most patients experience relapse or progression of their disease in the first 2 years after their diagnosis [24,25].

Data extraction and quality assessment

After removing duplicates, two authors independently screened titles and abstracts of the search results for eligibility (CNB and NH, AdJ, or HCWdV). The decision to include studies in the review was based on the full-text articles (CNB and AdJ or HCWdV). Extensive data extraction forms (available upon request) were developed which included the criteria from the methodological checklists for diagnostic accuracy studies (QUADAS-2) [26] and for prognostic studies (QUIPS) [27]. The forms were tested in a few articles and used independently by two review authors (CNB, AdJ). Consensus meetings (with three experts in nuclear medicine, hematology, and methodology, respectively) were organized to solve disagreements and to decide on eligibility of the final study selection. Besides general information about study design, patients, treatment, interim ^{18}F-FDG PET performance, and outcome measures (used for qualitative study descriptions and determination of eligibility) we extracted outcomes on two types of predictive parameters. For the first predictive meta-analysis we extracted univariate hazard ratios (HRs) and their corresponding 95% confidence intervals. If this data was not reported and not provided after contacting the authors, we used the methods of Tierney et al. [28] to deduce these from reported parameters or from the Kaplan-Meier (KM) curves, using numbers at risk when available. For the second predictive meta-analysis we used a diagnostic approach and constructed 2 × 2 contingency tables to calculate sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of interim ^{18}F-FDG PET for prediction of two-year PFS and - EFS. If no two-year survival percentages were reported we estimated the percentages from the KM curves at this time-point. If information was missing or unclear authors were contacted. A maximum of three reminders were sent. In case of no reply we used the information that was available from the original publication. Individual patient data was not requested for this meta-analysis.

Statistical analyses

Two approaches of meta-analysis

For the meta-analyses of the HRs, individual log hazard ratios (HRs) and standard errors (SE) were pooled using a random effects model (REML, restricted maximum likelihood). Together with the individual study results, the pooled effect estimate—expressed as HR and 95% confidence interval— was visualized in a Forest plot. Between-study heterogeneity was assessed by using Cochran's Q and I^2 statistics [29]. A 95% prediction interval around the HR was calculated to predict the expected range of the HR of a new (future) study [30]. A funnel plot was presented to visually assess if publication bias was likely [31]. For the diagnostic meta-analysis, the pooled sensitivity and specificity was obtained by Hierarchical Summary ROC curve (HSROC) models and ROC curves constructed in RevMan [32] using the input parameters of the HSROC models.

Influence of covariates

Several prespecified subgroup analyses—which included both clinical and methodological issues—were performed using univariate meta-regression models for the HRs and as covariate interaction term in the HSROC models. The following subgroup analyses were performed: study design (retrospective or prospective studies; blinded review or not reported; PFS or EFS), characteristics of patients (100% DLBCL or between 80 and 100%), treatments (ASCT upfront or not, preplanned or consolidative radiotherapy used or unknown), properties of scans (PET/CT or a combination of PET/CT and PET standalone systems, availability of a baseline PET or CT), and scoring issues (DS -, IHP -, or custom criteria, central review or local review).

Software

Statistical analysis was performed in R (version 3.2.5) [33] using the Metafor package and SAS Proc Nlmixed was used for the HSROC models. A *P* value of less than 0.05 was considered statistically significant.

Results

The search yielded 9,960 records after removing duplicates; 290 concerned studies on NHL and interim FDG-PET, the other 9,670 records were excluded because they did not report on NHL or I-PET. 85/290 were potentially eligible and full-text articles were retrieved. After checking detailed inclusion and exclusion criteria we included 20 eligible studies in the qualitative systematic review; 19 out of 20 were eligible for the HRs evaluations and 18 out of 20 for the HSROC analyses (Fig. 1).

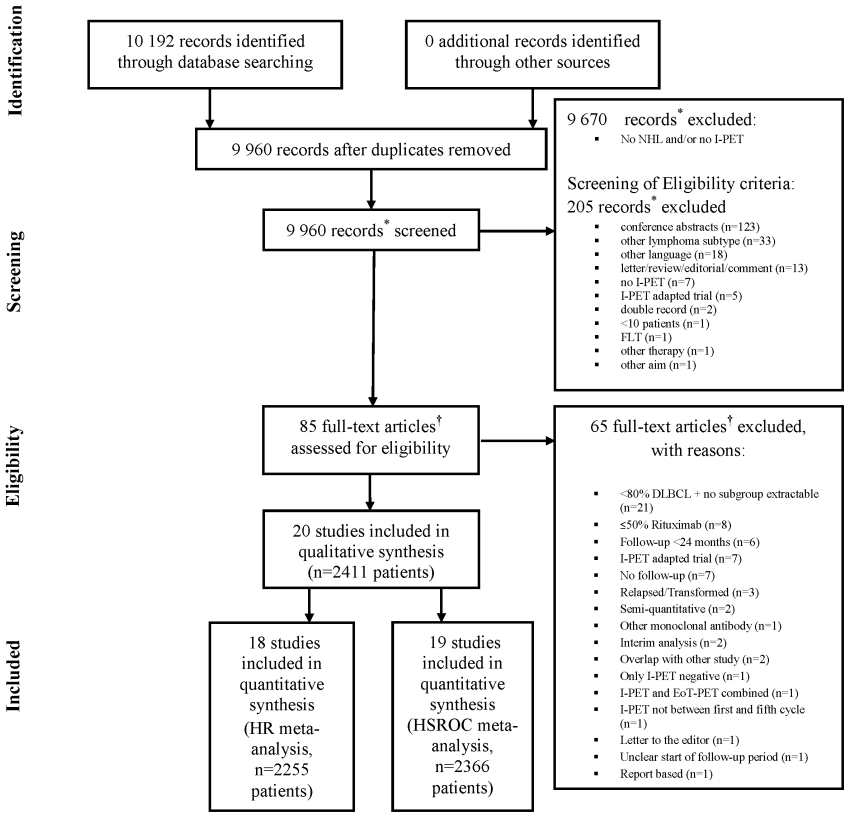


Fig. 1 PRISMA flow diagram.

*Records refer to the title and abstract screening of the search results

†Full-text articles refer to the full-text assessment of the selected articles from the title and abstract screening phase.

Abbreviations: I-PET= interim ^{18}F-FDG positron emission tomography, FLT= Fluorothymidine, DLBCL= diffuse large B-cell lymphoma, EoT-PET= end-of-treatment ^{18}F-FDG positron emission tomography, HR= hazard ratio, HSROC= hierarchical summary receiver operating curve

Table 1 Study- and patient characteristics

Study characteristics		Patient characteristics					Treatment characteristics		
First author, year	**Study design**	**No. of patients**	**DLBCL**	**Age median (range)**	**Male**	**Stage III-IV**	**First-line treatment**	**RT**	**ASCT**
Fan et al (2017)[34]	retrospective	119	100%	37% > 60	50·4%	52·9%	CHOP-like 88·2% R	NR	NR
Kim et al (2017)[35]	retrospective	150	100%	mean 58·5 SD 14	57·3%	65·3%	R-CHOP21[a]	NR	NR
De Oliveira Costa et al (2016)[36]	prospective	111[b]/147	100%	58.9 (16-86)[c]	45·0%	64·0%	Stage I/II: 4x R- CHOP21[a] + RT Stage III/IV: R-CHOP21[a]	43·2%	? refractory and relapsed pts: IVAC +ASCT
Kong et al (2016)[37]	retrospective	105	100%	56 (19-82)	54·3%	43·8%	R-CHOP	NR	NR
Mikhaeel et al (2016)[38]	retrospective	147	100%	57 (22-86)	49·7%	68·7%	R-CHOP Stage I/II non-bulky: 3-4x R-CHOP and IFRT	34·0%	not upfront (*n*=1 after clinical progression)
Mamot et al (2015)[39]	prospective	125[b]/138	100%	58.4(18-81)[c]	54·3%[‡]	53·6%[c]	R-CHOP14[d] + 2R	17·4%[c] =event	NR
Zhang et al (2015)[40]	retrospective	197	100%	46 (18-81)	60·4%	59·4%	14·2% R-CHOP14[d] 85·8% R-CHOP21[a]	18·8%	not upfront (*n*=1 progression at end-of-treatment)
Carr et al (2014)[41]	prospective	327[b]/361	DLBCL: 97·2% PMBCL: 2·8%	55 (IQR 44-64)	52·9%	64·2%	(R-)CHOP21[a] MACOP-B (n=1) R-CNOP (n=4) 86% R	20·2%	NR
Dabaja et al (2014)[42]	retrospective	294[b]/350	100%	49% > 61	55·4%	62·6%	82·0% R-CHOP 11·2% R-HCVAD 6·8% other	29·9%	NR
Mylam et al (2014)[43]	prospective	112	100%	62 (23-85)	52·7%	82·0%	84·8% R-CHOP 9·8% R-CHOEP 5·4% other (92·9% R 55% 14 day 45% 21 day)	in methods but no numbers	NR

Table 1 Continued

Study characteristics		Patient characteristics					Treatment characteristics		
First author, year	**Study design**	**No. of patients**	**DLBCL**	**Age median (range)**	**Male**	**Stage III-IV**	**First-line treatment**	**RT**	**ASCT**
Nols et al (2014)[44]	retrospective	73	100%	60 (18-85)	63·0%	68·5%	15·1% R-CHOP14[d] 49·3% RCHOP21[a] 11·0% R-mini-CHOP 23·3% R-ACVBP 1·4% CHOP	NR	8·2%
Fuertes et al (2013)[45]	prospective	50	100%	55 (21-79)	56·0%	44·0%	R-CHOP21[a]	NR	NR
Gonzalez-Barca et al (2013)[46]	prospective	69	100%	60 (18-78·9)	53·6%	65·2%	R-CHOP14[d]	5·8%	NR
Itti et al (2013)[47]	retrospective	114	100%	56 (23-80)	59·6%	82·5%	55·3% R-CHOP21[a] 44·7% R-CHOP14[d]/ R-ACVBP	3·5%	? in young high- risk pts as part of first-line consolidation or salvage
Lanic et al (2012)[48]	retrospective	45[b]/57	100%	65 (22-87)	48·9%	84·4%	75·5% R-CHOP 24·4% intensified R-CHOP	NR	8·9% frontline
Pregno et al (2012)[49]	retrospective	88	100%	55 (18-80)	46·6%	67·0%	35·2% R-CHOP21[a] 64·8% R-CHOP14[d]	15·9%	NR
Safar et al (2012)[50]	retrospective	112	100%	59 (20-79)	67·0%	81·3%	50·9% R-CHOP21[a] 21·4% R-CHOP14[d] 27·7% R-ACVBP	0%	16·1% consolidative HDT+ASCT (if <60 years + >1 aaIPI)
Cashen et al (2011)[51]	prospective	50	100%	mean 58 (29-80)	NR	100%	R-CHOP21[a]	upfront= excl. crit. *n*=1 after ASCT	not upfront (*n*=1 refractory disease and *n*=2 progressive disease)

Table 1 Continued

Study characteristics		Patient characteristics					Treatment characteristics		
First author, year	**Study design**	**No. of patients**	**DLBCL**	**Age median (range)**	**Male**	**Stage III-IV**	**First-line treatment**	**RT**	**ASCT**
Zinzani et al (2011)[52]	retrospective	91	DLBCL: 85·7% PMBCL: 14·3%	54 (17-90)	52·7%	67·0%	DLBCL: 16·7% R-VNCOP-B (>=60yr) 83·3% R-CHOP21[a] (<60yr) PMBCL: 92·3% R-MACOP-B 7·7% R-CHOP21[a]	NR	not upfront (7·7% with PR at end-of-treatment PET, 2·2% at time of relapse)
Zhao et al (2007)[53]	retrospective	32[e] /61	DLBCL/ PMBCL	57 (12[f]-85)[c]	60·7%[e]	75·4%[e]	90.6% R-CHOP21[g] 9.4% CHOP21[g]	preplanned, depending on stage and site	NR

[a] (R-)CHOP21: (rituximab,) cyclophosphamide, doxorubicin, vincristine, prednisone given with a 3 week interval between cycles.
[b]number of I-PET scans available for (central) review
[c]only available for complete patient cohort
[d](R-)CHOP14: (rituximab,) cyclophosphamide, doxorubicin, vincristine, prednisone given with a 2 week interval between cycles.
[e]data of DLBCL/PMBCL patients only (received from authors)
[f]authors replied that only 1 patient with DLBCL was younger than 18 years old.
[g]authors replied that 29/32 DLBCL/PMCL patients received R-CHOP21 and 3/32 received CHOP21.

Abbreviations: No=number, RT=radiotherapy, ASCT=autologous stem cell transplantation, (R-)CHOP= (rituximab,) cyclophosphamide, doxorubicin, vincristine, prednisone, R=rituximab, NR=not reported, SD=standard deviation, IVAC=ifosfamide, mesna, cytarabine, etoposide, IFRT= involved field radiotherapy, PMBCL=primary mediastinal B-cell lymphoma, IQR=interquartile range, MACOP-B= methotrexate, cytarabine, cyclophosphamide, vincristine, prednisone, bleomycin, CNOP=cyclophosphamide, mitoxantrone, vincristine, prednisone, HCVAD= hyperfractionated cyclophosphamide, doxorubicin, vincristine, dexamethasone, CHOEP= cyclophosphamide, doxorubicin, vincristine, etoposide, prednisone, R-ACVBP= rituximab, doxorubicin, vindesine, bleomycin, prednisone, pts=patients, HDT=high dose therapy, aaIPI=age adjusted international prognostic index, excl. crit=exclusion criterium, VNCOP-B= etoposide, mitoxantrone, cyclophosphamide, vincristine, prednisone, bleomycin, PR=partial response, CVP=cyclophosphamide, vincristine, prednisone, DHAP= dexamethasone, cytarabine, cisplatin, MTX-AraC= methotrexate, cytarabine.

A total of 2,411 newly diagnosed DLBCL patients from 20 studies were assessed for this analysis. Table 1 shows the main study-, patient-, and treatment characteristics of the included studies. The number of included patients per study ranged from 32 to 327 (median 112, interquartile range 70–142). Seven studies had a prospective study design. The median age of the patients ranged from 54 to 65 years, with the exception of one study with a median age of 46 [40], and 45–67% of the patients were of male gender. Most studies included patients with Ann Arbor stage I/II as well as stage III/IV; in two studies less than 50% of the patients had stage III or IV [37,45] and one study included patients with stage III and IV only [51]. First-line treatment regimens varied between and within the studies, but R-CHOP was the basic principle in all studies. Radiotherapy was given in most of the studies to selected patients (preplanned, e.g. in case of bulky disease or as a consolidation for residual lymphoma sites after treatment). Autologous stem cell transplantation had been planned upfront in three studies [44,48,50].

In Table 2 details of PET procedures, interpretation, and timing of interim PET between cycles are shown. Most studies performed an interim PET scan after two cycles of chemotherapy in all patients, one study made interim PET scans after only one course in all patients [43]; the remaining studies combined patient groups who had their interim assessment after a variable number of treatment cycles. The number of days after the previous treatment course at which the interim PET was acquired also varied between studies, mostly just before the next chemotherapy cycle, but the number of days after previous treatment was not reported by all studies. Twelve studies applied the Deauville scoring system and four the International Harmonization Project system [40,46,48,51]. The remaining studies used a custom scoring system [42,50,52,53].

The outcome measures of the included studies are shown in Table 3: 16 studies presented PFS and the other four studies reported EFS. The definitions of PFS and EFS for the different studies are presented in Supplemental Table 2. Percentages of positive interim PET scans ranged from 18.1 to 56.3%. Five original publications had reported univariate HRs, and four authors provided a (re)calculated HR upon our request. Two authors provided information about the number of events and *P*-values in order to use the method from Tierney et al. [28]. For one study we extracted the HR from the KM curves with numbers at risk provided by the

Table 2 Interim ^{18}F-FDG PET characteristics

	Timing		Interpretation		Scan Procedures				
First author, year	**Interim ^{18}F-FDG PET timing**	**Days after previous treatment median (range)**	**Criteria for positive scan**	**Interpreters**	**PET/CT or PET only**	**^{18}F-FDG dose (MBq)**	**Uptake interval (min)**	**Fasting period (hours)**	**Glucose**
Fan et al (2017)[34]	2	NR	Deauville 4-5	3 NM	PET/CT	3·7/kg	60 +- 10	≥6	<200 mg/dL
Kim et al (2017)[35]	2 or 3	19·7 +- 2·3	Deauville 4-5	2 NM	PET/CT	5·18/ kg	50	≥6	NR
De Oliveira Costa et al (2016)[36]	2	20	Deauville 4-5	1 NM + 1 RAD	Both	5/kg	60	≥6	<180 mg/dL
Kong et al (2016)[37]	2	NR	Deauville 4-5	≥2 NM	PET/CT	5·5/kg	40-60	≥6	<=8 mmol/L
Mikhaeel et al (2016)[38]	2	NR	Deauville 4-5	1 NM	PET/CT	370	90	6	NR
Mamot et al (2015)[39]	2 (if positive also after 4	(11-14)	Deauville 4-5[a]	NR	Both	370	60	≥4	Measured before tracer injection
Zhang et al (2015)[40]	2 (and 4)	NR	IHP	3 reviewers	PET/CT	4·4-7·4/kg	60-90	6	NR
Carr et al (2014)[41]	2, 3 or 4	18 (IQR 17-21)	Deauville[b]	4 NM	Both	NR	NR	NR	NR
Dabaja et al (2014)[42]	1, 2 or 3	Days from diagnosis 63 (21-126)	Custom[c]	2 physicians specializing in NM	PET/CT	Aver-age 630	mean 90 SD 22	≥6	<200 mg/dL
Mylam et al (2014)[43]	1	Min 10 days	Deauville 4-5	2 out of 5 (random), discrep: 3rd	NR	NR	NR	NR	NR
Nols et al (2014)[44]	3 or 4	NR	Deauville 4-5	2 NM from own dept	Both	200-365	97·5 +- 32·8	≥6	<175 mg/dL
Fuertes et al (2013)[45]	2 or 3	18 (16-21)	Deauville 4-5	2 NM	PET/CT	296-444	60	4-6	<180 mg/dL
Gonzalez-Barca et al (2013)[46]	2	2 days before 3rd cycle	IHP	NR	Both	3·7/kg	60-90	≥6	90-160 mg/dL
Itti et al (2013)[47]	2	R-CHOP14/R-ACVBP: 13+-2 R-CHOP 21: 19+-4	Deauville 4-5	3 NM	PET/CT	5.4/kg 393+-124	median 69, mean 70 +- 16	fasting	<=11 mmol/L
Lanic et al (2012)[48]	3 or 4	16 (3-27)	IHP	2 NM	PET/CT	5/kg	60	≥6	range: 3·4-15·8 mmol/L

Table 2 Continued

	Timing		Interpretation		Scan Procedures				
First author, year	**Interim ^{18}F-FDG PET timing**	**Days after previous treatment median (range)**	**Criteria for positive scan**	**Interpreters**	**PET/CT or PET only**	**^{18}F-FDG dose (MBq)**	**Uptake interval (min)**	**Fasting period (hours)**	**Glucose**
Pregno et al (2012)[49]	2, 3 or 4	13 (4-27)	Deauville 4-5	2 NM	PET/CT	37 /10 kg	60 (n=6 90 min)	≥6	90-160 mg/dL
Safar et al (2012)[50]	2	Median 14	Custom[d]	1 NM	Both	2-5/kg	60 +- 10	≥6	checked before the exam
Cashen et al (2011)[51]	2 or 3	3 weeks	IHP	2 nuclear radiologists	PET/CT	370-555	+- 60	fasting	<200 mg/dL (84-188)
Zinzani et al (2011)[52]	3 or mid-treatment[e]	Immediately before subsequent cycle	Custom[f]	3 experienced readers	PET/CT	5·3/kg	60-90	6	NR
Zhao et al (2007)[53]	3 or 4	Last day before new cycle	Custom[g]	2 NM	PET/CT	240-259	60	≥6	<7·8 mmol/L

[a]= results presented for central review results only, local review results are based on SUVmax lesion > SUVmax of mediastinal bloodpool.

[b]= 4 categories; Negative/CR= resolution of abnormal ^{18}F-FDG uptake at sites of disease identified on staging PET with any residual ^{18}F-FDG uptake less than or equal to the mediastinal blood pool, CR-MRU= residual low-level ^{18}F-FDG uptake at disease sites greater than mediastinum, but less than or equal to physiologic uptake in liver. Positive=residual or increased ^{18}F-FDG uptake with intensity greater than liver at a site of known disease. Mixed response= reduction in ^{18}F-FDG uptake at some disease sites, with increased ^{18}F-FDG uptake at other existing or new sites.

[c]= positive= according to a SUVmax measurement >2.5. Equivocal PET/CT or CT findings were then interpreted by using CT scan findings and clinical information to distinguish vascular/bowel activity from residual FDG uptake at the previously involved site.

[d]= a negative PET scan was defined as having no residual abnormal uptake or a minimal residual uptake. A positive scan was defined as having at least one residual site associated with an intensity markedly superior to local background as previously described (Haioun Blood 2005).

[e]= after 3 cycles R-CHOP21 or midtreatment in case of R-VNCOP-B or R-MACOP-B.

[f]= negative= no pathological tracer uptake was shown, unequivocally positive= areas of focal uptake localized to sites of previous disease (in this sense representing a residual disease or a disease relapse), within asymmetrical lymph nodes, or within lymph nodes unlikely to be affected by inflammation (mediastinal, except for hilar, and abdominal). Sites of known physiological uptake that showed symmetrical uptake were not described in the report.

[g]= negative was defined as no evidence of disease. MRU was defined as low-grade uptake of FDG (just above background) in a focus within an area of previously noted disease reported by the nuclear medicine physicians. Positive was defined as increased uptake suspicious for malignant diseases, which did not have a benign explanation.

Abbreviations: MBq=megabecquerel, min=minutes, NR=not reported, NM=nuclear medicine physician, RAD=radiologist, IHP= international harmonization project criteria, IQR=interquartile range, SD= standard deviation, dept=department, (R-)CHOP= (rituximab,) cyclophosphamide, doxorubicin, vincristine, prednisone, R=rituximab, R-ACVBP= rituximab, doxorubicin, vindesine, bleomycin, prednisone.

authors and for six studies we used the KM curves without numbers at risk. For two studies we could not extract the HRs, as there was insufficient data and no Kaplan-Meier curve [36,48].

In Fig. 2 the Forest plot with the 18 univariate HRs is shown. The pooled effect estimate was 3.13 (95% CI 2.52–3.89). The Cochran's Q test for heterogeneity was not statistically significant (P = 0.087) and between study heterogeneity was low (I^2 = 35.14%). The 95% prediction interval was 1.68–5.83, with one outlier [37].

The methodological quality was assessed based on the QUADAS-2 and QUIPS checklists. Subgroup analyses were performed on study design characteristics that were potential sources of bias.

Meta-regression showed that the outcomes did not differ between retrospective and prospective studies, studies with blinded review and studies that did not report whether they blinded the PET/CT assessment, or studies that used PFS or EFS as outcome measure. A statistically significant higher HR was found for studies with a combination of integrated PET/CT- and PET standalone systems compared to studies with integrated PET/CT systems only (HR 4.39 vs 2.85, P =0.0332) and a trend towards a higher HR in studies with 80–99% DLBCL compared to studies with 100% DLBCL (P =0.0577). Prespecified subgroups for different types of treatments and FDG-PET scoring systems showed no statistically significant differences (Supplemental Table 3). For the subgroups "availability of baseline PET or CT" and "central or local review procedure", insufficient information was reported to perform these analyses. Risk of publication bias as assessed with a Funnel plot was low (Supplemental Fig. 1).

Nineteen studies had data available for the calculation of PPV, NPV, sensitivity, and specificity of interim PET for prediction of two-year-PFS or -EFS. For one study we could not extract or calculate the diagnostic measures [48]. PPV and NPV ranged from 20 to 74% and 64 to 95%, respectively. Sensitivity and specificity ranged from 33 to 87% and 49 to 94%, respectively (Table 3, Supplemental Fig. 2).

Table 3 Study results; prognostic and diagnostic information

	Results			Prognostic information		Diagnostic information at 2 years[a]			
First author, year	**Outcome measure**	**Median follow-up in months (range)**	**Interim PET positive no (%)**	**HR univariate**	**95% CI**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Positive Predictive Value (95% CI)**	**Negative Predictive Value (95% CI)**
Fan et al (2017)[34]	PFS	26 (2-75)	42 (35·3%)	4·46[b]	2·42-8·23[b]	60% (46-74)	79% (69-87)	62% (47-75)	78% (67-86)
Kim et al (2017)[35]	PFS	31 (8-75)	43 (28·7%)	1·81[b]	0·95-3·45[b]	38% (25-54)	75% (66-82)	35% (22-50)	78% (69-84)
De Oliveira Costa et al (2016)[36]	PFS	41·5 (0·6- 71·1)	51 (45·9%)	NR	NR	59% (36-78)	56% (46-66)	20% (11-32)	88% (78-94)
Kong et al (2016)[37]	PFS	32 (9-59)	19 (18·1%)	9·74[c]	3·27 - 28·99[c]	64% (43-80)	94% (87-97)	74% (51-88)	91% (83-95)
Mikhaeel et al (2016)[38]	PFS	45·6 (15·6- 94·8)	65 (44·2%)	2·18[c]	1·11 - 4·27[c]	63% (48-76)	63% (54-72)	42% (30-54)	80% (71-88)
Mamot et al (2015)[39]	EFS	45 (4-64)[d]	58 (46·4%)	3·23[d]	1·78 - 5·89[d]	68% (54-79)	68% (57-77)	59% (46-70)	76% (65-85)
Zhang et al (2015)[40]	PFS	30 (5-94)	87 (44·2%)	2·74[c]	1·52 - 4·93[c]	70% (59-79)	72% (63-79)	60% (49-69)	80% (72-86)
Carr et al (2014)[41]	EFS	35 (75% at least 24 mo)	117(35·8%)	5·31[b]	3·29 - 8·56[b]	70% (58-79)	74% (68-79)	42% (33-51)	90% (85-93)
Dabaja et al (2014)[42]	PFS	36 (0·8-84)	54 (18·4%)	1·9[b]	1·1 - 3·2[b]	33% (20-46)	85% (80-89)	33% (22-47)	85% (79-89)
Mylam et al (2014)[43]	PFS	29 (2-80)	60 (53·6%)	1·45[e]	0·71 - 2·97[e]	64% (43-80)	49% (39-59)	23% (14-35)	85% (72-92)
Nols et al (2014)[44]	PFS	29 (3.3-86)	20 (27·4%)	5·26[f]	1·90 - 14·58[f]	58% (36-77)	83% (71-91)	55% (34-74)	85% (73-92)
Fuertes et al (2013)[45]	PFS	Surviving: 46·8 (2·4- 78)	12 (24·0%)	3·89[c]	1·07 - 14·22[c]	46% (23-71)	84% (69-92)	50% (25-75)	82% (67-91)
Gonzalez-Barca et al (2013)[46]	EFS	28·8 (5·8-52·6)	34 (49·3%)	2·70[c]	0·68 - 10·67[c]	75% (46-91)	56% (43-68)	26% (15-43)	91% (78-97)
Itti et al (2013)[47]	PFS	living: 39 (12-74) relapse: 6 (1-32) death: 17 (3-55)	51 (44·7%)	2·85[d]	1·38 – 5·87[d]	68% (49-82)	63% (52-72)	37% (25-51)	86% (75-92)
Lanic et al (2012)[48]	PFS	27 (7-73)	14 (31·1%)	NR	NR	NR	NR	NR	NR
Pregno et al (2012)[49]	PFS	26·2 (8-67)	25 (28·4%)	2·45[b]	1·01 - 5·93[b]	44% (23-67)	75% (64-84)	28% (14-47)	86% (75-92)
Safar et al (2012)[50]	PFS	living: 38 (4·4 -73)	42 (37·5%)	4·77[d]	2·26 - 10·05[d]	72% (54-85)	75% (64-83)	50% (35-64)	89% (79-94)

Table 3 Continued

	Results			Prognostic information		Diagnostic information at 2 years[a]			
First author, year	**Outcome measure**	**Median follow-up in months (range)**	**Interim PET positive no (%)**	**HR univariate**	**95% CI**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Positive Predictive Value (95% CI)**	**Negative Predictive Value (95% CI)**
Cashen et al (2011)[51]	PFS	Surviving: 33·9 (16-44)	24 (48·0%)	2·98[d]	1·16 - 7·67[d]	69% (42-87)	59% (43-74)	38% (21-57)	85% (66-94)
Zinzani et al (2011)[52]	EFS	No events: 50 (12-68)	35 (38·5%)	3·94[c]	1·69 - 9·19[c]	87% (68-95)	78% (67-86)	57% (41-72)	95% (85-98)
Zhao et al(2007)[53]	PFS	27 (9-45) [g]	18 (56·3%)	3·67[e]	1·61 - 8·35[e]	72% (49-88)	64% (39-84)	72% (49-88)	64% (39-84)

[a]=extracted predictive test accuracy measures at two years survival prediction in Kaplan-Meier curve.
[b]= reported in publication.
[c]= extracted from KM without numbers at risk.
[d]= reply on request for additional information.
[e]=Tierney method based on number of events and *P*-value.
[f]= extracted from KM with numbers at risk [g]=only available for complete patient cohort

Abbreviations: 95% CI= 95% confidence interval, PFS=progression-free survival, NR=not reported, EFS= event-free survival, mo=months.

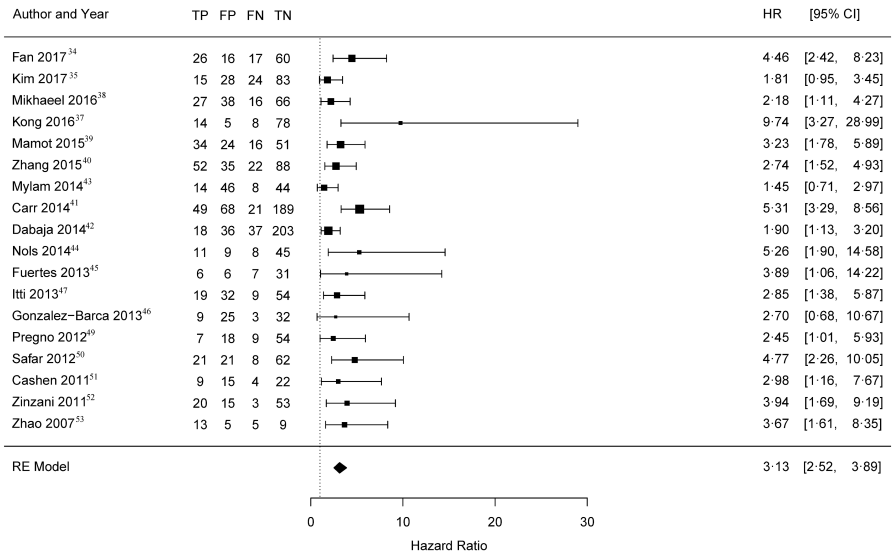


Fig. 2 Forest plot of univariate hazard ratios for interim PET scans in diffuse large B-cell lymphoma.

This plot shows the univariate hazard ratios (black squares, size based on study size), and 95% CI's (horizontal lines) of the individual studies sorted by publication year for PFS/EFS of the interim PET positive and negative patients. The estimated pooled effect estimation is shown with a diamond. For each study a 2x2 contingency table at 2 years follow-up is shown

In Fig. 3 the ROC curves of the different visual criteria are shown. The studies that were classified as "custom", did not have comparable scan positivity definitions and therefore no summary curve for this group was presented. We found no statistically significant differences between the curves for Deauville and IHP. There was a trend (P = 0.0503) towards a higher accuracy for studies with DLBCL 80–99% versus studies with 100% DLBCL patients.

Discussion

This systematic review and meta-analysis included 20 studies comprising a total of 2,411 DLBCL patients who underwent interim ^{18}F-FDG PET. Eighteen studies were eligible for the HR and 19 for the HSROC meta-analyses. We found a pooled estimated HR of 3.13 (95% CI 2.52–3.89) for interim PET in the prediction of PFS or EFS. The prediction interval ranged from 1.68 to 5.83, suggesting that a new study investigating the prognostic value of interim PET on PFS or EFS will find a HR in this range with 95% confidence. These results

confirm the predictive value of interim PET in DLBCL patients for PFS and EFS. Our pooled estimated HR was lower than reported in a previous meta-analysis (2013) [16] which reported a pooled estimated HR of 4.4 (95% CI 3.34–5.81) from nine studies investigating the prediction of PFS by interim PET. They used a similar approach to extract HRs; however, they had less strict inclusion criteria with regard to the NHL types and follow-up period, both visual and semi-quantitatively assessed PET scans were included, and no subgroup analyses were performed. Despite these differences, their HR result is within the range of our calculated 95% prediction interval and the amount of statistical heterogeneity (I^2 = 39%) amongst studies was comparable. Other meta-analyses did not compare the HRs between studies [15,17,18].

We have no explanation for the statistically significant higher HR for studies (*n* = 5) that used both PET/CT- and PET standalone systems compared to studies that used an integrated PET/CT system.

The trend towards a higher HR for the studies with both DLBCL and PMBCL patients compared to studies with only DLBCL patients could not directly be explained by the inclusion of both lymphoma subtypes. The fact that two out of three studies with both DLBCL and PMBCL patients [52,53] used custom criteria for the interpretation of the interim PET could possibly explain this. These meta-regression results should be interpreted with caution, as the number of studies per subgroup were relatively low (Supplemental Table 3) which precludes multivariate meta-regression analysis.

Diagnostic 2 × 2 contingency tables of interim PET showed wide ranges between studies for sensitivity, specificity, and positive predictive values at 2 years. The ranges reported in other systematic reviews and meta-analyses were hard to compare as they used the complete follow-up period for their calculations, included studies with follow-up periods less than 24 months, and used other statistical methods [15,17,18]. We decided to truncate at 2 years, as most clinically relevant events occur during this period. Moreover, the widely ranging complete follow-up periods of individual studies might introduce bias.

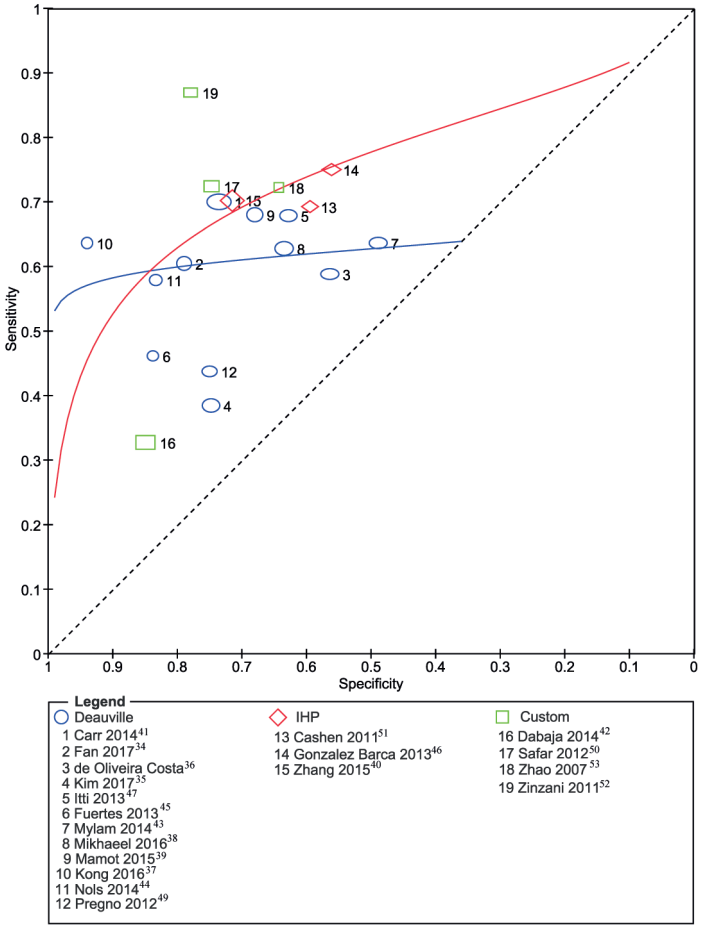


Fig. 3 Summary receiver operating curves (sROC) for different visual interim PET criteria.

Studies that assessed interim PET scans according to the Deauville's criteria are indicated with *blue circles*, studies that used the international harmonization project (IHP) criteria are indicated with *red diamonds* and studies that used custom visual criteria are indicated with *green squares*. The size of the *circles*, *diamonds*, and *squares* are based on the inverse standard error

Negative predictive values for 2-year progression-free status were generally above 80%, except in four studies [34,35,39,53]. In Mamot et al. [39], the somewhat lower negative predictive value could possibly be explained because radiotherapy (administered regardless of PET results) was counted as an event and resulted in a lower EFS rate compared to other clinical trials. Zhao et al. [53] had a low percentage of negative interim PET scans and a high number of events, which explains the lower NPV.

The higher sensitivity values seen in ROC analysis for both IHP and custom criteria vs. the Deauville system may be explained by the lower threshold of test positivity with IHP vs. Deauville (using liver and blood pool activity as the reference tissue, respectively). None of the studies using custom criteria defined a threshold comparable to or higher than hepatic uptake. We found widely ranging positivity rates between studies, which are mainly in agreement with the timing of interim PET between cycles and the criteria used. In an exploratory analysis on five studies [34,37–39,47] that performed interim PET strictly after 2 cycles of therapy and applied the Deauville scoring system we found a pooled estimated HR of 3.48 (95% CI 2.46–4.93) with a corresponding 95% prediction interval of 1.58–7.67 (Supplemental Fig. 3). The positivity rates for these studies ranged between 18 and 46%, PPV from 37 to 74% and NPV from 76 to 91%, comparable to the analysis including all studies.

We chose to present the methodological characteristics along the other characteristics of the study population and treatments (Table 1) and along characteristics (including timing between cycles) of the index test (Table 2).

QUADAS-2 and QUIPS criteria were applied to assess the quality of the studies from the perspective of risk of bias and applicability. In this review, the strict inclusion and exclusion criteria with regard to patient population (>80% DLBCL), index test (interim PET between one and five treatment cycles), and reference standard (PFS and EFS) guaranteed the applicability of the results to the review question. In the subgroup analyses we examined whether bias could have occurred because of methodological shortcomings. It appeared that none of these affected the results. Only characteristics of the population (< 100% DLBCL) and a combination of integrated and standalone systems seemed to have impact on the predictive value of interim PET.

We used a comprehensive search strategy and applied strict inclusion and exclusion criteria. We focused on DLBCL patients, and 2-year PFS. Moreover, we examined the influence of different design characteristics (retrospective and prospective, blinded review or not reported; PFS or EFS), characteristics of patients (100% DLBCL or between 80 and 100%), treatments (ASCT upfront or not, preplanned or consolidative radiotherapy used or unknown), availability of a baseline PET or CT, properties of scans (PET/CT or a combination of PET/CT and PET

standalone systems), and scoring issues (DS -, IHP -, or custom criteria, central review or local review). Only the patient characteristics and properties of scans affected the results. It appeared that the HR estimates of the included studies were quite homogeneous (I^2 = 35%).

By contacting the authors we were able to include most of the eligible studies in our meta-analysis and deducting data that was not presented by the authors directly. Some data though were hard to obtain from the studies.

First of all, the definition of the start of the progression-free survival and event-free survival differed amongst studies. Some studies started their follow-up period at the time from diagnosis and others from initiation of first-line treatment. Recently some data has shown that patients who have a more aggressive disease tend to be treated earlier, so there could be selection bias between studies that have a shorter period between time of diagnosis and initiation of treatment versus studies with a longer period [54]. For future studies it seems important to have a comparable start of the follow-up period and authors should report the interval between diagnosis and start of the treatment to prevent or adjust for this risk of bias.

Another issue is that timing of the interim PET scans between cycles was different between studies; not only did the timing after which cycle the scan is performed differ, but also the number of days between the previous treatment course and interim PET. Unfortunately, not all authors report on this, although it is recommended to perform the scan at least 10 days after the previous course of chemotherapy, because of possible effects on tumor metabolism and systemic effects by, for example, growth factors [55].

In systematic reviews, investigators need to make choices. We chose to use the univariate data. This choice was made because univariate data were available in most studies and because of the large heterogeneity in factors for which the HR was adjusted in the primary articles. The adjusted factors were limited by the low number of events in most studies and partially based on available information such as quantitative PET analyses, immunohistochemistry and collection of specific clinical data (e.g. bone marrow involvement). Fourteen of the 20 studies performed a multivariate analysis. Most articles adjusted for the IPI score [34–

39,41–43,46] or age-adjusted IPI [44,48,49], some dichotomized the score and others used the individual components. Results were varying widely; in some studies both interim PET and (aa)IPI showed an independent association with PFS or EFS [42,48], others only for interim PET [34,37,39,41,44,53], or (aa) IPI [43,49] or no independent associations were found for both interim PET or (aa)IPI [35,36,38,46]. One could argue that reporting univariate HRs instead of multivariate HRs could result in an overestimation of the predictive value of interim PET. Three studies reported both uni and multivariate HRs and differences between univariate and multivariate HRs were −0.99 [41], 0.0 [39],and + 0.2 [42], respectively.

We further decided to choose the DS threshold for the interim response criteria which is most commonly described (DS < = 3 versus DS > = 4), because presenting all thresholds would increase heterogeneity, influence effect sizes, and finally use the same patients data multiple times in the analyses. Four studies presented multiple scores. Mylam et al. [43] published data about positivity for Deauville scores 4 and 5 as well as for Deauville score 5 and for IHP. Kim et al. [35] and Itti et al. [47] presented data about different positivity cutoff values for Deauville scores. Fuertes et al. [45] published a regular Deauville score as well as a 3 point-scale. In this review, we focused on visual response assessment criteria, and the potential added value of quantitative PET metrics is currently being investigated. Recently, a large phase III PET-adapted trial showed in a post-hoc analysis that a SUVmax reduction strategy [56] seems to discriminate better between good and poor outcome compared to the Deauville scoring system [57]. Finally, it should be mentioned that the studies from Safar et al. [50] and Itti et al. [47] had a small overlap in patient inclusion (n = 7); however, this will presumably not bias our results due to the small number.

Conclusion

This systematic review and meta-analysis shows that interim PET in DLBCL patients has predictive value (HR 3.13). However, some diagnostic test characteristics are still too low, especially the positive predictive value should be improved, before a risk stratified treatment approach can be implemented in clinical practice.

Acknowledgements

This work was financially supported by an Alpe d'Huzes/KWF fund provided by the Dutch Cancer Society (# VU 2012-5848). The authors would like to thank the authors of the primary studies that replied to the requests and kindly provided additional information or analyses.

Funding

This work was financially supported by an Alpe d'Huzes/KWF fund provided by the Dutch Cancer Society (# VU 2012–5848).

References

1. Pfreundschuh M, Trümper L, Osterborg A, et al. CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera international trial (MInT) group. *Lancet Oncol.* 2006;7:379–391.
2. Pfreundschuh M, Schubert J, Ziepert M, et al. Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomized controlled trial (RICOVER-60). *Lancet Oncol.* 2008;9:105–116.
3. Vitolo U, Trněný M, Belada D, et al. Obinutuzumab or rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in previously untreated diffuse large B-cell lymphoma. *J Clin Oncol.* 2017;35:3529–3537.
4. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993;329:987–994.
5. Ziepert M, Hasenclever D, Kuhnt E, et al. Standard international prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28:2373–2380. Erratum in: *J Clin Oncol.* 2011; 29: 779.
6. Sehn LH, Berry B, Chhanabhai M, et al. The revised international prognostic index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood.* 2007;109:1857–1861.
7. Zhou Z, Sehn LH, Rademaker AW, et al. An enhanced international prognostic index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood.* 2014;123:837–42.
8. Stiff PJ, Unger JM, Cook JR, et al. Autologous transplantation as consolidation for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 2013;369:1681–1690.
9. 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. Erratum in: *J Clin Oncol.* 2016; 34: 2562.
10. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET study group. *Eur J Cancer.* 1999;35:1773–1782.
11. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50(Suppl 1):122S–50S.
12. Juweid ME, Stroobants S, Hoekstra OS, et al. Imaging Subcommittee of International Harmonization Project in lymphoma. 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.
13. 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.
14. André MPE, Girinsky T, Federico 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;35:1786–1794.
15. Terasawa T, Lau J, Bardet S, et al. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. *J Clin Oncol.* 2009;27:1906–1914.
16. Zhu Y, Lu J, Wei X, Song S, Huang G. The predictive value of interim and final [18F] fluorodeoxyglucose positron emission tomography after rituximab-chemotherapy in the treatment of non- Hodgkin's lymphoma: a meta-analysis. *Biomed Res Int.* 2013;275805 https://doi.org/10.1155/2013/275805.
17. Sun N, Zhao J, Qiao W, Wang T. Predictive value of interim PET/CT in DLBCL treated with R-CHOP: meta-analysis. *Biomed Res Int.* 2015;648572 https://doi.org/10.1155/2015/648572.
18. Adams HJ, Kwee TC. Prognostic value of interim FDG-PET in RCHOP-treated diffuse large B-cell lymphoma: systematic review and meta-analysis. *Crit Rev Oncol Hematol.* 2016;106:55–63.

19. 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.
20. Zimmermann M, Oehler C, Mey U, Ghadjar P, Zwahlen DR. Radiotherapy for non-Hodgkin's lymphoma: still standard practice and not an outdated treatment option. *Radiat Oncol.* 2016;11:110. https://doi.org/10.1186/s13014-016-0690-y.
21. Delarue R, Tilly H, Mounier N, et al. Dose-dense rituximab-CHOP compared with standard rituximab-CHOP in elderly patients with diffuse large B-cell lymphoma (the LNH03-6B study): a randomised phase 3 trial. *Lancet Oncol.* 2013;14:525–533.
22. 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.
23. 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.
24. Coiffier B, Thieblemont C, Van Den Neste E, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood.* 2010;116:2040–2045.
25. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28:4184–4190. Erratum in: *J Clin Oncol.* 2012; 30: 1896
26. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529–536.
27. Hayden JA, van derWindt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med.* 2013;158:280–286.
28. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into metaanalysis. *Trials.* 2007;8:16.
29. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327:557–560.
30. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects metaanalyses. *BMJ.* 2011;342:d549. https://doi.org/10.1136/bmj.d549.
31. Egger M, Davey Smith G, Schneider M, Minder C. Bias in metaanalysis detected by a simple, graphical test. *BMJ.* 1997;315:629–634.
32. Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.
33. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org.
34. Fan Y, Zhang Y, Yang Z, et al. Evaluating early interim fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography with the SUV(max-liver)-based interpretation for predicting the outcome in diffuse large B-cell lymphoma. *Leuk Lymphoma.* 2017;58:1–9.
35. Kim J, Song YS, Lee JS, Lee WW, Kim SE. Risk stratification of diffuse large B-cell lymphoma with interim PET-CT based on different cutoff Deauville scores. *Leuk Lymphoma.* 2018;59:340–347.
36. de Oliveira Costa R, Hallack Neto A, Siqueira S, et al. Interim fluorine-18 fluorodeoxyglucose PET-computed tomography and cell of origin by immunohistochemistry predicts progression-free and overall survival in diffuse large B-cell lymphoma patients in the rituximab era. *Nucl Med Commun.* 2016;37:1095–1101.
37. Kong Y, Qu L, Li Y, Liu D, Lv X, Han J. Predictive significance of a new prognostic score for patients with diffuse large B-cell lymphoma in the interim-positron emission tomography findings. *Medicine (Baltimore).* 2016;95:e2808. https://doi.org/10.1097/MD 0000000000002808.
38. 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.

39. 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. Erratum in: *J Clin Oncol.* 2015; 33: 3074.
40. Zhang X, FanW, Xia ZJ, et al. Use of subsequent PET/CT in diffuse large B-cell lymphoma patients in complete remission following primary therapy. *Chin J Cancer.* 2015;34:70–78.
41. Carr R, Fanti S, Paez D, et al. IAEA lymphoma study group. Prospective international cohort study demonstrates inability of interim PET to predict treatment failure in diffuse large B-cell lymphoma. *J Nucl Med.* 2014;55:1936–1944.
42. Dabaja BS, Hess K, Shihadeh F, et al. Positron emission tomography/computed tomography findings during therapy predict outcome in patients with diffuse large B-cell lymphoma treated with chemotherapy alone but not in those who receive consolidation radiation. *Int J Radiat Oncol Biol Phys.* 2014;89:384–391.
43. Mylam KJ, Kostakoglu L, Hutchings M, et al. (18)Ffluorodeoxyglucose- 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.
44. Nols N, Mounier N, Bouazza S, et al. Quantitative and qualitative analysis of metabolic response at interim positron emission tomography scan combined with international prognostic index is highly predictive of outcome in diffuse large B-cell lymphoma. *Leuk Lymphoma.* 2014;55:773–780.
45. Fuertes S, Setoain X, Lopez-Guillermo A, et al. Interim FDG PET/CT as a prognostic factor in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging.* 2013;40:496–504.
46. González-Barca E, Canales M, Cortés M, et al. GELTAMO (Grupo Español de Linfoma y Trasplante de Médula Ósea). Predictive value of interim 18F-FDG PET/CT for event-free survival in patients with diffuse large B-cell lymphoma homogenously treated in a phase II trial with six cycles of R-CHOP-14 plus pegfilgrastim as first-line treatment. *Nucl Med Commun.* 2013;34:946–952.
47. 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.
48. Lanic H, Mareschal S, Mechken F, et al. Interim positron emission tomography scan associated with international prognostic index and germinal center B cell-like signature as prognostic index in diffuse large B-cell lymphoma. *Leuk Lymphoma.* 2012;53:34–42.
49. Pregno P, Chiappella A, Bellò M, et al. Interim 18-FDG-PET/CT failed to predict the outcome in diffuse large B-cell lymphoma patients treated at the diagnosis with rituximab-CHOP. *Blood.* 2012;119:2066–2073.
50. Safar V, Dupuis J, Itti E, et al. Interim [18F]fluorodeoxyglucose positron emission tomography scan in diffuse large B-cell lymphoma treated with anthracycline-based chemotherapy plus rituximab. *J Clin Oncol.* 2012;30:184–190.
51. Cashen AF, Dehdashti F, Luo J, Homb A, Siegel BA, Bartlett NL. 18F-FDG PET/CT for early response assessment in diffuse large Bcell lymphoma: poor predictive value of international harmonization project interpretation. *J Nucl Med.* 2011;52:386–392.
52. Zinzani PL, Gandolfi L, Broccoli A, et al. Midtreatment 18Ffluorodeoxyglucose positron-emission tomography in aggressive non-Hodgkin lymphoma. *Cancer.* 2011;117:1010–1018.
53. Zhao J, Qiao W, Wang C, Wang T, Xing Y. Therapeutic evaluation and prognostic value of interim hybrid PET/CT with (18)F-FDG after three to four cycles of chemotherapy in non-Hodgkin's lymphoma. *Hematology.* 2007;12:423–430.
54. Maurer MJ, Ghesquières H, Link BK, et al. Diagnosis-to-Treatment Interval Is an Important Clinical Factor in Newly Diagnosed Diffuse Large B-Cell Lymphoma and Has Implication for Bias in Clinical Trials. *J Clin Oncol.* 2018;36:1603–1610.
55. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. European Association of Nuclear Medicine (EANM). FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328–354.
56. 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.
57. Dührsen U, Müller S, Hertenstein B, et al. PETAL Trial Investigators. Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL): a multicenter, randomized phase III trial. *J Clin Oncol.* 2018, May 11. https://doi.org/10.1200/JCO.2017.76.8093.

Supplemental materials

Supplemental Table 1A Pubmed/MEDLINE search strategy

Search	Query	Items found
#1	"Lymphoma, Non-Hodgkin"[Mesh] OR Lymphoma*[tiab] OR Non-Hodgkin*[tiab] OR Non Hodgkin*[tiab] OR nonhodgkin*[tiab] OR NHL[tiab] OR DLBCL[tiab] OR Lymphoma*[ot] OR Non-Hodgkin*[ot] OR Non Hodgkin*[ot] OR nonhodgkin*[ot] OR NHL[ot] OR DLBCL[ot]	182131
#2	("Tomography, Emission-Computed"[Mesh:NoExp] OR "Positron-Emission Tomography"[Mesh] OR deoxyglucose[MeSH] OR deoxyglucose[tiab] OR desoxyglucose[tiab] OR deoxy-glucose[tiab] OR desoxy-glucose[tiab] OR deoxy-d-glucose[tiab] OR desoxy-d-glucose[tiab] OR 2deoxyglucose[tiab] OR 2deoxy-d-glucose[tiab] OR fluorodeoxyglucose[tiab] OR fluorodesoxyglucose[tiab] OR fludeoxyglucose[tiab] OR fluordeoxyglucose[tiab] OR fluordesoxyglucose[tiab] OR 18fluorodeoxyglucose[tiab] OR 18fluorodesoxyglucose[tiab] OR Fluoro-d-glucose[tiab] OR Fludeoxyglucose[tiab] OR Fluordeoxyglucose[tiab] OR 18fluordeoxyglucose[tiab] OR fdg*[tiab] OR 18fdg*[tiab] OR 18f-dg*[tiab] OR 18f-fdg[tiab] OR SUV[tiab] OR ((fluor[tiab] OR 2fluor*[tiab] OR fluoro[tiab] OR fluorodeoxy[tiab] OR fludeoxy[tiab] OR fluorine[tiab] OR 18f[tiab] OR 18flu*[tiab] AND glucose[tiab])) AND (pet[tiab] OR pet/*[tiab] OR petscan*[tiab] OR "Tomography, Emission-Computed"[Mesh:NoExp] OR "Positron-Emission Tomography"[Mesh] OR (emission[tiab] AND tomograph[tiab] OR tomographs[tiab] OR tomographic*[tiab] OR tomography[tiab] OR tomographies[tiab] OR scan[tiab]))) OR (deoxyglucose[ot] OR desoxyglucose[ot] OR deoxy-glucose[ot] OR desoxy-glucose[ot] OR deoxy-d-glucose[ot] OR desoxy-d-glucose[ot] OR 2deoxyglucose[ot] OR 2deoxy-d-glucose[ot] OR fluorodeoxyglucose[ot] OR fluorodesoxyglucose[ot] OR fludeoxyglucose[ot] OR fluordeoxyglucose[ot] OR fluordesoxyglucose[ot] OR 18fluorodeoxyglucose[ot] OR 18fluorodesoxyglucose[ot] OR Fluoro-d-glucose[ot] OR Fludeoxyglucose[ot] OR Fluordeoxyglucose[ot] OR 18fluordeoxyglucose[ot] OR fdg*[ot] OR 18fdg*[ot] OR 18f-dg*[ot] OR 18f-fdg[ot] OR SUV[ot] OR ((fluor[ot] OR 2fluor*[ot] OR fluoro[ot] OR fluorodeoxy[ot] OR fludeoxy[ot] OR fluorine[ot] OR 18f[ot] OR 18flu*[ot] AND glucose[ot])) AND (pet[ot] OR pet/*[ot] OR petscan*[ot] OR "Tomography, Emission-Computed"[Mesh:NoExp] OR "Positron-Emission Tomography"[Mesh OR (emission[ot] AND tomograph[ot] OR tomographs[ot] OR tomographic*[ot] OR tomography[ot] OR tomographies[ot] OR scan[ot]))))	77819
#3	#1 AND #2	4042

Supplemental Table 1B Embase search strategy

Search	Query	Items found
#1	'nonhodgkin lymphoma'/exp OR lymphoma*:ab,ti OR (non NEXT/1 hodgkin*):ab,ti OR nonhodgkin*:ab,ti OR nhl:ab,ti OR dlbcl:ab,ti	258953
#2	'emission tomography'/de OR 'positron emission tomography'/exp OR 'whole body pet'/exp OR 'deoxyglucose'/exp OR deoxyglucose:ab,ti ORdesoxyglucose:ab,ti OR 'deoxy-glucose':ab,ti OR 'desoxy glucose':ab,ti OR 'deoxy-d-glucose':ab,ti OR 'desoxy-d-glucose':ab,ti OR2deoxyglucose:ab,ti OR '2deoxy-d-glucose':ab,ti OR fluorodeoxyglucose:ab,ti OR fluorodesoxyglucose:ab,ti OR fluordesoxyglucose:ab,ti OR 18fluorodeoxyglucose:ab,ti OR 18fluorodesoxyglucose:ab,ti OR 'fluoro d glucose':ab,ti OR fludeoxyglucose:ab,ti OR fluordeoxyglucose:ab,ti OR 18fluordeoxyglucose:ab,ti OR fdg*:ab,ti OR 18fdg*:ab,ti OR '18f-dg':ab,ti OR '18f fdg':ab,ti OR suv:ab,ti OR (fluor:ab,ti OR 2fluor*:ab,ti ORfluoro:ab,ti OR fluorodeoxy:ab,ti OR fludeoxy:ab,ti OR fluorine:ab,ti OR 18f:ab,ti OR 18flu*:ab,ti AND glucose:ab,ti) AND (pet*:ab,ti OR 'emission tomography'/de OR 'positron emission tomography'/exp OR 'whole body pet'/exp OR (emission:ab,ti AND (tomograph:ab,ti OR tomographs:ab,ti OR tomographic*:ab,ti OR tomography:ab,ti OR tomographies:ab,ti OR scan:ab,ti)))	136600
#3	#1 AND #2	9470
#4	#3 AND ('article'/it OR 'article in press'/it OR 'review'/it)	5648

Supplemental Table 1C Cochrane Library search strategy

Search	Query	Items found
#1	Lymphoma* or "Non-Hodgkin*" OR "Non Hodgkin*" OR nonhodgkin* OR NHL OR DLBCL:ti,ab,kw (Word variations have been searched)	7736
#2	Tomography OR deoxyglucose OR desoxyglucose OR deoxy-glucose OR desoxy-glucose OR deoxy-d-glucose OR desoxy-d-glucose OR 2deoxyglucose OR 2deoxy-d-glucose OR fluorodeoxyglucose OR fluorodesoxyglucose OR fludeoxyglucose OR fluordeoxyglucose OR fluordesoxyglucose OR 18fluorodeoxyglucose OR 18fluorodesoxyglucose OR Fluoro-d-glucose OR Fludeoxyglucose OR Fluordeoxyglucose OR 18fluordeoxyglucose OR fdg* OR 18fdg* OR 18f-dg* OR 18f-fdg OR SUV OR PET*:ti,ab,kw (Word variations have been searched)	26052
#3	#1 AND #2	502

Supplemental Table 2 PFS/EFS definitions.

First author (year)	PFS/EFS; definition
Fan et al (2017)[34]	PFS; was calculated *from the date of acquisition of biopsy results* to disease progression, relapse, death of patients from any causes, or the date of last follow-up for surviving ones.
Kim et al (2017)[35]	PFS; was calculated *from the start of the treatment* to disease progression, recurrence or death.
De Oliveira Costa et al (2016)[36]	PFS; was defined as the *time from the date of diagnosis* to the date of disease progression, relapse, or death as a result of any cause or last patient follow-up.
Kong et al (2016)[37]	PFS; was defined as the interval between the *date of diagnosis* and the date of lymphoma progression, first relapse, death from any cause, or the last follow-up date.
Mikhaeel et al (2016)[38]	PFS; defined as the *time from diagnosis* to the point of progression or death from any cause. Patients still alive were censored at the date of last contact.
Mamot et al (2015)[39]	EFS; The primary end point was 2-year EFS *from start of treatment*. Patients who progressed, relapsed, switched to other treatments (including concomitant radiotherapy), refused to continue trial treatment, or died within 2 years were considered to have treatment failure.
Zhang et al (2015)[40]	PFS; was defined as the *time from the start of treatment* to the progression of lymphoma, death from any cause, or last follow-up.
Carr et al (2014)[41]	EFS; Study events were relapse after complete remission; death from any cause; treatment escalation for progressive disease while on treatment, and disease progression or failure to achieve complete remission at end-chemotherapy based on the revised response criteria for PET, with confirmation by biopsy that residual or increased ^{18}F-FDG uptake was due to lymphoma. Cases lost to follow-up were censored at date of last known disease status. *Date of first treatment as origin*.
Dabaja et al (2014)[42]	PFS; was defined as the *time from diagnosis* until objective tumor progression or death.
Mylam et al (2014)[43]	PFS; was defined as the *time from diagnosis* to DLBCL progression or death from any cause. Patients who were still alive at the end of the study were censored at the date of data collection.
Nols et al (2014)[44]	PFS; was defined as the *time from study entry* until disease progression or death due to any cause.
Fuertes et al (2013)[45]	PFS; was defined as the interval *from the start of treatment* until disease progression of DLBCL, death from any cause or the most recent follow-up.
Gonzalez-Barca et al (2013)[46]	EFS; An event was defined as follows: nonachievement of a CR or Cru with treatment, relapse after achievement of complete remission, or death from any cause, whichever came first.
Itti et al (2013)[47]	PFS; calculated *from the date of diagnosis* until relapse with censoring at the time of last follow-up.
Lanic et al (2012)[48]	PFS; was calculated *from the date of enrollment* until disease progression, relapse or death (from any cause) or last patient follow-up.
Pregno et al (2012)[49]	PFS; was defined as the time *from the start of treatment* to death/progression as a result of any cause; patients still alive were censored at the date of last contact.
Safar et al (2012)[50]	PFS; was calculated *from the date of enrollment* until disease relapse, with censoring at the time of last follow-up.
Cashen et al (2011)[51]	PFS; defined to be the time interval *from enrollment in the study* to date of relapse.
Zinzani et al (2011)[52]	EFS; was defined as the interval *from the date of enrollment* to the first evidence of progression, disease relapse, death from any cause, treatment discontinuation due to any adverse event, or patient withdrawal.
Zhao et al (2007)[53]	PFS; was defined as the *time from diagnosis* to first evidence of progression or relapse, or to disease related death. Data were censored at other causes of death or if the patients were free of progression/relapse at follow-up.

Supplemental Table 3A Meta-regression analyses. Prognostic.

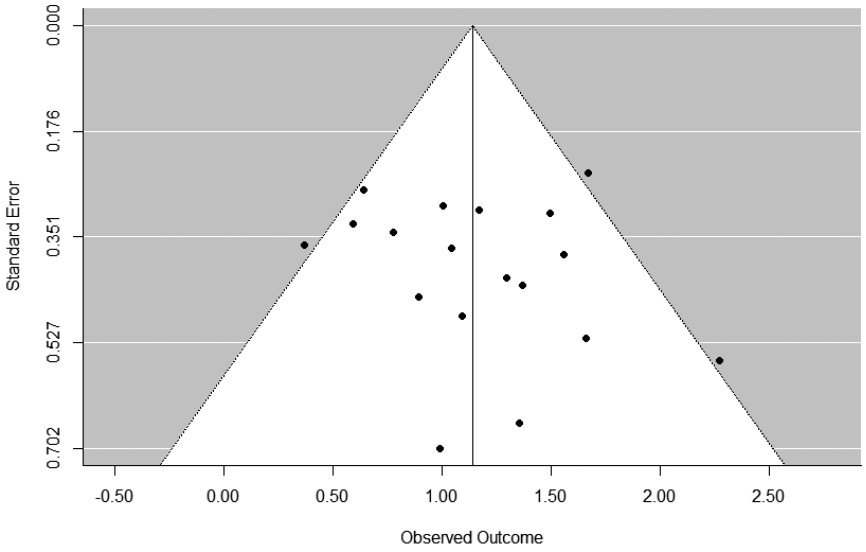
Subgroup	Reference	No of studies	HR of reference group	*P* value
	Control			
Study design	prospective	6	3·17 (95%CI 2·13-4·73)	0·9369
	retrospective	12		
% DLBCL	100% DLBCL	15	2·84 (95% CI 2·29-3·52)	0·0577
	80-99% DLBCL	3	4·55 (95% CI 2·94-7·03)	
Visual Criteria	Deauville	11	3·21 (95% CI 1·51 -5·19)	0·6997
	IHP	3		
Scanner*	PET/CT	12	2·85 (95%CI 2·28-3·56)	0·0332
	PET/CT + standalone	5	4·39 (95% CI 3·15-6·10)	
Radiotherapy	yes	9	2·88 (95%CI 2·13-3·89)	0·4069
	unknown/no	8/1		
ASCT upfront	yes	2	4·96 (95% CI 2·40-10·21)	
	no	5		0·2073
	unknown	11		0·2126
Blinded review	yes	9	3·12 (95%CI 2·31-4·21)	0·9653
	unknown/no	8/1		
Outcome measure	PFS	14	2·87 (95% CI 2·28-3·62)	0·1495
	EFS	4		

Supplemental Table 3B Meta-regression analyses. Diagnostic.

Subgroup	Reference	No of studies	Moderator	No of studies	*P* value
Study design	prospective	7	retrospective	12	NS
% DLBCL	100% DLBCL	16	80-99% DLBCL	3	0·0503 (accuracy)
Visual Criteria	Deauville	12	IHP	3	NS
Scanner*	PET/CT	12	PET/CT + standalone	6	NS
Radiotherapy	yes	10	unknown/no	9	NS
ASCT upfront	yes	2	no	5	NS
			unknown	12	NS
Blinded review	yes	10	unknown/no	9	NS
Outcome measure	PFS	15	EFS	4	NS

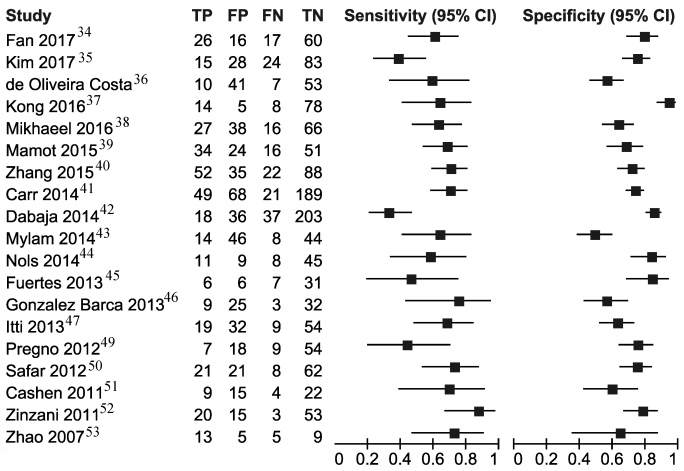
NS= not significant

*No information about type of PET system for one study

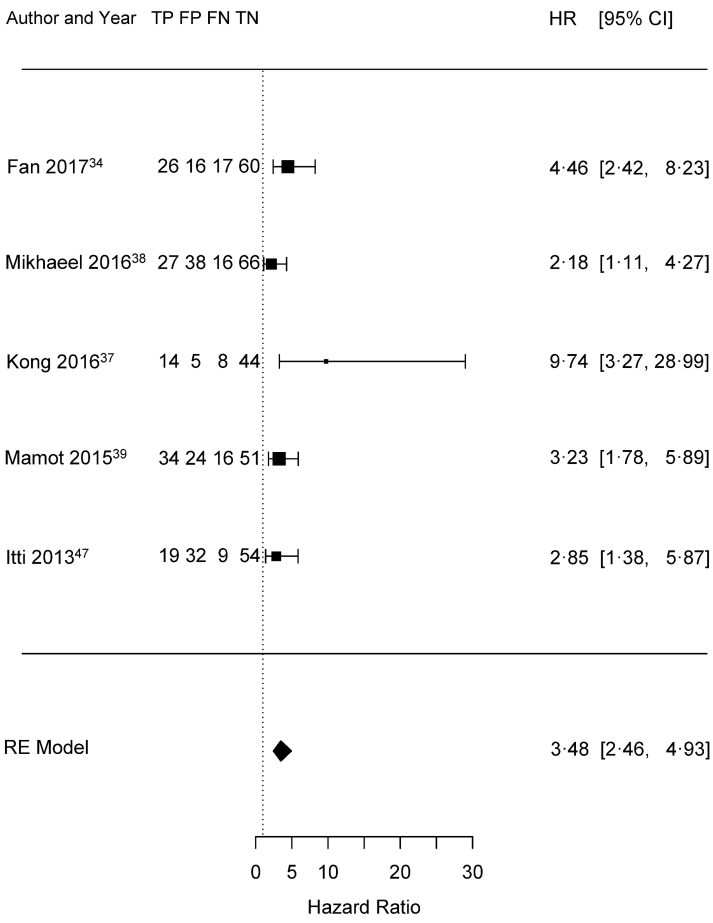


Supplemental Fig. 1 Funnel plot of studies investigating interim PET in DLBCL.

This plot shows the individual studies (black spheres) sorted by effect size (log HR) presented on the X-axis and standard error on the Y-axis. The solid vertical line corresponds to the estimated pooled log HR.



Supplemental Fig. 2 Forest plots of sensitivity and specificity at two years of follow-up.



Supplemental Fig. 3 Forest plot of univariate hazard ratios for interim PET scans after 2 cycles and assessed according to Deauville in diffuse large B-cell lymphoma.

PART II

Technical validation of PET in lymphoma

CHAPTER 4

Interobserver Agreement of Interim and End-of-Treatment ^{18}F-FDG PET/CT in Diffuse Large B-Cell Lymphoma: Impact on Clinical Practice and Trials.

Coreline N. Burggraaff
Otto S. Hoekstra
Bart de Keizer
Filiz Celik
Henrica C.W. de Vet
Alexander C. Cornelisse
Pieternella J. Lugtenburg
Anne I.J. Arens
Julia E. Huijbregts
Josée M. Zijlstra

On behalf of the HOVON Imaging Working Group

Journal of Nuclear Medicine 2018; 59: 1831-1836.

Abstract

We aimed to assess the interobserver agreement of interim PET (I-PET) and end-of-treatment PET (EoT-PET) using the Deauville score (DS) in first-line diffuse large B-cell lymphoma (DLBCL) patients.

Methods

I-PET and EoT-PET scans of DLBCL patients were performed in the HOVON84 study (2007–2012), an international multicenter randomized controlled trial. Patients received R-CHOP14 and were randomized to receive rituximab intensification in the first 4 cycles or not. I-PET was performed after 4 cycles (for observational purposes), and EoT-PET after 6 or 8 cycles. Two independent central reviewers retrospectively scored all scans according to the DS system, masked to clinical outcomes. Results were dichotomized as negative (DS of 1–3) or positive (DS of 4–5). Besides percentage overall agreement (OA), we calculated agreement for positive and negative scores, expressed as positive agreement (PA) and negative agreement (NA), respectively.

Results

465 I-PET and 457 EoT-PET scans were centrally reviewed; baseline ^{18}F-FDG PET or PET/CT was available in 75%–77%, and CT in the remaining cases. Percentage OA for I-PET and EoT-PET were 87.7% and 91.7% (P = 0.049), with NA of 92.0% and 95.0% (P = 0.091), and PA of 73.7% and 76.3% (P = 0.656), respectively.

Conclusion

Interobserver agreement using DS in DLBCL patients in I-PET and EoT-PET yields high OA and NA. The lower PA suggests that EoT-PET/CT treatment evaluation in daily practice and I-PET–adapted trials may benefit from dual reads and central review, respectively.

Key Words

observer variation; positron emission tomography; DLBCL; Deauville score; Lugano criteria

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of malignant lymphoma, accounting for 30%–40% of non-Hodgkin lymphomas [1]. Current international guidelines [2,3] recommend ^{18}F-FDG PET before therapy in typically ^{18}F-FDG– avid lymphoma types—for example, Hodgkin lymphoma and DLBCL [4]—and to apply the Lugano response classification based on the Deauville score (DS) on a 5-point scale at the end of treatment. Application of ^{18}F-FDG PET during therapy (interim-PET, or I-PET) allows PET-guided patient management, with success in Hodgkin lymphoma [5–8]. In DLBCL, the value of I-PET is less clear [9]: most I-PET–adapted trials in DLBCL did not demonstrate a strategy that overcomes treatment resistance [10], except for a phase II study with intensification after rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) at a 14-d cycle (R-CHOP14) in I-PET–positive patients to R-ICE and Z-BEAM autologous stem cell transplantation [11]. Therefore, I-PET is currently not used in clinical practice. An important prerequisite for these PET-guided studies is a consistent classification of the I-PET scans into a positive or negative category.

Similar to other disciplines, observer variation is the Achilles' heel of radiology [12]. In the DS scoring system, the ^{18}F-FDG uptake in potentially malignant tissue is rated versus normal ^{18}F-FDG distribution in mediastinal blood pool and liver. Such a semiquantitative approach is less prone to observer variation than visual readings purely based on perception, knowledge, experience, and pattern recognition [12], possibly influenced by optical illusion effects [13]. There are few studies on interobserver agreement of DS in PET scans in DLBCL patients treated with rituximab-containing chemotherapy, reporting a 0.4–0.8 range of κ-values for I-PET [14–16] and 0.5 for DS in end-of-treatment PET (EoT-PET) [15].

In clinical practice and trials, it is essential to know the specific agreement, that is, the absolute probability of obtaining the same test result by different reviewers rating the same scan. In cases of I-PET–driven treatment escalation on a positive I-PET scan or I-PET– driven treatment deescalation on a negative I-PET scan, observer variation–driven misclassification might compromise the results of clinical trials or induce overtreatment or undertreatment, respectively.

Our primary objective was to assess the interobserver agreement of I-PET and EoT-PET using DS in a large randomized clinical trial in DLBCL patients. Secondary objectives were to identify potential sources of observer variation (timing of PET— that is, I-PET vs. EoT-PET; baseline imaging modality— CT vs. PET or PET/CT, and the site of residual tracer uptake). The results are reported in accordance with Guidelines for Reporting Reliability and Agreement studies [17].

Materials and methods

Study Population

I-PET and EoT-PET scans were collected from the HOVON84 study, an international multicenter phase 3 trial in DLBCL (EudraCT 2006-005174-42). Patients were enrolled from 69 hospitals in The Netherlands, Belgium, and Denmark between November 2007 and April 2012. The main inclusion criteria were newly diagnosed, histologically proven, CD-20–positive DLBCL patients with Ann Arbor stage II–IV, age 18–80 y, and World Health Organization performance status 0–2. Exclusion criteria were primary central nervous system, testicular, transformed indolent, and primary mediastinal B-cell lymphoma, as well as posttransplant lymphoproliferative disease. Patients were randomized between standard R-CHOP14 or R-CHOP14 with intensification of rituximab in the first 4 cycles (R2- CHOP14). Administration of granulocyte colony-stimulating factor was mandatory and served to oppose the neutropenic side effects of the R-CHOP14 scheme. Six milligrams of pegfilgrastim were injected subcutaneously on day 2 of each R-CHOP cycle. I-PET and EoT-PET were performed after 4 cycles of therapy for observational purposes only, and after 6 (patients aged >65 y) or 8 cycles (patients ≤60 y), respectively. Baseline PET was recommended but not mandatory. HOVON84 has been approved by the institutional review board, and all subjects signed an informed consent form for use of their data for scientific purposes.

Image Analysis

DS was used for central image review [2]. Between 2013 and 2016, each I-PET and EoT-PET scan was read independently by 2 reviewers from a pool of 10, who randomly drew scans from the image warehouse. All PET and CT scans

were anonymized and uploaded to a database server hosted by Keosys (Imagys), allowing reviewers to read the images in their own workspaces. Seven percent of the PET scans performed in the HOVON84 trial were done with dedicated PET scanners, but this analysis of interobserver agreement was limited to PET/CT examinations. Reviewers were experienced nuclear medicine physicians (>5 y of experience with response evaluation of lymphoma in academic or large peripheral hospitals), actively participating in the HOVON Imaging Working Group. They were masked to clinical follow-up and randomization arm. Reviewers had access to all baseline imaging data (electronic case records containing clinical and imaging staging information provided by local clinicians and image reviewers). For the trial, discrepancies between the 2 reviewers were adjudicated by a third reviewer.

Reviewers used an electronic case record with prespecified nodal localizations (specifying regions as Waldeyer's ring, cervical, supraclavicular, axillary, mediastinum, hilar, paraaortic, mesenteric, spleen, iliac, inguinal, and other) and extranodal locations (gastrointestinal, central nervous system, skin, liver, lung, pleural, skeletal, and other). Open text fields were available for explanation of difficulties in reading. Reviewers assigned a DS for individual nodal and extranodal localizations together with a final patient-based score (highest lesional DS). We analyzed the DS of I-PET and EoT-PET as ordinal as well as dichotomized scores (DS 1–3 considered negative, DS 4–5 positive) [2].

Statistical Analysis

We performed patient- and region-based analyses. Besides the percentage overall agreement (OA), we calculated the percentage specific agreement, separating positive agreement (PA) from negative agreement (NA). PA and NA were defined as the probability that, if one reviewer assigns a positive or negative score, respectively, a second reviewer scores positive or negative as well [18]. The prevalence of positive scans was calculated as the sum of the number of scans in which both reviewers scored positive and half the scans with discrepancies divided by the total number of scans. We analyzed the following potential sources of observer variation: I-PET and EoT-PET; availability of a baseline PET, PET/CT, or CT scan for reference; and residual ^{18}F-FDG uptake in different nodal and extranodal localizations. Discrepancies in these specific sites were related to baseline lymphoma prevalence, to assess which localizations were most difficult to read. In addition, we checked the assumption that there was no difference in

observer variation between the control and intervention arms. For comparison of the percentage OA, PA, and NA between groups, the $\chi 2$ test was used. A *P* value of less than 0.05 was considered statistically significant. Statistical analyses were performed using SPSS Statistics (IBM, version 20).

Results

Study Population

In total, 575 patients were eligible for the final analysis of the main trial [19]. I-PET and EoT-PET evaluation was performed in 534 and 517 patients, respectively (Supplemental Figs. 1 and 2; supplemental materials are available at http://jnm.snmjournals.org). For 7 patients, no PET data were received from the hospitals, 38 I-PET and 34 EoT-PET scans were performed on a stand-alone PET scanner, and 11 I-PET and 7 EoT-PET scans were not accessible in DICOM format or contained incomplete series.

I-PET

Table 1 summarizes the results of the interobserver agreement for I-PET with dichotomized DS. We obtained 465 evaluable scan results, because in 13 I-PET scans one of the reviewers did not provide a DS rating. The median time interval of I-PET scanning after the last chemotherapy cycle was 11 d (interquartile range, 9– 13 d). In 408 of 465 scans, the reviewers agreed on the final conclusion (negative or positive), yielding a percentage OA of 87.7% (95% confidence interval [95%CI], 84.7–90.8). The prevalence of positive I-PET scans was 23.3%. The NA, at 92% (95%CI, 89.1–95.0), was markedly higher than the PA, at 73.7% (95%CI, 65.0–82.5).

A baseline ^{18}F-FDG PET or PET/CT scan was available in 77% (*n* = 349 integrated PET/CT scan and *n* = 8 PET stand-alone with a separate CT scan), and diagnostic CT in the remaining cases (*n* = 108). Percentage OA, NA, and PA were not statistically significant between these groups (percentage OA, 88% and 87%, *P* = 0.799; NA, 92% and 92%, *P* = 0.947; PA, 75.7% and 65%; *P* = 0.347). Percentage OA was similar in both treatment arms (*P* = 0.606).

Table 1. Interobserver agreement on dichotomized DS, by baseline modality.

	I- PET/CT				**EoT- PET/CT**			
Agreement	Total (*n*=465)	Baseline CT only (*n*=108)	Baseline ^{18}F-FDG PET or PET/CT (*n*=357)	*P**	Total (*n*=457)	Baseline CT only (*n*=114)	Baseline ^{18}F-FDG PET or PET/CT (*n*=343)	*P**
Positivity†	23.3	18.5	24.8		17.5	18.0	17.3	
Percentage OA	87.7 (84.7-90.8)	87.0 (80.2-93.8)	88.0 (84.4-91.5)	0.799	91.7 (89.0-94.3)	93.9 (89.0-98.7)	91.0 (87.8-94.1)	0.332
Percentage PA	73.7 (65.0-82.5)	65.0 (41.6-88.4)	75.7 (66.2-85.2)	0.347	76.3 (66.3-86.2)	82.9 (59.2-90.8)	73.9 (62.0-85.9)	0.486
Percentage NA	92.0 (89.1-95.0)	92.0 (85.8-98.3)	92.0 (88.6-95.4)	0.947	95.0 (92.6-97.3)	96.3 (89.5-97.9)	94.5 (91.7-97.4)	0.605

DS 1-3= negative; DS 4-5= positive.

**P*-values of $\chi 2$ test refer to comparison of baseline CT vs. baseline PET or PET/CT.

†Prevalence of positive scans was calculated as sum of number of scans in which both reviewers scored positively and half of scans with discrepancies divided by total number of scans.

Data are percentages, with 95%CIs in parentheses.

For ordinal DSs, the reviewers agreed in 214 of 465 cases (Supplemental Table 1), resulting in 46% exact agreement (95%CI, 41.4–50.7). Percentage agreement was 78.3% (95%CI, 74.4–82.1) when we allowed a 1-point difference—except for a discrepancy between scores 3 and 4—between the reviewers' scorings.

Table 2 presents the percentage OA for the specific nodal and extranodal localizations for dichotomized DS, related to the baseline prevalence. Gastrointestinal, Waldeyer's ring, skeletal, spleen, and mesenteric sites showed a relatively large number of discrepancies. An example of a discrepancy in the interim assessment of a mesenteric bulky lesion is shown in Figure 1.

Table 2. Interobserver agreement on specific nodal and extranodal localizations on I-PET.

Location	Number baseline positive	Number of discrepancies on I-PET	Agreement on negativity (absolute)	Agreement on positivity (absolute)	Percentage OA	Related to baseline prevalence
Nodal						
Paraaortic*	414	17	899	14	98.2	4.1%
Cervical*†	302	8	915	6	99.1	2.6%
Iliac*	272	6	917	7	99.4	2.2%
Supraclavicular*	228	6	920	4	99.4	2.6%
Axillary*	225	9	920	1	99.0	4.0%
Mediastinal†	212	12	445	6	97.4	5.7%
Inguinal*	210	3	926	1	99.7	1.4%
Mesenteric	189	16	433	16	96.6	8.5%
Hilar*†	147	7	918	3	99.2	4.8%
Spleen†	115	11	442	6	97.6	9.6%
Other	105	7	457	1	98.5	6.7%
Waldeyer†	53	8	456	0	98.3	15.1%
Extranodal						
Other extranodal†	124	17	436	8	96.3	13.7%
Skeletal†	95	12	447	4	97.4	12.6%
Gastrointestinal†	61	12	441	7	97.4	16.7%
Lung†	55	3	455	6	99.4	5.5%
Liver	37	3	461	1	99.4	8.1%
Pleura	25	1	464	0	99.8	4.0%
Skin	11	0	465	0	100.0	0.0%
Central nervous system	0	0	465	0	100.0	0.0%

*Right and left are summed and presented together.
†Totals not 465 or 930, because of missing values or localization scored as unclear.
Percentage OA = (number of agreement on positivity + number of agreement on negativity) / (number of discrepancies + number of agreement on positivity + number of agreement on negativity) x 100%; related to baseline prevalence = (number of discrepancies/number baseline positive) x 100%.

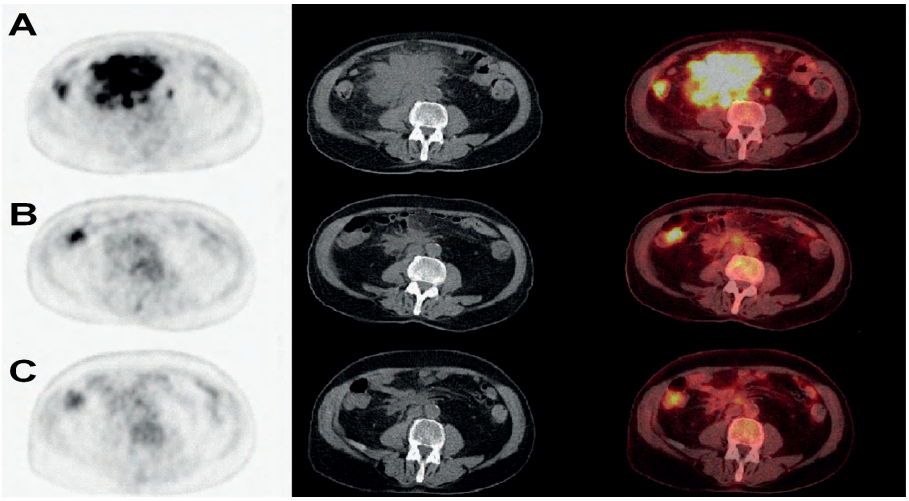


Figure 1. Example of discrepancy between reviewers' assessment of mesenteric lymph nodes on, from left to right, axial attenuation-corrected PET, low-dose CT, and fused PET/CT images.

(A) Baseline ^{18}F-FDG PET/CT with mesenteric bulky mass. (B) I-PET/CT after 4 cycles of R-CHOP14. One reviewer scored scan negatively (DS 1) and the other reviewer scored DS 4 for residual uptake in mesenteric mass. (C) EoT-PET/CT after 6 cycles of R-CHOP14. Both reviewers scored scan negatively (DS 1 and DS 2, respectively).

EoT-PET

Because in 10 EoT-PET scans one reviewer, and in 2 scans both, did not give a final conclusion, 457 scans were evaluable (Table 1). The median interval of EoT-PET scanning after the last chemotherapy cycle was 31 d (interquartile range, 22.5–48). The prevalence of positive EoT-PET scans was 17.5%. In 419 of 457 scans, the reviewers agreed on the final conclusion (negative or positive), yielding a percentage OA of 91.7% (95%CI, 89.0–94.3), a PA of 76.3% (95%CI, 66.3–86.2), and an NA of 95% (95%CI, 92.6–97.3).

Baseline ^{18}F-FDG PET or PET/CT was available in 75% (n = 333 integrated PET/CT, and n = 10 PET stand-alone with a separate CT scan), and diagnostic CT was available in the remaining cases (n = 114). Percentage OA, NA, and PA did not significantly differ between these groups (percentage OA, 91% and 93.3%, P = 0.332; NA, 94.5% and 96.3%, P = 0.605; PA, 73.9% and 82.9%; P = 0.486). Percentage OA for R2-CHOP14 compared with RCHOP14 was 93.9 versus 89.4% (P = 0.082).

For ordinal DSs, the reviewers agreed in 220 of 457 cases (Supplemental Table 2), resulting in a percentage of exact agreement of 48.1% (95%CI, 43.4–52.8). Percentage agreement was 83.4% (95%CI, 79.8–86.9) when we allowed a 1-point difference— except for a discrepancy between scores 3 and 4— between the reviewers' scorings.

Supplemental Table 3 presents the percentage OA of the specific nodal and extranodal locations for dichotomized DS, related to the baseline prevalence. Gastrointestinal, skeletal, and mesenteric sites relatively showed the greatest number of discrepancies. Observer variation at EoT-PET in spleen and Waldeyer's ring was less than at I-PET.

Comparison I-PET and EoT-PET Interobserver Agreement

PA did not significantly differ between I-PET and EoT-PET assessments (P = 0.656), but percentage OA was lower for I-PET (87.7% vs. 91.7%, respectively, P = 0.049), and there was a trend toward lower NA (92.0% vs. 95.0%, respectively, P = 0.091).

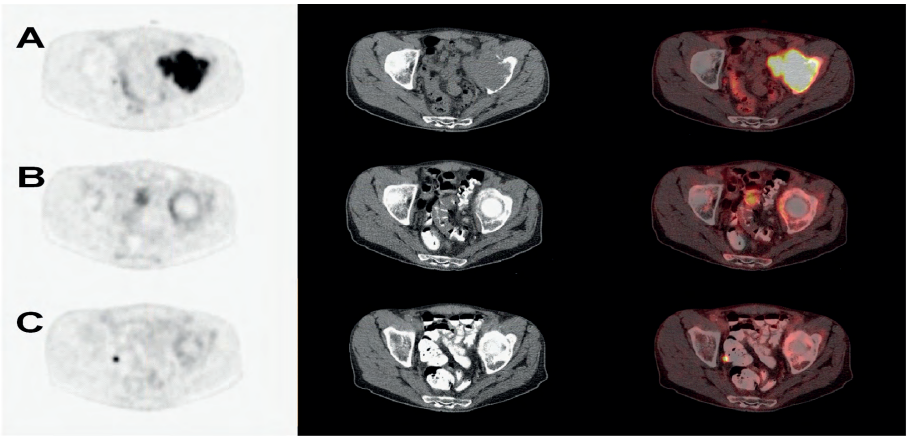


Figure 2. Example of discrepancy between reviewers' assessment of skeletal lesion on, from left to right, axial attenuation-corrected PET, low-dose CT, and fused PET/CT images.

(A) Baseline ^{18}F-FDG PET/CT with skeletal lesion in left acetabulum. (B) I-PET/CT after 4 cycles of R-CHOP14 showing rim of uptake scored by one reviewer as DS 4 and by other reviewer as unclear. (C) EoT-PET/CT after 8 cycles of R-CHOP14 showing residual uptake scores by one reviewer as DS 4 and by other reviewer as unclear.

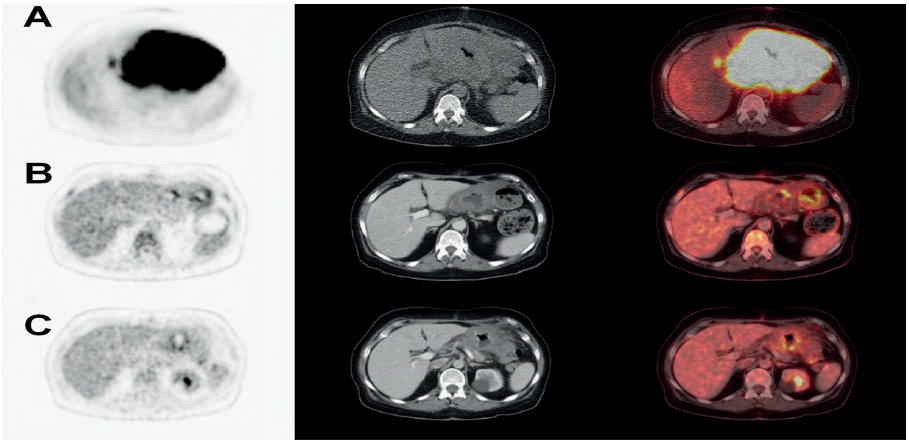


Figure 3. Example of discrepancy between reviewers' assessment of stomach on, from left to right, axial attenuation-corrected PET, low-dose CT, and fused PET/CT images.

(A) Baseline ^{18}F-FDG PET/CT with clear localization of lymphoma in stomach. (B) I-PET/CT after 4 cycles of R-CHOP14. Reviewer 1 did not give final DS score and commented that stomach was "DS 4 but could be physiologic uptake". Reviewer 2 scored this scan negatively (DS 2). (C) EoT-PET/CT after 6 cycles of R-CHOP14. Reviewer 1 still commented on stomach but now scored negatively. Reviewer 2 again scored scan negatively (DS 2).

Discussion

Our study presents the interobserver agreement of DS results for I-PET and EoT-PET from a central review of a large multicenter randomized clinical trial in DLBCL. We found high percentages of OA (88%–92%) and NA (92%–95%) for both I-PET and EoT-PET using a DS of at least 4 for test positivity, at a lower (74%–76%) PA.

Most studies on interobserver agreement primarily report Cohen's κ and some present percentage of OA in addition. Cohen's κ is a relative measure, and the values are low in relatively low-prevalence situations (e.g., of residual lymphoma sites). Therefore, we report percentage OA, which is independent of differences in prevalence. In addition, we report specific agreement measures, which reflect the absolute probability that another reviewer gives the same conclusion as a colleague, specified for positive and negative test results [18]. In other words: 74% PA implies that if one reviewer rates an I-PET scan as positive, the probability that another reviewer will provide the same result is 74%.

Similar studies presented different agreement measures [14–16]. Itti et al. [14] (*n* = 114, 3 readers) reported pairwise Cohen's κ-values (0.53–0.80) for I-PET after 2 cycles of R-CHOP or R-ACVBP in a retrospective cohort but did not report specific agreement measures. From their presented data, we calculated OA of 77%–90% between observer pairs, subdivided into a NA of 81%–91% and PA of 72%–89%. Horning et al. [16] (*n* = 38 patients, 3 readers) reported a Fleiss' κ of 0.50 and percentage OA of 71% for I-PET after 3 cycles of R-CHOP, but specific agreement measures could not be extracted. Han et al. [15] reported κ-values of 0.41–0.52 and OAs of 82%–88% for I-PET (*n* = 55, after 3 cycles of R-CHOP) and EoT-PET (*n* = 57), respectively, as assessed by 2 readers. NA and PA as extracted from their presented data were 89% and 50% for I-PET and 92% and 59% for EoT-PET, respectively.

Taken together, it appears that NA was generally above 80% in all studies (probably at least partly related to the high prevalence of negative scans). However, PA seemed to have a wider range between studies. In our study, we found a Cohen's κ -value of 0.65 and 0.71 for I-PET and EoT-PET, respectively.

Our data suggest that I-PET is more difficult to assess than EoT-PET. We found that the percentage of OA was lower for I-PET than for EoT-PET. The trend toward a lower NA for I-PET than for EoT-PET, could (in part) be caused by the higher number of negative scans at the end of treatment. In the study from Han et al., agreement measures also seemed generally higher for EoT-PET than for I-PET [15]. Treatment-related inflammation shortly after chemotherapy might hamper the identification of lymphoma-related ^{18}F-FDG uptake.

In addition, we explored observer agreement as a function of disease location. Related to initially involved sites, we found the lowest percentages of OA for mesenteric, gastrointestinal, and skeletal sites in I-PET and EoT-PET. In these tissues, the local background of ^{18}F-FDG varies between and within patients over time; uptake due to intercurrent inflammation and, for example (healing), pathologic fractures needs to be accounted for, and this is not always covered by the Lugano criteria. In I-PET, discrepancies in spleen and Waldeyer's ring were more common. The short interval between the I-PET exams after the previous R-CHOP14 course [20] and the recent administration of granulocyte colony stimulating factor [21] in our study could cause false-positive uptake in these

organs. In an intra- and interobserver agreement study of baseline PET/CT from a mix of lymphoma subtypes, the lowest weighted κ-values were observed in hilar nodes, infraclavicular nodes, and bowel [22]. However, these sites were also those sites that were least frequently involved with lymphoma in their cohort, thus κ-values could have been low because of the low prevalence of these sites. For some specific nodal and extranodal localizations, only a few positive cases were identified; therefore, we decided to report on numbers of discrepancies and percentage OA only.

Baseline PET/CT provides more accurate staging than CT only [2,3] and serves as a reference to quantify tumor response (SUV, metabolically active tumor volume). In our study, in which baseline PET/CT was not mandatory according to prevailing guidelines at the start of the study [23,24], we found that observer variability in treatment evaluation was independent of the baseline imaging modality.

A strength of this study is that the I-PET and EoT-PET scans were assessed by 2 reviewers from a pool of 10, in contrast to previous studies with small, fixed numbers of reviewers. Scoring by a pool of reviewers represents the normal situation of ^{18}F-FDG PET/CT lymphoma response assessment in clinical practice. A limitation is that scans rated as unclear were excluded from our main analyses; in 13 I-PET and 10 EoT-PET scans one of the reviewers rated the final conclusion as unclear, specifying in the free-text section that they were not certain that the residual ^{18}F-FDG uptake was lymphoma-related. A similar conclusion was drawn by both observers in 2 EoT-PET scans. In analysis of a best-case (both reviewers agreed on negative or positive scores) and worst-case (discrepancy) scenario, we found that these results only slightly affected observer agreement: for I-PET and EoT-PET, percentages of OA were 88.1% and 91.9% in the best-case scenario, versus 85.4% and 89.7% in the worst-case scenario. In most of these unclear scans—13 of 13 I-PET scans and 6 of 12 EoT-PET scans residual ^{18}F-FDG uptake in extranodal lymphoma sites caused the uncertainty, especially skeletal lesions in I-PET scans (perhaps because of enhanced bone marrow background uptake due to granulocyte colony-stimulating factor, or healing fractures or bone remodeling in previous lymphoma locations). Other reasons mentioned for unclear reads were missing baseline PET status, no contrast-enhanced CT scan available, inferior quality of a CT scan, a possible sarcoid like response, or uncertainty about a nonresponding lesion while all other lesions responded (Figs. 2 and 3). Another

limitation is the use of older generation PET/CT systems (Supplemental Table 4), which could influence the generalizability of our results.

Our findings, especially the suboptimal PA, have implications for trial design and clinical practice. A PA of 74% at I-PET clearly emphasizes the need for central review procedures in clinical trials investigating intensified therapy in I-PET–positive patients. The 76% PA at EoT-PET reinforces the recommendation to discuss patients during multidisciplinary tumor board meetings, allowing for a second read of the test result, aiming for optimal patient management (e.g., confirmatory scan or biopsy).

Our data indicate that reviewers are especially uncertain in cases of extranodal lymphoma involvement, which is common in DLBCL patients, with baseline frequencies of up to 20% depending on the site (Table 2 and Supplemental Table 3). These results could be helpful in focusing the training of nuclear medicine physicians, such as by using the harmonization approach of Ceriani et al. [25]. During the last 10–15 y, ^{18}F-FDG PET/CT systems' quality continued to evolve, and guidelines therefore need to be updated on a regular basis [26,27].

Conclusion

Interobserver agreement among experienced nuclear medicine physicians using DS for I-PET and EoT-PET response assessment in DLBCL has high percentages of OA (88%–92%) and NA (92%–95%). The lower (74%–76%) PA suggests that the accuracy of EoT-PET/CT treatment evaluation in daily practice and I-PET–adapted trials may benefit from dual reads and central review, respectively.

Disclosure

This work was financially supported by an Alpe d'HuZes/KWF fund, provided by the Dutch Cancer Society (VU 2012-5848). Pieternella Lugtenburg has received research funding from Roche, Servier, and Takeda and honoraria for advisory boards from Roche, Servier, Sandoz, BMS, Genmab, Takeda, and Celgene. No other potential conflict of interest relevant to this article was reported.

References

1. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016;127:2375–2390.
2. 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.
3. 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.
4. Weiler-Sagie M, Bushelev O, Epelbaum R, et al. 18F-FDG avidity in lymphoma readdressed: a study of 766 patients. *J Nucl Med.* 2010;51:25–30.
5. Borchmann P, Goergen H, Kobe C, et al. PET-guided treatment in patients with advanced-stage Hodgkin's lymphoma (HD18): final results of an open-label, international, randomised phase 3 trial by the German Hodgkin Study Group. *Lancet.* 2018;390:2790–2802.
6. Radford J, Illidge T, Counsell N, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. *N Engl J Med.* 2015;372:1598–1607.
7. André MPE, Girinsky T, Federico 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;35:1786–1794.
8. Johnson P, Federico M, Kirkwood A, et al. Adapted treatment guided by interim PET-CT scan in advanced Hodgkin's lymphoma. *N Engl J Med.* 2016;374:2419–2429.
9. Terasawa T, Lau J, Bardet S, et al. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. *J Clin Oncol.* 2009;27:1906–1914.
10. Zijlstra JM, Burggraaff CN, Kersten MJ, Barrington SF; EHA Scientific Working Group on Lymphoma. FDG-PET as a biomarker for early response in diffuse large B-cell lymphoma as well as in Hodgkin lymphoma? Ready for implementation in clinical practice? *Haematologica* [editorial]. 2016;101:1279–1283.
11. Hertzberg M, Gandhi MK, Trotman J, et al.; Australasian Leukaemia Lymphoma Group (ALLG). Early treatment intensification with R-ICE and 90Y-ibritumomab tiuxetan (Zevalin)-BEAM stem cell transplantation in patients with high-risk diffuse large B-cell lymphoma patients and positive interim PET after 4 cycles of R-CHOP-14. *Haematologica.* 2017;102:356–363.
12. Robinson PJ. Radiology's Achilles' heel: error and variation in the interpretation of the Röntgen image. *Br J Radiol.* 1997;70:1085–1098.
13. 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.
14. 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 DSUVmax. *Eur J Nucl Med Mol Imaging.* 2013;40:1312–1320.
15. Han EJ, O JH, Yoon H, et al. FDG PET/CT response in diffuse large B-cell lymphoma: reader variability and association with clinical outcome. *Medicine (Baltimore).* 2016;95:e4983.
16. 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.
17. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64:96–106.
18. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's k. *BMJ.* 2013;346:f2125.

19. Lugtenburg PJ, de Nully Brown P, van der Holt B, 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 [abstract]. *J Clin Oncol.* 2016;34(suppl):7504.

20. Hüttmann A, Müller S, Jöckel KH, Dührsen U. Pitfalls of interim positron emission tomography scanning in diffuse large B-cell lymphoma. *J Clin Oncol.* 2010;28:e488–e489.

21. Jacene HA, Ishimori T, Engles JM, Leboulleux S, Stearns V, Wahl RL. Effects of pegfilgrastim on normal biodistribution of 18F-FDG: preclinical and clinical studies. *J Nucl Med.* 2006;47:950–956.

22. Hofman MS, Smeeton NC, Rankin SC, Nunan T, O'Doherty MJ. Observer variation in interpreting 18F-FDG PET/CT findings for lymphoma staging. *J Nucl Med.* 2009;50:1594–1597.

23. Cheson BD, Pfistner B, Juweid ME, et al.; International Harmonization Project on Lymphoma. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007;25:579–586.

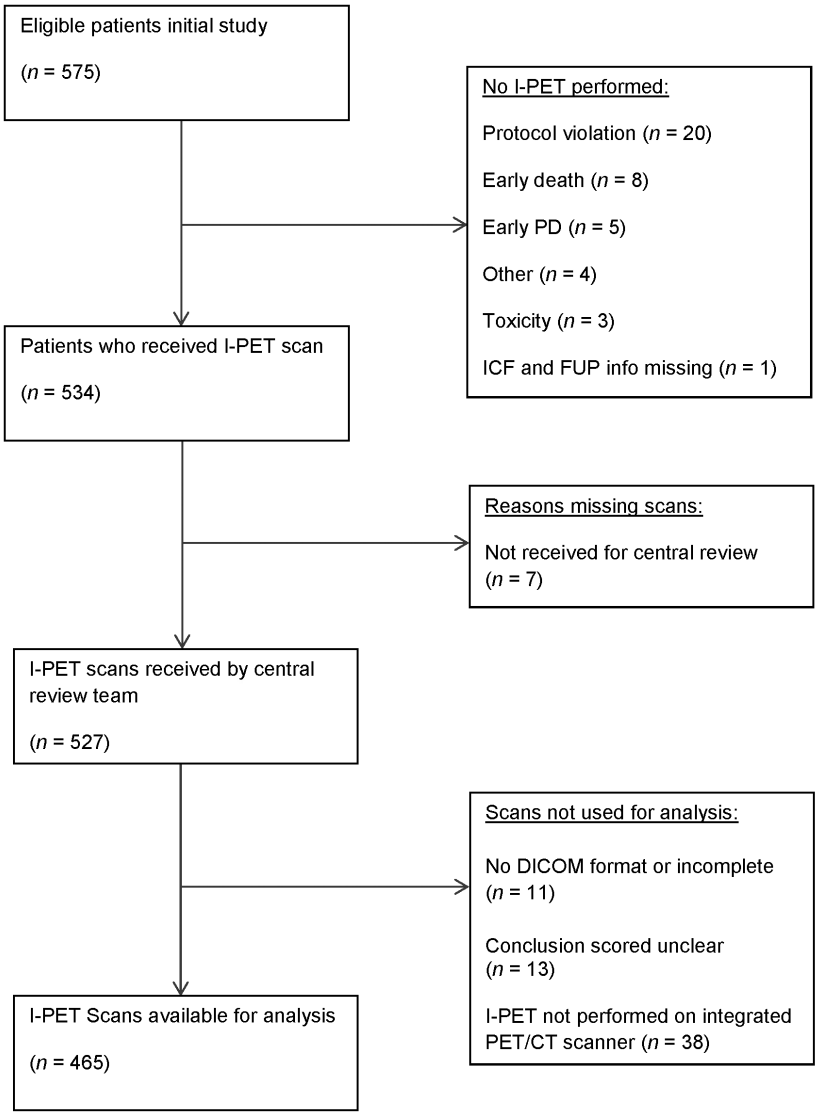
24. Juweid ME, Stroobants S, Hoekstra OS, et al.; Imaging Subcommittee of International Harmonization Project in Lymphoma. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *J Clin Oncol.*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

25. Ceriani L, Barrington S, Biggi A, et al. Training improves the interobserver agreement of the expert positron emission tomography review panel in primary mediastinal B-cell lymphoma: interim analysis in the ongoing International Extranodal Lymphoma Study Group-37 study. *Hematol Oncol.* 2017;35:548–553.

26. Boellaard R, Oyen WJ, Hoekstra CJ, et al. The Netherlands protocol for standardization and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging.* 2008;35:2320–2333.

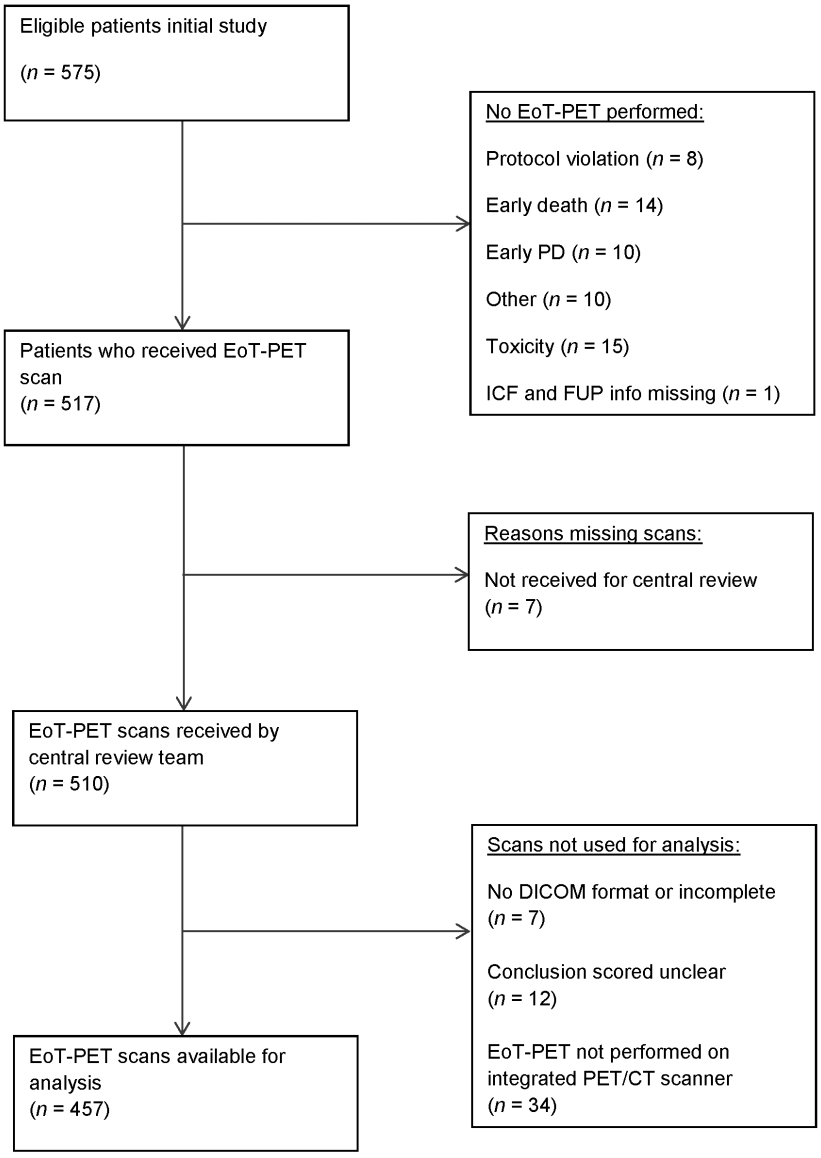
27. Boellaard R, Delgado-Bolton R, Oyen WJ, et al.; European Association of Nuclear Medicine (EANM). FDG PET/CT: EANM procedure guidelines for tumour imaging—version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328–354.

Supplemental materials



Supplemental figure 1. Flowchart of I-PET available for qualitative central review.

Abbreviations: FUP= follow-up; ICF= informed consent form; I-PET= interim ^{18}F-FDG PET; PD= progressive disease.



Supplemental Figure 2. Flowchart of EoT-PET available for qualitative central review.

Abbreviations: EoT-PET= end-of-treatment ^{18}F-FDG PET; FUP= follow-up; ICF= informed consent form; PD= progressive disease.

Supplemental Table 1.

Interobserver agreement of ordinal DS in I-PET					
	DS 1	**DS 2**	**DS 3**	**DS 4**	**DS 5**
DS 1	88	40	25	7	3
DS 2	46	43	26	7	1
DS 3	19	20	21	8	0
DS 4	10	9	11	38	10
DS 5	1	0	0	8	24

Percentage exact agreement = ((88 + 43 + 21 + 38 + 24)/465)*100% = (214/465)*100% = 46.0%.
Percentage agreement (+1, -1) = ((88 + 43 + 21 + 38 + 24 + 46 + 40 + 20 + 26 + 8 + 10)/465)*100% = (364/465)*100% = 78.3%.

Supplemental Table 2.

Interobserver agreement of ordinal DS in EoT-PET					
	DS 1	**DS 2**	**DS 3**	**DS 4**	**DS 5**
DS 1	128	49	22	3	4
DS 2	57	37	16	1	0
DS 3	16	17	16	7	0
DS 4	5	7	11	16	8
DS 5	0	0	0	14	23

Percentage exact agreement = ((128 + 37 + 16 + 16 + 23)/457)*100% = (220/457)*100% = 48.1%.
Percentage agreement (+1, -1) = ((128 + 37 + 16 + 16 + 23 + 57 + 49 + 17 + 16 + 14 + 8)/457)*100% = (381/457)*100% = 83.4% .

Supplemental Table 3. Interobserver agreement of specific nodal and extranodal localizations in EoT-PET.

	Number baseline positive	Number of discrepancies at EoT-PET	Agreement on negativity (absolute)	Agreement on positivity (absolute)	Percentage overall agreement‡	Related to baseline prevalence§
Nodal						
Para-aortic†	397	17	884	13	98.1	4.3%
Cervical†	286	6	903	5	99.3	2.1%
Iliac†	267	8	899	7	99.1	3.0%
Axillary†	220	1	909	4	99.9	0.5%
Supraclavicular†	213	2	908	4	99.8	0.9%
Inguinal†	204	5	908	1	99.5	2.5%
Mediastinal*	202	5	442	8	98.9	2.5%
Mesenteric	188	17	429	11	96.3	9.0%
Hilar*†	142	8	897	1	99.1	5.6%
Spleen*	114	6	442	8	98.7	5.4%
Other	105	6	450	1	98.7	5.7%
Waldeyer	48	1	456	0	99.8	2.1%
Extranodal						
Other extranodal*	123	13	431	10	97.1	10.6%
Skeletal*	90	8	441	5	98.2	8.9%
GI*	62	6	444	6	98.7	9.7%
Lung*	54	3	445	6	99.3	5.6%
Liver*	38	2	453	1	99.6	5.3%
Pleura*	25	0	456	0	100.0	0.0%
Skin	13	1	456	0	99.8	7.7%
CNS	0	0	456	1	100.0	0.0%

Abbreviations: CNS= central nervous system; EoT-PET= end-of-treatment positron emission tomography; GI= gastrointestinal

* Totals not 457 or 914, because of missing values or localization scored as unclear.

† Right and left are summed and presented together.

‡Percentage overall agreement: (number of agreement on positivity + number of agreement on negativity) / (number of discrepancies + number of agreement on positivity + number of agreement on negativity)*100%.

§Related to baseline prevalence: (number of discrepancies/number baseline positive)*100%.

Supplemental Table 4. Overview of PET/CT scanner types used in the HOVON84 study.

Manufacturer	PET/CT Model	I-PET (*n* = 465)	EoT-PET (*n* =457)
GE Medical Systems	Discovery RX	*n* = 3	*n* = 1
	Discovery ST	*n* = 8	*n* = 8
	Discovery STE	*n* = 24	*n* = 26
	Discovery 690	*n* = 1	*n* = 3
Philips	Allegro Body (C)	*n* = 5	NA
	Gemini TF TOF 16	*n* = 37	*n* = 48
	Gemini TF TOF 64	*n* = 62	*n* = 71
	Gemini TF (C)	*n* = 16	*n* = 10
	Gemini GXL 10	*n* = 5	*n* = 3
	Gemini GXL 16	*n* = 23	*n* = 22
	Guardian Body	*n* = 1	NA
Siemens	Biograph 6	*n* = 17	*n* = 16
	Biograph 16	*n* = 6	*n* = 6
	Biograph 40	*n* = 112	*n* = 100
	Biograph 64	*n* = 72	*n* = 79
	Biograph 128	*n* = 2	*n* = 2
CTI PET Systems	Biograph mCT	*n* = 71	*n* = 62

Abbreviations: EoT-PET= end-of-treatment positron emission tomography; I-PET= interim positron emission tomography; NA: not applicable.

Supplemental Table 5. GRRAS checklist for reporting reliability and agreement studies.

		Reported on page number(s):
Title and abstract	1. Identify in title or abstract that interrater/intrarater reliability or agreement was investigated.	1,3
Introduction	2. Name and describe the diagnostic or measurement device of interest explicitly.	5,6
	3. Specify the subject population of interest.	5
	4. Specify the rater population of interest (if applicable).	Methods 6
	5. Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	5
Methods	6. Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	6
	7. Describe the sampling method.	5,6
	8. Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	6
	9. State whether measurement/ratings were conducted independently.	6
	10. Describe the statistical analysis.	7
Results	11. State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	7-9
	12. Describe the sample characteristics of raters and subjects (e.g. training, experience).	7,8 + methods
	13. Report estimates of reliability and agreement including measures of statistical uncertainty.	8,9
Discussion	14. Discuss the practical relevance of results.	10-14, esp 13,14
Auxiliary material	15. Provide detailed results if possible (e.g. online).	Supplementals

"Reprinted from *J Clin Epidemiol,* 64 (1), Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed, Table 1, Page 98, Copyright 2011 with permission from Elsevier."

Chapter 5

Optimizing Workflows for Fast and Reliable Metabolic Tumor Volume Measurements in Diffuse Large B Cell Lymphoma.

Coreline N. Burggraaff
Isabelle Kaβner
Sally F. Barrington
Gerben J.C. Zwezerijnen
Otto S. Hoekstra
Henrica C.W. de Vet
Fareen Rahman
Simone Pieplenbosch
Yvonne W.S. Jauw
Stefan Müller
Josée M. Zijlstra
Ronald Boellaard
On behalf of the PETRA Consortium

Molecular Imaging Biology 2020; 22: 1102-1110.

Abstract

Purpose:
This pilot study aimed to determine interobserver reliability and ease of use of three workflows for measuring metabolic tumor volume (MTV) and total lesion glycolysis (TLG) in diffuse large B cell lymphoma (DLBCL).

Procedures:
Twelve baseline [^{18}F]FDG PET/CT scans from DLBCL patients with wide variation in number and size of involved organs and lymph nodes were selected from the international PETRA consortium database. Three observers analyzed scans using three workflows. Workflow A: user-defined selection of individual lesions followed by four automated segmentations (41%SUVmax, A50%SUVpeak, SUV≥2.5, SUV≥4.0). For each lesion, observers indicated their "preferred segmentation." Individually selected lesions were summed to yield total MTV and TLG. Workflow B: fully automated preselection of [^{18}F]FDG-avid structures (SUV≥4.0 and volume≥3ml), followed by removing non-tumor regions with single mouse clicks. Workflow C: preselected volumes based on Workflow B modified by manually adding lesions or removing physiological uptake, subsequently checked by experienced nuclear medicine physicians. Workflow C was performed 3 months later to avoid recall bias from the initial Workflow B analysis. Interobserver reliability was expressed as intraclass correlation coefficients (ICC).

Results:
Highest interobserver reliability in Workflow A was found for SUV≥2.5 and SUV≥4.0 methods (ICCs for MTV 0.96 and 0.94, respectively). SUV≥4.0 and A50%Peak were most and SUV≥2.5 was the least preferred segmentation method. Workflow B had an excellent interobserver reliability (ICC = 1.00) for MTV and TLG. Workflow C reduced the ICC for MTV and TLG to 0.92 and 0.97, respectively. Mean workflow analysis time per scan was 29, 7, and 22 min for A, B, and C, respectively.

Conclusions:
Improved interobserver reliability and ease of use occurred using fully automated preselection (using SUV≥4.0 and volume≥3ml, Workflow B) compared with individual lesion selection by observers (Workflow A). Subsequent manual

modification was necessary for some patients but reduced interobserver reliability which may need to be balanced against potential improvement on prognostic accuracy.

Key Words:
Diffuse large B cell lymphoma, Metabolic tumor volume, PET/CT, Total lesion Glycolysis

Introduction

In young patients with diffuse large B cell lymphoma (DLBCL), a large maximum tumor diameter is an indicator of poor prognosis [1]. Recent progress in lymphoma care has recommended exploration of the prognostic value of volumetric tumor bulk measured on staging 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) PET/CT, with methods combining metabolic activity and volume [2]. In lung cancer patients, studies have focused on finding the most reliable tumor segmentation method [3–5]. However, compared with lung cancer, lymphoma segmentation is more challenging due to higher number of lesions, multiple anatomical locations, and inter- and intratumoral [^{18}F]FDG uptake heterogeneity.

Preliminary data suggest that baseline metabolic tumor volume (MTV) has a prognostic value in DLBCL [6–9] and predict outcome better than bulky disease measured by maximum tumor diameter [7]. Total lesion glycolysis (TLG)—defined as SUVmean in a volume multiplied by the corresponding MTV—seems to perform similarly [7] or inferiorly [6,8] in predicting outcome of DLBCL patients. Various segmentation methods to measure MTV and TLG are being used in clinical lymphoma studies [10]: most use a fixed SUV threshold (e.g., SUV≥2.5 [7,9] or SUV≥4.0 [11]) or a percentage of SUVmax (e.g., 41 % of SUVmax [6,8,12]) to define MTV. An important finding from earlier studies in DLBCL is that optimal cutoff values range widely (220–550 ml), probably because of using different methodologies, small patient cohorts, differences in patient risk factors, and therapies [13]. Moreover, these data-driven cutoff values should be interpreted with caution, as they depend highly on acquisition and reconstruction protocols.

Segmentation methods in these studies are generally derived from phantom experiments [4,12], or correlation with pathological specimens in lung cancer [4]. Limited data are available about the differences in ease of use in the lymphoma clinical setting and interobserver reliability of these tumor segmentation methods [10]. Previous studies in DLBCL [10], T cell [14], and Hodgkin lymphoma [15] showed that different segmentation methods, despite having different cutoff values, show comparable accuracy for predicting survival. Therefore, for future use in practice and clinical trials a robust, reliable and easy—i.e., with least required observer interaction—segmentation workflow is necessary. To the best of our knowledge, this is the first pilot study in DLBCL that compares interobserver reliability and ease of use of three workflows for measuring MTV and TLG, and that assesses the effect of manual modification on interobserver reliability.

Materials and Methods

Study Population

Twelve baseline [^{18}F]FDG PET/CT scans from newly diagnosed DLBCL patients with wide variation in number and size of involved organs and lymph nodes lesions were selected from the international PETRA database (http://www.petralymphoma.org). The use of all data within the PETRA imaging database has been approved by the Medical Ethics Review Committee of the VU University Medical Center (JR/20140414) after patients' consent to participate in the studies included in the database.

Image Analysis Workflows A and B

Two semi-automated workflows (Workflows A and B) were performed in the same week, by three independent observers using the ACCURATE software tool [16]. Manual modifications of the semi-automatically generated volumes of interest (VOIs) were not allowed initially. The workflow with the best interobserver reliability and ease of use was selected as starting point for manual modification in Workflow C.

Workflow A comprised a user-defined selection of individual lesions. The observers had to select individual lesions (by a single mouse click in the "hottest" part of

each lesion), followed by automated segmentation in the tool using four separate frequently published segmentation methods:

1. 41 % of SUVmax (41%MAX)
2. A50% of SUVpeak, i.e., 50 % of SUVpeak with local background correction [17] (A50%P)
3. fixed SUV threshold of 2.5 (SUV≥2.5)
4. fixed SUV threshold of 4.0 (SUV≥4.0).

The four segmentation methods were initiated from one single click by the observer, to avoid introduction of extra variability by repeated clicking. Moreover, the tool first calculated a robust local maximum (using a region growing method applying a 70 % threshold of the point clicked) in order to be less dependent on the exact point clicked by the observer. Generated VOIs were summed for all lesions selected by each observer to calculate MTV and TLG according to each of the four segmentation methods.

To explore the use and performance of consensus methods, two methods were added afterwards, which use the delineations found with the above four standard methods as input for a majority vote (MV) approach [18]. MV volumes were defined by all voxels included in the MTV or TLG by at least two (MV2) or three (MV3) of the input methods.

Workflow B consisted of a fully automated preselection of [^{18}F]FDG-avid structures defined by an SUV≥4.0 and a volume threshold of ≥3 ml. These preselected regions resulted into an identical starting point for all observers but could include non-tumor regions with normal increased [^{18}F]FDG uptake, such as the brain or bladder. From this starting point, the observers decided on the removal of non-tumor regions by using a clearing option (i.e., single click(s)) or spatial limits to reduce the analyzed field of view (e.g., using a slider option to exclude superior slices including the brain or inferior slices including the bladder); after this, only lymphoma lesions remain. Therefore, a region is defined as any preselected 3D-VOI with uptake above the SUV≥4.0 threshold, whereas a lesion is defined as a 3D-VOI identified by the observer as lymphoma.

To determine ease of use for both workflows, each observer noted the total analysis time per patient (including loading of the scan, performing the analysis, and saving results).

In addition, the success of all semi-automatically generated VOIs was rated by each observer according to the following definitions:

- Failed: generated VOI is unrealistic or does not contain complete lesion
- Poor: generated VOI takes into account physiological uptake or contains a lot of background and manual modification is needed
- Acceptable: only minimal manual modification needed for good VOI
- Good: generated VOI is comparable to what you consider to be lymphoma

A mean "success rate" (all acceptable and good ratings) was calculated for each method. Finally, observers had to choose one "preferred segmentation" for the generated VOIs. The MV2 and MV3 consensus methods were rated by one experienced observer according to the same success definitions. As these MV methods were assessed afterwards, they could not be chosen as "preferred segmentation."

Image Analysis Workflow C

The observers used the fully automated method as in Workflow B for the analyses on the same twelve scans (Workflow C1). These analyses were performed 3 months later to minimize recall bias. In addition to the interactive deletion of physiological uptake regions similar to Workflow B, the observers were allowed in Workflow C to manually modify the generated VOIs by adding missed lesions (with the A50%P option or manually) and removing of physiological uptake with an "eraser" tool. The manually modified MTVs and TLGs were checked for correct delineation and identification of tumor sites (and changed if needed) by independent nuclear medicine physicians (NM, one per observer) with more than 10 years of experience with [^{18}F]FDG PET/CT evaluation in lymphoma (OSH, SFB, SM; Workflow C2).

Statistical Analysis

Success rates of generated VOIs were analyzed descriptively. Interobserver reliability was expressed as intraclass correlation coefficients (ICCs) and coefficients of variation (CoVs). ICC estimates and their 95 % confidence

intervals (95%CIs) were calculated with a two-way random-effects model for absolute agreement [19]. The 95%CIs of the ICC values were interpreted as poor (< 0.5), moderate (0.5–0.75), good (0.75–0.9), and excellent (> 0.9) [20,21]. CoV was calculated as the ratio of the standard deviation (over three observers) of MTVs or TLGs divided by the mean values per patient. Mean CoVs are presented, i.e., CoVs averaged over all patients. Bland-Altman plots were drawn to visually assess potential bias of the mean differences between the workflows and to estimate 95 % limits of agreement [22]. Normality of MTV and TLG differences before and after manual modification was checked with the Shapiro-Wilkinson (SW) test, in which $P < 0.05$ was an indication of a non-normal distribution. Statistical analyses were performed using SPSS Statistics (IBM, v.20).

Results

Workflow A; Individual Lesion Selection

Lesion Selection

The total number of selected lesions for observer 1, 2, and 3 was 162, 117, and 118, respectively, which was due to the fact that observer 1 separately selected small lesions close to larger lesions, which were ignored by observers 2 and 3. It resulted in larger volumes for the A50%P and the 2 MV consensus methods for observer 1 (Supplemental Fig. 1). In total, 76 lesions were selected by all observers; of which, 35 showed identical segmentation results, and 18 lesions had a difference in volume between observers of < 1 ml. Twenty-three non-identical lesions were caused by clicking in different parts of a heterogeneous lesion, which resulted in missing the SUVmax or SUVpeak of the lesion.

Interobserver Reliability

ICC values for semi-automated MTVs were 0.43, 0.86, 0.96, and 0.94 for the 41%MAX, A50%P, SUV≥2.5, and SUV≥4.0 thresholds, respectively. Mean CoVs were 65.5 %, 36.7 %, 13.3 %, and 13.8 %, respectively (Table 1). When considering the 95%CIs of ICCs, only SUV≥2.5 and SUV≥4.0 showed excellent and good to excellent reliability, respectively. For the MV2 and MV3 consensus methods, the mean CoVs were 22.7 % and 33.5 % and ICCs were 0.92 and 0.91, respectively. Overall, fixed SUV threshold methods (SUV≥2.5 and SUV≥4.0) showed least

interobserver variability for MTV assessment in Workflow A. TLG showed similar ICCs and CoVs for these two methods.

Ease of Use

Mean analysis time in Workflow A was 28.7 min per patient (range 5–63, Table 2). The most preferred method differed per patient and between observers (Table 3). A50%P and SUV≥4.0 were most often chosen as "preferred segmentation" on a patient-level with success rates (rated as acceptable or good segmentations of visible tumor) ranging from 33 to 87 % and 35–76 %, respectively. The mean success rate for the 41%MAX method ranged from 31 to 86 % between observers. The success rates for the MV2 and MV3 methods, as scored by one observer, were 84 % and 87 %, respectively. Although SUV≥2.5 showed the highest observer reliability, this method was chosen only in 2 patients as the most preferred method by 1 observer. The mean success rate for this method ranged between 27 and 39 % between observers. This method tended to overestimate the tumor volumes (Supplemental Figs. 1–2). Therefore, we decided to focus on the SUV≥4.0 method as preselection criterion.

Workflow B; Preselection Strategy

Lesion Selection

The total number of selected tumor regions for observer 1, 2, and 3 was 76, 76, and 77, respectively. Seventy-two identical tumor regions were selected by all three observers.

Interobserver Reliability

Workflow B is based on the SUV≥4.0 threshold and showed good correlation with SUV≥4.0 threshold of Workflow A with a Pearson correlation of 0.812 (after removing 4 volumes as outliers in 2 patients 0.995, Fig. 1). Outliers were caused by one patient with many lesions, in whom the SUV≥4.0 threshold failed (large parts of the liver and spleen were included in this segmentation) and another with a large abdominal lesion that was interpreted as non-lymphoma by one observer. Complete agreement of the preselected volumes on a patient-level between all observers was found in six patients. The ICC value for generated MTVs in this workflow was excellent (1.00, 95%CI 1.00–1.00) and the mean CoV was 2.3 % (range 0–10.4 %, Table 1), with similar results for TLG.

Ease of Use

Time to complete Workflow B ranged from 1 to 15 min (mean 7.3, Table 2). Preselected MTVs were rated as successful in seven, three, and four patients by the observers, respectively. They were classified as failures in zero, four, and six patients respectively.

Workflow C; Manual Modification

Effect of Manual Modification

After manual modification of the preselected volumes the ICC of the final MTV was 0.92 (95%CI 0.82–0.98, Table 1). Mean CoV for the final MTV was 16.7 %. Results for TLG again were similar, with excellent ICC values and good to excellent ICC values for MTV. The total time to perform this workflow ranged from 5 to 62 min (mean 22.2, Table 2).

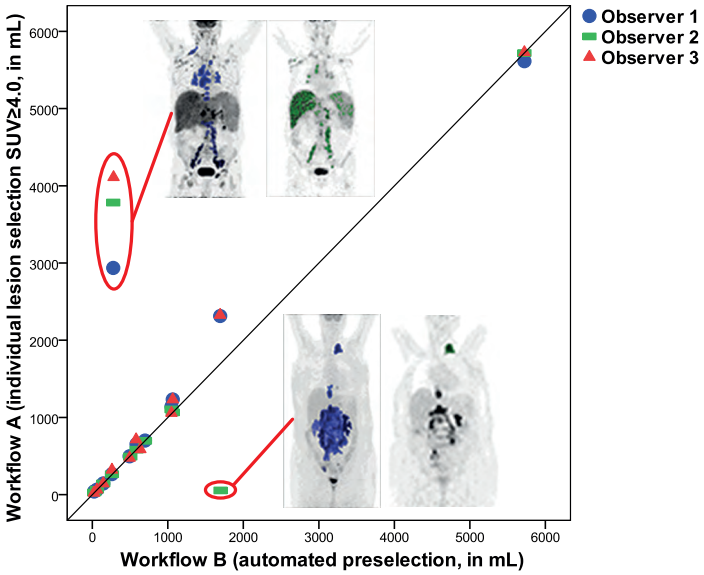


Fig. 1. Scatterplot of MTV for Workflow A (user-defined selection with SUV≥4.0) and Workflow B (automated preselection).

PET images represent examples of different MTV interpretations between the workflows. Top left images (patient 10): Workflow B contains only lymphoma lesions around the large vessels (left), while in Workflow A, the liver and spleen were also included in the lesion selection (right). Bottom right images (patient 8): in Workflow B, the large lesion was selected (left), while it was interpreted as not being lymphoma in Workflow A (right).

Table 1. Interobserver reliability of semi-automated MTV and TLG assessment for the different workflows.

	MTV			TLG		
	Mean (range)	**Mean CoV (range)**	**ICC (95%CI)**	**Mean (range)**	**Mean CoV (range)**	**ICC (95%CI)**
Workflow A (individual lesion selection)						
41%MAX	1106 (33-4991)	65.54 (0-164.38)	0.43 (0.08-0.76)	6236 (471-21431)	54.57 (0-151.84)	0.37 (0.02-0.72)
A50%P	550 (34-4153)	36.74 (0-139.73)	0.86 (0.68-0.95)	5736 (245-45441)	26.76 (0-118.26)	0.93 (0.82-0.98)
SUV≥2.5	2399 (73-7404)	13.34 (0-54.21)	0.96 (0.91-0.99)	15902 (347-55588)	7.11 (0-33.81)	0.99 (0.98-1.00)
SUV≥4.0	1289 (30-5688)	13.78 (0-83.59)	0.94 (0.86-0.98)	13617 (220-50068)	11.32 (0-82.52)	0.97 (0.93-0.99)
MV2	1505 (59-6258)	22.68 (0-83.59)	0.92 (0.80-0.97)	14422 (301-51908)	15.84 (0-82.52)	0.97 (0.91-0.99)
MV3	927 (33-4654)	33.54 (0-154.17)	0.91 (0.79-0.97)	12181 (229-43669)	24.91 (0-135.92)	0.96 (0.89-0.99)
Workflow B (automated preselection						
SUV≥4.0, Volume≥3ml	1004 (23-5723)	2.32 (0-10.43)	1.00 (1.00-1.00)	8446 (189-50779)	1.85 (0-7.49)	1.00 (1.00-1.00)
Workflow C (automated preselection with manual modification)						
Final MTV	1115 (53-5589)	16.71 (0-109.46)	0.92 (0.82-0.98)	8610 (284-48079)	13.33 (0-111.83)	0.97 (0.93-0.99)

MV, Majority Vote; MTV, metabolic tumor volume; CoV, coefficient of variation; ICC, intraclass correlation coefficient; CI, confidence interval; TLG, total lesion glycolysis.

Table 2. Mean analysis time for the different workflows in minutes (mean ± standard deviation (range)).

Workflow	**A individual lesion selection (*n*=12)**	**B automated preselection (*n*=12)**	**C with manual modification (*n*=12)**
Observer 1	29.1±20.8(5-63)	7.2±3.7(3-15)	23.3±13.4(5-45)
Observer 2	Not reported*	Not reported*	26.7±15.6(10-62)
Observer 3	28.2±13.7(15-60)	7.3±3.5(1-12)	16.7±9.7(8-42)
Mean	28.7†	7.3†	22.2

*Observer 2 summed the total time for workflow A + B; mean 27.3±19.2 (7-75) minutes.
†Mean value based on 2 observers.

Table 3. Most preferred method* per observer for Workflow A.

Patient	Observer 1	Observer 2	Observer 3
1	41%MAX	41%MAX	SUV≥4.0
2	41%MAX	41%MAX/A50%P/SUV≥4.0	SUV≥2.5
3	A50%P	41%MAX	SUV≥4.0
4	SUV≥4.0	A50%P	SUV≥4.0
5	SUV≥4.0	A50%P	SUV≥4.0
6	SUV≥4.0	41%MAX/A50%P	SUV≥4.0
7	A50%P	A50%P	SUV≥2.5
8	A50%P	41%MAX/A50%P	SUV≥4.0
9	A50%P	41%MAX	SUV≥4.0
10	A50%P	A50%P	A50%P
11	SUV≥4.0	41%MAX	A50%P
12	41%MAX/SUV≥4.0	41%MAX	A50%P

*Each observer indicated their "preferred segmentation" for individual lesions. The most preferred method per patient was defined as the method most often noted as "preferred segmentation".

Figure 2 shows the modified MTVs approved by a nuclear medicine physician (final MTV). Figure 3 shows a scatterplot of the correlation between the preselected and final MTV in Workflow C. Interestingly, the same outlier (patient 10) occurred as in Fig. 1, but contrary to this, two observers now decided to keep the entire liver in the preselection of Workflow C while they removed the liver uptake in Workflow B. For the final MTV, observer 2 had to remove the liver uptake after the check by the experienced NM physician. In another patient (patient 11), the preselection missed many small bone lesions, which were added manually. Figure 4 shows the Bland-Altman plot of the preselected and final MTV in Workflow C. The 95 % limits of agreement ranged widely (– 525 to 458). The differences between preselected and final MTV did not have a normal distribution according to the SW test (P = 0.002). After excluding patients 10 and 11 (Figs. 3 and 4) described as outliers, the mean difference had a normal distribution (P = 0.106). The plot shows both the original—as well as the recalculated 95 % limits of agreement after exclusion of the outliers.

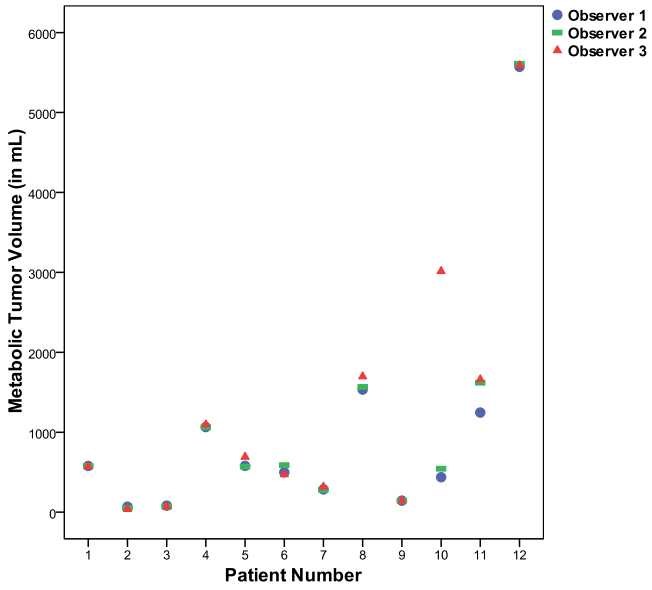


Fig. 2. Scatterplot of final MTV assessment in Workflow C.

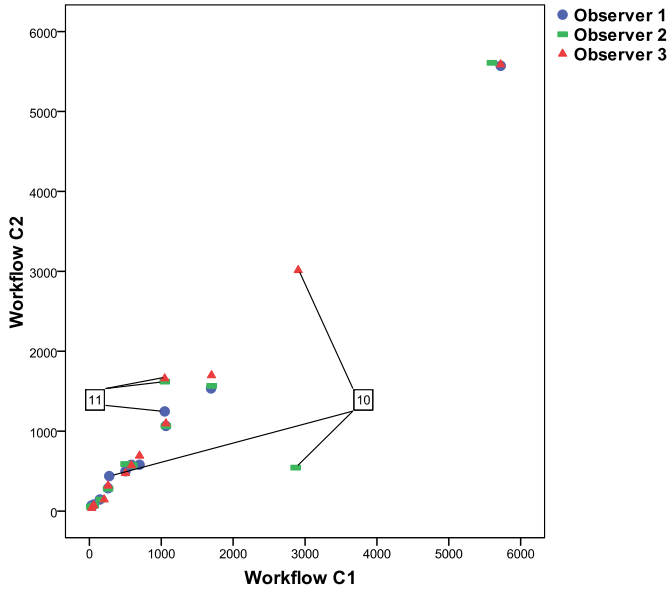


Fig. 3. Scatterplot of MTV assessment in Workflow C (automated preselection before (C1) – and final MTV after manual modification (C2), in milliliters).

Datapoints from two challenging patients (patients 10 and 11) are indicated by lines. The numbers in the boxes refer to the patient numbers described in the main text.

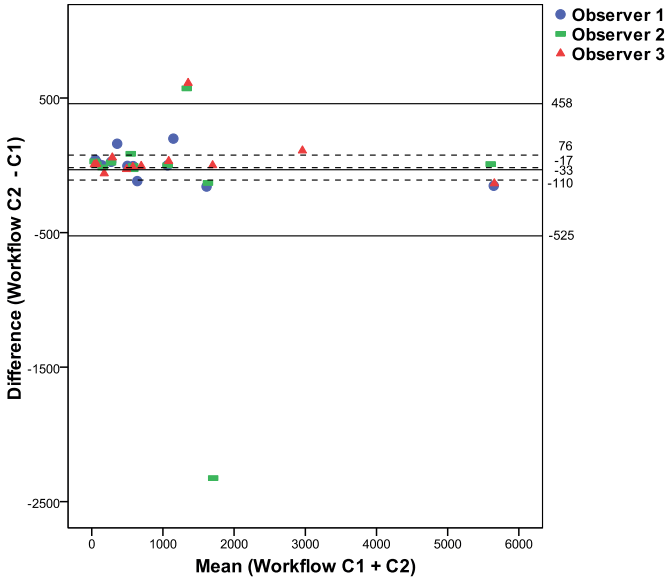


Fig. 4. Bland-Altman plot showing effect of manual modification of MTV assessment in Workflow C (automated preselection before (C1) – and final MTV after manual modification (C2)).

Solid line: mean value, upper- and lower limit of agreement without exclusion of outliers. Dashed line: mean value, upper- and lower limit of agreement after exclusion of patients 10 and 11.

Discussion

We assessed the interobserver reliability and ease of use of three workflows for measuring MTV and TLG in 12 DLBCL patients and found that both improved when using a fully automated preselection approach to measure MTV and TLG (using SUV≥4.0 and volume≥3ml).

Ilyas et al. [10] compared three MTV segmentation methods (SUV≥2.5, 41%MAX and PERCIST) in patients with DLBCL and concluded that data-driven optimal cutoff values for separation of patients into a good and a poor prognosis group were largely dependent on the method used, but these data-driven cutoff values had comparable prognostic accuracy. In a subset of 50 patients evaluated by two observers, they found that interobserver reliability was excellent (ICC > 0.98). They further reported a mean analysis time ranging between 2.7 and 6.2 min for the 3 methods [10]. The data-analysis in our study took more time, possibly due to less experience of the observers with the software and the datasheets that had

to be completed, which was not included in the time per patient reported in Ilyas study. Yet, also in our study, we found that when total metabolic tumor volume was derived using the preselection and when unwanted normal tissue uptake could be removed and missed lesions could be added by single mouse clicks, the overall processing time was typically less than 5 min. In cases where manual corrections or manual definitions of the VOIs were needed, processing time could well exceed 20 min.

Another important finding in the Ilyas study and our study is that the SUV≥2.5 method showed the highest interobserver reliability. Interestingly though, the observers in our study considered that SUV≥2.5 often overestimated the volume compared with other methods and was almost never chosen as their preferred method on a patient-level.

However, a recent study (partly by the same authors) showed that a slightly higher threshold (SUV≥4.0)outperformed the SUV≥2.5 in terms of success rate [23].

A recent phantom and patient study in primary mediastinal B cell lymphoma that compared four different MTV methods found that SUV≥2.5 resulted in an overestimation, particularly at high SUV values and 41%MAX underestimated MTV when there were high levels of heterogeneity [24].

In a publication by Meignan et al. [12], two observers used two percentage-based methods for MTV assessment in DLBCL (41%MAX and a variable SUVmax threshold that visually resulted in optimal segmentations). They found substantial reliability of 0.99 for the 41%MAX threshold and poor reliability of 0.86 for the variable percentage of SUVmax according to Lin's concordance correlation coefficient. This study also suggests that reliability decreased with an increasing level of user interaction.

Based on the ratings of individual lesions it could be argued that no single semi-automated segmentation method performed well for every patient and within every lesion of that patient. Lymphoma sites can be difficult to segment because of heterogeneity within and between lesions. Some patients have many lesions, making it almost impossible to delineate each lesion. Besides that, it should be noted that a visual check of the generated segmentation by an experienced nuclear

medicine physician or radiologist is necessary if a semi-automated method is applied, as was illustrated by the outliers in this pilot study. For example, patient 10 showed a large difference between the three workflows (Figs. 1 to 3). It appeared that the decision whether the liver was involved or not was the main reason for the large differences in the assessments. Both the observers and the NM physicians did not agree on the question of whether the liver was involved or not. In clinical practice, access to additional clinical information (e.g., physical examination or lab results) may help to support the decision whether a site is involved or not. This situation illustrates the importance of the development of clear clinical criteria, definitions, and guidelines for lesion selection in PET/CT studies of patients with different lymphoma types [25].

We also compared the results of the observers (who were clinicians, but not NM physicians) before and after the check of the NM physician. It appeared that only small lesions were added, and in a few patients, physiological uptake was erroneously included in MTV, again supporting the need for checking of results by a NM physician.

This study has strengths and limitations that should be taken into account when interpreting the results. First, we deliberately selected patients with a large variation in number and size of lesions. This might be a strength because it represents examples of different challenges that can occur when analyzing MTV in lymphoma, but this could give a higher prevalence of difficult cases compared with the general DLBCL cohort. However, according to the three experienced nuclear medicine physicians, the dataset was representative of a general DLBCL cohort, even though we selected a relatively small number of patients.

Another strength is the comparison of different workflows for MTV and TLG assessment and their impact on interobserver reliability. Most studies acknowledge the difficulties in the assessment of multiple lymphoma lesions. Some used boxes or VOIs to constrain individual tumors [6,8,12], or limited segmentation to a representative maximum of 5 lymphoma lesions [26], but none of these studies compared such strategies with another workflow.

A limitation is the dependency of the ICC values on the range of MTV values in the population [21]. This is present in other MTV studies as well and hampers

comparability of ICCs within and between studies. Therefore, we also presented CoVs and Bland-Altman plots which are not dependent on the variability of MTV values among patients.

Finally, a preselection strategy as suggested in this study is not yet widely available in other commercially available (clinical) software tools but could be implemented relatively easily after validation in a larger patient cohort.

Future research should focus on the comparison of a preselection strategy in a larger patient cohort with different segmentation methods, their success rates, and the effect on the prognostic value of MTV and TLG measurements. A possible solution for the problem that none of the methods will be satisfactory in each patient and for each lesion could be the use of a MV approach, which should be investigated further. In addition, the effect of reconstruction settings, different uptake times, and effect of adding small lesions on the accuracy of MTV and TLG measurements should be addressed.

Conclusions

A semi-automated workflow based on individual lesion selection (Workflow A) is not recommended, because of the large differences observed in lesion selection. Using a fully automated preselection (SUV≥4.0 and volume≥3ml, Workflow B) of lesions improved interobserver reliability and ease of use of MTV and TLG assessment in DLBCL patients. Subsequent manual modification (Workflow C) is necessary for some patients, but this reduced interobserver reliability which may need to be balanced against any potential improvement of prognostic accuracy.

Acknowledgements

The authors thank all the patients who took part in the trials and the collaborating investigators that kindly supplied their trial data.

Funding Information.

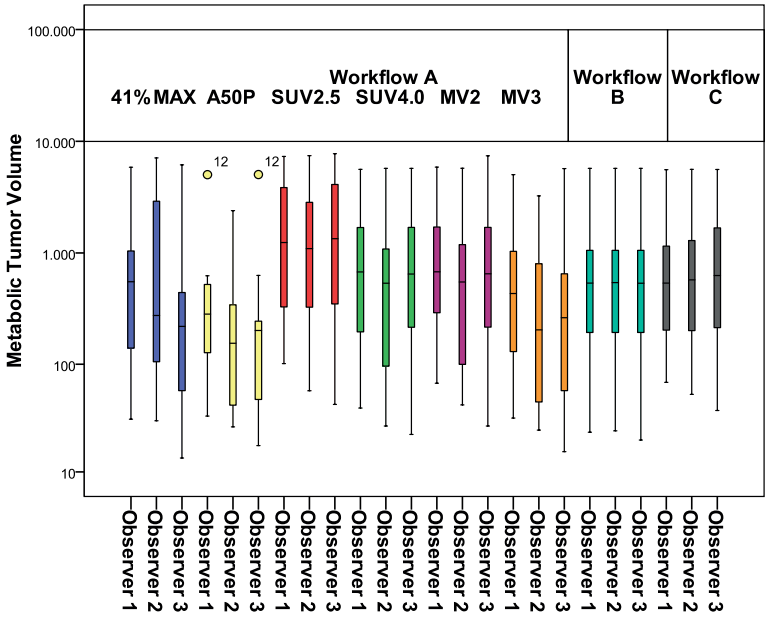
The PETRA study is supported by the Alpe d'HuZes/KWF fund, provided by the Dutch Cancer Society (#VU2012-5848). Alpe d'HuZes/KWF did not have any role in the study design, data collection, analysis, interpretation, and writing of the manuscript.

References

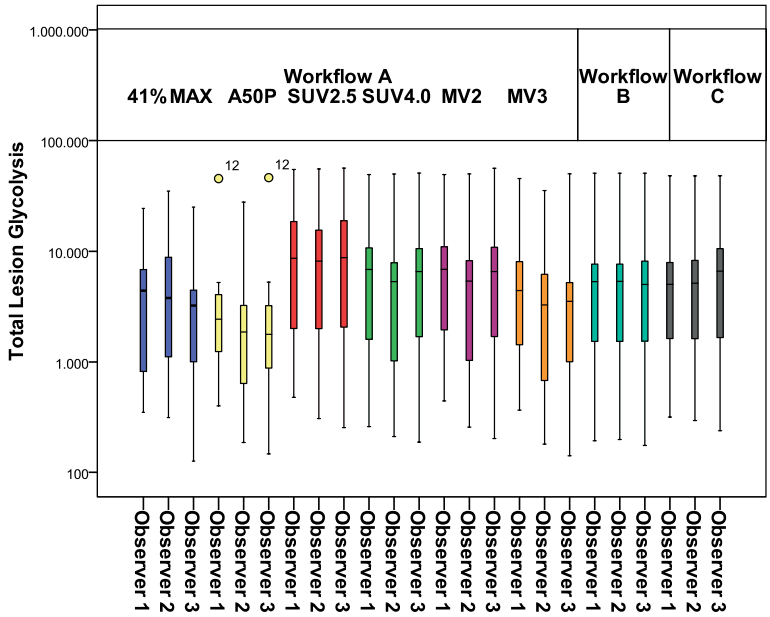
1. Pfreundschuh M, Ho AD, Cavallin-Stahl E et al (2008) Prognostic significance of maximum tumour (bulk) diameter in young patients with good-prognosis diffuse large-B-cell lymphoma treated with CHOP-like chemotherapy with or without rituximab: an exploratory analysis of the MabThera International Trial Group (MInT) study. *Lancet Oncol* . 9:435–444.
2. Barrington SF, Mikhaeel NG, Kostakoglu L et al (2014) 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.* 32:3048–3058.
3. Cheebsumon P, Boellaard R, de Ruysscher D et al (2012) Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology. *EJNMMI Res.* 2:56.
4. Frings V, de Langen AJ, Smit EF et al (2010) Repeatability of metabolically active volume measurements with 18F-FDG and 18FFLT PET in non-small cell lung cancer. *J Nucl Med.* 51:1870–1877.
5. Wang XY, Zhao YF, Liu Y, Yang YK, Zhu Z, Wu N (2017) Comparison of different automated lesion delineation methods for metabolic tumor volume of 18F-FDG PET/CT in patients with stage I lung adenocarcinoma. *Medicine (Baltimore).* 96(51):e9365.
6. Cottereau AS, Lanic H, Mareschal S et al (2016) Molecular profile and FDG-PET/CT total metabolic tumor volume improve risk classification at diagnosis for patients with diffuse large B-cell lymphoma. *Clin Cancer Res.* 22:3801–3809.
7. Mikhaeel NG, Smith D, Dunn JT et al (2016) Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL. *Eur J Nucl Med Mol Imaging.* 43:1209–1219.
8. Sasanelli M, Meignan M, Haioun C et al (2014) Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging.* 41:2017–2022.
9. Song MK, Chung JS, Shin HJ et al (2012) 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.* 91:697–703.
10. Ilyas H, Mikhaeel NG, Dunn JT et al (2018) Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging.* 45:1142–1154.
11. Kurtz DM, Green MR, Bratman SV et al (2015) Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. *Blood.* 125:3679–3687.
12. Meignan M, Sasanelli M, Casasnovas RO et al (2014) Metabolic tumour volumes measured at staging in lymphoma: methodological evaluation on phantom experiments and patients. *Eur J Nucl Med Mol Imaging.* 41:1113–1122.
13. Kostakoglu L, Chauvie S (2018) Metabolic tumor volume metrics in lymphoma. *Semin Nucl Med.* 48:50–66.
14. Cottereau AS, Hapdey S, Chartier L et al (2017) Baseline total metabolic tumor volume measured with fixed or different adaptive thresholding methods equally predicts outcome in peripheral T cell lymphoma. *J Nucl Med.* 58:276–281.
15. Kanoun S, Tal I, Berriolo-Riedinger A et al (2015) 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.* 10:e0140830.
16. Boellaard R (2018) Quantitative oncology molecular analysis suite: ACCURATE [abstract]. *J Nucl Med.* 59(suppl.1):1753.
17. Frings V, van Velden FH, Velasquez LM et al (2014) Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. *Radiology.* 273:539–548.
18. Schaefer A, Vermandel M, Baillet C et al (2016) Impact of consensus contours from multiple PET segmentation methods on the accuracy of functional volume delineation. *Eur J Nucl Med Mol Imaging.* 43:911–924.

19. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1:30–46.
20. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 15:155–163.
21. Portney LG, Watkins MP (2009) Intraclass correlation coefficient (ICC). In: Foundations of clinical research: applications to practice. Pearson Prentice Hall, Upper Saddle River, New Jersey, pp 588–598.
22. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1:307–310.
23. Barrington SF, de Vet HCW, Mikhaeel NG, et al. (2018) Automated segmentation of total tumour burden in DLBCL: which method is most successful? 7thPILM, Menton [Abstract V1] https://d r i v e . g o o g l e . c o m / d r i v e / f o l d e r s /1a0q5uYLBx6HqcUjBxTryiWPUOWFTJGLG;
24. Ceriani L, Milan L, Johnson PWM et al (2019) Baseline PET features to predict prognosis in primary mediastinal B cell lymphoma: a comparative analysis of different methods for measuring baseline metabolic tumour volume. *Eur J Nucl Med Mol Imaging.* 46:1334– 1344.
25. Barrington SF, Meignan MA (2019) Time to prepare for risk adaptation in lymphoma by standardising measurement of metabolic tumour burden. *J Nucl Med.* 60:1096–1102.
26. Parvez A, Tau N, Hussey D, Maganti M, Metser U (2018) (18)F-FDG PET/CT metabolic tumor parameters and radiomics features in aggressive non-Hodgkin's lymphoma as predictors of treatment outcome and survival. *Ann Nucl Med.* 32:410–416.

Supplemental materials



Supplemental Figure 1.



Supplemental Figure 2.

PART III

Clinical validation of PET in lymphoma

CHAPTER 6a

^{18}F-FDG PET Improves Baseline Clinical Predictors of Response in Diffuse Large B-cell Lymphoma: The HOVON-84 Study

Coreline N. Burggraaff
Pieternella J. Lugtenburg
Anne I.J. Arens
Martijn W. Heymans
Sanne E. Wiegers
Ronald Boellaard
Josée M. Zijlstra
Jakoba J. Eertink
Otto S. Hoekstra
Bart de Keizer
Bronno van der Holt
Simone Pieplenbosch
Henrica C.W. de Vet
On behalf of the HOVON Imaging Working Group and HOVON Lymphoma Working Group

Journal of Nuclear Medicine 2022; 63: 1001-1007.

Abstract

We aimed to determine the added value of baseline metabolic tumor volume (MTV) and interim PET (I-PET) to the age-adjusted international prognostic index (aaIPI) to predict 2-y progression-free survival (PFS) in diffuse large B-cell lymphoma. Secondary objectives were to investigate optimal I-PET response criteria (using Deauville score [DS] or quantitative change in SUVmax [ΔSUVmax] between baseline and I-PET4 [observational I-PET scans after 4 cycles of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone administered in 2-wk intervals with intensified rituximab in the first 4 cycles [R(R)-CHOP14]).

Methods:
I-PET4 scans in the HOVON-84 (Hemato-Oncologie voor Volwassenen Nederland [Haemato Oncology Foundation for Adults in the Netherlands]) randomized clinical trial (EudraCT 2006-005174-42) were centrally reviewed using DS (cutoff 4-5). Additionally, ΔSUVmax (prespecified cutoff, 70%) and baseline MTV were measured. Multivariable hazard ratio (HR), positive predictive value (PPV), and negative predictive value (NPV) were obtained for 2-y PFS.

Results:
In total, 513 I-PET4 scans were reviewed according to DS, and ΔSUVmax and baseline MTV were available for 367 and 296 patients. The NPV of I-PET ranged between 82% and 86% for all PET response criteria. Univariate HR and PPV were better for ΔSUVmax (4.8 and 53%, respectively) than for DS (3.1 and 38%, respectively). AaIPI and ΔSUVmax independently predicted 2-y PFS (HR, 3.2 and 5.0, respectively); adding MTV brought about a slight improvement. Low or low-intermediate aaIPI combined with a ΔSUVmax of more than 70% (37% of patients) yielded an NPV of 93%, and the combination of high-intermediate or high aaIPI and a ΔSUVmax of 70% or less yielded a PPV of 65%.

Conclusion:
In this study on diffuse large B-cell lymphoma, I-PET after 4 cycles of R(R)-CHOP14 added predictive value to aaIPI for 2-y PFS, and both were independent response biomarkers in a multivariable Cox model. We externally validated that ΔSUVmax outperformed DS in 2-y PFS prediction.

Keywords:
DLBCL; PET; Deauville score; ΔSUVmax; metabolic tumor volume

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of non-Hodgkin lymphoma, characterized by an aggressive clinical course. Standard first-line treatment consists of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) generally administered at 2-wk (R-CHOP14) or 3-wk (R-CHOP21) intervals.

No significant benefits were shown for R-CHOP14 versus R-CHOP21 in 2 large randomized clinical trials [1,2]. Approximately 25%-40% of DLBCL patients experience relapse or progression in the first years after diagnosis. This problem underlines the need for early stratification between good and poor responders [3,4]. An early switch to second-line treatment in poor responders might improve patient outcomes.

The international prognostic index (IPI) and age-adjusted IPI (aaIPI), both consisting of baseline clinical characteristics, have retained prognostic value after the introduction of rituximab [5]. However, these prognostic indices are not widely used for individual treatment adaptation except for research purposes [6], do not inform about chemosensitivity, and are unable to identify a subgroup with survival clearly below 50%. Therefore, a powerful biomarker (e.g. imaging characteristics during treatment reflecting chemosensitivity) of early response is needed. Recently, measurement of baseline metabolic tumor volume (MTV) was reported to have prognostic value in DLBCL and was suggested as an alternative to IPI [7,8]. Combining MTV with early response assessment at ^{18}F-FDG interim PET (I-PET) further improved prediction of progression-free survival (PFS) [7,8]. Several operationalizations of I-PET response criteria have been proposed, such as the visual 5-point Deauville score (DS, with various possible cutoffs) [9] and quantitative changes in ^{18}F-FDG uptake between baseline and I-PET [10,11].

In the HOVON-84 study (Hemato-Oncologie voor Volwassenen Nederland [Haemato Oncology Foundation for Adults in the Netherlands]), DLBCL patients were randomized between R-CHOP14 and RR-CHOP14 (R-CHOP14 with intensified rituximab in the first 4 cycles) [12]. In both arms, observational I-PET was performed after 4 cycles (I-PET4). To our knowledge, this was the first DLBCL randomized clinical trial in which I-PET4 results did not lead to treatment modification, which enables examination of its predictive value.
Our primary objective was to use prespecified cutoffs and methodologies from previous DLBCL studies to validate the potential added predictive value of baseline MTV and I-PET4 response to baseline clinical characteristics (aaIPI) for 2-y PFS in DLBCL in an independent study. A secondary objective was to determine the optimal I-PET4 response criteria.

Materials and methods

Study population
Newly diagnosed DLBCL patients included in the HOVON-84 NHL study (EudraCT2006-005174-42, NTR1014) with I-PET4 were eligible. For this analysis, we combined the R-CHOP14 and RR-CHOP14 study arms, as there were no statistically significant outcome differences between the arms [12]. Randomization was stratified for aaIPI score. The main eligibility criteria of the clinical study are described elsewhere [12,13]. The HOVON-84 study was approved by the institutional review board of all centers, and participants signed an informed consent form.

Study design
Patients at least 66 y old received 6 cycles of R-CHOP14 followed by 2 additional doses of rituximab; patients aged 65 y or less received 8 cycles of R-CHOP14. Baseline PET was highly recommended but not mandatory. I-PET was performed after 4 cycles R-CHOP14 or RR-CHOP14 (without treatment modifications, I-PET4).

Qualitative and quantitative image analysis
Baseline PET scans were analyzed with the semiautomatic ACCURATE tool [14] (Fig. 1) to obtain MTV using a fixed SUV of at least 4.0 [15,16]. Continuous

MTV values had a nonnormal distribution and were log-transformed using the natural logarithm. We used both the continuous and the dichotomized MTV with a prespecified cutoff adopted from the PETAL study to identify a high-MTV (>345 cm^3) and a low-MTV group (MTV ≤345 cm^3) [8].

I-PET4 scans were centrally reviewed by 2 independent reviewers from a pool of 10 reviewers [13] according to DS criteria [9,17]. Discrepancies were resolved by adjudication. DS4-5 was categorized as no complete metabolic response (PET-positive), and DS1-3 was categorized as complete metabolic response (PET-negative) [9,17]. DS4 was assigned when tumor SUVmax exceeded hepatic SUVmax by fewer than 3 times, and DS5 was assigned when there were new lymphoma lesions or when tumor SUVmax was 3 or more times hepatic SUVmax [9]. The accuracy of other DS cutoffs (i.e. 1 vs. 2-5, 1-2 vs. 3-5, and 1-4 vs. 5) for I-PET4 were evaluated in sensitivity analyses.

In patients with a baseline PET scan and an I-PET4 scan with DS2-5, we measured the change in SUVmax between baseline and I-PET4 (ΔSUVmax). For DS1, ΔSUVmax was set at 100% reduction [9]. We applied a prespecified ΔSUVmax cutoff of 70% reduction between baseline and I-PET4 to define a positive (≤70%) or negative (>70%) I-PET result [10].

Statistical analysis

The primary outcome measure was 2-y PFS, defined as time from randomization to disease progression, relapse, or death from any cause within 2 y [18]. Survival curves were obtained with Kaplan-Meier analyses for PFS stratified by dichotomized PET response criteria and compared with log-rank tests. We used univariate and multivariable Cox proportional hazards regression models to assess the effects of baseline clinical factors (aaIPI, age, B-symptoms, MTV, sex, treatment arm) and I-PET4 response criteria (DS, ΔSUVmax) on 2-y PFS. A backward Wald elimination procedure was used to test which prognostic factors were independently associated with 2-y PFS. In addition, 2x2 contingency tables were constructed to calculate diagnostic measures (i.e. sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) to predict 2-y PFS. Sensitivity, specificity, predictive values, univariate hazard ratio (HR) and receiver operating-characteristic curve were used to define the optimal I-PET4 response criteria to predict 2-y PFS. We examined whether the addition of

baseline MTV to the multivariable Cox model improved prediction. Statistical analyses were performed using SPSS Statistics (version 22; IBM) and R (version 3.6.3). A *P*-value of less than 0.05 was considered statistically significant.

Results

Study population

In total, 574 eligible DLBCL patients were included in the HOVON-84 study; 534 (93%) underwent I-PET4. Twenty-one I-PET4 scans were not evaluable (Fig. 1). The distribution of baseline characteristics and 2-y PFS were similar for patients with and without baseline MTV, I-PET4, and ΔSUVmax evaluations (Table 1).

Prognostic value of baseline aaIPI and MTV

After a median follow-up of 91 mo (interquartile range, 84-101 mo) the estimated 2-y PFS was 79% (95% CI, 76%-83%). Most patients belonged to the low-intermediate or high-intermediate aaIPI groups (35% and 50%, respectively, Table 1). In the Kaplan-Meier analysis both low and low-intermediate aaIPI survival curves and high-intermediate and high aaIPI survival curves crossed each other without statistically significant differences (Supplemental Fig. 1a; supplemental materials are available at http://jnm.snmjournals.org). Dichotomization into low or low-intermediate and high-intermediate or high yielded a 2-y PFS of 91% (95%CI, 87%-95%) and 71% (95%CI, 66%-76%), respectively, with a corresponding univariate HR of 3.6 (95%CI, 2.2-5.9; Supplemental Fig. 1b; Table 2).

Of 384 patients who underwent baseline PET, baseline MTV was measurable in 296 (52%; Fig. 1). The continuous log-transformed MTV had a univariate HR of 1.4 (95%CI, 1.2-1.8; Supplemental Table 1). Patients in the low-MTV group (MTV≤345 cm^3, *n*=137; 46%) had a 2-y PFS of 86% (95%CI, 80%-92%) versus 75% (95%CI, 68%-81%) in the high-MTV group (MTV>345 cm^3, *n*=159; 54%), with a corresponding univariate HR of 2.0 (95%CI, 1.1-3.4; Table 2). I-PET and end-of-treatment PET scans were both available in 474 patients (Supplemental Table 2), with an overall agreement of 87% (95%CI, 84%-90%).

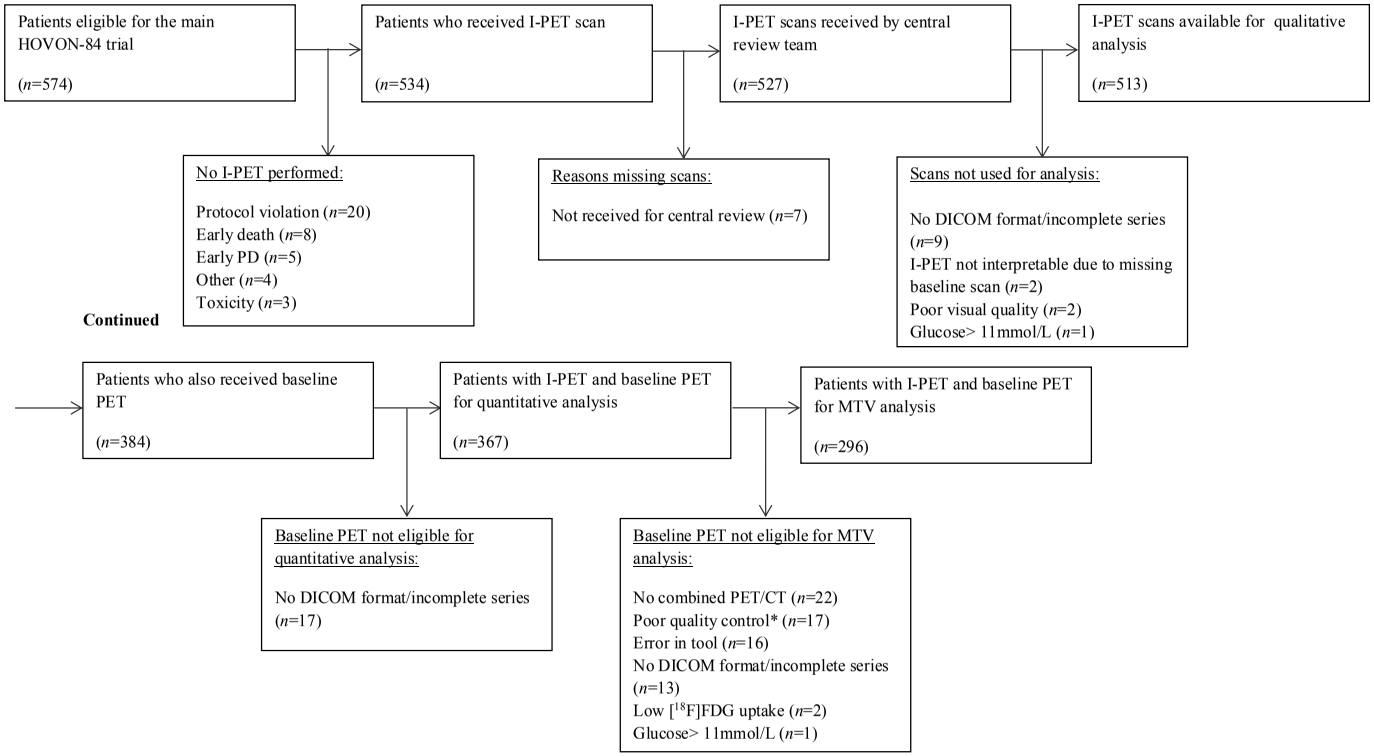


Figure 1. Flowchart of PET scans available for I-PET4, ΔSUVmax, and baseline MTV analyses.

*PET quality was acceptable when liver SUVmean was 1.3-3.0 and total image activity was between 50%-80% of total injected dose.

PD=progressive disease.

Table 1. Baseline Patient Characteristics.

Characteristic	I-PET4	ΔSUVmax	MTV
Number of patients	513 (100)	367 (100)	296 (100)
Age at diagnosis (y)			
Median (range)	65 (23-80)	65 (23-80)	65 (23-80)
≤60	172 (33.5)	123 (33.5)	96 (32.4)
>60	341 (66.5)	244 (66.5)	200 (67.6)
Sex			
Male	267 (52.0)	192 (52.3)	150 (50.7)
Female	246 (48.0)	175 (47.7)	146 (49.3)
WHO performance status			
0	266 (51.9)	201 (54.8)	165 (55.7)
1	183 (35.7)	118 (32.2)	92 (31.1)
2	61 (11.9)	46 (12.5)	37 (12.5)
Unknown	3 (0.6)	2 (0.5)	2 (0.7)
Ann Arbor Stage			
II	97 (18.9)	61 (16.6)	52 (17.6)
III	163 (31.8)	113 (30.8)	90 (30.4)
IV	253 (49.3)	193 (52.6)	154 (52.0)
LDH			
Normal	171 (33.3)	124 (33.8)	98 (33.1)
>Normal	342 (66.7)	243 (66.2)	198 (66.9)
aaIPI			
Low	36 (7.0)	23 (6.3)	21 (7.1)
Low-intermediate	177 (34.5)	127 (34.6)	97 (32.8)
High-intermediate	255 (49.7)	181 (49.3)	150 (50.7)
High	45 (8.8)	36 (9.8)	28 (9.5)
B symptoms			
No	297 (57.9)	211 (57.5)	169 (57.1)
Yes	216 (42.1)	156 (42.5)	127 (42.9)
Treatment Arm			
R-CHOP14	252 (49.1)	186 (50.7)	150 (50.7)
RR-CHOP14	261 (50.9)	181 (49.3)	146 (49.3)
Diagnosis-treatment interval (d)			
Median(IQR)	20 (13-28)	20 (13-28)	20 (14-28)
Range	1-112	1-81	1-81
Baseline PET	384 (74.9)	367 (100)	296 (100)

IQR=interquartile range; LDH=lactate dehydrogenase; WHO=world health organization.
Data are number followed by percentage in parentheses, unless indicated otherwise.

I-PET4 analyses

Of 513 I-PET4 scans, 113 (22%) were rated as PET-positive (no complete metabolic response). Dichotomization of I-PET4 results into DS4-5 (positive) versus DS1-3 (negative) yielded a 2-y PFS of 61% (95%CI, 52%-70%) for I-PET4-positive patients and 84% (95%CI, 81%-88%) for I-PET4-negative patients (P <0.001), with a corresponding univariate HR of 3.1 (95%CI, 2.1-4.5; Table 2; Fig. 2a). Among the patients who experienced a relapse, the median time to relapse for I-PET4-positives was 8.1 mo (interquartile range, 4.4-23.2), versus 18.1 mo

Table 2. Diagnostic and Prognostic Measures for aaIPI, Baseline MTV, Different Cutoffs of Deauville 5-Point Scale at I-PET4, and ΔSUVmax for 2-Year PFS.

			Diagnostic information				Prognostic information		
Measure	**Parameter**	**Patients (*n*)**	**NPV**	**PPV**	**Sensitivity**	**Specificity**	**Univariate HR**	***P***	**Discrimination (AUC)**
aaIPI	**L/LI vs. HI/H**	213 vs. 300	91.1 (86.5-94.2)	28.7 (23.9-34.0)	81.9 (73.5-88.1)	47.6 (42.8-52.4)	3.59 (2.18-5.90)	<0.0001	0.63 (0.58-0.68)
Baseline MTV	**≤345 vs. >345 cm^3**	137 vs. 159	86.1 (79.4-90.9)	25.2 (19.2-32.4)	67.8 (55.1-78.3)	49.8 (43.5-56.1)	1.96 (1.13-3.38)	0.0161	0.58 (0.52-0.65)
I-PET4	**DS1 vs. DS2-5**	178 vs. 335	82.0 (75.7-87.0)	21.8 (17.7-26.5)	69.5 (60.2-77.5)	35.8 (31.3-40.5)	1.26 (0.83-1.91)	0.275	0.53 (0.48-0.57)
	DS1-2 vs. DS3-5	290 vs. 223	84.5 (79.9-88.2)	26.9 (21.5-33.1)	57.1 (47.6-66.2)	60.1 (55.2-64.7)	1.95 (1.32-2.87)	<0.0001	0.59 (0.54-0.64)
	DS1-3 vs. DS4-5	400 vs. 113	84.5 (80.6-87.7)	38.1 (29.6-47.3)	41.0 (32.0-50.5)	82.8 (78.9-86.2)	3.07 (2.08-4.54)	<0.0001	0.62 (0.58-0.66)
	DS1-4 vs. DS5	488 vs. 25	82.0 (78.3-85.1)	68.0 (48.4-82.8)	16.2 (10.4-22.4)	98.0 (96.2-99.0)	7.40 (4.39-12.48)	<0.0001	0.57 (0.56-0.59)
ΔSUVmax	**>70% vs. ≤70%**	329 vs. 38	82.7 (78.2-86.4)	52.6 (37.3-67.5)	26.0 (17.5-36.7)	93.8 (90.4-96.0)	4.80 (2.88-8.00)	<0.0001	0.60 (0.57-0.63)

AUC=area under the receiver operating curve; L=low-risk group; LI=low- to intermediate-risk group; HI=high- to intermediate-risk group; H=high-risk group. Diagnostic information is percentage: data in parentheses are 95% CIs.

(interquartile range, 8.3-46.3) for I-PET-negatives. The corresponding PPV and NPV for 2-y PFS were 38% (95%CI, 30%-47%) and 85% (95%CI, 81%-88%), respectively.

Optimal I-PET4 response criterion

For various DS cutoffs, NPVs ranged between 82% and 85% for I-PET4 (Table 2). PPVs varied widely for different cutoffs (22%-68%); the highest PPV was seen for the DS5 cutoff in I-PET4 (68%). Also, the univariate HR of 7.4 was highest for the DS1-4 cutoff versus DS5, yielding the best separation between good and poor outcome (Supplemental Fig. 2). However, only 25 of 513 patients (5%) had a DS5.

ΔSUVmax analysis was feasible in 367 of 574 patients (64%; Fig. 1). In patients with no more than a 70% ΔSUVmax reduction between baseline and I-PET4 (n=38, 10%) the 2-y PFS was 47% (95%CI, 31%-63%), versus 83% (95%CI, 78%-87%) for patients with more than a 70% reduction (Fig. 2b, P<0.001) with a univariate HR of 4.8 (95%CI, 2.9-8.0). Corresponding PPVs and NPVs for 2-y PFS were 53% (95%CI, 37%-68%) and 83% (95%CI, 78%-86%), respectively (Table 2). Repeating these comparisons in the 296 patients with complete metrics on baseline MTV yielded similar results (Supplemental Table 3).

PPV and HRs were better for ΔSUVmax than for the most commonly used cutoff, DS4-5 (53% vs. 38% and 4.8 vs. 3.1, respectively). NPV was above 80% for all applied criteria. When ΔSUVmax was compared with the most commonly used DS cutoff, DS4-5, ΔSUVmax was preferred for prediction of 2-y PFS, but the highest PPV and HR were found for the DS5 cutoff.

Combined baseline and I-PET4 analysis

Statistically significant prognostic factors for 2-y PFS in univariate Cox regression analyses were a ΔSUVmax of 70% or less, a high-intermediate or high aaIPI, and B-symptoms. In multivariable analysis, a high-intermediate or high aaIPI and no more than a 70% reduction of ΔSUVmax were independently associated with 2-y PFS (Supplemental Table 4). A Low or low-intermediate aaIPI and a ΔSUVmax of more than 70% (37% of patients) resulted in an NPV of 93% (95%CI, 87%-96%), whereas a high-intermediate or high aaIPI and a ΔSUVmax of 70% or less (6% of patients) resulted in a PPV of 65% (95%CI, 45%-81%, Supplemental Fig. 3).

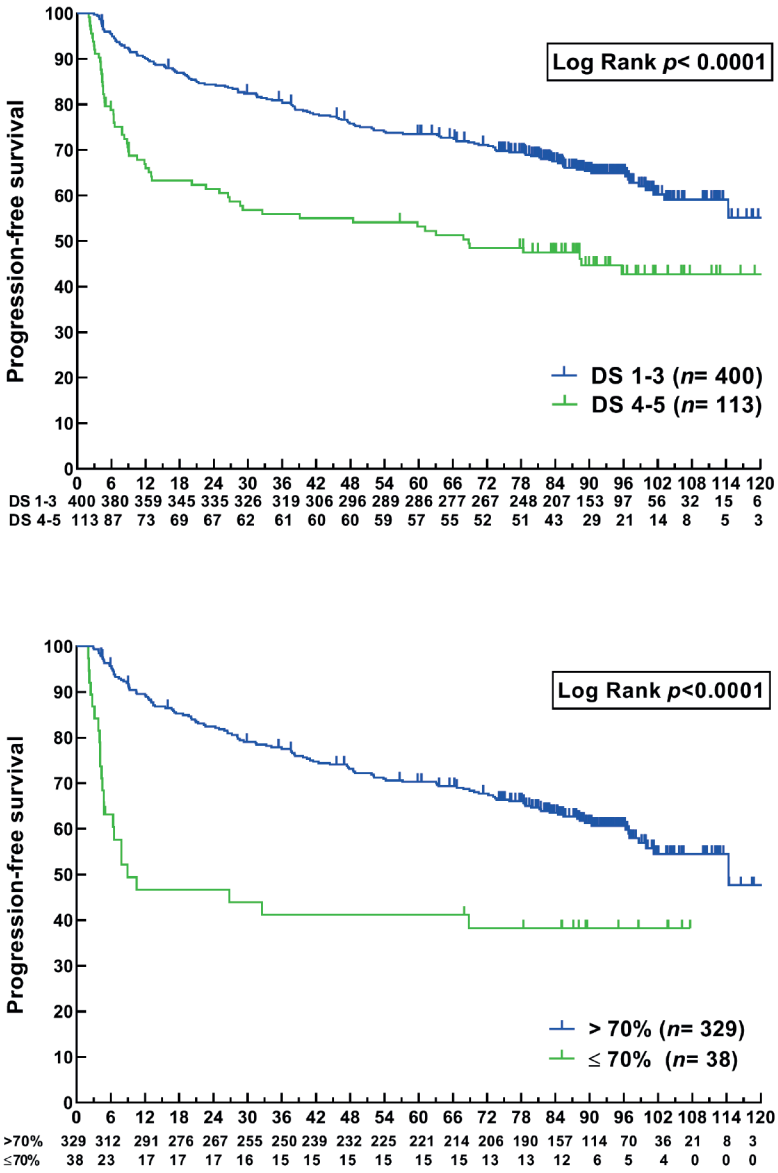


Figure 2. Kaplan-Meier curves with numbers at risk for PFS in months stratified by I-PET4 result according to DS (A) and according to ΔSUVmax result (B).

Dichotomized baseline MTV did not add prognostic value to ΔSUVmax and aaIPI for prediction of 2-y PFS. When continuous log-transformed MTV was added to the multivariable Cox model, aaIPI was eliminated by backward

elimination, yielding log-transformed MTV, an age of more than 60 y, B symptoms, and ΔSUVmax as factors independently associated factors with 2-y PFS (Supplemental Table 1)

Overall survival analyses

The results of the response criteria and uni- and multivariable analyses for 2-y overall survival are presented in Supplemental Tables 5-7 and Supplemental Figure 4.

Discussion

In this multicenter study, DLBCL I-PET after 4 cycles of R(R)-CHOP14 added predictive value to baseline clinical characteristics (aaIPI) for 2-y PFS, with high NPVs (82%-86%) independent of all I-PET response criteria. However, the PPV was still relatively low. Combining clinical and PET data showed that aaIPI and ΔSUVmax were independently associated with 2-y PFS, with HRs of 3.2 and 5.0, respectively. Adding log-transformed baseline MTV only slightly improved the predictive value combined with the ΔSUVmax response criteria. As a secondary objective, we compared the most commonly used visual and semiquantitative criteria and externally validated that ΔSUVmax criteria were the optimal I-PET4 criteria to predict 2-y PFS, with a HR of 4.8 and a PPV of 53%.

On the basis of the PPV and univariate HR in I-PET, the DS5 cutoff performed best, with a PFS clearly below 50% for the DS5 group. However, the percentage of DS5-positive patients was low (5%), but this group could be of interest for future new therapy strategies. The univariate HR for 2-y PFS with DS4-5 cutoff in I-PET4 was 3.1 (95%CI, 2.1-4.5), which is similar to the pooled HR of 3.1 (95%CI, 2.5-3.9) in a systematic review, even though in that review I-PET was performed after 1-4 cycles of treatment and less strict I-PET response criteria were applied [19]. The NPV for 2-y PFS in our study was 85%, which is in line with these previous studies generally reporting NPVs above 80% (range, 64%–95%, [19]).

Two recent retrospective DLBCL studies analyzed the value of I-PET after 4 cycles [20,21], and both concluded that ΔSUVmax had a higher accuracy and PPV than DS in predicting PFS. The retrospective study from Itti et al. (*n*=114,

I-PET after 2 cycles), who analyzed different cutoffs for DS after 2 cycles, reported PPVs for DS4-5 and ΔSUVmax that were remarkably identical to our study (39% vs. 38% and 52% vs. 53%, respectively) [22]. A DLBCL subgroup analysis of the PETAL study also reports a more favorable PPV for ΔSUVmax I-PET assessment than for Deauville assessment [23].

Baseline clinical characteristics and chemoimmunotherapy sensitivity are both relevant factors in outcome prediction. This relevancy was demonstrated in our multivariable analysis, in which aaIPI and ΔSUVmax (reflecting chemosensitivity) were both independent predictors of 2-y PFS. Again, the subgroup with both high-intermediate or high aaIPI and a ΔSUVmax of 70% or less had a PFS clearly below 50% but was relatively small (6% of all patients). Selection of a poor-risk group of only 6% is justified both from a cost awareness perspective and because it is the group most likely not be cured by standard treatment. These patients can be treated within clinical trials investigating the efficacy of new drugs.

Several relatively small retrospective studies reported inconsistent results regarding associations of clinical characteristics and I-PET results (DS or ΔSUVmax) with survival in multivariable Cox models [7,22,24]. Two prospective studies concluded that only I-PET and not IPI was independently associated with event-free survival [25,26]. The randomized phase III trials PETAL (I-PET after 2 cycles of R-CHOP21) and CALGB-50303 (I-PET after 2 cycles R-CHOP21 or DA-EPOCH-R [dose-adjusted etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin, and rituximab]) also concluded that I-PET with ΔSUVmax (cutoff 66%) and IPI were independent predictors for event-free survival and PFS [11,27], respectively.

Baseline MTV assessment was not a strong predictor of 2-y PFS in our study (Table 2; Supplemental Table 1,3,5, and 7). We used a segmentation method applying a fixed SUV of at least 4.0, on the basis of a recent study showing that this method performed best and had a discriminative power similar to that of other segmentation methods [16]. Addition of dichotomized baseline MTV (345cm^3 cutoff) to ΔSUVmax did not improve the predictive value, but log-transformed continuous MTV added some independent predictive value when combined with ΔSUVmax. In a secondary analysis of the PETAL randomized clinical trial (DLBCL subset, I-PET after 2 cycles, same MTV software and methodology as

in our study), baseline MTV and ΔSUVmax were the only independent outcome predictors [8,28]. We could not confirm these findings; possible explanations are the different PET-timing (HOVON-84: I-PET4) or patient characteristics (HOVON-84: median age 3 y higher; advanced stage, 82% vs. 58% in PETAL). We chose a higher ΔSUVmax because the PET timing was different (I-PET4 vs. I-PET2) and to validate a formerly presented cutoff [10,20]. This choice does not explain the difference in added value of MTV, since the positivity percentages were the same (10.4% vs. 9.6% in PETAL), as was the 2-y PFS for the positive (46.9% and 46.7%) and negative (80.2% and 82.5%) groups according to the ΔSUVmax criteria for HOVON-84 and PETAL, respectively. Recently, Vercellino et al. showed that a combination of high baseline MTV and high performance status (≥2) identifies an ultra-risk DLBCL population [29]. We could not confirm this extra risk in our study.

There were several strengths to our study. First, to our knowledge, there are no other large, randomized trials with a homogeneous first-line treatment regimen and an observational I-PET after 4 R-CHOP14 cycles. Another strength was the central review procedure for Deauville scoring, with 2 independent reviewers and a strict DS5 definition, which allowed for an analysis to determine the optimal I-PET4 response criteria [13].

On the basis of the relatively low values for PPV, escalation of treatment for the I-PET4 positive group is not yet recommend for clinical practice, but evidence for I-PET adapted treatment is clearly growing (11,30-32). The GAINED randomized clinical trial [30] enrolled 670 DLBCL patients (aged 18-60 y, aaIPI ≥1); I-PET2-positive/I-PET4-negative patients (*n*=87) were scheduled to receive high-dose chemotherapy with autologous stem cell transplantation and had no statistically different PFS from the I-PET2-negative/I-PET4-negative patients (*n*=401) who continued standard treatment. However, no firm conclusions can be made, because there was no randomization within these I-PET-adapted groups.

Because the NPV is acceptable (>80% for all criteria), reduction of treatment based on I-PET4 could be of interest, especially for low-risk and elderly patients. The randomized FLYER trial showed that in a group of 592 DLBCL patients (aged 18-60 y, no aaIPI risk factors, no bulky disease), 4 cycles of R-CHOP21+ 2 cycles of rituximab was not inferior to 6 cycles of R-CHOP21 [6], and in an

exploratory analysis the international GOYA randomized clinical trial found no PFS benefit with 8 cycles of R-CHOP21 compared with 6 cycles of R-CHOP21 +2 cycles of rituximab [31]. The S1001 study presented 4 cycles R-CHOP as the new standard for most patients with limited-stage disease [32].

Conclusions

In this large DLBCL study, I-PET after 4 cycles of R(R)-CHOP14 added predictive value to aaIPI for 2-y PFS, and both were independent response biomarkers in a multivariable Cox model, yielding a high NPV of 93% for 2-y PFS. Comparing the most commonly used DS and ΔSUVmax cutoffs, the optimal response criterion for I-PET4 to predict 2-y PFS was ΔSUVmax.

Disclosure

This work was supported by the Alpe d'HuZes/KWF fund, provided by the Dutch Cancer Society (VU2012-5848). Pieternella Lugtenburg receives research funding from Roche, Takeda, and Servier and honoraria for advisory boards from Roche, Takeda, Servier, Genmab, Celgene, and Genentech. Josée Zijlstra receives research funding from, and is on the advisory board for, Roche. No other potential conflict of interest relevant to this article was reported.

Key points

Question:
What value do baseline MTV and I-PET add to aaIPI in predicting 2-y PFS in DLBCL, and what are the optimal I-PET response criteria?

Pertinent findings:
aaIPI and ΔSUVmax were independent predictors for 2-y PFS in DLBCL. Six percent of patients had a high PPV of 65% resulting in poor survival outcome. ΔSUVmax outperformed Deauville score in 2-y PFS prediction

Implications for patient care:
The subgroup comprising the 6% of patients having a high or high-intermediate aaIPI and a 70% or less ΔSUVmax reduction at I-PET is of interest for testing new therapy strategies in DLBCL

References

1. 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.
2. Delarue R, Tilly H, Mounier N, et al. Dose-dense rituximab-CHOP compared with standard rituximab-CHOP in elderly patients with diffuse large B-cell lymphoma (the LNH03-6B study): a randomised phase 3 trial. *Lancet Oncol.* 2013;14:525–533.
3. van Imhoff GW, McMillan A, Matasar MJ, et al. Ofatumumab versus rituximab salvage chemoimmunotherapy in relapsed or refractory diffuse large B-cell lymphoma: the ORCHARRD study. *J Clin Oncol.* 2017;35:544–551.
4. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28:4184–4190.
5. Ziepert M, Hasenclever D, Kuhnt E, et al. Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28:2373–2380.
6. Poeschel V, Held G, Ziepert M, et al. FLYER Trial Investigators; German Lymphoma Alliance. Four versus six cycles of CHOP chemotherapy in combination with six applications of rituximab in patients with aggressive B-cell lymphoma with favourable prognosis (FLYER): a randomised, phase 3, non-inferiority trial. *Lancet.* 2020;394:2271–2281.
7. 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.
8. 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.
9. 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.
10. 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.
11. Dührsen U, Müller S, Hertenstein B, et al. PETAL Trial Investigators. Positron emission tomography-guided therapy of aggressive non-Hodgkin lymphomas (PETAL): a multicenter, randomized phase III trial. *J Clin Oncol.* 2018;36:2024–2034.
12. Lugtenburg PJ, de Nully Brown P, van der Holt B, et al. Rituximab-CHOP with early rituximab intensification for diffuse large B-cell lymphoma: a randomized phase 3 trial of the HOVON and the Nordic Lymphoma Group (HOVON-84). *J Clin Oncol.* 2020;38:3377–3387.
13. Burggraaff CN, Cornelisse AC, Hoekstra OS, et al. HOVON Imaging Working Group. Interobserver agreement of interim and end-of-treatment (18)F-FDG PET/CT in diffuse large B-cell lymphoma (DLBCL): impact on clinical practice and trials. *J Nucl Med.* 2018;59:1831–1836.
14. Boellaard R. Quantitative oncology molecular analysis suite: ACCURATE [abstract]. *J Nucl Med.* 2018;59(suppl.1):1753.
15. Burggraaff CN, Rahman F, Kaßner I, et al. PETRA Consortium. Optimizing workflows for fast and reliable metabolic tumor volume measurements in diffuse large B cell lymphoma. *Mol Imaging Biol.* 2020;22:1102–1110.
16. Barrington SF, Zwezerijnen BGJC, de Vet HCW, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method Is most successful? A study on behalf of the PETRA consortium. *J Nucl Med.* 2021;62:332-337.
17. 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.

18. Maurer MJ, Habermann TM, Shi Q, et al. Progression-free survival at 24 months (PFS24) and subsequent outcome for patients with diffuse large B-cell lymphoma (DLBCL) enrolled on randomized clinical trials. *Ann Oncol.* 2018;29:1822–1827.
19. Burggraaff CN, de Jong A, Hoekstra OS, et al. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2019;46:65–79.
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. Li X, Sun X, Li J, Liu Z, et al. Interim PET/CT based on visual and semiquantitative analysis predicts survival in patients with diffuse large B-cell lymphoma. *Cancer Med.* 2019;8:5012–5022.
22. 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.
23. 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 ΔSUVmax method. *J Nucl Med.* 2021;62:37–42.
24. Nols N, Mounier N, Bouazza S, et al. Quantitative and qualitative analysis of metabolic response at interim positron emission tomography scan combined with international prognostic index is highly predictive of outcome in diffuse large B-cell lymphoma. *Leuk Lymphoma.* 2014;55:773– 780.
25. Carr R, Fanti S, Paez D, et al. IAEA Lymphoma Study Group. Prospective international cohort study demonstrates inability of interim PET to predict treatment failure in diffuse large B-cell lymphoma. *J Nucl Med.* 201;55:1936–1944.
26. 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.
27. Schöder H, Polley MY, 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.
28. Schmitz C, Hüttmann A, Müller SP, et al. Supporting data for positron emission tomography based risk modelling using a fixed-instead of a relative thresholding method for total metabolic tumor volume determination. *Data Brief.* 2019;28:104976.
29. 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.
30. Le Gouill S, Ghesquières H, Oberic L, et al. Obinutuzumab vs rituximab for advanced DLBCL: a PET-guided and randomized phase 3 study by LYSA. *Blood.* 2021;137:2307-2320.
31. Sehn LH, Congiu AG, Culligan DJ, et al. No added benefit of eight versus six cycles of CHOP when combined with rituximab in previously untreated diffuse large B-cell lymphoma patients: results from the international phase III GOYA study. *Blood.* 2018;132:783.
32. Persky DO, Li H, Stephens DM. et al. Positron emission tomography–directed therapy for patients with limited-stage diffuse large B-cell lymphoma: results of intergroup national clinical trials Network Study S1001. *J Clin Oncol.* 2020;38:3003-3011.

Supplemental materials

Supplemental Table 1. Uni- and multivariable Cox Proportional Hazard analyses including baseline MTV for 2-year PFS (*n*=296).

	2-year PFS			
	Univariate HR (95%CI)	***P*-value**	**Multivariable HR (95%CI)**	***P*-value**
Age (≤60 vs >60)	1·44 (0·80-2·59)	0·222	1·83 (1·01-3·32)	0·046
aaIPI (low/low-intermediate vs high-intermediate/high)	2·83 (1·50-5·34)	0·001*		
B symptoms (no vs yes)	1·97 (1·18-3·30)	0·010*	1·75 (1·04-2·98)	0·036
Baseline MTV log-transformed	1·43 (1·16-1·76)	0·001*	1·32 (1·07-1·62)	0·010
ΔSUVmax (>70% vs ≤70%)	7·44 (4·29-12·92)	<0·0001*	7·87 (4·48-13·83)	<0·0001*
Gender (male vs female)	0·73 (0·44-1·23)	0·240		
Treatment arm (R-CHOP14 vs RR-CHOP14	0·85 (0·51-1·42)	0·539		

* Statistically significant difference

Abbreviations: 95%CI= 95% confidence interval; aaIPI= age-adjusted international prognostic index; HR= Hazard Ratio; LDH= lactate dehydrogenase; MTV=metabolic tumor volume; PFS= progression-free survival; WHO= world health organization.

Supplemental Table 2. I-PET4 and EoT-PET 2x2 contingency table.

	EoT-PET positive (DS 4-5)	**EoT-PET negative (DS 1-3)**	**Total**
I-PET4 positive (DS 4-5)	54	42~	96
I-PET4 negative (DS 1-3)	20^	358	378
Total	74 *	400 †	474

Abbreviations: DS= Deauville 5-point scale; EoT-PET= end-of-treatment ^{18}F-FDG PET(/CT); I-PET4= interim ^{18}F-FDG PET(/CT) after four treatment cycles.

*No I-PET4 was performed in 5 patients with positive EoT-PET (reasons unknown).
†No I-PET4 was performed in 14 patients with negative EoT-PET(reasons unknown), in 3 patients I-PET4 was not available for qualitative analysis (high glucose, poor visual quality and not interpretable due to missing baseline scan, respectively).
^ Twenty patients (4.2%) switched from a negative I-PET4 to a positive EoT-PET, sixteen of these patients had a high-intermediate or high aaIPI and had a 2-year PFS of 40% (95%CI 18-62%).
~ Forty-two patients (8.9%) had a positive I-PET4 and turned negative at EoT-PET, of these only 4 patients had progressive disease within 2 years after randomization of whom 2 died within this period. These converting patients had a 2-year PFS of 90% (95%CI 81-99%).

Supplemental Table 3. Diagnostic and prognostic measures for aaIPI, baseline MTV, for different cutoff values of the Deauville 5-point scale at I-PET4, and ΔSUVmax for 2-year PFS for subset of patients with baseline MTV analysis (*n*=296).

			Diagnostic information				Prognostic information		Discrimination
		Number of patients (*n*)	Negative Predictive Value %(95%CI)	Positive Predictive Value %(95%CI)	Sensitivity %(95%CI)	Specificity %(95%CI)	Univariate Hazard Ratio (95%CI)	*P*-value	AUC (95%CI)
AaIPI	**L/LI vs HI/H**	118 vs 178	89·8 (83·1-94·1)	26·4 (20·5-33·3)	79·7 (67·7-88·0)	44·7 (38·5-51·2)	2·83 (1·50-5·34)	0·0013	0·61 (0·55-0·67)
Baseline MTV	**≤345ml vs >345ml**	137 vs 159	86·1 (79·4-90·9)	25·2 (19·2-32·4)	67·8 (55·1-78·3)	49·8 (43·5-56·1)	1·96 (1·13-3·38)	0·0161	0·58 (0·52-0·65)
I-PET4	**DS1 vs DS2-5**	88 vs 208	84·1 (75·1-90·7)	21·6 (16·6-27·7)	76·3 (64·0-85·3)	31·2 (25·7-37·4)	1·42 (0·78-2·59)	0·252	0·54 (0·48-0·59)
	DS1-2 vs DS3-5	159 vs 137	86·8 (80·7-91·2)	27·7 (20·9-35·8)	64·4 (51·7-75·4)	58·2 (51·9-64·3)	2·39 (1·40-4·07)	0·0014	0·61 (0·55-0·67)
	DS1-3 vs DS4-5	226 vs 70	86·7 (81·7-90·5)	41·4 (30·6-53·1)	49·2 (36·8-61·6)	82·7 (77·4-87·0)	3·99 (2·39-6·66)	<0·0001	0·65 (0·60-0·70)
	DS1-4 vs DS5	280 vs 16	83·2 (78·4-87·1)	75·0 (50·5-89·8)	20·3 (12·0-32·3)	98·3 (95·7-99·3)	9·49 (5·00-18·01)	<0·0001	0·59 (0·57-0·62)
ΔSUVmax	**>70% vs ≤70%**	266 vs 30	85·0 (80·2-88·8)	63·3 (45·5-78·1)	32·2 (21·7-44·9)	95·4 (91·9-97·4)	7·46 (4·30-12·95)	<0·0001	0·64 (0·61-0·67)

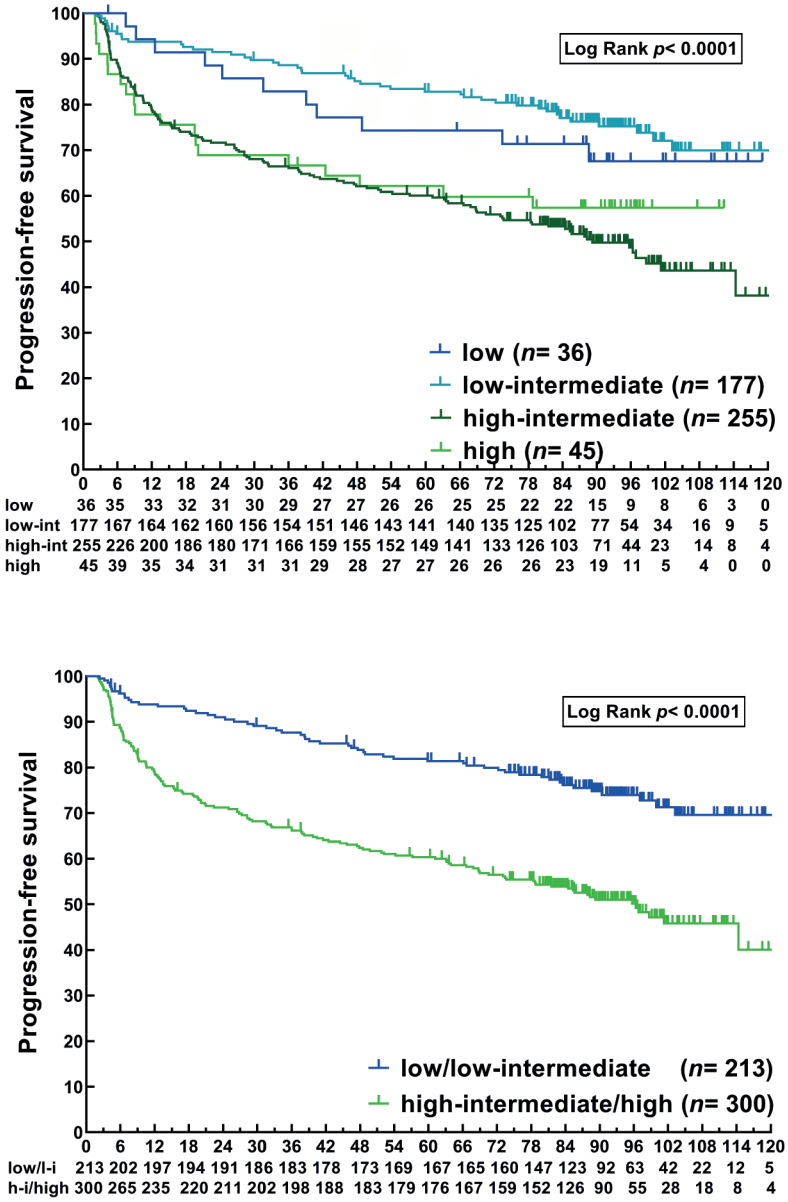
Abbreviations: 95%CI= 95% confidence interval; AaIPI= age-adjusted international prognostic index; AUC= area under the receiver operating curve; DS= Deauville 5-point scale; ΔSUVmax= reduction of maximum standardized uptake value between baseline and interim ^{18}F-FDG PET(/CT); H=high risk group; HI= high-intermediate risk group; I-PET= interim ^{18}F-FDG PET(/CT) after four cycles; MTV= metabolic tumor volume; PFS= progression-free survival.

Supplemental Table 4. Uni- and multivariable Cox Proportional Hazard analyses of ΔSUVmax analysis-group for 2-year PFS (*n*=367).

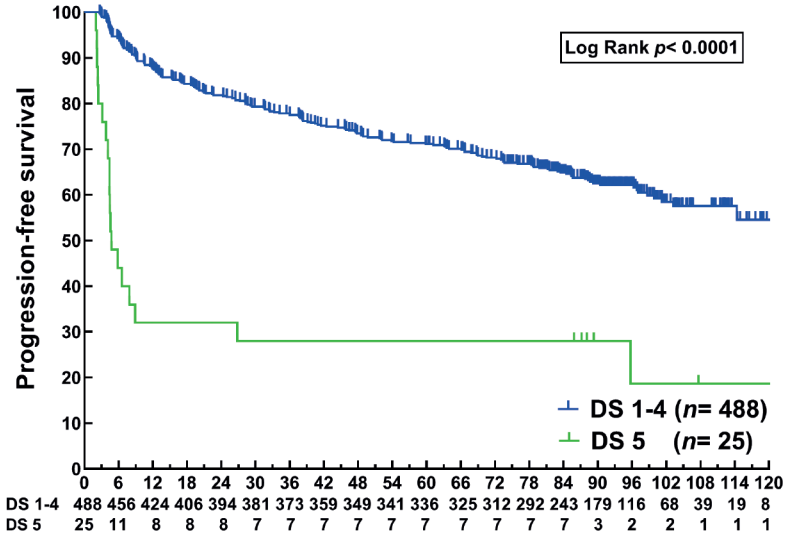
	2-year PFS			
	Univariate HR (95%CI)	***P*-value**	**Multivariable HR (95%CI)**	***P*-value**
Age (≤60 vs >60)	1·60 (0·95-2·69)	0·075		
aaIPI (low/low-intermediate vs high-intermediate/high)	3·16 (1·80-5·55)	<0·0001*	3·27 (1·86-5·75)	<0·0001*
B symptoms (no vs yes)	1·67 (1·07-2·61)	0·025*		
ΔSUVmax (>70% vs ≤70%)	4·80 (2·88-8·00)	<0·0001*	5·01 (3·00-8·36)	<0·0001*
Gender (male vs female)	1·25 (0·80-1·96)	0·335		
Treatment arm (R-CHOP14 vs RR-CHOP14	0·99 (0·63-1·54)	0·957		

* Statistically significant difference

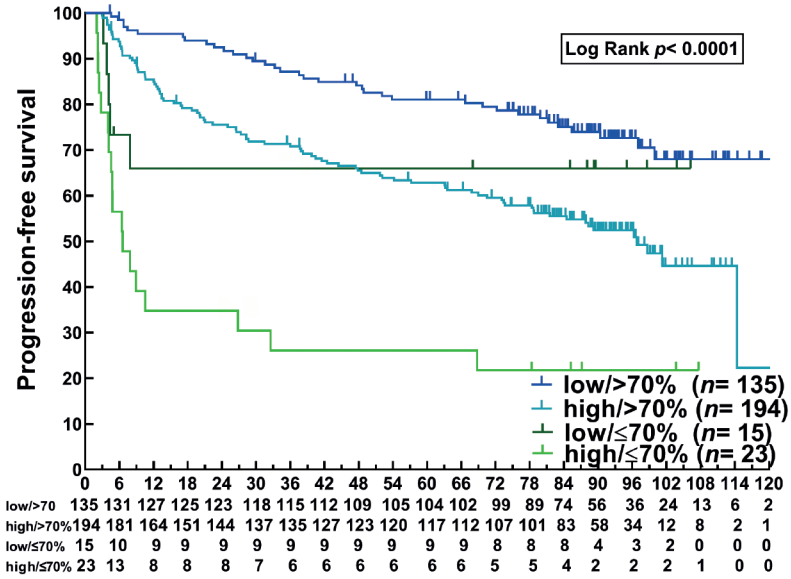
Abbreviations: 95%CI= 95% confidence interval; aaIPI= age-adjusted international prognostic index; ΔSUVmax= reduction of maximum standardized uptake value between baseline and interim ^{18}F-FDG PET(/CT); HR= Hazard Ratio; LDH= lactate dehydrogenase; PFS= progression-free survival; WHO= world health organization.



Supplemental Figure 1. Kaplan-Meier curves for progression-free survival in months stratified by ordinal aaIPI (1a) and dichotomized aaIPI (1b).



Supplemental Figure 2. Kaplan-Meier curves for I-PET4 with numbers at risk for progression-free survival in months stratified by DS1-4 vs DS5 result.

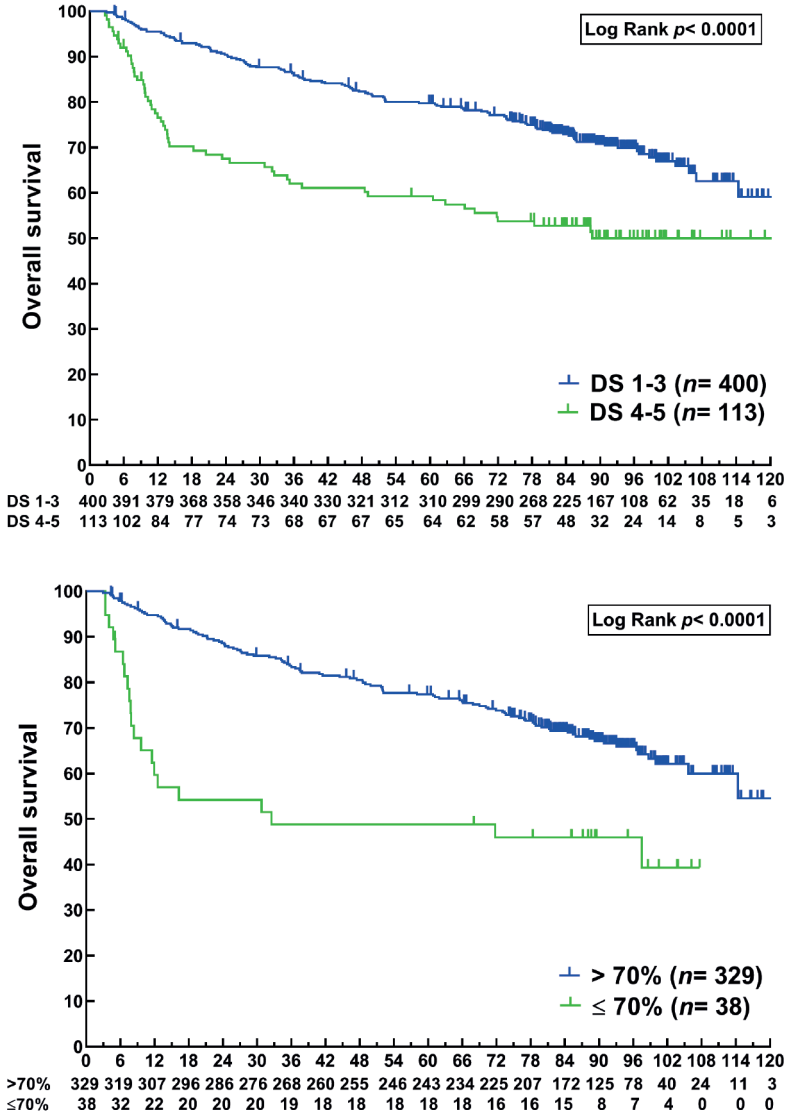


Supplemental Figure 3. Kaplan-Meier curves for progression-free survival in months stratified by combined aaIPI and ΔSUVmax subgroups.

Secondary outcome measures

Definitions:

Overall survival (OS) was defined as time from randomisation to death, patients still alive were censored at date of last contact.



Supplemental Figure 4. Kaplan-Meier curves with numbers at risk for overall survival in months stratified by I-PET4 result according to DS (4a) and according to ΔSUVmax result (4b).

Supplemental Table 5. Diagnostic and prognostic measures for baseline MTV, for different cutoff values for the Deauville 5-point scale at I-PET4, and ΔSUVmax for 2-year OS.

		Diagnostic information				Prognostic information	Discrimination
		Negative Predictive Value %(95%CI)	Positive Predictive Value %(95%CI)	Sensitivity %(95%CI)	Specificity %(95%CI)	Univariate Hazard Ratio (95%CI)	AUC (95%CI)
Baseline MTV	**≤345ml vs >345ml**	90·5 (84·4-94·4)	20·1 (14·6-27·0)	71·1 (56·6-82·3)	49·9 (43·3-55·6)	2·23 (1·17-4·24)	0·59 (0·52-0·66)
I-PET4	**DS1 vs DS2-5**	87·6 (82·0-91·7)	15·5 (12·0-19·8)	70·3 (59·1-79·5)	35·5 (31·2-40·1)	1·29 (0·79-2·13)	0·53 (0·47-0·64)
	DS1-2 vs DS3-5	90·3 (86·4-93·2)	20·6 (15·8-26·4)	62·2 (50·8-72·4)	59·7 (55·0-64·2)	2·35 (1·47-3·75)	0·61 (0·55-0·66)
	DS1-3 vs DS4-5	90·5 (87·2-93·0)	31·6 (24·0-40·9)	48·7 (37·6-59·8)	82·5 (78·6-85·7)	4·02 (2·55-6·35)	0·65 (0·61-0·70)
	DS1-4 vs DS5	88·3 (85·2-90·9)	68·0 (48·4-82·8)	23·0 (14·9-33·7)	98·2 (96·5-99·1)	9·85 (5·69-17·03)	0·60 (0·58-0·62)
ΔSUVmax	**>70% vs ≤70%**	88·8 (84·9-91·7)	44·7 (30·2-60·3)	31·5 (20·7-44·7)	93·3 (90·0-95·6)	5·52 (3·10-9·83)	0·62 (0·59-0·66)

Abbreviations: 95%CI= 95% confidence interval; AUC= area under the receiver operating curve; DS= Deauville 5-point scale; ΔSUVmax= reduction of maximum standardized uptake value between baseline and interim ^{18}F-FDG PET(/CT); I-PET= interim ^{18}F-FDG PET(/CT) after four cycles; MTV= metabolic tumor volume; OS= overall survival.

Supplemental Table 6. Uni- and multivariable Cox Proportional Hazard analyses of ΔSUVmax analysis-group for 2-year OS (*n*=367).

	2-year OS			
	Univariate HR (95%CI)	***P*-value**	**Multivariable HR (95%CI)**	***P*-value**
Age (≤60 vs >60)	1·65 (0·88-3·08)	0·116	1·92 (1·01-3·62)	0·046*
aaIPI (low/low-intermediate vs high-intermediate/high)	2·85 (1·47-5·52)	0·0002*	2·42 (1·24-4·76)	0·010*
B symptoms (no vs yes)	2·12 (1·23-3·65)	0·0007*	1·82 (1·01-3·16)	0·036*
ΔSUVmax (>70% vs ≤70%)	5·52 (3·10-9·83)	<0·0001*	6·03 (3·36-10·81)	<0·0001*
Gender (male vs female)	0·68 (0·40-1·18)	0·172	0·55 (0·31-0·95)	0·034*
Treatment arm (R-CHOP14 vs RR-CHOP14	1·01 (0·59-1·72)	0·969		

* Statistically significant difference

Abbreviations: 95%CI= 95% confidence interval; aaIPI= age-adjusted international prognostic index; ΔSUVmax= reduction of maximum standardized uptake value between baseline and interim ^{18}F-FDG PET(/CT); HR= Hazard Ratio; OS= overall survival.

Supplemental Table 7. Uni- and multivariable Cox Proportional Hazard analyses including baseline MTV for 2-year OS (*n*=296).

	2-year OS			
	Univariate HR (95%CI)	***P*-value**	**Multivariable HR (95%CI)**	***P*-value**
Age (≤60 vs >60)	1·36 (0·70-2·62)	0·367		
aaIPI (low/low-intermediate vs high-intermediate/high)	2·43 (1·20-4·91)	0·013*		
B symptoms (no vs yes)	2·15 (1·19-3·91)	0·012*		
Baseline MTV log-transformed	1·62 (1·25-2·08)	0·0002*	1·55 (1·20-2·00)	0·001*
ΔSUVmax (>70% vs ≤70%)	7·33 (3·97-13·55)	<0·0001*	6·75 (3·63-12·55)	<0·0001*
Gender (male vs female)	0·67 (0·37-1·21)	0·182		
Treatment arm (R-CHOP14 vs RR-CHOP14	0·97 (0·54-1·74)	0·923		

* Statistically significant difference

Abbreviations: 95%CI= 95% confidence interval; aaIPI= age-adjusted international prognostic index; HR= Hazard Ratio; MTV=metabolic tumor volume; OS= overall survival.

CHAPTER 6b

Rituximab-CHOP With Early Rituximab Intensification for Diffuse Large B-Cell Lymphoma: A Randomized Phase III Trial of the HOVON and the Nordic Lymphoma Group (HOVON-84)

Pieternella J. Lugtenburg
Bronno van der Holt
Harry R. Koene
Rob Fijnheer
Lara H. Böhmer
Gregor E. Verhoef
Memis Y. Bilgin
Nicole C. van der Burg-de Graauw
Kon-Siong G. Jie
Marjolein W. van der Poel
Matthijs H. Silbermann
Aart Beeker
Jeanette K. Doorduijn
Rolf E. Brouwer
Coreline N. Burggraaff
Anne I. Arens
Otto S. Hoekstra
Peter de Nully Brown
Francesco A. D'Amore
Eva de Jongh
Joost W. van Esser
Johannes F. Pruijt
Mels Hoogendoorn
Marcel Nijland
Margreet Oosterveld
Thomas Stauffer Larsen
Maria B. Leijs
Marinus van Marwijk Kooy
Marie J. Kersten
Lidwine W. Tick
King H. Lam
Bart de Keizer
Daphne de Jong
Josée M. Zijlstra-Baalbergen

Journal of Clinical Oncology. 2020; 38: 3377-3387.

Abstract

Purpose

Immunochemotherapy with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) has become standard of care for patients with diffuse large B-cell lymphoma (DLBCL). This randomized trial assessed whether rituximab intensification during the first 4 cycles of R-CHOP could improve the outcome of these patients compared with standard R-CHOP.

Patients and methods

A total of 574 patients with DLBCL age 18 to 80 years were randomly assigned to induction therapy with 6 or 8 cycles of R-CHOP-14 with (RR-CHOP-14) or without (R-CHOP-14) intensification of rituximab in the first 4 cycles. The primary end point was complete remission (CR) on induction. Analyses were performed by intention to treat.

Results

CR was achieved in 254 (89%) of 286 patients in the R-CHOP-14 arm and 249 (86%) of 288 patients in the RR-CHOP-14 arm (hazard ratio [HR], 0.82; 95% CI, 0.50 to 1.36; $P = .44$). After a median follow-up of 92 months (range, 1-131 months), 3-year failure-free survival was 74% (95% CI, 68% to 78%) in the R-CHOP-14 arm versus 69% (95% CI, 63% to 74%) in the RR-CHOP-14 arm (HR, 1.26; 95% CI, 0.98 to 1.61; $P = .07$). Progression-free survival at 3 years was 74% (95% CI, 69% to 79%) in the R-CHOP-14 arm versus 71% (95% CI, 66% to 76%) in the RR-CHOP-14 arm (HR, 1.20; 95% CI, 0.94 to 1.55; $P = .15$). Overall survival at 3 years was 81% (95% CI, 76% to 85%) in the R-CHOP-14 arm versus 76% (95% CI, 70% to 80%) in the RR-CHOP-14 arm (HR, 1.27; 95% CI, 0.97 to 1.67; $P = .09$). Patients between ages 66 and 80 years experienced significantly more toxicity during the first 4 cycles in the RR-CHOP-14 arm, especially neutropenia and infections.

Conclusion

Early rituximab intensification during R-CHOP-14 does not improve outcome in patients with untreated DLBCL.

Context

Key Objective

Diffuse large B-cell lymphoma (DLBCL) is a curable disease. However, 40% of patients are refractory to or relapse after treatment with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP). Several single-arm phase II studies in elderly patients with DLBCL have explored variations of the rituximab schedule in combination with CHOP and have reported a better outcome for patients with poor prognosis. Our randomized study examined whether rituximab intensification during the first 4 cycles of 2-week R-CHOP could improve the outcome of untreated patients with DLBCL compared with standard 2-week R-CHOP.

Knowledge Generated

Intensification of rituximab during the first 4 cycles of 2-week R-CHOP did not improve complete remission rate, progression-free survival, or overall survival. Patients between ages 66 and 80 years experienced more neutropenia and infections during rituximab intensification.

Relevance

R-CHOP remains the standard treatment for DLBCL. Novel therapies are needed to improve the outcome of these patients.

Introduction

The overall survival (OS) of patients with diffuse large B-cell lymphoma (DLBCL) has improved significantly since the addition of rituximab to standard 3-week cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP-21) or dose-dense 2-week CHOP (R-CHOP-14) [1,2]. No significant benefits have been shown for R-CHOP-14 versus R-CHOP-21, and these regimens are currently standard treatments worldwide [3,4]. However, approximately 40% of patients experience primary refractory disease or relapse, which is often fatal [5,6]. Therefore, further improvement of first-line therapy is needed.

The dose and schedule of rituximab in the R-CHOP combination are largely empirically determined on historical grounds. Few phase II studies have explored variations of the rituximab schedule in combination with CHOP in elderly patients with DLBCL [7,8]. In a single study in which patients were treated with rituximab administered in shorter intervals at the beginning of treatment and over a prolonged period of time, a better outcome for patients with poor prognosis with International Prognostic Index (IPI) score of 3 to 5 compared with historical controls was reported [8]. The same group reported significantly reduced rituximab clearance in elderly women compared with elderly men [9]. During standard R-CHOP-14 treatment, serum levels of rituximab show a gradual increase up to cycle 5, reaching a plateau thereafter [10]. The lag time of 5 cycles may result in suboptimal rituximab serum levels, especially early during treatment. Therefore, treatment outcome may be improved through intensification of rituximab during the first 4 cycles by providing a steeper increase to the optimal therapeutic serum level as well as reaching a higher serum concentration within the large therapeutic window of rituximab [11,12].

To assess the efficacy of early rituximab intensification during first-line treatment in patients with DLBCL, we performed a prospective randomized phase III study to compare standard R-CHOP-14 with R-CHOP-14 combined with 4 extra administrations of rituximab during the first 4 induction cycles. Patients in complete remission (CR) after induction treatment were randomly assigned a second time between observation and rituximab maintenance. Here, we present the final analysis of the induction random assignment, including long-term follow-up data with a data cutoff of October 16, 2019.

Patients and methods

Patient Population

The HOVON-84 (Haemato Oncology Foundation for Adults in the Netherlands) study was an investigator-initiated prospective randomized phase III study conducted among 68 participating centers in the Netherlands, Denmark, and Belgium. The study was approved by the institutional review boards at all centers. Eligibility included previously untreated, biopsy-confirmed, CD20+ DLBCL according to local pathology and Ann Arbor stage II to IV. Patients between age 18 and 65

years and with an age-adjusted IPI score of 1 to 3 and patients between age 66 and 80 years and an age-adjusted IPI score of 0 to 3 were eligible. Central pathology review was performed as part of quality control (HOVON Pathology Facility and Biobank). CNS involvement, testicular DLBCL, primary mediastinal B-cell lymphoma, transformed indolent lymphoma, any solid malignancy in the preceding 5 years, and illnesses precluding study treatment rendered patients ineligible.

Computed tomography (CT) scanning and bone marrow biopsies were minimum mandatory staging procedures. Baseline ^{18}F-fluorodeoxyglucose positron emission tomography (PET) scans were recommended but not mandated.

Random Assignment

After providing written informed consent, patients were randomly allocated to receive either R-CHOP-14 (arm A) or R-CHOP-14 with intensification of rituximab in the first 4 cycles (RR-CHOP-14; arm B). Random assignment was stratified by center, age group (18-65 *v* 66-80 years), and age-adjusted IPI score using a minimization procedure, ensuring balance within each stratum and overall balance.

Treatment and Response Assessment

The R-CHOP-14 regimen consisted of 14-day cycles of intravenous cyclophosphamide 750 mg/m^2, doxorubicin 50 mg/m^2, vincristine 1.4 mg/m^2 (maximum, 2 mg), and rituximab 375 mg/m^2 on day 1 and prednisone 100 mg once daily on days 1 to 5, for a total of 8 cycles [13]. Pegfilgrastim was administered on day 2 of each cycle. Patients randomly assigned to arm B received extra intravenous rituximab 375 mg/m^2 on day 8 of the first 4 cycles (RR-CHOP-14). Initially, inclusion was limited to elderly patients (age 66-80 years). In July 2009, the protocol was amended to also include patients age 18 to 65 years. At the same time, because of the results of the RICOVER-60 trial, the number of CHOP-14 cycles for patients age 66 to 80 years was reduced to 6, whereas the number of rituximab cycles was maintained at 8[2]. Details regarding prephase and supportive measures during treatment are provided in the Appendix (online only). Consolidation radiotherapy was not allowed.

Response at the end of induction treatment was assessed using PET-CT scans [14,15]. Patients with progressive disease on CT scan after 4 cycles went off

protocol. The interim PET scan after 4 cycles was performed for observational purposes only. All PET-CT scans were centrally reviewed by the HOVON Imaging Group according to standard procedures as previously described [16] using Deauville score (DS) for visual assessment[15]. Scores of 1 to 3 were interpreted as complete metabolic response, and scores of 4 to 5 were consistent with partial metabolic response or progressive disease. CT scans of neck, chest, abdomen, and pelvis were required at 6, 12, 18, and 24 months after completion of induction treatment. Severity of adverse events was defined according to the National Cancer Institute Common Terminology Criteria for Adverse Events (version 3.0).

Sample Size Calculation and Statistical Analysis

This trial was designed to compare CR rates on induction treatment between R-CHOP-14 and RR-CHOP-14 (first randomization; R1) and compare failure-free survival (FFS) from second randomization (R2) between no further treatment and rituximab maintenance. The sample size for R1 was 575 patients, accrued over 5 years, with a power of 86% to detect an improvement in CR rate from 77% to 87%. Additional sample size calculation details are provided in the Appendix. The primary end point for R1 was CR on induction. Logistic regression analysis with adjustment for age group (18-65 *v* ≥ 66-80 years) and age-adjusted IPI score (0 *v* 1 *v* 2 *v* 3; categorical) was applied for the primary analysis, and odds ratios and 95% CIs were determined, with *P* values < .05 considered statistically significant. Secondary end points were best response on protocol treatment, adverse events, FFS, progression-free survival (PFS), and OS from R1 and disease-free survival (DFS) from CR. For the survival end points, the hazard ratios (HRs) and 95% CIs were determined using univariable and multivariable Cox regression analyses. Kaplan-Meier curves by treatment arm were generated to illustrate survival.

All analyses were performed according to the intention-to-treat (ITT) principle. However, patients initially randomly assigned but considered ineligible in retrospect based on information that should have been available before random assignment were excluded from the respective analyses (modified ITT). The proportion of patients with specific adverse events was compared between arms post hoc using the χ^2 test or Fisher's exact test, whichever was appropriate. All reported *P* values are 2 sided and were not adjusted for multiple testing. Additional details on statistical methods and survival end point definitions are provided in the Appendix.

Results

Study Patients

Between November 14, 2007, and April 6, 2012, 600 patients were enrolled. Twenty-six patients (R-CHOP-14 arm, *n* = 14; RR-CHOP-14 arm, *n* = 12) were considered ineligible in hindsight and excluded from all analyses because of diagnosis other than DLBCL at study entry according to local pathology (*n* = 12), stage I disease (*n* = 4), absence of age-adjusted IPI risk factors (*n* = 4), CNS involvement (*n* = 2), absence of measurable disease (*n* = 1), heart disease (*n* = 1), administrative error (*n* = 1), or missing data (*n* = 1). Of 574 patients included in the modified ITT analysis, 286 individuals were allocated to the R-CHOP-14 arm and 288 were assigned to the RR-CHOP-14 arm (Fig 1). Central pathology review was available for 522 (91%) of 574 eligible patients, and diagnosis of CD20+ DLBCL according to the 2008 WHO classification was confirmed for 492 (94%) of 522 patients. Baseline characteristics of patients were well balanced between arms (Table 1; Appendix Table A1, online only).

Treatment

At least 6 cycles were received by 269 (94%) of 286 patients in the R-CHOP-14 arm and 261 (91%) of 288 patients in the RR-CHOP-14 arm; 151 patients (53%) received 7 to 8 cycles of R-CHOP-14, compared with 158 (55%) in the RR-CHOP-14 arm (Fig 1). The median total dose received and median relative dose-intensities achieved for cyclophosphamide (98%) and doxorubicin (98%) were similar in the R-CHOP-14 and RR-CHOP-14 arms. However, for vincristine, in patients age 66 to 80 years, the median total dose and median relative dose-intensities were 12.0 versus 10.0 mg (P = .015) and 92% versus 85% (P = 0.083) for the R-CHOP-14 and RR-CHOP-14 arms, respectively.

Efficacy Outcomes

There was no statistically significant difference in the primary end point of CR rate on induction between the 2 treatment arms. CR was achieved in 254 patients (89%) in the R-CHOP-14 arm and in 249 (86%) in the RR-CHOP-14 arm (HR, 0.82; 95% CI, 0.50 to 1.36; P = .44; adjusted for age and age-adjusted IPI score). Also, CR rates for patients age < 66 years (90% *v* 85%) and patients age ≥ 66 years (88% *v* 88%) were not different per treatment arm.

Table 1. Baseline patient demographic and clinical characteristics.

	No. (%)	
Characteristic	**R-CHOP-14 (*n* = 286)**	**RR-CHOP-14 (*n* = 288)**
Sex		
Male	145 (51)	154 (53)
Female	141 (49)	134 (47)
Age, years		
Median	66	65
Range	18-80	31-80
≤ 65	140 (49)	149 (52)
> 65	146 (51)	139 (48)
WHO performance status		
0-1	254 (89)	251 (87)
2	30 (10)	36 (13)
Unknown	2 (1)	1 (0)
Ann Arbor stage		
II	53 (18)	61 (21)
III	88 (31)	89 (31)
IV	145 (51)	138 (48)
B symptoms	112 (39)	120 (42)
LDH > ULN	183 (64)	196 (68)
Bulky disease (> 10 cm)	83 (29)	85 (30)
BM involvement	30 (10)	36 (13)
Age-adjusted IPI risk group		
Low	22 (8)	24 (8)
Low-intermediate	107 (37)	93 (33)
High-intermediate	132 (46)	147 (51)
High	25 (9)	24 (8)
Histology (central review)		
DLBCL	251 (88)	244 (85)
Other diagnosis or unclassified[a]	11 (4)	16 (6)
Not reviewed	25 (8)	28 (10)
Phenotype[b]		
Germinal center	124 of 200 (62)	107 of 177 (60)
Nongerminal center	76 of 200 (38)	70 of 177 (40)
MYC rearrangement	14 of 104 (13)	5 of 73 (7)
MYC SH	4 of 14	1 of 5
MYC plus BCL2 and/or BCL6[c]	10 of 14	4 of 5

Abbreviations: BM, bone marrow; DLBCL, diffuse large B-cell lymphoma; IPI, International Prognostic Index; LDH, lactate dehydrogenase; R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B); SH, single hit; ULN, upper limit of normal.

[a]Appendix Table A1. [b]Based on standard Hans criteria. [c]According to WHO classification 2016; now classified as high grade B-cell lymphoma with MYC and BCL2 and/or BCL6 rearrangements.

After a median follow-up of 92 months (range, 1-131 months) in the 364 patients still alive, the median FFS and median PFS were not reached in the R-CHOP-14 arm and were both 101 months in the RR-CHOP-14 arm, and the median DFS

and OS had not been reached in either arm. The 3-year FFS rate was 74% (95% CI, 68% to 78%) in the R-CHOP-14 arm versus 69% (95% CI, 63% to 74%) in the RR-CHOP-14 arm (HR, 1.26; 95% CI, 0.98 to 1.61; *P* = .07; adjusted for age group and age-adjusted IPI score; Fig 2A); FFS rates at 5 years were 68% (95% CI, 62% to 73%) and 62% (95% CI, 56% to 67%), respectively. PFS at 3 years

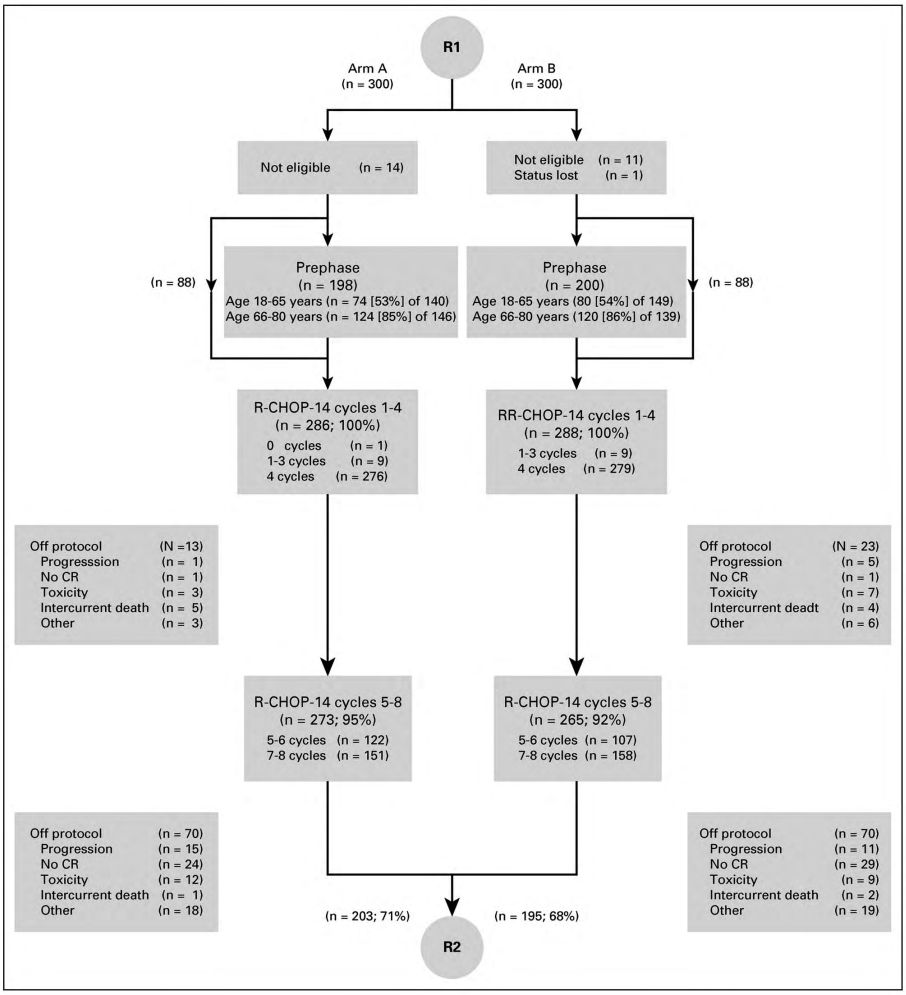


Fig. 1. CONSORT diagram of induction treatment of patients with diffuse large B-cell lymphoma in the HOVON-84 non-Hodgkin lymphoma trial by treatment arm.

CR, complete remission; R1, induction randomization; R2, maintenance randomization; R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, prednisone (arm B).

was 74% (95% CI, 69% to 79%) in the R-CHOP-14 arm versus 71% (95% CI, 66% to 76%) in the RR-CHOP-14 arm (HR, 1.20; 95% CI, 0.94 to 1.55; *P* = .15; adjusted for age group and age-adjusted IPI score; Fig 2B); the 5-year PFS rates were 69% (95% CI, 63% to 74%) and 64% (95% CI, 58% to 69%), respectively.

Among patients who had achieved CR on protocol treatment, the 3-year DFS rate from date of CR was 81% (95% CI, 76% to 85%) in the R-CHOP-14 arm versus 76% (95% CI, 70% to 81%) in the RR-CHOP-14 arm (HR, 1.24; 95% CI, 0.93 to 1.65; *P* = .15; adjusted for age group and age-adjusted IPI score; Fig 2C); the 5-year DFS rates were 75% (95% CI, 69% to 80%) and 70% (95% CI, 64% to 75%), respectively. OS at 3 years was 81% (95% CI, 76% to 85%) in the R-CHOP-14 arm versus 76% (95% CI, 70% to 80%) in the RR-CHOP-14 arm (HR, 1.27; 95% CI, 0.97 to 1.67; *P* = .09; adjusted for age group and age-adjusted

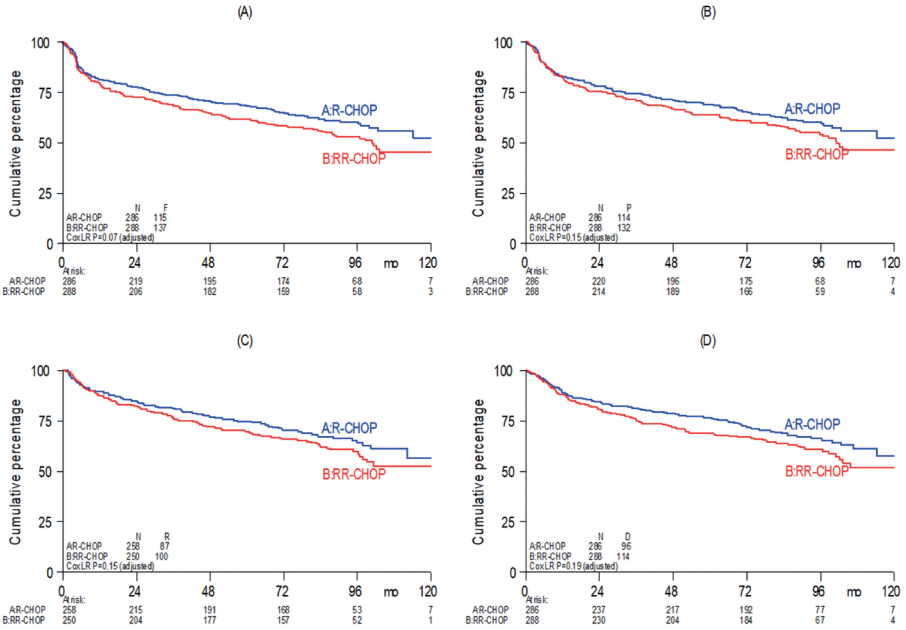


Fig. 2. Kaplan-Meier survival curves according to assigned treatment arm. (A) Failure-free survival (FFS), (B) progression-free survival (PFS), (C) disease-free survival (DFS) from complete remission, and (D) overall survival (OS).

All *P* values by Cox logistic regression (adjusted). D, death; F, no complete remission, relapse, or death; P, progression, relapse, or death; R, relapse or death; R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on day 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B).

IPI score; Fig 2D); the 5-year OS rates were 77% (95% CI, 71% to 81%) and 69% (95% CI, 63% to 74%), respectively.

A total of 210 patients died, 96 in the R-CHOP-14 arm (lymphoma related, *n* = 41; treatment related, *n* = 9; intercurrent death, *n* = 8; secondary malignancies, *n* = 11; other reasons, *n* = 15; and unknown causes, *n* = 12) and 114 in the RR-CHOP-14 arm (lymphoma related, *n* = 56; treatment related, *n* = 10; intercurrent death, *n* = 10; secondary malignancies, *n* = 11; other reasons, *n* = 11; and unknown causes, *n* = 16).

Planned subgroup analyses showed that the impact of RR-CHOP-14 versus R-CHOP-14 on FFS, PFS, DFS, and OS was not different between subgroups of age (18-65 *v* 66-80 years), sex (male *v* female), or age-adjusted IPI score (low *v* low-intermediate *v* high-intermediate *v* high). Post hoc analyses showed similar results for subgroups according to DLBCL phenotype. Figure 3 and Appendix Figures A1 and A2 (online only) show the Kaplan-Meier PFS curves for these subgroups.

Results of the multivariable analyses of individual prognostic factors for the survival end points FFS, PFS, and OS are listed in Table 2 (and for DFS in Appendix Table A2, online only). The HRs for both treatment arms were similar compared with those in the analyses with adjustment for only age group and age-adjusted IPI score, confirming that survival was not improved in either subgroup in the RR-CHOP-14 arm. The only statistically significant prognostic factor was age 66 to 80 years.

PET-CT Assessment

PET-CT scans were visually assessed using the 5-point DS; DSs 1 to 3 were regarded as negative and DSs 4 to 5 as positive. A total of 496 end-of-treatment (EOT) PET scans were centrally reviewed. In 417 patients (84%), the EOT PET-CT scans were negative, and 79 patients (16%) had positive EOT PET scans. The estimated 2-year PFS rate in patients with EOT PET–positive scans was 46% (95% CI, 36% to 57%) versus 88% (95% CI, 85% to 92%) in those with EOT PET–negative scans ($P < .001$). The 2-year OS rate was 58% (95% CI, 47% to 69%) for patients with EOT PET–positive scans and 94% (95% CI, 91% to 96%) for those with EOT PET–negative scans. Corresponding positive and negative

Table 2. Multivariable analysis of prognostic factors for FFS, PFS, and OS.

	FFS			PFS			OS		
Factor	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
RR-CHOP-14 arm	1.25	0.98 to 1.61	.08	1.20	0.93 to 1.54	.16	1.25	0.95 to 1.65	.10
Age ≥ 66 years	1.57	1.21 to 2.03	.001	1.58	1.21 to 2.04	.001	1.78	1.33 to 2.36	< .001
Age-adjusted IPI score[a]	1.09	0.81 to 1.46	.57	1.12	0.84 to 1.51	.44	1.17	0.85 to 1.61	.35
Female sex	0.80	0.63 to 1.04	.09	0.85	0.66 to 1.09	.20	0.80	0.61 to 1.05	.11
WHO performance score[b]	1.02	0.82 to 1.28	.84	1.04	0.83 to 1.30	.76	1.12	0.88 to 1.44	.35
LDH > ULN	1.51	1.00 to 2.30	.051	1.46	0.96 to 2.23	.08	1.49	0.94 to 2.37	.09
B symptoms	1.12	0.86 to 1.46	.42	1.13	0.87 to 1.49	.36	1.08	0.80 to 1.44	.62
Bulky disease	1.05	0.79 to 1.38	.75	0.93	0.70 to 1.24	.63	0.85	0.62 to 1.15	.29
BM involvement	1.21	0.84 to 1.75	.30	1.19	0.82 to 1.72	.36	0.98	0.65 to 1.49	.93

Abbreviations: BM, bone marrow; FFS, failure-free survival; HR, hazard ratio; IPI, International Prognostic Index; LDH, lactate dehydrogenase; OS, overall survival; PFS, progression-free survival; RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B); ULN, upper limit of normal.

[a] Analyzed as low v low-intermediate v high-intermediate v high.

[b] Analyzed as WHO 0 v 1 v 2.

predictive values for 2-year PFS were 53% (95% CI, 42% to 64%) and 89% (95% CI, 85% to 91%) for EOT PET scans, respectively.

Rituximab Pharmacokinetics

Rituximab trough serum levels increased after each subsequent treatment cycle during the first 4 cycles and reached a plateau at cycles 5 to 8 in both treatment arms. Rituximab trough serum levels were systematically higher in the RR-CHOP-14–treated patients than in R-CHOP–treated patients (Appendix Figure A3, online only).

Adverse Events

We analyzed safety for all patients who received at least 1 administration of study treatment. The proportion of patients with at least 1 adverse or serious adverse event did not differ between the R-CHOP-14 and RR-CHOP-14 arms. The most common grade 3 and 4 adverse events were cytopenias and infections (Table 3). During the first 4 cycles, patients between ages 66 and 80 years experienced significantly more toxicity in the RR-CHOP-14 arm, especially neutropenia and infections (Table 4).

Table 3. Grade 3-4 adverse events during cycles 1-8 in all patients.

	No. (%)			
	R-CHOP-14 (n = 285)		RR-CHOP-14 (n = 288)	
Adverse Event	**Grade 3**	**Grade 4**	**Grade 3**	**Grade 4**
All toxicity	70 (25)	127 (45)	70 (24)	146 (51)
Neutropenia	23 (8)	91 (32)	29 (10)	107 (37)
Febrile neutropenia	—	3 (1)	—	1 (0)
Anemia	44 (15)	11 (4)	49 (17)	5 (2)
Thrombocytopenia	13 (5)	19 (7)	20 (7)	16 (6)
Infection	57 (20)	13 (5)	64 (22)	7 (2)
Neurologic toxicity	38 (13)	2 (1)	37 (13)	3 (1)
GI	36 (13)	4 (1)	31 (11)	6 (2)
Cardiac toxicity	11 (4)	—	11 (4)	3 (1)

NOTE. Data are No. of patients (%) with an event. Patients could have the same type of event more than once.

Abbreviations: R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B).

Table 4. Grade 3-4 adverse events during cycles 1-4 in patients age 18-65 versus 66-80 years.

	No. (%)							
	Age 18-65 years				Age 66-80 years			
Adverse Event	R-CHOP-14 (n = 140)		RR-CHOP-14 (n = 149)		R-CHOP-14 (n = 145)		RR-CHOP-14 (n = 139)	
	Grade 3	**Grade 4**	**Grade 3**	**Grade 4**	**Grade 3**	**Grade 4**	**Grade 3**	**Grade 4**
All toxicity[a]	30 (21)	42 (30)	32 (21)	46 (31)	26 (18)	58 (40)	23 (17)	78 (56)
Neutropenia[b]	12 (9)	33 (24)	10 (7)	38 (26)	6 (4)	41 (28)	14 (10)	48 (35)
Febrile neutropenia	—	1 (1)	—	—	—	1 (1)	—	1 (1)
Anemia	11 (8)	3 (2)	11 (7)	2 (1)	21 (14)	4 (3)	18 (13)	3 (2)
Thrombocytopenia	2 (1)	5 (4)	5 (3)	2 (1)	7 (5)	6 (4)	7 (5)	5 (4)
Infection[c]	13 (9)	3 (2)	17 (11)	1 (1)	23 (16)	4 (3)	30 (22)	4 (3)
Neurologic toxicity	5 (4)	—	8 (5)	—	11 (8)	—	8 (6)	—
GI	15 (11)	1 (1)	5 (3)	—	14 (10)	—	17 (12)	3 (2)
Cardiac toxicity	1 (1)	—	—	—	2 (1)	—	7 (5)	2 (1)

NOTE. Data are No. of patients (%) with an event. Patients could have the same type of event more than once.

Abbreviations: R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B).

[a] In patients age 66-80 years: grade 4, 40% v 56% (P = .007); grade 3-4, 58% v 73% (P = .009).

[b] In patients age 66-80 years: grade 4, 28% v 35% (P = .26); grade 3-4, 32% v 45% (P = .04).

[c] In patients 66-80 years: grade 4, 3% v 3% (P = 1.00); grade 3-4, 19% vs 25% (P = .23).

Seventeen grade 5 adverse events were reported during induction, 9 in the R-CHOP-14 arm and 8 in the RR-CHOP-14 arm. The main cause of death was infection (4 patients in each arm). Other causes of death in the R-CHOP-14 arm were small-bowel perforation (*n* = 2), sudden death (*n* = 2), and progressive multifocal leukoencephalopathy (*n* = 1). In the RR-CHOP-14 arm, other causes of death were myocardial infarction (*n* = 1), GI bleeding (*n* = 1), small-bowel perforation (*n* = 1), and cardiac arrhythmia (*n* = 1).

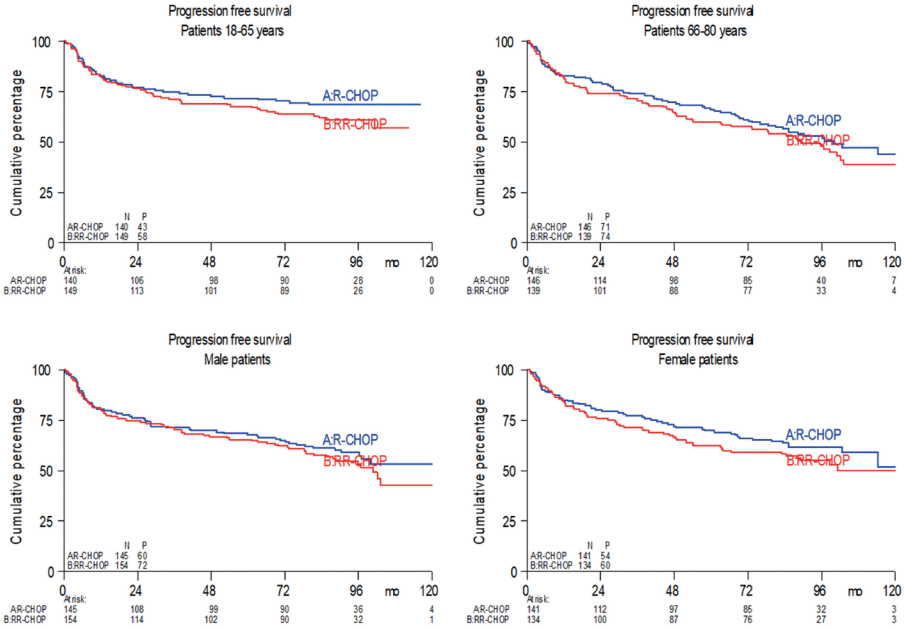


Fig 3. Progression-free survival (PFS) by treatment arm within subgroups: (A) age 18 to 65 years, (B) age 66 to 80 years, (C) male patients, and (D) female patients.

P, progression, relapse, or death; R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, prednisone (arm B).

Discussion

The primary objective of achieving a significantly superior CR rate with RR-CHOP-14 treatment as compared with standard R-CHOP-14 treatment was not met. RR-CHOP-14 treatment also did not improve FFS, PFS, DFS, or OS. In

DLBCL, rapid tumor control is critical to improve outcome by avoiding development of refractory disease on or after R-CHOP, because patients with refractory disease have poor prognosis [17]. Several phase II studies have explored optimization of rituximab for the treatment of DLBCL. In the DENSE-R-CHOP-14 trial, early dose-intensification of rituximab in combination with R-CHOP-14 was tested in 124 elderly patients with DLBCL [7]. In this study, 4 additional rituximab administrations were added during the first 3 weeks. Compared with a historical control population (RICOVER-60 population), no differences in outcome were observed for the whole population. Subgroup analysis revealed that patients with high-intermediate and high IPI scores had higher CR/unconfirmed CR (CRu) rates after rituximab intensification, but this did not translate into better survival outcome. A high rate of grade 3 and 4 infectious complications was reported, which improved after mandatory prophylaxis with acyclovir and cotrimoxazole was instituted. In the SMARTE-R-CHOP-14 study, a prolonged exposure time of rituximab using a loading schedule of 2 rituximab administrations before the first CHOP cycle and 3 additional rituximab administrations after completion of R-CHOP was investigated in 189 elderly patients with DLBCL [8]. Compared with the RICOVER-60 population, survival outcome was not significantly better for the complete study population, and subgroup analysis showed that patients with high-intermediate and high IPI scores had higher CR/CRu rates and better 3-year PFS (71% *v* 59%) and OS (80% *v* 67%) rates. Elderly male patients showed a significantly faster rituximab clearance than elderly female patients, resulting in a shorter rituximab serum elimination half-life, lower serum levels, and shorter rituximab exposure times [8,10]. Because in the RICOVER-60 study elderly male patients seemed to benefit to a lesser extent from addition of rituximab to CHOP than elderly female patients, an increased dose of 500 mg/m^2 of rituximab for male patients and the standard dose of 375 mg/m^2 for female patients were investigated in 271 elderly patients with DLBCL in the SEXIE-R-CHOP-14 study [2,18]. No survival differences were found, and the authors concluded that the increased rituximab dose may have abrogated the negative effect in elderly male patients. These phase II studies in elderly patients with DLBCL supported the notion that patients with DLBCL with poor prognosis would be most likely to benefit from adapted rituximab schedules.

In our study, trough rituximab levels were indeed consistently higher during the first 4 cycles in the RR-CHOP-14 arm than in the R-CHOP-14 arm, and they

remained higher during further treatment. However, this did not translate into better short- or long-term outcome for the complete study population. Also, exploratory subgroup analyses for different age groups, age-adjusted IPI risk groups, and sexes could not identify any subgroup that might benefit from rituximab intensification. Our randomized phase III study differs in some essential aspects from the phase II studies. The study populations were not comparable; in our study, both young and elderly patients with DLBCL were included, whereas the phase II studies included elderly patients only and included a broader spectrum of aggressive B-cell lymphoma diagnoses. In our phase III study, staging and response evaluation was based on PET-CT, whereas it was based on CT scanning only in the phase II studies. Lastly, the schedules for rituximab intensification differed to some extent. However, from these studies, it may be concluded that dose-intensification within a standard R-CHOP regimen is insufficient to improve outcome for patients with DLBCL. Tout et al [19] demonstrated that rituximab exposure is influenced by baseline metabolic tumor volume (MTV) and suggest that outcome might improve when the rituximab dose is individualized according to the MTV. This interesting hypothesis needs to be confirmed in a prospective trial.

For the past 2 decades, R-CHOP has remained the standard treatment for previously untreated DLBCL, and it has proven exceedingly difficult to improve on this baseline [20]. To date, neither next-generation anti-CD20 monoclonal antibodies, such as obinutuzumab or ofatumumab, nor approaches adding targeted therapy based on molecular subtypes of DLBCL, such as bortezomib, ibrutinib, or lenalidomide in ABC/non-GCB subgroups, have proven successful [6,21-24]. More recent developments in chemo-immunotherapy using antibody-drug conjugates (eg, polatuzumab vedotin), bispecific antibodies (eg, anti-CD3 × anti-CD20), immune checkpoint inhibitors, and CAR T-cell therapy may reveal new opportunities, and novel insights into DLBCL biology may provide essential information for meaningful patient selection for such treatments [25,26]. Our phase III study shows that early rituximab intensification in patients with untreated DLBCL during R-CHOP-14 does not improve outcome.

References

1. Coiffier B, Lepage E, Briere J, et al: CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med.* 346:235-242, 2002.
2. Pfreundschuh M, Schubert J, Ziepert M, et al: Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: A randomised controlled trial (RICOVER-60). *Lancet Oncol.* 9:105-116, 2008.
3. Delarue R, Tilly H, Mounier N, et al: Dose-dense rituximab-CHOP compared with standard rituximab-CHOP in elderly patients with diffuse large B-cell lymphoma (the LNH03-6B study): A randomised phase 3 trial. *Lancet Oncol.* 14:525-533, 2013.
4. 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.* 381:1817-1826, 2013.
5. Vaidya R, Witzig TE: Prognostic factors for diffuse large B-cell lymphoma in the R(X)CHOP era. *Ann Oncol.* 25:2124-2133, 2014.
6. van Imhoff GW, McMillan A, Matasar MJ, et al: Ofatumumab versus rituximab salvage chemoimmunotherapy in relapsed or refractory diffuse large B-cell lymphoma: The ORCHARRD study. *J Clin Oncol.* 35:544-551, 2017.
7. Murawski N, Pfreundschuh M, Zeynalova S, 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.* 25:1800-1806, 2014.
8. Pfreundschuh M, Poeschel V, Zeynalova S, et al: Optimization of rituximab for the treatment of diffuse large B-cell lymphoma (II): Extended rituximab exposure time in the SMARTE-R-CHOP-14 trial of the German high-grade non-Hodgkin lymphoma study group. *J Clin Oncol.* 32:4127-4133, 2014 [Erratum: *J Clin Oncol* 33:1991, 2015].
9. Pfreundschuh M, Müller C, Zeynalova S, et al: Suboptimal dosing of rituximab in male and female patients with DLBCL. *Blood.* 123:640-646, 2014.
10. Müller C, Murawski N, Wiesen MHJ, et al: The role of sex and weight on rituximab clearance and serum elimination half-life in elderly patients with DLBCL. *Blood.* 119:3276-3284, 2012.
11. Keating M, O'Brien S: High-dose rituximab therapy in chronic lymphocytic leukemia. *Semin Oncol.* 27:86-90, 2000 (suppl 12).
12. O'Brien SM, Kantarjian H, Thomas DA, et al: Rituximab dose-escalation trial in chronic lymphocytic leukemia. *J Clin Oncol*. 19:2165-2170, 2001.
13. Pfreundschuh M, Trümper L, Kloess M, et al: Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of elderly patients with aggressive lymphomas: Results of the NHL-B2 trial of the DSHNHL. *Blood.* 104:634-641, 2004.
14. 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.* 32:3059-3068, 2014.
15. 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.* 32:3048-3058, 2014.
16. Burggraaff CN, Cornelisse AC, Hoekstra OS, et al: Interobserver agreement of interim and end-of-treatment 18F-FDG PET/CT in diffuse large B-cell lymphoma (DLBCL): Impact on clinical practice and trials. *J Nucl Med.* 59:1831-1836, 2018.
17. Crump M, Neelapu SS, Farooq U, et al: Outcomes in refractory diffuse large B-cell lymphoma: Results from the international SCHOLAR-1 study. *Blood.* 130:1800-1808, 2017. [Erratum: *Blood.* 131:587-588, 2018].
18. Pfreundschuh M, Murawski N, Zeynalova S, et al: Optimization of rituximab for the treatment of DLBCL: Increasing the dose for elderly male patients. *Br J Haematol.* 179:410-420, 2017.

19. Tout M, Casasnovas O, Meignan M, et al: Rituximab exposure is influenced by baseline metabolic tumor volume and predicts outcome of DLBCL patients: A Lymphoma Study Association report. *Blood.* 129:2616-2623, 2017.

20. Tilly H, Gomes da Silva M, Vitolo U, et al: Diffuse large B-cell lymphoma (DLBCL): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 26:v116-v125, 2015 (suppl 5).

21. Vitolo U, Trněný M, Belada D, et al: Obinutuzumab or rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in previously untreated diffuse large B-cell lymphoma. *J Clin Oncol.* 35:3529-3537, 2017.

22. Davies A, Cummin TE, Barrans S, et al: Gene-expression profiling of bortezomib added to standard chemoimmunotherapy for diffuse large B-cell lymphoma (REMoDL-B): An open-label, randomised, phase 3 trial. *Lancet Oncol.* 20:649-662, 2019.

23. Younes A, Sehn LH, Johnson P, et al: Randomized phase III trial of ibrutinib and rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in non-germinal center B-cell diffuse large B-cell lymphoma. *J Clin Oncol.* 37:1285-1295, 2019.

24. Vitolo U, Witzig TE, Gascoyne RD, et al: Robust: First report of phase III randomized study of lenalidomide/R-CHOP (R2-CHOP) vs placebo/R-CHOP in previously untreated ABC-type diffuse large B-cell lymphoma. *Hematol Oncol.* 37, 2019 (suppl 2; abstr 005).

25. 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.* 24:679-690, 2018. [Errata: *Nat Med.* 24:1290-1291, 2018; *Nat Med* 24:1292, 2018].

26. Schmitz R, Wright GW, Huang DW, et al: Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med.* 378:1396-1407, 2018.

Supplemental materials

Prephase Treatment and Supportive Measures During R-CHOP-14 Treatment
Prephase treatment
A prephase treatment before the start of study treatment was mandatory in all elderly patients (age 66-80 years) and was left at the discretion of the treating physician in young patients (age 18-65 years). The prephase treatment consisted of a 5-day course of 100 mg of prednisone once daily.

Allopurinol
Allopurinol was applied according to local practices. The dose should have been adapted if the creatinine clearance was decreased.

Prednisone tapering
A gradual reduction of the prednisone dose was recommended to prevent marked fatigue after prompt discontinuation of prednisone. Prednisone 50 mg could be administered on day 6, 25 mg on day 7, and 10 mg on day 8. For patients complaining of fatigue after tapering of prednisone, hydrocortisone 20 mg orally in the morning and 10 mg orally at 1200 was recommended.

Prophylaxis of infection
Pneumocystis jiroveci and herpes infection prophylaxis was mandatory in all patients. This consisted of oral cotrimoxazol 480 mg once daily and oral valaciclovir 500 mg twice per day, starting with the prephase treatment until 4 weeks after the last rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP-14) cycle.

Intrathecal prophylaxis for CNS relapse was at the discretion of the treating physician.

Sample Size Calculation
The sample size was calculated to have a sufficient number of patients available for the second randomization (R2); thereafter, the statistical power for the first randomization (R1) was determined. To detect with 80% power an improvement in failure-free survival (FFS) from R2 with a hazard ratio (HR) of 0.60 (2-sided significance level, $\alpha = 0.05$), 126 events were required. Assuming a proportional

hazard for young versus elderly patients of 0.62, an accrual period of 5 years, and 2 years of follow-up after the last patient was included in the maintenance randomization, this would require 395 patients (young, n = 174; elderly, n = 221). Therefore, 575 patients should be included in this trial, resulting in a power of 86% to detect an improvement in complete remission (CR) rate from 77% to 87%.

Statistical Methods

The primary end point for R1 was CR on induction. Patient treatment was considered a success if CR was achieved during or after induction treatment. All other patient treatments were considered a failure. Logistic regression analysis with adjustment for age group (18-65 *v* ≥ 66-80 years) and age-adjusted International Prognostic Index (IPI) score (0 *v* 1 *v* 2 *v* 3; categorical) was applied for the primary analysis, and odds ratios and 95% CIs were determined, with P values < .05 considered statistically significant.

Secondary end points were best response on protocol treatment, adverse events, FFS, progression-free survival (PFS) and overall survival (OS) from R1, and disease-free survival (DFS) from CR. FFS was defined as time from R1 to no CR on protocol, relapse, or death, whichever came first. PFS was calculated from R1 to progression, relapse, or death, whichever came first. OS was determined from R1 to death resulting from any cause. Patients still alive at last contact were censored. DFS was measured from date of CR to relapse or death, whichever came first.

The proportion of patients with specific adverse events was compared between arms post hoc using the χ^2 test or Fisher's exact test, whichever was appropriate.

For the survival end points, the HRs and 95% CIs were determined using univariable and multivariable Cox regression analyses. Multivariable Cox regression analysis was primarily aimed at evaluating the impact of adjustment on the HRs and 95% CIs of treatment arms, rather than at evaluating the prognostic value of individual covariates, and included: treatment arm, age (18-65 *v* ≥ 66-80 years), sex (male *v* female), age-adjusted IPI stage (low *v* low-intermediate *v* high-intermediate *v* high; continuous), WHO performance (0 *v* 1 *v* 2; continuous), lactate dehydrogenase (normal *v* elevated), B symptoms (no *v* yes), bulky mass (no *v* yes), and bone marrow involvement (no *v* yes), as specified in the statistical analysis plan. Because the number of patients with missing data was low (ie, 3

of 574 eligible patients [1%]), the multivariable Cox regression analyses were restricted to patients with complete data. Kaplan-Meier curves by treatment arm were generated to illustrate survival.

All analyses were performed according the intention-to-treat (ITT) principle. However, patients initially randomly assigned but considered ineligible in retrospect based on information that should have been available before random assignment were excluded from the respective analyses (modified ITT).

Two interim analyses were planned after the inclusion of 200 and 400 evaluable patients, primarily to guard against unfavorable results in the experimental arm, and the results were presented confidentially to an independent data and safety monitoring board. All reported *P* values are 2 sided and were not adjusted for multiple testing.

Rituximab Pharmacokinetics

Rituximab pharmacokinetics were evaluated in 6 patients in the R-CHOP-14 arm and 4 patients in the RR-CHOP-14 (R-CHOP-14 with intensification of rituximab in the first 4 cycles) arm during the induction phase. Thirty to 60 minutes before each rituximab infusion, 5 mL of blood was drawn, and samples were centrifuged at 1,000 *g* for 10 minutes at room temperature and stored at −20°C until shipping on dry ice for analysis. Rituximab serum levels were measured by enzyme-linked immunosorbent assay at Xendo Laboratories (Groningen, the Netherlands).

Acknowledgement

We acknowledge all local data managers and the HOVON Data Center trial team for trial management and central data management. We thank the members of the data and safety monitoring board, E. Brusamolino (Pavia, Italy), B. Coiffier (deceased; Lyon, France), and statistician W.C.J. Hop (deceased; Rotterdam, the Netherlands) for their contribution to the conduct of the study and also thank all collaborators and patients who participated in this study.

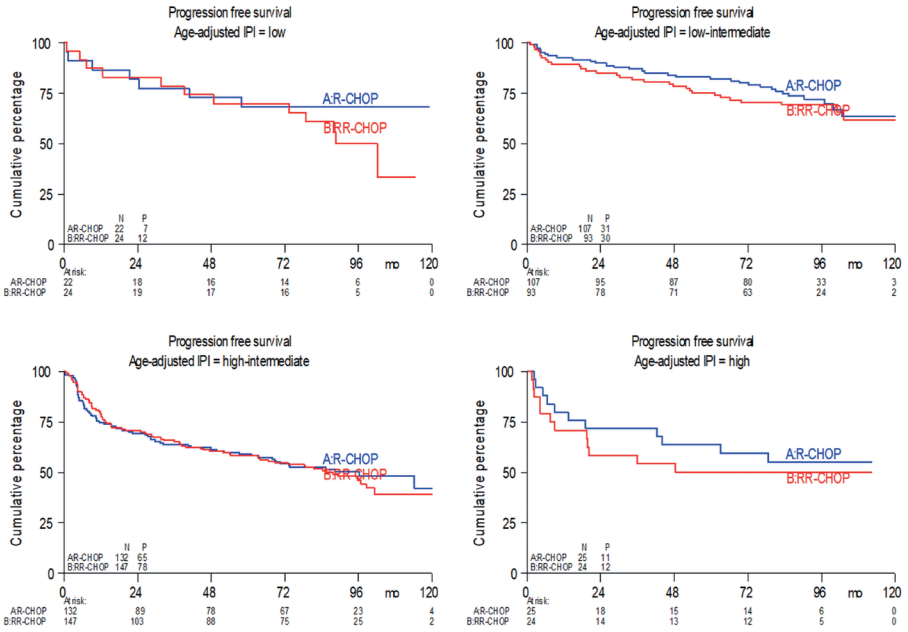


Fig A1. Progression-free survival (PFS) by treatment arm for age-adjusted International Prognostic Index score.

(A) low, (B), low-intermediate, (C) high-intermediate, and (D) high. P, progression, relapse, or death; R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B).

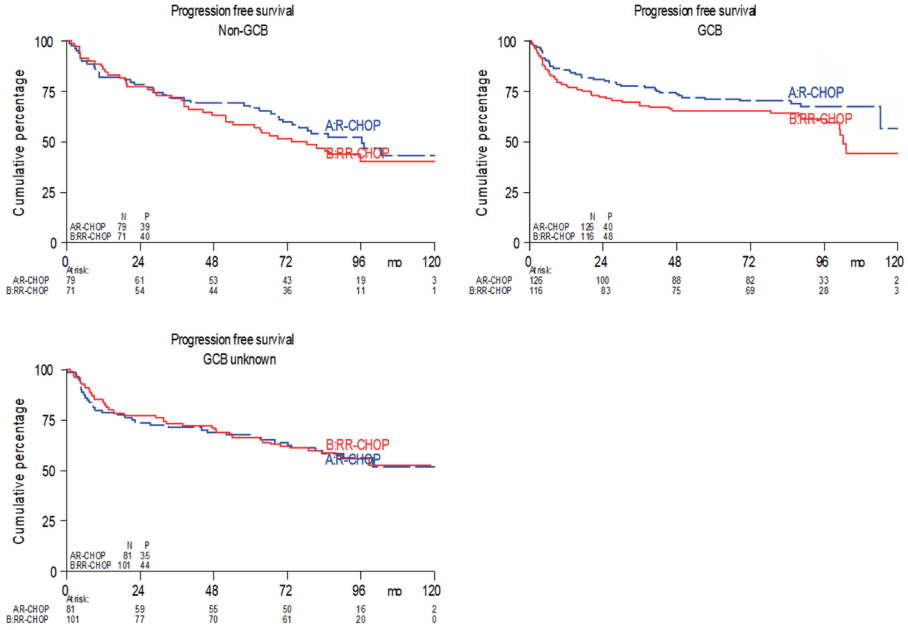


Fig A2. Progression-free survival (PFS) by treatment arm for diffuse large B-cell lymphoma phenotype.

(A) non–germinal center B cell (GCB), (B) GCB, and (C) GCB unknown. P, progression, relapse, or death; R-CHOP, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B).

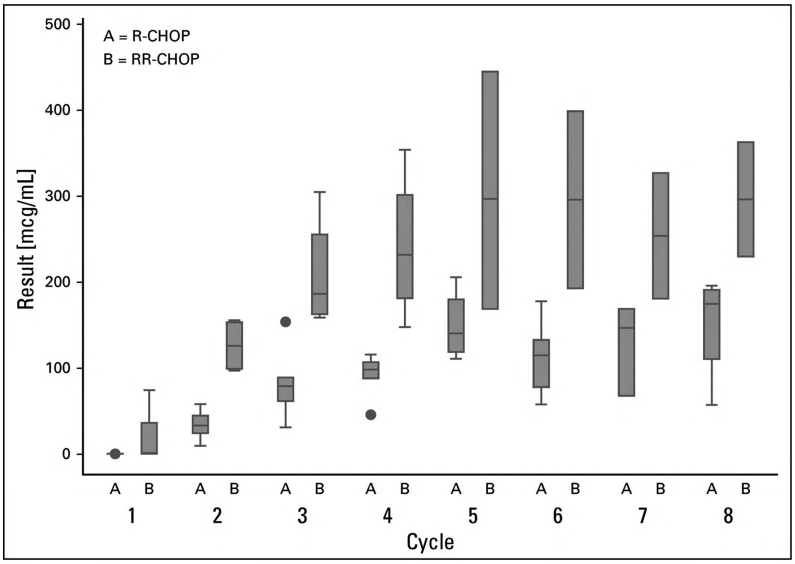


Fig A3. Rituximab trough serum levels.

A, rituximab on day 1 of each cycle plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm A); B, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B).

Table A1. Central Pathology Review Category for Other Diagnosis or Unclassifiable.

Histology	No. of Patients
Indolent B-cell lymphoma	8
Transformed follicular lymphoma	6
B-cell lymphoma unclassifiable	6
Angioimmunoblastic T-cell lymphoma	2
Transformed nodular lymphocyte-predominant Hodgkin lymphoma	1
Poor-quality sample	4

Table A2. Multivariable Analysis of Prognostic Factors for DFS From CR.

Factor	HR	95% CI	*P*
RR-CHOP-14 arm	1.24	0.93 to 1.66	.14
Age ≥ 66 years	1.77	1.30 to 2.40	<.001
Age-adjusted IPI score[a]	1.14	0.81 to 1.60	.45
Female sex	1.02	0.76 to 1.37	.89
WHO performance score[b]	0.99	0.76 to 1.28	.92
LDH > ULN	1.50	0.93 to 2.42	.10
B symptoms	1.04	0.76 to 1.42	.82
Bulky disease	0.94	0.68 to 1.31	.72
BM involvement	1.34	0.89 to 2.03	.16

Abbreviations: BM, bone marrow; CR, complete remission; DFS, disease-free survival; HR, hazard ratio; IPI, International Prognostic Index; LDH, lactate dehydrogenase; RR-CHOP, rituximab on days 1 and 8 of first 4 cycles and day 1 of remaining cycles plus cyclophosphamide, doxorubicin, vincristine, and prednisone (arm B); ULN, upper limit of normal.

[a]Analyzed as low v low-intermediate v high-intermediate v high.

[b]Analyzed as WHO 0 v 1 v 2.

CHAPTER 7

Treatment of patients with MYC rearrangement positive large B-cell lymphoma with R-CHOP plus lenalidomide: results of a multicenter HOVON phase II trial

Martine E.D. Chamuleau
Marcel Nijland
Rogier Mous
Daan Dierickx
Joost S.P. Vermaat
Otto Visser
Yavuz M. Bilgin
Mark F. Durian
Lara H. Bohmer
Rinske S. Boersma
Harry C. Schouten
Eva de Jongh
Arjan Diepstra
Anne I.J. Arens
Otto Hoekstra
Daphne de Jong
Coreline N. Burggraaff
Katerina Bakunina
Pieternella J. Lugtenburg
Gustaaf W. van Imhoff
Erik A.F. Marijt
Caroline Mandigers
Aart Beeker
Bas van Rees
Lidwine W. Tick
Tjeerd J.F. Snijders
Harry R. Koene
Nathalie Hijmering
Anke van den Berg
Julia Huijbregts
Josée M. Zijlstra
Marie José Kersten

Haematologica. 2020; 105: 2805-2812.

Abstract

Patients with *MYC*-rearrangement positive large B-cell lymphoma (*MYC*+ LBCL) have an inferior prognosis following standard first-line therapy with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone (R-CHOP) compared to patients without *MYC* rearrangement. Although intensive chemotherapy regimens yield higher remission rates, toxicity remains a concern. Lenalidomide is an oral immunomodulatory drug which downregulates MYC and its target genes thereby providing support using lenalidomide as additional therapeutic option for *MYC*+ LBCL. A phase II trial was conducted evaluating the efficacy of lenalidomide (15 mg day 1-14) in combination with R-CHOP (R2CHOP) in newly diagnosed *MYC*+ LBCL patients identified through a nationwide *MYC*-FISH screening program. The primary endpoint was complete metabolic response (CMR) on centrally reviewed ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography (PET)-computed tomography (CT)-scan at end-of-treatment. Secondary endpoints were overall survival (OS), disease-free survival (DFS) and event-free survival (EFS). Eighty-two patients with stage II-IV *MYC*+ LBCL were treated with six cycles of R2CHOP. At end of treatment, 67% (95% Confidence interval [CI]: 58-75) of the patients reached CMR. With a median follow-up of 25.4 months, 2-year estimates for OS, DFS, EFS were 73% (95% CI: 62-82), 75% (95% CI: 63-84) and 63% (95% CI: 52-73) respectively. In this prospective trial for newly diagnosed *MYC*+ LBCL patients, we found that administering R2CHOP was safe, and yields comparable CMR and survival rates as in studies applying more intensive chemotherapy regimens. Hence, these findings offer new prospects for *MYC*+ LBCL patients and warrant comparison in prospective randomized clinical trials. This trial was registered at *www.clinicaltrialsregister.eu* (#2014-002654-39).

Introduction

Diffuse large B-cell lymphoma (DLBCL) comprises about 35% of all non-Hodgkin lymphomas (NHL) and is the most common lymphoma subtype [1]. The outcome of patients treated with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone (R-CHOP) is heterogeneous for which the IPI score and cell-of-origin (COO) are the most well-known denominators [2,3]. *MYC* rearrangement status is an independent prognostic factor, and is reported in 10-15% of DLBCL patients (hereafter *MYC*+ LBCL) [4-7]. In about 30% of

these patients, only a single *MYC* rearrangement is found (single hit [SH]), while in 70% *MYC* rearrangement is detected together with either a *BCL2* or *BCL6* rearrangement (double hit [DH]: *MYC+/BCL2+* or *MYC+/BCL6+*) or with both (triple hit [TH]: *MYC+/BCL2+/BCL6+*) [4]. It has been shown that in patients with *MYC+* LBCL, standard first-line therapy with R-CHOP results in an inferior prognosis compared to those without *MYC* rearrangement (2-year OS 35% *vs.* 61% [8] and 5-year OS 31% *vs.* 66% [6]). Moreover, patients with *MYC+* LBCL have an increased risk of central nervous system (CNS) relapse [5,6]. Recently, Rosenwald *et al*. demonstrated that the inferior prognosis of *MYC* rearranged patients is however largely observed in patients with DH/TH lymphoma [7]. In the revised World Health Organisation (WHO) 2017 classification, SH is not recognized as a separate entity in contrast to DH/TH lymphoma [1].

In search for improvement, intensified chemotherapy regimens, such as hyper-CVAD and R-CODOX-M/R-IVAC, have been investigated. Data mainly come from subanalyses of *MYC* rearrangement positive patients in trials designed for unselected DLBCL patients. These studies indicate that intensified treatment results in improvement of progression free survival (PFS), but not OS [9-11]. Only recently, a prospective, multicenter, single arm phase II study specifically designed for *MYC+* LBCL patients showed that DA-EPOCH-R resulted in a promising CMR rate at end of treatment (EOT) of 74% and 4 year EFS and OS of 71% and 77%, respectively [12].

Lenalidomide is an oral immunomodulatory drug with direct antitumor effects and indirect effects on the tumor microenvironment [13]. *In vitro* studies have demonstrated that lenalidomide exposure results in down-regulation of MYC and its target genes *via* cereblon and IRF4 in lymphoid cells, thereby providing the rationale for introducing lenalidomide as a therapeutic option in *MYC+* LBCL [14]. Two phase II studies in ABC/non-GCB-subtype DLBCL have demonstrated that the addition of lenalidomide to R-CHOP (R2CHOP) is indeed feasible and may contribute to a favorable outcome by decreasing CNS relapse [15,16]. Against expectation, R2CHOP did not result in a survival advantage in ABC-subtype DLBCL, as has recently been shown in a phase III study (ROBUST) [17].

The present study reports the results of a prospective single-arm phase II trial for *MYC+* LBCL patients treated with R2CHOP. Patients were identified through a nationwide molecular biomarker diagnostics program. We report outcome based on

the primary endpoint, which was CMR by centrally reviewed ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography (PET)-computer tomography (CT) scan at EOT, as well as 2-year OS, DFS and EFS rates.

Methods

Screening program and patient eligibility

To support timely diagnosis of *MYC*+ LBCL and optimal enrolment in the present clinical trial, a nationwide diagnostic support program for *MYC* rearrangement assessment by fluorescence *in situ* hybridization (FISH) was implemented [18]. Patients ≥18 years with newly diagnosed DLBCL or with B-cell lymphoma, unclassifiable, with features intermediate between DLBCL and Burkitt lymphoma (BLC-U) according to the WHO 2008 classification with a proven *MYC* rearrangement by FISH analysis including SH (not Burkitt lymphoma), DH or TH DLBCL were eligible. During the screening period one cycle of R-CHOP, a short course of steroids, or irradiation to control local symptoms was allowed. Patients with Ann Arbor stage II-IV, a WHO performance status (PS) of 0-3, ≥ one lesion of ≥1.5 cm on a contrast-enhanced CT scan and ≥one positive lesion on PET-CT scan were eligible. Patients diagnosed with any other subtype of aggressive B-cell lymphoma, a history of follicular lymphoma, proven CNS localization or HIV positivity were excluded.

Treatment

Treatment consisted of six cycles of standard R-CHOP every 3 weeks plus lenalidomide 15 mg orally on day 1-14 (R2CHOP; *Online Supplementary Table S1*), followed by two additional rituximab administrations. Prophylactic intrathecal methotrexate or cytarabine (≥4 administrations), pegfilgrastim, venous thromboembolism prophylaxis (with aspirin or low-molecular-weight-heparin), and *Pneumocystis* prophylaxis were mandatory.

Safety assessments

Adverse events were graded according to National Cancer Institute Common Terminology Criteria for Adverse Events (AE), version 4.03. AE grade 1 were not reported.

Study overview

This multicenter, phase II study was designed by investigators of HOVON and was approved by the medical ethics committee of the Amsterdam UMC. All patients provided written informed consent. The trial was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. An independent data and safety monitoring board conducted a review during the planned interim analysis.

Endpoints

The primary endpoint of the study was CMR on EOT PET-CT scan as determined by central review plus EOT bone marrow (BM) examination in case of BM localization at diagnosis. In case a BM examination was not repeated at EOT in patients with baseline BM localization and the EOT-PET scan showed no BM uptake localization, the response was classified as CMR (based on recent findings that CMR on PET-CT has a high negative predictive value for BM localization) [23,24]. Secondary endpoints were: OS defined as time from registration to death; DFS defined as time from achievement of first CMR on protocol until relapse or death whichever comes first; EFS defined as the time from registration to lack of CMR on EOT PET-CT, relapse or death; and positive predictive value (PPV) and negative predictive value (NPV) of iPET-CT for EOT result.

Statistical analyses

An optimal Simon two-stage design was used with a response rate of 45% as the null hypothesis, and 60% as the alternative hypothesis. With a statistical significance level of 5% and a power of 80%, the required number of patients was 77, with an interim analysis for futility involving the first 26 included patients. In order to overcome dropouts due to ineligibility, 85 patients were enrolled. All efficacy analyses were restricted to eligible patients, while safety analyses included all enrolled patients. Data cut-off was June 28, 2019. For the clinical protocol, central pathology review, central PET-CT review and additional statistical information, see the *Online Supplementary Data*.

Results

Clinical characteristics

From April 2015 to February 2018, 85 patients were included from 20 hospitals in the Netherlands and Belgium. Three patients were declared ineligible (two because the *MYC*+ status was based on immunohistochemistry and not on FISH and one because of a transformed lymphoma), leaving 82 patients for efficacy and 85 patients for safety analyses. Baseline patient and disease characteristics are shown in Table 1. The median age was 63 years (range: 28-82 years). 49 of 81 patients (60%) had a WHO performance status (PS) of 0; 58 of 71 patients (71%) had stage IV disease, and 42 of 82 patients (51%) had ≥2 extranodal localizations. The IPI score was high-intermediate and high in 65% of patients. During treatment 12 of 82 patients went off protocol before completion (progressive disease [*n*=7], toxicity [*n*=2]; pulmonary embolism and diarrhea, other reasons [*n*=3]; new diagnosis of colon cancer, patient refusal, and vertebral fracture), see Figure 1.

Table 1. Patient demographics and disease characteristics.

	N	**%**
Patients completed treatment	82	100
Median age (range) in years	63 (28-82)	
Sex		
male	56	68
female	26	32
WHO performance status		
0	49	60
1	26	32
2	5	6
3	2	2
Prior treatment		
no	13	16
1 course of R-CHOP	68	83
only corticosteroids	1	1
Ann Arbor stage		
II	12	15
III	12	15
IV	58	71
Extranodal localisations		
0	31	38
1	9	11
≥2	42	51
LDH > ULN		
yes	57	70
no	20	24
unknown	5	6

Table 1. Continued

	N	%
Bone Marrow involvement		
yes	16	20
no	44	54
not done	22	27
IPI		
Low	12	15
Low-intermediate	17	21
High-intermediate	32	39
High	21	26
Morphology (WHO2008)		
DLBCL	65	79
BCL-U	12	15
indecisive between DLBCL or BCL-U	5	6
COO IHC (Hans classification)		
GCB subtype	63	77
Non-GCB subtype	8	10
Not evaluable	11	13
COO GEP (Nanostring) n=38		
GCB subtype	29	76
ABC subtype	7	18
Intermediate	2	5
FISH analysis		
single hit	20	24
double hit	44	54
MYC+/*BCL2*+	31*	
MYC+/*BCL6*+	13**	
Triple hit	9	11
MYC+ (*BCL2* and *BCL6* status unknown)	9	11

Demographics and disease characteristics of 82 *MYC*+ LBCL patients treated with R2CHOP. LBCL: large B-cell lymphoma; DLBCL: diffuse large B-cell lymphoma; WHO: World Health Organisation; R-CHOP: rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone; FISH: fluorescence *in situ* hybridization, LDH: lactate dehydrogenase, ULN: upper limit of normal; GCB: germinal center B-cell subtype; COO: cell-of-origin; IHC: immune-histochemistry; GEP: gene expression profiling; BCL-U; B-cell lymphoma, unclassifiable, with features intermediate between DLBCL and Burkitt lymphoma
*from 4 of these patients *BCL6* status is unknown, **from 1 of these patients *BCL2* status is unknown.

Pathology review

Diagnostic biopsy samples of all 85 patients were available for pathology review. Results of all 82 eligible patients are summarized in Table 1 and the *Online Supplementary Table S2*. A diagnosis of DLBCL according to the WHO 2008 classification was confirmed in 65 of 82 patients (79%) and BCL-U in 12 of 82 patients (15%) and morphology was indecisive between DLBCL and BCL-U in 5 of 82 patients (6%). For classification according to the WHO classification 2017 see the *Online Supplementary Table S2*. In 81 of 82 patients *MYC* rearrangement was confirmed at central review. Based on the intention to treat principle, the one patient in whom *MYC* rearrangement could not be confirmed was included

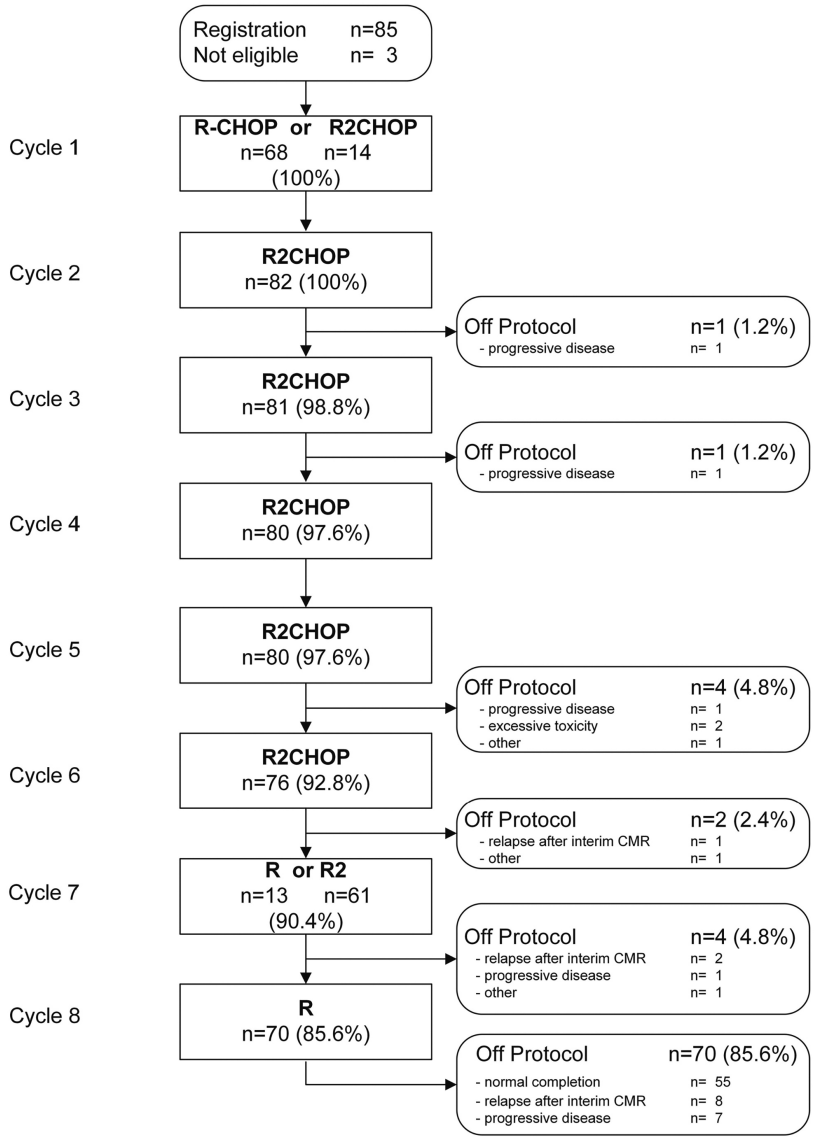


Figure 1. Disposition of the patients.

Eighty-two patients were included. In 14 patients, *MYC* fluorescence *in situ* hybridization (FISH) was performed immediately at diagnosis, these patients started with R2CHOP (lenalidomide in combination with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone) in cycle 1. In 68 patients, *MYC* results became available during the first cycle of R-CHOP; these patients were registered after the first cycle of R-CHOP and started with R2CHOP in the second cycle and continued lenalidomide for 14 days after the sixth cycle of R-CHOP. During treatment 13 patients went off protocol (progressive disease [n=7], toxicity [n=2; pulmonary embolism and diarrhea], other reasons [n=3; new diagnosis of colon cancer, patient refusal, and vertebral fracture]). R: rituximab, R2: rituximab + lenalidomide.

in all analyses. In 9 of 82 cases, insufficient material was available to perform additional *BCL2* and *BCL6* rearrangement. In 73 of 82 cases, data on *BCL2* and *BCL6* rearrangement were available: 20 of 82 (26%) had a single *MYC* rearrangement (SH); 44 of 82 (54%) had DH lymphoma (31 patients had *MYC/BCL2* rearrangements and 13 patients MYC/BCL6 rearrangements), and 9 of 82 (11%) had all three rearrangements (TH). COO classification using a standard Hans algorithm showed GCB phenotype in 63 of 71 (89%) and non-GCB phenotype in 8 of 71 (11%). Lymph2Cx classification was performed in 38 cases showing GCB-subtype in 29 of 38 patients (76%), ABC-subtype in 7 of 38 patients (18%), and intermediate subtype in 2 of 38 patients (5%). Out of the 24 DH of TH patients, 21 showed GCB-subtype and 3 ABC-subtype. Out of the 12 SH patients, 8 showed GCB-subtype and 4 ABC-subtype.

Treatment

Most patients (*n*=68) started with lenalidomide in the second cycle and continued lenalidomide for 14 days after the sixth cycle of R-CHOP. When *MYC* FISH results were available at diagnosis, R2CHOP was started in the first cycle (*n*=14) (Figure 1). Patients received a median (interquartile range [IQR]) dose of the planned drugs in the R-CHOP regimen as follows: cyclophosphamide 99.9% (99.0-101); vincristine 100% (72.5-100); doxorubicin 99.5% (97.7-101); prednisone 100% (100-100); rituximab 98.1 (95.1-100); pegfilgrastim 100% (100-100). Lenalidomide was given at a median dose intensity of 100% (range: 85.7-100). 57 of 82 patients (70%) received the planned ≥4 intrathecal prophylactic administrations.

Primary endpoint: CMR at EOT

At EOT PET-CT, 55 of 82 patients (67%) reached the primary endpoint of CMR (95% CI: 58-75, *P*<0.001), 5 of 82 patients (6%) reached a partial metabolic response (PMR), and 21of 82 patients (26%) had progressive metabolic disease (PMD) (Table 2). One patient went off protocol due to toxicity after cycle 5 without EOT PET-CT (response unknown). Univariate logistic regression analysis of baseline characteristics (BM localization, WHO performance status, stage, B symptoms, IPI, number of extranodal sites and age) did not reveal any significant predictors for reaching CMR. Exploratory descriptive subgroup analyses revealed no differences between SH and DH/TH patients regarding achievement of the primary endpoint: CMR rate in both groups was 70% and 66% respectively

(nine patients with unknown *BCL2* and *BCL6* rearrangement not included).

Table 2. Response rates on interim and EOT PET-CT scan.

	EOT	CMR	PMR	PMD	Unknown[$]	Total
Interim		N	n	n	n	*n*
CMR	n	45	0	11	1	57
PMR	n	10	4	9	0	23
PMD	n	0	1	1	0	2
Total	n	55	5	21	1	82

Response rates on interim and EOT PET-CT scan. Correlation of interim and end of treatment (EOT) response rates by centrally reviewed positron emission tomography (PET)- computed tomography (CT) scan.
[$]One patient was in complete metabolic response (CMR) at interim scan but went off protocol due to toxicity without an EOT scan. PMR: partial metabolic response, PMD: progressive metabolic disease.

Secondary endpoints: survival analyses

With a median follow-up of 25.4 months (IQR 18.3- 30.3), 1-year OS was 85% (95% CI: 76-91), DFS 77% (95% CI: 65-85) and EFS 66% (95% CI: 54-75). 2-year estimates for OS, DFS, EFS were and 73% (95% CI: 62-82), 75% (95% CI: 63-84) and 63% (95% CI: 52-73) respectively (Figure 2A-C). Baseline patient characteristics (BM localization, WHO performance status, stage, B symptoms, IPI, number of extranodal sites and age) were not significantly predictive for prolonged OS in a univariate analysis at the 5% significance level. Univariate regression analyses indicated that SH and DH/TH patients had comparable EFS and DFS, however DH/TH patients had a tendency for a higher risk of death compared to SH patients (Hazard ratio [HR]4.18, P=0.055; 95% CI: 0.97-18.02) (*Online Supplementary Figure S1A-C*). Separate analyses of DH *MYC/BCL2* and DH *MYC/BCL6* and TH in comparison to SH revealed no significant differences in OS (*Online Supplementary Figure S2A-B*). In univariate analyses with response as time dependent covariate we found that patients who had achieved CMR at EOT PET experienced a reduced risk of death compared to patients who had not achieved CMR (HR 0.1, 95% CI: 0.03–0.33, P<0.001), (Figure 3). EOT PET-CT predicted relapse within 12 months, with a positive predictive value (PPV) of 81% and a negative predictive value (NPV) of 93% (*Online Supplementary Table S3A*). In total, 29 patients showed progressive disease (11 without achieving CMR, 18 after achieving CMR [(at interim or EOT PET-CT)] including one patient with a CNS relapse.

Safety

Grade 2, 3 and 4 AE were seen in 27 (32%), 33 (39%) and 14 (16%) of all 85 registered patients respectively (Table 3). The most common grade 3–4 AE were neutropenia (18%), infections (14%) and gastrointestinal disorders (14%). Four patients experienced deep venous thrombosis (grade 2), and two patients pulmonary embolism (grade 3). Two of these patients (one with deep venous thrombosis and one with pulmonary embolism) had not received the mandatory thrombosis prophylaxis (protocol violation). One patient went off protocol due to grade 3 diarrhea. 71 serious AE were reported in 36 patients; 66 were due to hospitalization (42% infections, 26% gastrointestinal disorders), four to other conditions [two second primary malignancies, two recurrence of previously diagnosed (>5 year) malignancies]. One patient died during treatment due to progression. There were no treatment related deaths.

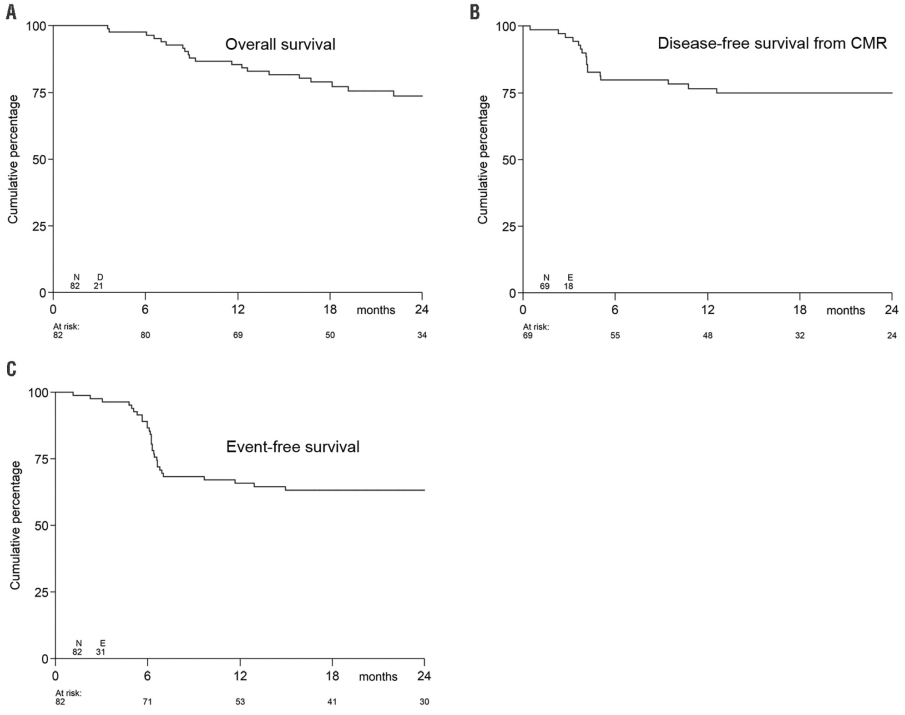


Figure 2. Survival analyses.

(A) Overall survival (OS; time from registration to death, n=82); (B) disease-free survival (DFS; time from achievement of first complete metabolic response [CMR] on protocol until relapse or death whichever comes first, n=69); (C) event-free survival (EFS; defined as the time from registration to lack of CMR on end of treatment [EOT] positron emission tomography [PET]-computer tomography [CT] scan, relapse or death, n=82) of *MYC*+ LBCL patients.

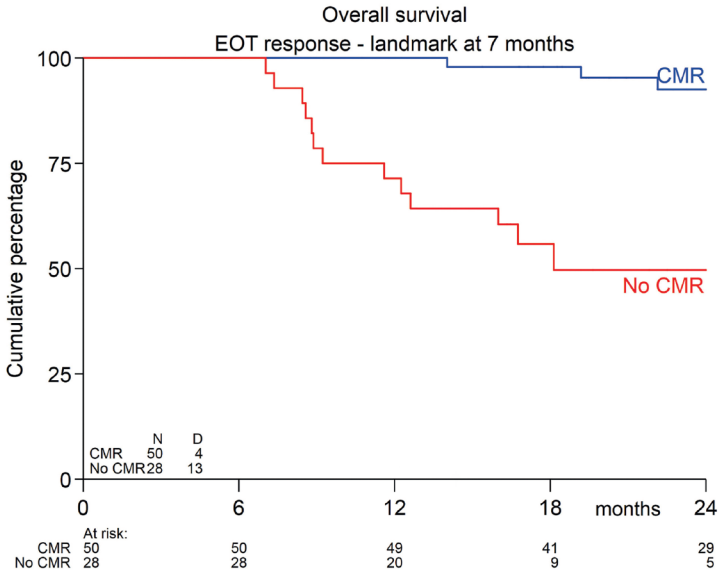


Figure 3. Survival according to end-of-treatment PET-CT scan result.

Patients who have achieved complete metabolic response (CMR) at the end of treatment (EOT) positron emission tomography (PET)-computer tomography (CT) scan experienced a reduced risk of death compared to patients who have not yet achieved CMR (Hazard ratio [HR] 0.1, 95% Confidence interval [CI]: 0.03– 0.33, *P*<0.001). Response was simplified to "CMR" *versus* "no-CMR".

Table 3. Adverse events.

	Grade 2		Grade 3		Grade 4	
	n	%	*n*	%	*n*	%
Hematologic						
Neutropenia	1	1	5	6	10	12
Febrile neutropenia			6	7		
Thrombocytopenia	2	2	2	2	4	5
Anemia	6	7	5	6		
Infectious	15	18	12	14		
Vascular disorders						
Pulmonary embolism			2	3		
Deep venous thrombosis	4	5				
Superficial thrombophlebitis	4	5				
Nervous system disorders (PNP)	25	29	9	11		
Gastrointestinal disorders	17	20	12	14		
Hepatobiliary disorders (ALT, AST increased)	5	6	1	1	2	2
General disorder	11	13	5	6	1	1
Any*	27	32	33	39	14	16

Adverse events (AE) were graded per patient (maximum grade per cycle) according to National Cancer Institute Common Terminology Criteria for AE. AE events grade 1 were not reported.
*In this row each patient is counted only once with the highest grade AE experienced. PNP: polyneuropathy, ALT: alanine transaminase, AST: aspartate transaminase.

Observational analysis: predictive value of iPET-CT

At iPET-CT after three cycles of R2CHOP, 57 of 82 patients (70%) were in CMR; of these 45 of 57 (79%) were still in CMR and 11 of 57 (19%) showed PMD at EOT PET-CT, and one missed EOT evaluation (Table 2). 23 of 82 patients (28%) were in PMR at iPET-CT; 10 of 23 (43%) of these converted to CMR, 4 of 23 (17%) remained in PMR, 9 of 23 (39%) showed PMD at EOT. The PPV of iPET-CT for predicting EOT PET-CT result was 60% (15 of 25), the NPV 79% (45 of 57) (*Online Supplementary Table S3B*).

Discussion

From retrospective series it is clear that first-line R-CHOP therapy is not sufficiently effective for patients with *MYC*+ LBCL with CR rates of 40-50% and 3-year OS rates of 35% only [6,8]. Intensified chemotherapy regimens such as R-CODOX-M/R-IVAC or autologous stem cell transplantation have not improved OS, and result in increased toxicity [5,9-11].

We designed a prospective clinical trial for *MYC*+ LBCL patients, in which a time window of one cycle of R-CHOP was allowed to perform molecular diagnostics. This approach permitted high risk patients to start treatment immediately and overcame the bias of inclusion of mainly lower risk patients due to enrolment delays [25]. In this trial we show that the addition of lenalidomide to R-CHOP resulted in EOT CMR rate of 67% CMR and 2-year survival rates of 73%, 75%, and 63% for OS, DFS and EFS respectively. To our knowledge, this is the second prospective trial especially designed for *MYC*+ LBCL patients. Recently, Dunleavy and colleagues reported a single arm phase II study in which the efficacy of DA-EPOCH-R for *MYC*+ LBCL patients was explored [12]. Results for EOT CMR and survival rates are largely comparable between both approaches with EOT CMR rate of 74%, 4-year OS rates of 77% and EFS of 71% in the study of Dunleavy. When compared to the trial of Dunleavy, our patient population was larger (82 *vs.* 53 patients), comparable in age (median 63 *vs.* 61 years) but included more patients with IPI ≥3 (65% *vs.* 49%) and more patients with DH/TH (65% *vs.* 45%). Regarding safety, grade 3/4 infections were seen in 24% of cycles with DA-EPOCH-R *versus* grade 3 (and no grade 4 infections) in only 2,8 % of cycles (18 episodes) with R2CHOP. DA-EPOCH-R resulted in three treatment related deaths *vs.* none with R2CHOP.

Lenalidomide penetrates the CNS, and thereby may aid to prevent CNS relapses as has been suggested for nongerminal center B-cell (GCB) subtype lymphoma patients treated with R2CHOP [15]. Indeed, in this study, which combined lenalidomide and intrathecal prophylaxis, a remarkably low rate of CNS relapse at a median follow-up of 25.4 months was seen (*n*=1).

Several remarks regarding our study can be made. First, although correlation CMR at EOT PET-CT with survival in *MYC*+ LBCL has been described in a retrospective study [26], one might argue that it is not an ideal primary endpoint. Given the high FDG-avidity of *MYC*+ LBCL and the fact that CMR at EOT PET-CT in our study was highly predictive for DFS (NPV of 93%), we feel that using CMR at EOT PET-CT as a surrogate endpoint for highly aggressive B-cell lymphomas such as *MYC*+ LBCL is justified.

Second, clinical prognostic markers, including age, stage, IPI score, as well as COO were not significantly correlated to CMR on EOT PET-CT and survival, which might be explained by the inclusion of high risk patients; 65% of our patients have an IPI score of ≥3, *versus* only 27% in Ziepert's meta-analysis of the value of IPI in the rituximab era [2].

Furthermore, our patient population included patients with SH lymphoma (24%) based on previous reports demonstrating poor prognosis of these patients following R-CHOP [5,6,8,27]. However, in the revised WHO 2017 classification, SH is not recognized as a separate entity in contrast to DH/TH lymphoma. Recently, Rosenwald *et al*. demonstrated that the inferior prognosis of *MYC* rearranged patients is largely observed in patients with DH lymphoma [7]. However, SH patients still have a worse prognosis compared to patients without a *MYC* rearrangement, although this is not statistically significant when regarding OS (*P*=0.077). Our trial was not powered to study the prognostic impact of SH versus DH/TH in the R2CHOP setting.

Finally, we explored the role of iPET-CT scanning as a tool for early identification of refractory *MYC*+ LBCL. In non-selected cases of DLBCL, CMR on iPET-CT after two to four R-CHOP cycles has a high NPV for 2-year PFS, but the PPV varies widely [28]. In our study, the PPV of iPET-CT for achievement of CMR on EOT PET-CT was only 60% and therefore does not support the use

of an interim PET-CT scan as interpreted with the current standard criteria to identify primary refractory cases treated with R2CHOP.

In this prospective trial for newly diagnosed *MYC*+ LBCL patients, we found that administering R2CHOP was safe, and yielded comparable CMR and survival rates as in studies applying intensive chemotherapy regimens. Moreover, R2CHOP can be delivered on an outpatient basis in contrast to Burkitt schemes and is easier to deliver than DA-EPOCH-R, since it does not require placement of a central line. These findings offer new prospects for *MYC*+ LBCL patients and warrant comparison in prospective randomized trials.

Acknowledgements

The authors wish to acknowledge: the trial managers and central data managers of the study (N. Lamers and A. Elsinghorst, R. Sewsaran and H. Hofwegen), HOVON data Center in Rotterdam; Dr Staudt's Laboratory at NCI/NIH Bethesda, MD, USA, for online analysis of Lymph2Cx raw data for COO characterization; and the DSMB members M. Jerkeman, K. Dunleavy and E. van Werkhoven for their advice following the interim analysis.

Funding

This work was supported by Celgene who funded the research, provided lenalidomide free of charge, approved the original protocol and all amendments as written, but had no role in either the design of the study, the collection, analysis, and interpretation of the data, or the preparation of the manuscript. The Dutch Cancer Society provided funding for the screening program and the clinical trial.

References

1. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016;127(20):2375-2390.
2. Ziepert M, Hasenclever D, Kuhnt E, et al. Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28(14):2373-2380.
3. Lenz G, Wright GW, Emre NC, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A.* 2008;105(36):13520-13525.
4. Aukema SM, Kreuz M, Kohler CW, et al. Biologic characterization of adult MYC translocation positive mature B-cell lymphomas other than molecular Burkitt lymphoma. *Haematologica.* 2014;99(4):726-735.
5. Oki Y, Noorani M, Lin P, et al. Double hit lymphoma: the MD Anderson Cancer Center clinical experience. *Br J Haematol.* 2014;166(6):891-901.
6. Savage KJ, Johnson NA, Ben-Neriah S, et al. MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood.* 2009;114(17):3533-3537.
7. Rosenwald A, Bens S, Advani R, et al. Prognostic significance of MYC rearrangement and translocation partner in diffuse large B-cell lymphoma: a study by the Lunenburg Lymphoma Biomarker Consortium. *J Clin Oncol.* 2019;37(35):3359-3368.
8. Barrans S, Crouch S, Smith A, et al. Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab. *J Clin Oncol.* 2010;28(20):3360-3365.
9. Howlett C, Snedecor SJ, Landsburg DJ, et al. Front-line,dose-escalated immunochemotherapy is associated with a significant progression-free survival advantage in patients with double-hit lymphomas: a systematic review and meta-analysis. *Br J Haematol.* 2015; 170(4):504-514.
10. Petrich AM, Gandhi M, Jovanovic B, et al. Impact of induction regimen and stem cell transplantation on outcomes in double-hit lymphoma: a multicenter retrospective analysis. *Blood.* 2014;124(15):2354-2361.
11. Landsburg DJ, Falkiewicz MK, Maly J, et al. Outcomes of patients with double hit lymphoma who achieve first complete remission. *J Clin Oncol.* 2017;35(20):2260-2267.
12. Dunleavy K, Fanale MA, Abramson JS, et al. Dose-adjusted EPOCH-R (etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin, and rituximab) in untreated aggressive diffuse large B-cell lymphoma with MYC rearrangement: a prospective, multicentre, single-arm phase 2 study. *Lancet Haematol.* 2018;5(12):e609-e617.
13. Witzig TE, Wiernik PH, Moore T, et al. Lenalidomide oral monotherapy produces durable responses in relapsed or refractory indolent non-Hodgkin's Lymphoma. *J Clin Oncol.* 2009;27(32):5404-5409.
14. Lopez-Girona A, Heintel D, Zhang LH, et al. Lenalidomide downregulates the cell survival factor, interferon regulatory factor-4, providing a potential mechanistic link for predicting response. *Br J Haematol.* 2011; 154(3):325-336.
15. Ayed AO, Chiappella A, Pederson L, et al. CNS relapse in patients with DLBCL treated with lenalidomide plus R-CHOP (R2CHOP): analysis from two phase 2 studies. *Blood Cancer J.* 2018;8(7):63.
16. Castellino A, Chiappella A, LaPlant BR, et al. Lenalidomide plus R-CHOP21 in newly diagnosed diffuse large B-cell lymphoma (DLBCL): long-term follow-up results from a combined analysis from two phase 2 trials. *Blood Cancer J.* 2018;8(11):108.
17. Vitolo U, Witzig TE, Gascoyne RD, et al. ROBUST: first report of phase III randomized study of lenalidomide/R-CHOP (R2- CHOP) vs placebo/R-CHOP in previously untreated ABC-type diffuse large B-cell lymphoma. *Haematol Oncol.* 2019;37(52):36-37.
18. Chamuleau M, Nijland M, Lamers N, et al. First Report on a successful screening program for MYC rearrangements and a prospective clinical trial based on MYC rearrangement in newly diagnosed DLBCL patients in the Netherlands. *Blood.* 2017; 130(Supplement 1):4144.

19. Scott DW, Wright GW, Williams PM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin embedded tissue. *Blood.* 2014;123(8):1214-2117.
20. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *J Clin Oncol.* 2014;32(27):3048-3058.
21. 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(27):3059-3068.
22. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42(2):328-354.
23. Teagle AR, Barton H, Charles-Edwards E, Dizdarevic S, Chevassut T. Use of FDG PET/CT in identification of bone marrow involvement in diffuse large B cell lymphoma and follicular lymphoma: comparison with iliac crest bone marrow biopsy. *Acta Radiol.* 2017;58(12):1476-1484.
24. Berthet L, Cochet A, Kanoun S, et al. In newly diagnosed diffuse large B-cell lymphoma, determination of bone marrow involvement with 18F-FDG PET/CT provides better diagnostic performance and prognostic stratification than does biopsy. *J Nucl Med.* 2013; 54(8):1244-1250.
25. Maurer MJ, Ghesquieres H, Link BK, et al. Diagnosis-to-treatment interval is an important clinical factor in newly diagnosed diffuse large B-cell lymphoma and has implication for bias in clinical trials. *J Clin Oncol.* 2018;36(16):1603-1610.
26. Cohen JB, Geyer SM, Lozanski G, et al. Complete response to induction therapy in patients with Myc-positive and double-hit non-Hodgkin lymphoma is associated with prolonged progression-free survival. *Cancer.* 2014;120(11):1677-1685.
27. Landsburg DJ, Falkiewicz MK, Petrich AM, et al. Sole rearrangement but not amplification of MYC is associated with a poor prognosis in patients with diffuse large B cell lymphoma and B cell lymphoma unclassifiable. *Br J Haematol.* 2016;175(4): 631-640.
28. Burggraaff CN, de Jong A, Hoekstra OS, et al. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2019; 46(1):65-79.

Supplemental materials

Study protocol

Complete protocol is available as supplemental file and on line at the HOVON website: http://www.hovon.nl/studies/studies-per-ziektebeeld/nhl.html?action=showstudie&studie_id=107&categorie_id=1

National screening program to support implementation of FISH screening in pathology practice

To support timely diagnosis of *MYC*+ LBCL and optimal enrolment in the present clinical trial, a nationwide diagnostic support program for *MYC* rearrangement assessment by fluorescence *in situ* hybridization (FISH) was implemented [1].

In brief, at registration of de novo aggressive B-cell lymphoma in the program, limited financial support was provided for FISH diagnostics. With this support, pathology labs, who did not have these assays available in-house were invited to submit cases to dedicated regional reference laboratories to guarantee access to standard FISH testing for *MYC*, *BCL2* and *BCL6*. An initiating quality control validation was performed prior to acceptance as reference or "in-house" lab (August 2013, coordinators D. de Jong, P.M. Kluin). Both technical quality and scoring reproducibility were monitored. Validation was repeated as more labs implemented FISH diagnostics over time during trial accrual. At initial quality control validation, labs performed FISH according to routine procedures with standard commercial probes: MYC Break-apart provided by Vysis/Abbott (*n*=7) DAKO (*n*=7) and Kreatech (*n*=1); BCL2 Break-apart provided by Vysis/Abbott (*n*=6), DAKO (*n*=9); BCL6 break-apart provided by Vysis/Abbott (*n*=4), DAKO (*n*=6) and Kreatech (*n*=1). Initially, 10/15 labs were accepted as reference or "in-house" lab based on optimal performance and 5 labs were rejected based on insufficient quality (high false negative and/or false positive rate). During trial accrual, 7 additional labs passed quality assessment criteria and were accepted. It should be noted, that over time *MYC* Break-apart from DAKO was replaced for Vysis/Abbot by most labs based on the results of the validation round.

Central pathology review

Central pathology review included classification according to the criteria of the WHO classification 2008 and 2017, including appropriate immunohistochemistry (IHC) for at least CD20, CD10, BCL6 and BCL2 and confirmation of *MYC*

rearrangement status based on complete pathology/molecular reports. In case of equivocal documentation, FISH assays were repeated at the HOVON Pathology Facility. *BCL2* and *BCL6* FISH results were completed when sufficient material was available. In cases with sufficient material COO classification was determined by IHC (Hans algorithm) and by using gene expression profiling (Nanostring Lymph2CX assay: raw counts obtained by Nanostring gene expression analysis were uploaded at the Lymphoma/Leukemia Molecular Profiling Project website [2].

Imaging assessments and central PET-CT review

Contrast-enhanced CT scans and ^{18}F-FDG PET scans combined with low-dose CT scans (PET-CT) were performed at baseline, after 3 cycles of treatment (interim PET (iPET-CT)), and at EOT. The EOT PET-CT scan was scheduled 6-8 weeks after the last lenalidomide administration. Treatment response at iPET-CT and EOT PET-CT was assessed according to the Lugano criteria using the visual Deauville 5-point scoring system [3,4]. Deauville scores of 1- 3 were interpreted as CMR, while scores 4 and 5 indicated stable or progressive disease. PET-CT scans were anonymized and uploaded to a Keosys (Imagys) web-based viewing and reporting system and centrally reviewed by two independent experienced nuclear medicine physicians of the HOVON Imaging Working Group who were blinded for survival outcome. In case of discordance, a third reviewer performed adjudication. PET-CT scans were performed and reviewed in compliance with EANM guidelines [5]. Patients with CMR at iPET-CT but with a positive EOT PET-CT scan were classified as progressive metabolic disease (PMD) at EOT, even when the EOT scan was in partial metabolic response (PMR) compared to the pre-treatment PET-CT scan.

Statistical analyses

In order to take the two-stage sampling nature intrinsic to the study design into account, the primary study endpoint was estimated using the method proposed by Jung [6] , which uses the design parameters and the interim analysis results. The design poses a one-sided hypothesis that the response rate is larger or equal to 60%, which we evaluated at a 5% significance level. For the construction of the corresponding two-sided 90% CI the method of Koyama was followed [7]. Both methods are implemented in the R software package "OneArmPhaseTwoStudy" [8]. The secondary survival endpoints were evaluated using the Kaplan-Meier method. Univariate logistic and Cox proportional hazards regression models were

used to assess the effect on EOT response rate and the survival endpoints of the following baseline characteristics: BM involvement, WHO PS categorized as 0, 1, 2 or 3, disease stage I-II versus III-IV, presence of B symptoms, presence of concomitant diseases, IPI, number of extranodal localizations categorized as 0, 1, 2 or more, and age as continuous variable. The predictive value of CMR at interim response evaluation for CMR at EOT was assessed through positive predictive value (PPV) and negative predictive value (NPV), where response was simplified to "CMR" versus "no-CMR". PPV was defined as the proportion of patients without EOT PET-CT CMR among the patients without CMR on iPET, and NPV was defined as the proportion of patients with EOT PET-CT CMR among the patients with CMR on iPET. The effect of CMR at EOT on OS was independently evaluated using achievement of CMR as a time-dependent covariate in a Cox proportional hazards regression model, and visualized using the Kaplan-Meier method with a landmark at 7 months.

Exploratory analyses consisted of descriptive subgroup analyses based on rearrangement group (SH versus DH and TH) as determined by central pathology review. Analyses were performed by tabulation of response rate and Kaplan-Meier curves for OS by rearrangement group. All analyses, except analysis of the primary endpoint for which R software was used, were performed using Stata software, version 15. Data cut-off was June 28, 2019.

References

1. Chamuleau M, Nijland M, Lamers N, et al. First Report on a Successful Screening Program for MYC Rearrangements and a Prospective Clinical Trial Based on MYC Rearrangement in Newly Diagnosed DLBCL Patients in the Netherlands. *Blood.* 2017; 130(Suppl 1): 4144.
2. Scott DW, Wright GW, Williams PM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood.* 2014; 123(8): 1214-1217.
3. 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(27): 3048-3058.
4. 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(27): 3059-3068.
5. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015; 42(2): 328-354.
6. Jung SH, Kim KM. On the estimation of the binomial probability in multistage clinical trials. *Statistics in medicine.* 2004; 23(6): 881-896.
7. Koyama T, Chen H. Proper inference from Simon's two-stage designs. *Statistics in medicine.* 2008; 27(16): 3145-3154.
8. Kieser M WM, Englert S, Kunz CU and Rauch G. "OneArmPhaseTwoStudy: An R Package for Planning, Conducting, and Analysing Single-Arm Phase II Studies.". *Journal of Statistical Software.* 2017; 81(8): 1-28

Table S1. Treatment schedule of R2CHOP.

Agent	Dose/day	Route of administration	Days
Day 1	Cyclophosphamide	750 mg/m^2	i.v.
	Vincristine	1.4 mg/m^2 (max 2 mg)	i.v.
	Doxorubicin	50 mg/m^2	i.v.
	Rituximab	375 mg/m^2	i.v.
Day 1-5	Prednisone	100 mg	p.o.
Day 1-14	Lenalidomide	15 mg	p.o.
Day 2	Pegfilgrastim	6 mg	s.c.

The R2CHOP schema consist of R-CHOP21 with lenalidomide 15 mg on 1-14. Additionally, patients received at least 4 intrathecal administrations of methotrexate or cytarabine.

Table S2. Central pathology review data on 85 *MYC*+ LBCL patients treated with R2CHOP.

		Immunohistochemistry				
Patient number	**Eligible**	**CD20 0= negative (<95%), 1=positive (>95%)**	**BCL2 0=negative (<50%), 1=positive (50%), 9=not available**	**MYC-IHC 0=negative (<40%), 1=positive (>40%), 9=not available**	**CD10 0=negative, 1=positive, 9=not available**	**BCL6 0=negative (<40%), 1=positive (>40%), 9=not available**
1	yes	1	0	1	1	1
2	yes	1	1	9	0	1
3	yes	1	1	9	1	1
4	yes	1	1	0	1	1
5	yes	1	1	9	1	1
6	yes	1	1	1	1	1
7	yes	1	0	9	1	1
8	no					
9	no					
10	yes	1	1	1	1	1
11	yes	1	0	1	1	9
12	yes	1	0	1	1	1
13	yes	1	1	1	1	1
14	yes	1	1	9	1	1
15	yes	1	1	1	9	9
16	yes	1	0	0	1	1
17	yes	1	1	1	1	1
18	yes	1	1	1	1	1
19	yes	1	1	1	1	1
20	yes	1	0	1	1	1
21	yes	1	1	1	1	
22	yes	1	1	0	1	1
23	yes	1	1	9	1	1
24	yes	1	1	1	0	1
25	yes	1	1	1	0	1
26	yes	1	1	9	0	1
27	yes	1	1	9	1	1
28	yes	1	0	1	1	1
29	yes	1	0	1	0	1
30	yes	1	1	0	1	1
31	yes	1	1	0	0	1
32	yes	1	1	1	1	1
33	yes	1	1	1	1	0
34	yes	1	0	1	1	1
35	yes	1	1	9	1	1
36	yes	1	0	1	1	1
37	yes	1	0	9	1	1
38	yes	1	0	1	1	0

		FISH			GEP	Classification	
MUM1 0=negative, (<40%), 1=positive (>40%), 9=not available	**GCB/ non GCB, 9=not available**	**MYC-BA 0=neg 1=pos**	**BCL2-BA 0=neg 1=pos**	**BCL6-BA 0=neg 1=pos**	**ABC/GCB/ undeterminate, 9=not available**	**WHO 2008**	**WHO 2017**
9	GCB	1	0	1	GCB	DLBCL	HGBCL, DH/TH
9	9	1	1	1	ABC	DLBCL	HGBCL, DH/TH
9	GCB	1	1	1	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	1	1	poor quality	BCL-U	HGBCL, DH/TH
0	GCB	1	0	1	9	DLBCL	HGBCL, DH/TH
9	GCB	1	1	0	9	BCL-U	HGBCL, DH/TH
9	GCB	1	9	1	9	DLBCL	HGBCL, DH/TH
		0					
		0					
9	GCB	1	1	0	9	BCL-U	HGBCL, DH/TH
9	GCB	1	0	0	GCB	dd DLBCL or BCL-U	HGBCL, NOS
9	GCB	1	0	0	9	DLBCL	DLBCL
9	GCB	1	1	0	9	dd DLBCL or BCL-U	HGBCL, DH/TH
9	GCB	1	1	0	9	DLBCL	HGBCL, DH/TH
9	9	1	1	0	GCB	DLBCL	HGBCL, DH/TH
0	GCB	1	0	0	9	DLBCL	DLBCL
9	GCB	1	1	9	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	1	1	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	1	9	9	DLBCL	HGBCL, DH/TH
9	GCB	1	0	0	GCB	DLBCL	DLBCL
9	GCB	1	1	0	poor quality	DLBCL	HGBCL, DH/TH
0	GCB	1	1	0	GCB	BCL-U	HGBCL, DH/TH
9	GCB	1	1	1	9	DLBCL	HGBCL, DH/TH
1	non-GCB	1	0	1	9	DLBCL	HGBCL, DH/TH
9	9	1	0	9	9	DLBCL	dd DLBCL or HGBCL, DH/TH
0	GCB	1	9	9	9	DLBCL	dd DLBCL or HGBCL, DH/TH
9	GCB	1	1	0	9	DLBCL	HGBCL, DH/TH
1	GCB	1	1	9	9	dd DLBCL or BCL-U	HGBCL, DH/TH
1	non-GCB	0	0	1	9	DLBCL	HGBCL, DH/TH
9	GCB	1	1	1	9	DLBCL	HGBCL, DH/TH
1	non-GCB	1	0	1	9	DLBCL	HGBCL, DH/TH
9	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	0	0	9	DLBCL	DLBCL
0	GCB	1	9	9	9	BCL-U	dd HGBCL, NOS or HGBCL, DH/TH
9	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	9	9	9	DLBCL	dd DLBCL or HGBCL, DH/TH
9	GCB	1	0	0	GCB	DLBCL	DLBCL

Table S2. Continued

				Immunohistochemistry		
Patient number	**Eligible**	**CD20 0= negative (<95%), 1=positive (>95%)**	**BCL2 0=negative (<50%), 1=positive (50%), 9=not available**	**MYC-IHC 0=negative (<40%), 1=positive (>40%), 9=not available**	**CD10 0=negative, 1=positive, 9=not available**	**BCL6 0=negative (<40%), 1=positive (>40%), 9=not available**
39	yes	1	0	0	0	1
40	yes	1	1	1	1	1
41	yes	1	1	0	0	1
42	yes	1	1	1	1	1
43	yes	1	0	9	0	1
44	yes	1	9	1	9	9
45	yes	1	0	1	1	1
46	yes	1	1	0	1	1
47	yes	1	1	9	1	1
48	yes	1	1	1	9	9
49	yes	1	1	9	1	1
50	yes	1	1	1	1	1
51	yes	1	1	1	9	1
52	yes	1	1	1	0	1
53	yes	1	0	1	1	1
54	yes	1	1	1	1	1
55	yes	1	1	1	1	1
56	yes	1	1	1	1	1
57	yes	1	1	1	1	1
58	yes	1	0	1	1	1
59	yes	1	0	1	9	9
60	yes	1	1	1	1	1
61	yes	1	1	1	1	0
62	yes	1	1	9	1	1
63	yes	1	1	1	1	0
64	no					
65	yes	1	1	9	9	9
66	yes	1	9	1	1	1
67	yes	1	1	9	1	1
68	yes	1	9	9	1	9
69	yes	1	1	9	9	0
70	yes	1	1	1	1	9
71	yes	1	1	1	1	1
72	yes	1	1	0	0	1
73	yes	1	0	9	1	1
74	yes	1	1	1	1	1
75	yes	1	1	1	9	9
76	yes	1	0	1	1	1

		FISH			GEP	Classification	
MUM1 0=negative, (<40%), 1=positive (>40%), 9=not available	**GCB/ non GCB, 9=not available**	**MYC-BA 0=neg 1=pos**	**BCL2-BA 0=neg 1=pos**	**BCL6-BA 0=neg 1=pos**	**ABC/GCB/ undeterminate, 9=not available**	**WHO 2008**	**WHO 2017**
9	9	1	0	1	9	DLBCL	HGBCL, DH/TH
9	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
1	non-GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
1	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
1	non-GCB	1	0	1	9	DLBCL	HGBCL, DH/TH
1	9	1	1	0	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	9	9	9	BCL-U	dd HGBCL, NOS or HGBCL, DH/TH
9	GCB	1	1	0	GCB	BCL-U	HGBCL, DH/TH
0	GCB	1	1	0	9	DLBCL	HGBCL, DH/TH
9	9	1	0	0	ABC	dd DLBCL or BCL-U	dd DLBCL or HGBCL, NOS
1	GCB	1	0	0	9	BCL-U	HGBCL, NOS
0	GCB	1	0	0	poor quality	DLBCL	DLBCL
0	GCB	1	1	0	poor quality	DLBCL	HGBCL, DH/TH
0	GCB	1	0	0	ABC	DLBCL	DLBCL
0	GCB	1	1	1	GCB	DLBCL	HGBCL, DH/TH
0	GCB	1	0	1	poor quality	DLBCL	HGBCL, DH/TH
0	GCB	1	1	1	9	DLBCL	HGBCL, DH/TH
0	GCB	1	1	9	9	BCL-U	HGBCL, DH/TH
0	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
0	GCB	1	0	0	GCB	DLBCL	DLBCL
9	9	1	0	0	9	BCL-U	HGBCL, NOS
9	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	0	1	GCB	DLBCL	HGBCL, DH/TH
0	GCB	1	1	0	9	BCL-U	HGBCL, DH/TH
0	GCB	1	0	0	GCB	DLBCL	DLBCL
		1				synchronous follicular lymphoma	
9	9	1	0	0	9	DLBCL	DLBCL
9	GCB	1	0	0	GCB	DLBCL	DLBCL
0	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
9	GCB	1	9	9	9	DLBCL	dd DLBCL or HGBCL, DH/TH
0	9	1	9	9	9	DLBCL	DLBCL
9	GCB	1	1	0		DLBCL	HGBCL, DH/TH
0	GCB	1	1	0	9	DLBCL	HGBCL, DH/TH
1	non-GCB	1	0	1	ABC	DLBCL	HGBCL, DH/TH
0	GCB	1	0	0	9	DLBCL	DLBCL
9	GCB	1	1	0	GCB	DLBCL	HGBCL, DH/TH
9	9	1	0	0	unclassified	DLBCL	DLBCL
1	GCB	1	0	1	GCB	DLBCL	HGBCL, DH/TH

Table S2. Continued

		Immunohistochemistry				
Patient number	**Eligible**	**CD20 0= negative (<95%), 1=positive (>95%)**	**BCL2 0=negative (<50%), 1=positive (50%), 9=not available**	**MYC-IHC 0=negative (<40%), 1=positive (>40%), 9=not available**	**CD10 0=negative, 1=positive, 9=not available**	**BCL6 0=negative (<40%), 1=positive (>40%), 9=not available**
77	yes	1	0	1	9	9
78	yes	1	1	1	0	1
79	yes	1	1	1	1	1
80	yes	1	1	1	0	1
81	yes	1	0	1	1	1
82	yes	1	1	1	1	1
83	yes	1	1	1	0	1
84	yes	1	0	0	1	1
85	yes	1	1	1	1	1

BA= break apart, NOS= not otherwise specified.

Table S3. Positive and negative predictive value of PET results.

Table S3A: Positive and negative predictive values of EOT PET-CT scan for progression within 1 year.

	Progression within 1 year	
CMR at EOT	no	yes
no	5	22
yes	51	4

PPV 81%
NPV 93%

Table S3B: Positive and negative predictive values of interim PET-CT scan for EOT result. Response was simplified to "CMR" versus "no-CMR".

	CMR EOT *	
CMR at interim	no	yes
no	15	10
yes	12	45

PPV 60%
NPV 79%

*2 patients missing PET (due to progression) and 1 patient missing EOT PET-CT (off-protocol due to toxicity) were counted as failures.

		FISH			GEP	Classification	
MUM1 0=negative, (<40%), 1=positive (>40%), 9=not available	GCB/ non GCB, 9=not available	MYC-BA 0=neg 1=pos	BCL2-BA 0=neg 1=pos	BCL6-BA 0=neg 1=pos	ABC/GCB/ undeterminate, 9=not available	WHO 2008	WHO 2017
9	9	1	0	0	ABC	DLBCL	DLBCL
1	non-GCB	1	0	0	ABC	dd DLBCL or BCL-U	dd DLBCL or HGBCL, NOS
0	GCB	1	9	9	9	DLBCL	dd DLBCL or HGBCL, DH/TH
1	non-GCB	1	0	1	ABC	DLBCL	HGBCL, DH/TH
1	GCB	1	0	0	GCB	DLBCL	DLBCL
nd	GCB	1	1	0	poor quality	BCL-U	HGBCL, DH/TH
0	GCB	1	0	1	unclassified	DLBCL	HGBCL, DH/TH
0	GCB	1	0	0	GCB	DLBCL	DLBCL
9	GCB	1	1	1	9	DLBCL	HGBCL, DH/TH

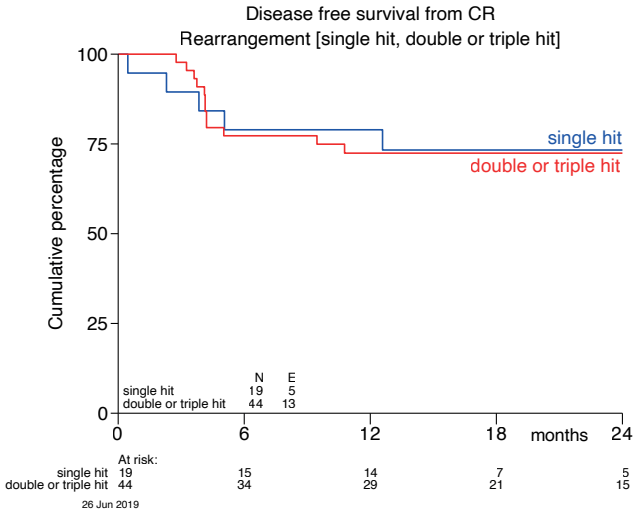


Figure S1A

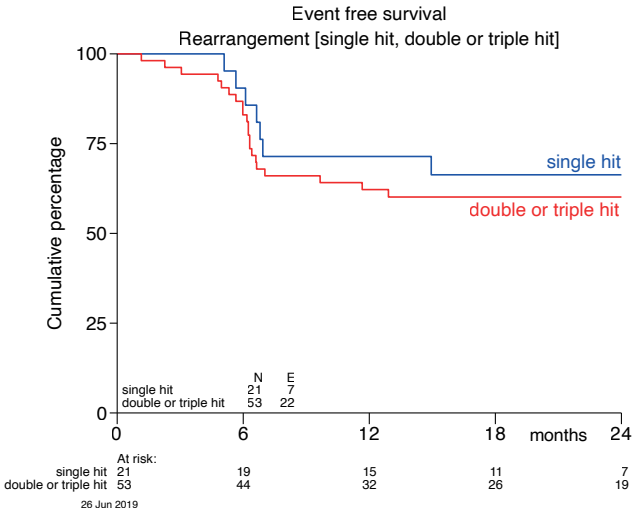


Figure S1B

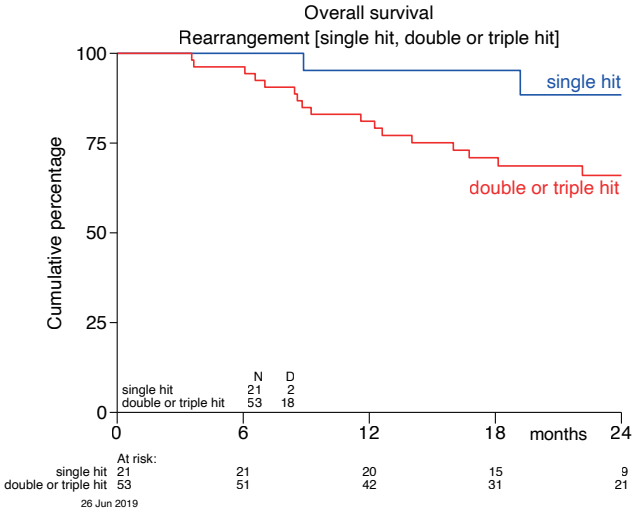


Figure S1C

Figure S1. Survival according to rearrangement status.

Figure S1A: Disease Free Survival of SH vs DH/TH *MYC*+ LBCL patients revealed no significant differences. Figure S1B: Event Free Survival of SH vs DH/TH *MYC*+ LBCL patients revealed no significant differences. Figure S1C: Overall survival analysis indicated that DH/TH patients had a tendency for higher risk of death compared to SH patients (HR 4.18, p=0.055; 95% CI 0.97-18.02). Eight patients with unknown *BCL2* and *BCL6* rearrangement were not included in this analysis.

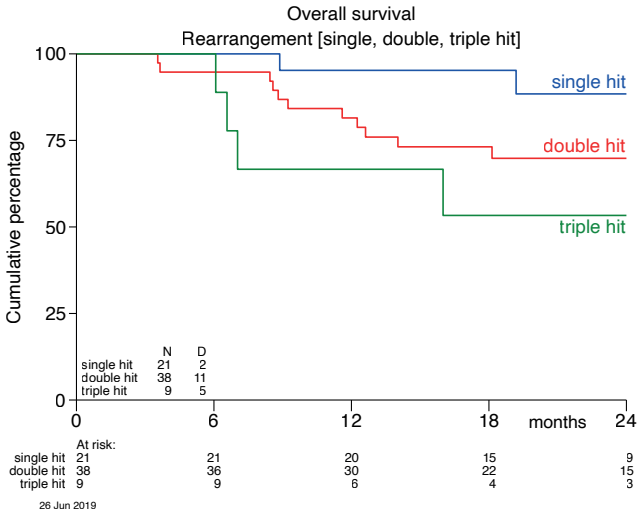


Figure S2A

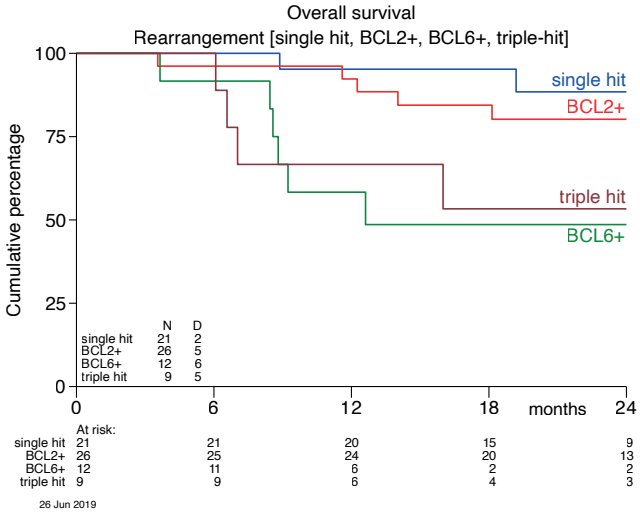


Figure S2B

Figure S2. Overall survival according to rearrangement status (SH vs DH vs TH).

Figure S2A: Overall survival of *MYC*+ LBCL patients according rearrangement status SH vs DH vs TH revealed no significant differences. Figure S2B: Overall survival of *MYC*+ LBCL patients according rearrangement status SH vs DH MYC/BCL2 vs DH MYC/BCL6 vs TH revealed no significant differences. Eight patients with unknown *BCL2* and *BCL6* rearrangement were not included in this analysis.

CHAPTER 8

Optimal timing and criteria of interim PET in DLBCL: a comparative study of 1692 patients

Jacoba J. Eertink
Martijn W. Heymans
Andreas Hüttmann
Stefan Müller
Sally F. Barrington
Robert Carr
Tamas Györke
Emanuele Zucca
Lale Kostakoglu
Stefano Fanti
Simone Pieplenbosch
Otto S. Hoekstra
Henrica C.W. de Vet
Coreline N. Burggraaff
Ulrich Dührsen
Christine Schmitz
Pieternella J. Lugtenburg
N. George Mikhaeel
Sandor Czibor
Luca Ceriani
Martin Hutchings
Annika Loft
Sanne E. Wiegers
Ronald Boellaard
Josée M. Zijlstra

Blood advances. 2021; 5: 2375-2384.

Abstract

Interim ^{18}F-fluorodeoxyglucose positron emission tomography (Interim-^{18}F-FDG-PET, hereafter I-PET) has the potential to guide treatment of patients with diffuse large B-cell lymphoma (DLBCL) if the prognostic value is known. The aim of this study was to determine the optimal timing and response criteria for evaluating prognosis with I-PET in DLBCL. Individual patient data from 1692 patients with de novo DLBCL were combined and scans were harmonized. I-PET was performed at various time points during treatment with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) therapy. Scans were interpreted using the Deauville score (DS) and change in maximum standardized uptake value (ΔSUVmax). Multilevel Cox proportional hazards models corrected for International Prognostic Index (IPI) score were used to study the effects of timing and response criteria on 2-year progression-free survival (PFS). I-PET after 2 cycles (I-PET2) and I-PET4 significantly discriminated good responders from poor responders, with the highest hazard ratios (HRs) for I-PET4. Multivariable HRs for a PET-positive result at I-PET2 and I-PET4 were 1.71 and 2.95 using DS4-5, 4.91 and 6.20 using DS5, and 2.93 and 4.65 using ΔSUVmax, respectively. ΔSUVmax identified a larger proportion of poor responders than DS5 did. For all criteria, the negative predictive value was >80%, and positive predictive values ranged from 30% to 70% at I-PET2 and I-PET4. Unlike I-PET1, I-PET3 discriminated good responders from poor responders using DS4-5 and DS5 thresholds (HRs, 2.94 and 4.67, respectively). I-PET2 and I-PET4 predict good response equally during R-CHOP therapy in DLBCL. Optimal timing and response criteria depend on the clinical context. Good response at I-PET2 is suggested for de-escalation trials, and poor response using ΔSUVmax at I-PET4 is suggested for randomized trials that are evaluating new therapies.

Key points

- Good response at I-PET2 and poor response at I-PET4 may qualify for randomized trials evaluating treatment de-escalation or new therapies.
- The best response criterion at I-PET was ΔSUVmax, with higher discriminative power and predictive values than currently used DS4-5 criteria.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is characterized by an aggressive clinical course. Standard first-line therapy consists of rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP). Up to one-third of patients relapse or fail to achieve complete remission. These patients have a poor prognosis and low response rates to salvage treatment [1,2]. Early identification of patients with poor prognosis is an important step toward testing alternative treatment options. Interim ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) can be used to differentiate good and poor responders during treatment to modify therapy and improve outcome for poor responders and de-escalate treatment for good responders [3].

Current tools for predicting outcome in DLBCL, such as the International Prognostic Index (IPI) [4], which captures pretreatment clinicopathologic features, have limited precision. Many studies have investigated the potential of metabolic imaging with ^{18}F-FDG PET in the context of treatment evaluation using end-of-treatment PET/CT scans [5,6] or of prediction of therapy success using on-treatment (interim ^{18}F-FDG PET [I-PET]) evaluation. End-of-treatment PET is the current clinical standard, but the impact of I-PET is less clear. A recent systematic review and meta-analysis concluded that I-PET has predictive value in DLBCL patients, but small sample sizes, use of different response criteria, different timings, and other methodologic variations among studies hamper the ability to draw firm conclusions [3].

Analyzing the individual patient data (IPD) from various studies made it possible to re-analyze clinical data and ^{18}F-FDG PET scans, which reduced variability and thus allows for a statistically more robust analysis of prediction or prognosis, subgroup analyses, and identification of potential effect modifiers. To this end, we established the PETRA database (www.petralymphoma.org), for collecting individual patient data and PET/CT scans from high quality international clinical studies. The aim of this IPD meta-analysis was to determine the optimal timing and PET response criteria for I-PET in DLBCL.

Materials and methods

Database

This IPD meta-analysis included 1692 patients with de novo DLBCL from the PETRA database with I-PET scans after 1 to 4 cycles of chemotherapy for those who were treated with R-CHOP. This database was established by the PETRA consortium and contains patient-level data for 2539 patients with non-Hodgkin lymphoma who were enrolled in the Bologna [7], HOVON-84 [8], IAEA [9], GSTT15 [10], NCRI [11], Nordic-US Intergroup [12], PETAL [13], and SAKK 38/07 [14] studies.

The following are eligibility criteria for the PETRA database: adult patients age 18 years or older who had first-line treatment for non-Hodgkin lymphoma and had received an ^{18}F-FDG I-PET scan. The trial had to have a prospective design or retrospective design with consecutive patients, at least 40 patients with progression-free survival (PFS) and overall survival (OS) data, and a full-ring PET system. Individual study protocols were approved by local institutional review boards, and written informed consent was provided by all participants in each study. After signing a data sharing agreement, data were made available to PETRA. Data remained the property of contributing investigators. The use of all data within the PETRA imaging database has been approved by the institutional review board of the Vrije Universiteit Medical Center (JR/20140414).

Data collection, harmonization, and re-analysis

Each study was checked for missing data and for data that were consistent with those in published reports. Trial investigators were contacted about discrepancies or missing information. Patient numbers were recoded to PETRA identification numbers that consisted of study-specific and patient-specific parts. Patient data from original studies were merged into an online database and harmonized using the PETRA coding for all studies. All PET images were given a new pseudonym and were uploaded to an online database [15].

Survival rates were recalculated by using the date of baseline PET at the start of follow-up. If the date of baseline PET was not available, we used the baseline CT date or date of diagnosis (supplemental Data). Missing variables were completed, whenever possible, by reviewing scans. Scans were reviewed to provide Deauville

scores (DS) for the IAEA study, to assess extranodal involvement to determine the IPI score for the HOVON-84 study, and to measure the change in maximum standardized uptake value (ΔSUVmax) for the Bologna, IAEA, HOVON-84, and SAKK studies [7-9,14]. Follow-up was updated for the GSTT15 and Bologna studies [7,10]. Patients were divided into 4 prognostic IPI score subgroups (low, low-intermediate, high-intermediate, and high) [4].

PET/CT review

All I-PET scans were reviewed according to the 5-point DS [5,6] by individual PETRA study groups. To harmonize DS5 scores between studies, we re-analyzed all DS5 patients, assigning DS5 if the lesional SUVmax exceeded 3 times the liver SUVmax and/or in the case of new lymphomatous lesions. We applied 2 different cutoffs for PET response assigning DS4-5, as recommended in international guidelines, and DS5 as PET-positive, respectively. Patients with a negative PET were considered to have a complete metabolic response. We also validated alternative criteria: ΔSUVmax between baseline and I-PET assessing response as ≥66% SUVmax reduction for I-PET scans after 1, 2, or 3 cycles [16], and ≥70% SUVmax reduction after 4 cycles of therapy [17], respectively.

Quality assessment

Two independent researchers (J.J.E., C.N.B., or H.C.W.d.V.) rated the quality of included studies by scoring all relevant items with the Quality In Prognosis Studies (QUIPS) tool (ie, a risk of bias tool for prognosis studies) [18]. Quality was rated as high, low, or unclear risk of bias on the following aspects: study participation, study attrition, prognostic factor measurement, and outcome measurement. Differences in quality assessments were resolved by consensus.

Statistical analysis

A statistical plan was created before data were pooled and statistically analyzed. The primary end point of this study was 2-year PFS, defined as time from baseline PET to progression, relapse, or death as a result of any cause. Secondary end points were 2-year time to progression (TTP), defined as time from baseline PET to progression at which time patients dying within 2 years were censored, and 2-year OS, defined as time from baseline PET to death. Patients still alive were censored at date of last contact or the end of study period.

Survival curves of individual studies were obtained with Kaplan-Meier analyses for PFS. We used multivariable Cox proportional hazards models and multilevel Cox regression models to study the effects of timing and PET response criteria on PFS, TTP, and OS. Multilevel analyses were used to account for clustering of data within studies. To adjust for different inclusion criteria applied in the original studies, survival curves were corrected for IPI score. Corresponding hazard ratios (HRs) and their 95% confidence intervals (CIs) were obtained by Cox regression. For each variable included in the Cox regression model (timing I-PET, PET response criteria, and IPI score), the assumption of proportional hazards was assessed on the basis of Schoenfeld residuals [19], which was not violated. Univariable HRs were calculated for the DS4-5 response criterion and IPI score. To compare the discriminative ability of IPI score (low and low-intermediate vs high-intermediate and high) and age-adjusted IPI score (aaIPI; low and low-intermediate vs high-intermediate and high), univariable HRs of both prognostic scores in patients age 60 years or younger were calculated.

Diagnostic measures (positive predictive value [PPV] and negative predictive value [NPV]) were estimated from the Cox regression model probabilities of the event outcome (PPV) or survival probabilities (NPV) stratified for I-PET timing for DS4-5, DS5, and ΔSUVmax response criteria on 2-year PFS, TTP, and OS. Statistical analysis was performed using IBM SPSS version 24 and R version 3.6.3. A *P* value of < .05 was considered statistically significant.

Results

Patients' characteristics

There were 2122 treatment-naïve DLBCL patients in the PETRA database, and 1692 of them were included in this IPD analysis (Figure 1). Patients who were treated with regimens other than R-CHOP (*n* = 107), who were ineligible for the original study (*n* = 101), or who had an I-PET after 5 cycles (*n* = 11) were ineligible for this study. To avoid duplication, we excluded the Bologna patients from the IAEA study (*n* = 40). Other reasons for exclusion were missing I-PET results (*n* = 99), survival data (*n* = 38), or clinical data (*n* = 32) and were younger than age 18 years (*n* = 2). Descriptive statistics for the main patient and I-PET outcome variables are presented in supplemental Table 1. There was low risk of bias for individual studies

(supplemental Table 2). After correcting the survival curves for IPI scores, studies had similar 2-year PFS, 2-year TTP, and 2-year OS survival rates (using the largest study [PETAL] as the reference; Figure 2; supplemental Table 3).

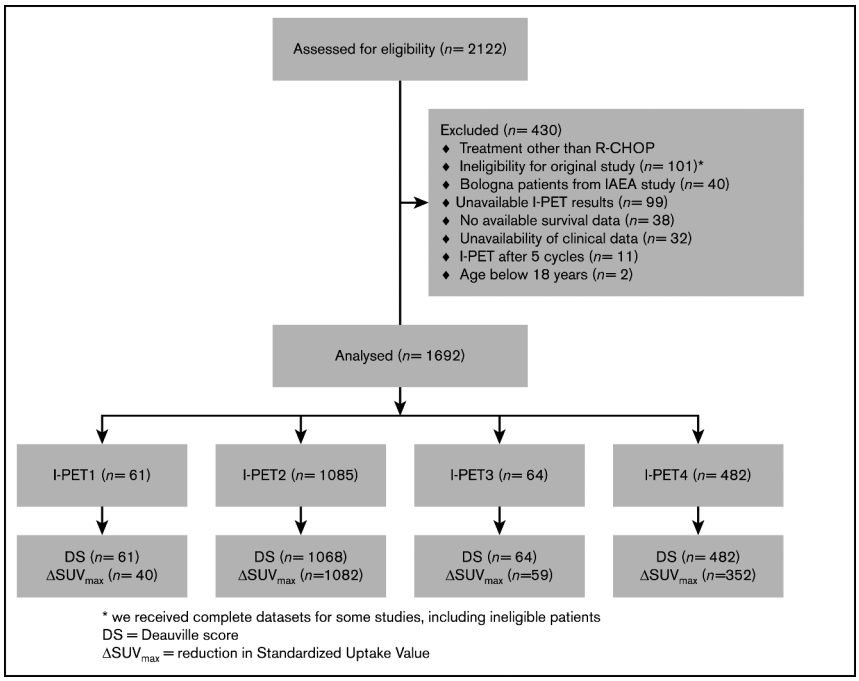


Figure 1. CONSORT diagram of patient inclusion.

I-PET response criteria

In total, 1085 patients had scans after 2 cycles of I-PET (I-PET2) scans and 482 had I-PET4 scans. There were relatively few patients with I-PET1 and I-PET3 scans (Figure 1). The prevalence of positive I-PET scans was lower when I-PET scans were performed later during treatment, independent of PET response criteria (Table 1). A total of 1675 patients were assessed according to DS, and 1533 patients also had a baseline scan available to calculate the ΔSUVmax. There were no differences in baseline characteristics between these groups (supplemental Table 4). ΔSUVmax identified a larger proportion of poor responders at I-PET scanning than did DS5 (supplemental Table 5).

In a multivariable Cox regression analysis, IPI score and I-PET scans (all PET response criteria) were independent predictors of outcome. The univariable HR

of I-PET scans using DS4-5 response criteria was 2.20 (95% CI, 1.79-2.69). The univariable HR of IPI scores for the entire study population was 2.91 (95% CI, 2.34-3.61). When selecting patients age 60 years or younger, the univariable HR of IPI score was 2.46 (95% CI, 1.76-3.44) vs 2.58 (95% CI, 1.83-3.66) for the aaIPI prognostic score.

I-PET2 and I-PET4 significantly discriminated good responders and poor responders (Table 1), with higher HRs for I-PET4 for a PET positive result using DS4-5–positive (HR, 1.71 and 2.95; Figure 3), DS5-positive (HR, 4.91 and 6.20; Figure 4), and ΔSUVmax (HR, 2.93 and 4.65; Figure 5) criteria. Unlike I-PET1, I-PET3 discriminated good responders and poor responders using DS4-5 and DS5 PET response criteria (HR, 2.94 and 4.67) but not ΔSUVmax.

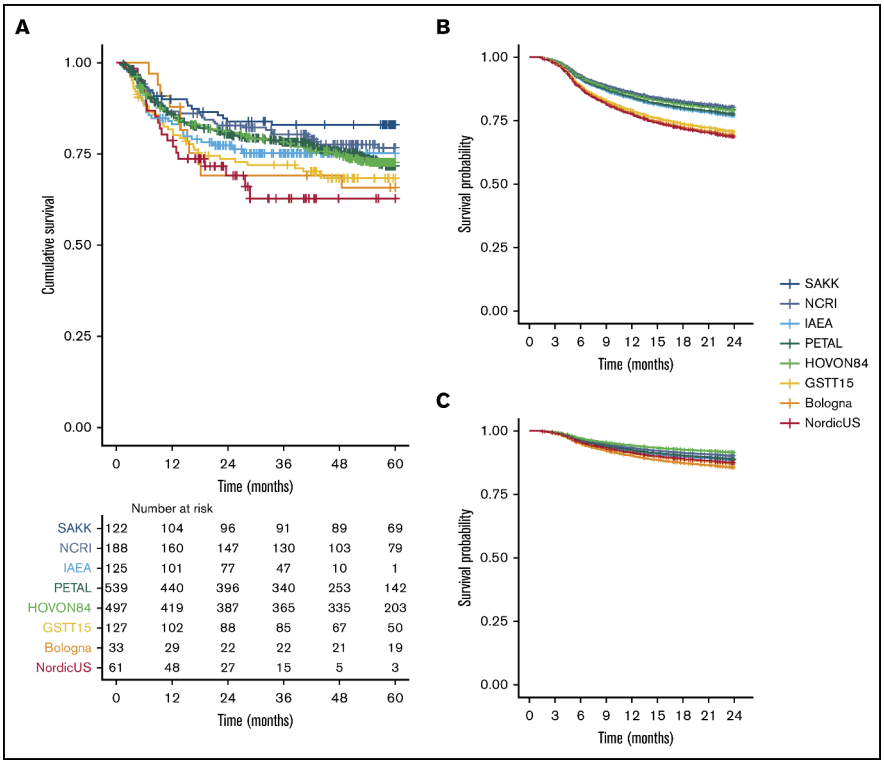


Figure 2. PFS from day of baseline scan for individual studies included in our analysis.

(A) Kaplan-Meier survival curve for 5-year PFS for all studies. (B) Uncorrected Cox regression 2-year PFS for all studies. (C) Cox regression corrected for IPI score for 2-year PFS for all individual studies.

Table 1. Percentage of PET-positive scans and HRs of I-PET using DS4-5 or DS5 or ΔSUVmax to assign a PET-positive result with 2-year PFS as outcome.

	DS1-3 vs DS4-5		DS1-4 vs DS5		ΔSUVmax	
Timing	**I-PET positive**	**HR (95% CI)**	**I-PET positive**	**HR (95% CI)**	**I-PET positive**	**HR (95% CI)**
I-PET1	38 (62.3)	1.22 (0.46-3.20)	12 (19.7)	2.33 (0.88-6.13)	15 (37.5)	1.46 (0.45-4.80)
I-PET2	442 (41.4)	1.71 (1.32-2.22)	60 (5.6)	4.91 (3.46-6.97)	137 (12.7)	2.93 (2.18-3.93)
I-PET3	14 (21.9)	2.94 (1.08-7.96)	4 (6.3)	4.67 (1.52-14.37)	9 (15.3)	2.27 (0.73-7.04)
I-PET4	102 (21.2)	2.95 (1.98-4.40)	24 (5)	6.20 (3.62-10.61)	36 (10.2)	4.65 (2.76-7.83)

Data are presented as n (%), unless otherwise labeled.

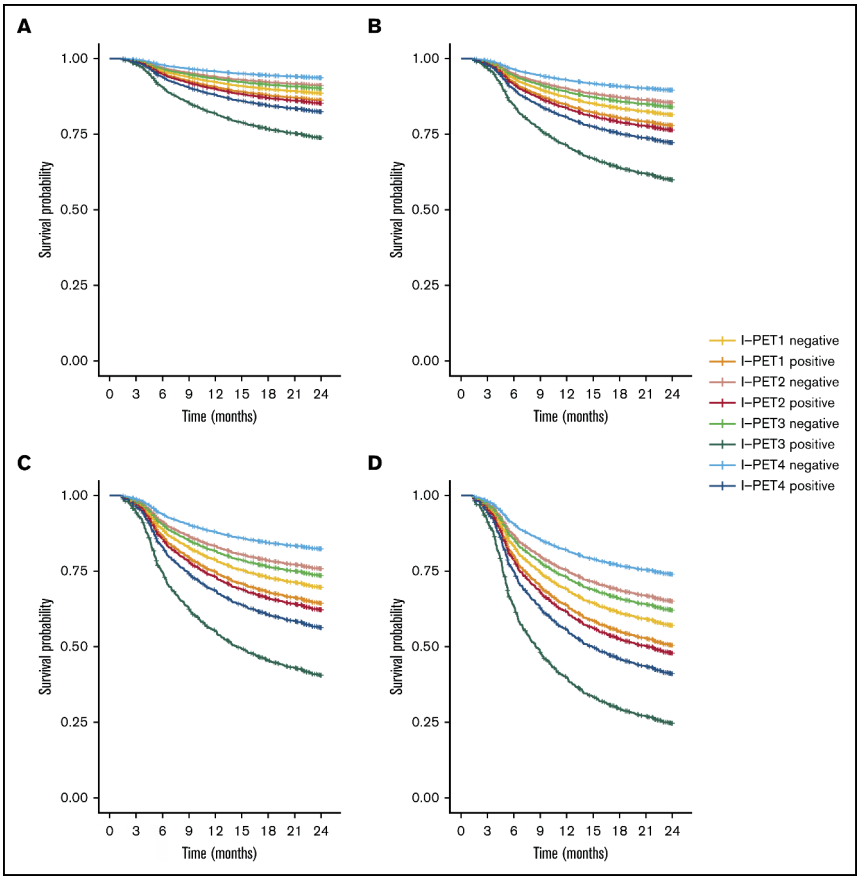


Figure 3. Two-year PFS Cox regression stratified for DS4-5 I-PET-positive patients and timing. Corrected for low-risk (A), low-intermediate risk (B), high-intermediate risk (C), and high-risk (D) IPI groups.

Optimal timing for I-PET scans

HRs were lowest for I-PET1 and increased for later PET scans with highest HRs at I-PET4 for all criteria. HRs at I-PET3 were lower than HRs for I-PET2 using ΔSUVmax and DS5 to define PET response. NPV was high for all criteria at both I-PET 2 and I-PET4 (range, 80.0% to 84.7%; Table 2). The PPV was higher at I-PET4 than at I-PET2 for DS4-5 (42.6% and 30.5%), DS5 (70.0% and 68.5%), and ΔSUVmax (57.4% and 45.7%) criteria.

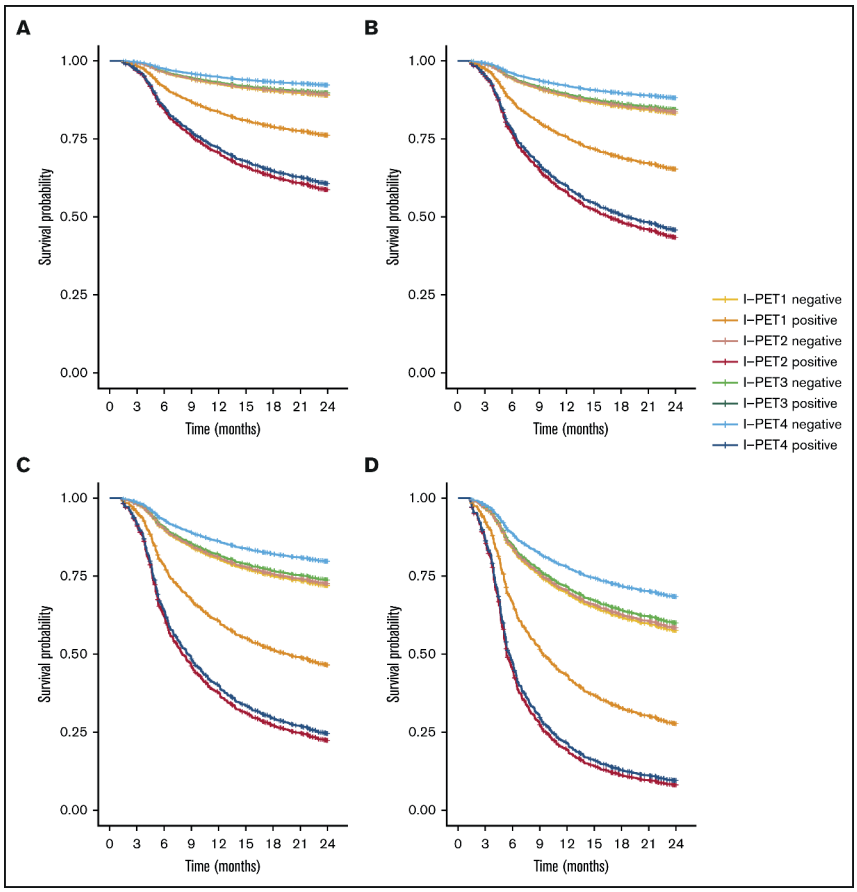


Figure 4. Two-year PFS Cox regression stratified for DS5 I-PET-positive patients and timing.

Corrected for low-risk (A), low-intermediate risk (B), high-intermediate risk (C), and high-risk (D) IPI groups. For all risk groups, I-PET2-positive and I-PET3-positive regression curves are superimposed.

Good responders at I-PET4 had a significantly higher survival compared with good responders at I-PET2 for all PET positivity criteria (DS4-5 negative: HR, 0.70; DS5 negative: HR, 0.74; and ΔSUVmax negative: HR, 0.72; supplemental Table 6). There were no significant differences in PFS between good responders at other time points compared with I-PET2. There were no significant differences between poor responders at all time points compared with I-PET2 (supplemental Table 6). The tables for TTP and OS as outcome parameters are similar (supplemental Tables 7-12).

Discussion

I-PET was predictive in all IPI risk groups in this meta-analysis of individual patient data. PET criteria that applied ΔSUVmax and DS5 positivity discriminate good responders from poor responders better than DS4-5 positivity criteria. But the DS5 criterion identified only a very small number of patients. Performing I-PET scans at later time points during therapy improved patient stratification. Limited data for I-PET1 and I-PET3 timings precluded firm conclusions from being drawn about these time points.

We found a univariable HR of 2.20 (95% CI, 1.79-2.69) for I-PET using DS4-5 positivity criteria, confirming the predictive value of I-PET scans in patients with DLBCL. In this study, the univariable HR was lower than the pooled univariable HR of 3.13 (95% CI, 2.52-3.89) reported in a recent meta-analysis [3]. This difference in HRs can be explained partly because different definitions of outcome parameters and different response criteria were used in various studies. Moreover, we included recent, larger trials (PETAL, HOVON84) that reported lower HRs than the ones included in this meta-analysis. The higher HR using ΔSUVmax positivity criteria is in line with other recent studies, which reported that ΔSUVmax positivity criteria better discriminated poor responders and good responders at I-PET scans compared with currently used DS4-5–positivity criteria [11,20,21].

Our results showed that I-PET scans of patients with DLBCL have an NPV >80% for 2-year PFS, which is in line with previously published results [3,22,23]. In the literature, PPVs at I-PET2 ranged between 37% and 74% using DS4-

5 criteria [3], which were higher than the PPV of 30.5% for DS4-5–positive patients at I-PET2 in this study. This difference can be explained by the fact that 40.3% of the patients with an I-PET scan after 2 cycles were from the PETAL trial, and the PPV of DS4-5 in the PETAL trial was 26.4% (data not shown). Nyilas et al [23] retrospectively included mainly patients with an I-PET after 4 cycles and reported a PPV of 48%, which is slightly higher than our PPV of 42.6% at I-PET4. The PPV of ΔSUVmax–positivity criteria was higher in our study, which confirms the higher PPV using ΔSUVmax positivity criteria in the PETAL trial [24]. For all PET-positivity criteria, PPVs are rather low, but both PPV and NPV are dependent on the prevalence of the outcome. Because the prior probability (ie, prevalence) of progression is 21.9% in our data, it is hard to reach a high PPV. After I-PET scans, the posterior probability increases for poor responders (ie, increase in PPV) and decreases for good responders (ie, high NPV), further stratifying risk groups.

These results show that I-PET scans have the potential to guide risk adapted therapy. By detecting suboptimal response, therapy can be adapted earlier, potentially leading to higher cure rates and lower toxicity. DS5 patients have the worst response and can be identified as early as I-PET2 because PPV at I-PET2 is similar to that at I-PET4. For ΔSUVmax positive and DS4 patients, I-PET4 would be the optimal timing, because the discriminative power is higher at I-PET4. However, I-PET4 is quite late for an I-PET–based strategy, so the importance of a high predictive value should be balanced with the reduced potential for early treatment escalation in the case of ineffective chemotherapy. In clinical trials, I-PET can be used to power new trials that investigate the potential of new drugs for treating patients who have DLBCL and a DS5 at I-PET2 or with a poor response using ΔSUVmax criteria at I-PET4. So far, I-PET2–based treatment escalation has not been effective in DLBCL [13]. When detecting good response from I-PET scans, de-escalation of therapy might be considered. For treatment de-escalation, I-PET2 seems to be the optimal timing. A recent trial showed that treatment de-escalation seems feasible for patients with DLBCL between ages 18 and 60 years who have a favorable prognosis, because treatment with 4 cycles of R-CHOP plus 2 cycles of rituximab was noninferior to 6 cycles of R-CHOP [25]. Moreover, interim-PET–guided treatment in limited-stage nonbulky de novo DLBCL resulted in high survival rates for poor responders and good responders at I-PET [26]. Similar approaches could be considered for all DLBCL patients with a good response at I-PET2.

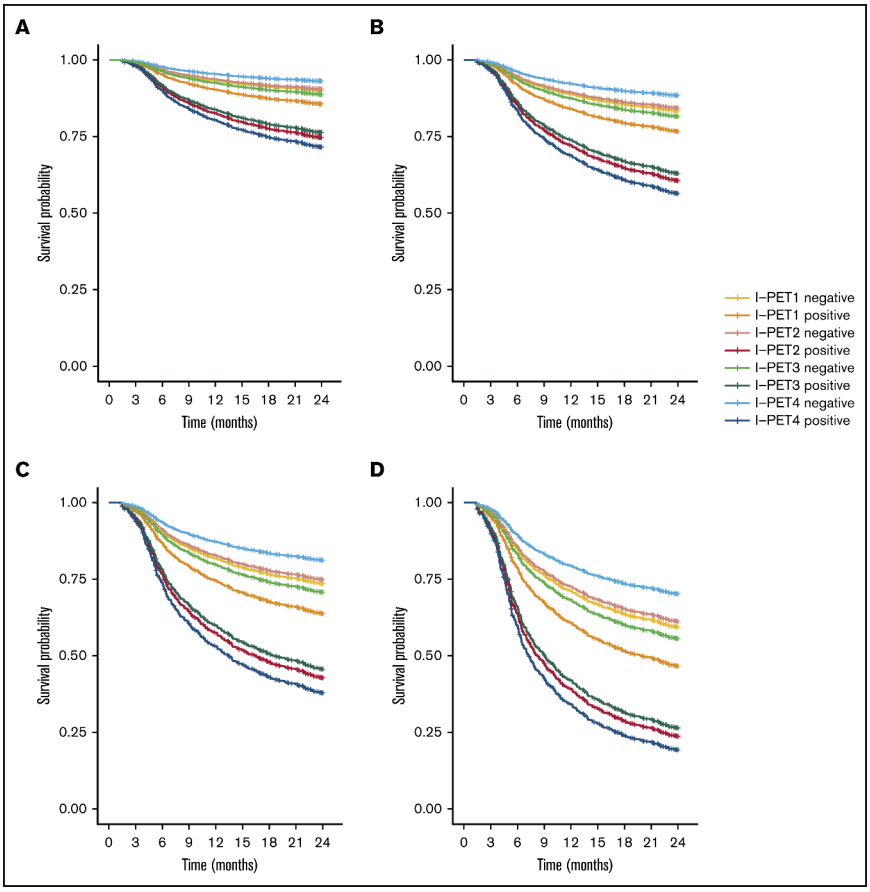


Figure 5. Two-year PFS Cox regression stratified for ΔSUVmax I-PET-positive patients and timing. Corrected for low-risk (A), low-intermediate risk (B), high-intermediate risk (C), and high-risk (D) IPI groups.

This study had several strengths. By collecting individual patient data from high-quality studies that performed an I-PET at multiple time points and by collecting both DS and ΔSUVmax data, our analysis enabled us to determine the optimal timing and make firmer conclusions on optimal criteria for I-PET in DLBCL. Furthermore, survival data were harmonized by re-calculating the follow-up between original studies. All available I-PETs without DS or ΔSUVmax data were re-reviewed. The lack of standardization between I-PET response criteria was overcome by re-classifying DS5 patients on a semi-quantitative basis. This recalculation of variables in the PETRA database allowed for a statistically more

robust analysis of effect modifiers. We were also able to correct for differences in baseline patient characteristics between individual patients. Moreover, there was low risk of bias in our included studies according to QUIPS screening criteria. We decided to truncate survival at 2 years, because most clinically relevant events occur during this period. A recent IPD analysis showed that patients who are alive without progression at 2 years have survival rates similar to those of the age-, sex-, and country-matched population 7 years after this time [27].

Table 2. PPV and NPV using DS4-5, DS5, or ΔSUVmax to assign a PET-positive result at I-PET2 and I-PET4.

	I-PET criteria	**PPV (95% CI)**	**NPV (95% CI)**
I-PET2	DS1-3 vs DS4-5	30.5 (26.2-33.8)	82.9 (80.0-85.7)
	DS1-4 vs DS5	68.5 (56.6-80.3)	80.0 (77.5-82.5)
	ΔSUVmax	45.7 (37.3-54.1)	80.6 (78.1-83.2)
I-PET4	DS1-3 vs DS4-5	42.6 (33.0-52.3)	84.7 (81.1-88.3)
	DS1-4 vs DS5	70.0 (51.7-88.3)	81.5 (77.9-85.0)
	ΔSUVmax	57.4 (41.2-73.5)	82.2 (78.0-86.4)

A limitation of this study was that for some patients, the baseline PET scan was not performed, which precluded calculation of ΔSUVmax. However, this should not bias our results because the DS was not different between the patients with and without a baseline PET/CT scan. We decided to use PFS as the primary outcome parameter because it is widely accepted. However, PFS is affected by age [28]. Outcome of older patients is determined not only by lymphoma but also by age-related comorbidities, adverse treatment effects, and limited life expectancy in general. Note that all findings were consistent when considering TTP and OS instead of PFS.

Future studies should focus on improving the PPV by further stratifying patients into risk groups based on baseline PET characteristics such as metabolic tumor volume [29] and dissemination [30] and by improving the criteria for assigning a PET-positive result at I-PET. In addition, the effect of therapy on I-PET criteria requires further study, because all patients in our analysis were treated with R-CHOP.

In conclusion, the best response criterion at I-PET was ΔSUVmax, which had higher discriminative power and predictive values than DS4-5 criteria. Although

the DS5 criterion had a higher discriminative power than ΔSUVmax, it identified a smaller group of poor responders. The optimal timing for identifying good responders is after 2 cycles. Good response at I-PET2 may qualify as a starting point for de-escalation trials. Poor response at I-PET4 using ΔSUVmax response criteria may work best for randomized trials evaluating new therapy regimens. However, optimal timing and response criteria may vary, depending on the clinical context of the study.

Acknowledgements

The authors thank all the patients who took part in the trials, the collaborating investigators who kindly supplied their data, all data managers, and the central and local review groups who scored the I-PET scans. This work was supported by a grant from the Alpe d'HuZes (#VU 2012-5848) provided by the Dutch Cancer Society. The sponsor had no role in gathering, analyzing, or interpreting the data. S.F.B. acknowledges support from the National Institute for Health Research and Social Care (NIHR) (RP-2-16-07-001). King's College London and University College London Comprehensive Cancer Imaging Center is funded by Cancer Research United Kingdom and Engineering and Physical Sciences Research Council (EPSRC) in association with the Medical Research Council and Department of Health and Social Care. S.F.B. was also supported by the Wellcome/EPSRC Centre for Medical Engineering at King's College London (WT 203148/Z/16/Z). The views expressed are those of the author(s) and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care.

References

1. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28(27):4184-4190.
2. Crump M, Neelapu SS, Farooq U, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. *Blood.* 2017;130(16):1800-1808.
3. Burggraaff CN, de Jong A, Hoekstra OS, et al. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2019;46(1):65-79.
4. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993;329(14):987-994.
5. Cheson BD, Fisher RI, Barrington SF, et al; United Kingdom National Cancer Research Institute. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol.* 2014;32(27):3059-3068.
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(27):3048-3058.
7. Zinzani PL, Gandolfi L, Broccoli A, et al. Midtreatment 18F-fluorodeoxyglucose positron-emission tomography in aggressive non-Hodgkin lymphoma. *Cancer.* 2011;117(5):1010-1018.
8. Lugtenburg PJ, de Nully Brown P, van der Holt B, et al. Rituximab-CHOP with early rituximab intensification for diffuse large B-cell lymphoma: A randomized phase III trial of the HOVON and the Nordic Lymphoma Group (HOVON-84). *J Clin Oncol.* 2020;38(29):3377-3387.
9. Carr R, Fanti S, Paez D, et al; IAEA Lymphoma Study Group. Prospective international cohort study demonstrates inability of interim PET to predict treatment failure in diffuse large B-cell lymphoma. *J Nucl Med.* 2014;55(12):1936-1944.
10. 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(7):1209-1219.
11. Mikhaeel NG, Cunningham D, Counsell N, et al. FDG-PET/CT after two cycles of R-CHOP in DLBCL predicts complete remission but has limited value in identifying patients with poor outcome - final result of a UK National Cancer Research Institute prospective study. *Br J Haematol.* 2021;192(3):504-513.
12. Mylam KJ, Kostakoglu L, Hutchings M, et al. (18)F-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(7):2005-2012.
13. Dührsen U, Müller S, Hertenstein B, et al; PETAL Trial Investigators. Positron emission tomography-guided therapy of aggressive non-Hodgkin lymphomas (PETAL): A multicenter, randomized phase III trial. *J Clin Oncol.* 2018;36(20):2024-2034.
14. 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(23):2523-2529.
15. Health-RI. Online database. 2020. Available at: https://www.health-ri.nl/services?portfolio57&research_process5All.
16. 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(10):1626-1632.
17. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al; Groupe d'etude 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(1):37-43.
18. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med.* 2006;144(6):427-437.
19. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika.* 1982;69(1):239-241.

20. Schöder 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(25):2224-2234.
21. Rekowski J, Hüttmann A, Schmitz C, et al. Interim PET evaluation in diffuse large B-cell lymphoma using published recommendations: Comparison of the Deauville 5-point scale and the DSUV max method. *J Nucl Med.* 2021;62(1):37-42.
22. Györke T, Carr R, Cerci JJ, et al. Combined visual and semiquantitative evaluation improves outcome prediction by early mid-treatment 18 F-FDG PET in diffuse large B-cell lymphoma. *J Nucl Med.* 2020;61(7):999-1005.
23. Nyilas R, Farkas B, Bicsko RR, et al. Interim PET/CT in diffuse large B-cell lymphoma may facilitate identification of good-prognosis patients among IPI-stratified patients. *Int J Hematol.* 2019;110(3):331-339.
24. Kurch L, Andreas H, Georgi TW, et al. Interim positron emission tomography in diffuse large B-cell lymphoma. *J Nucl Med.* 2020;jnumed.120.255034.
25. Poeschel V, Held G, Ziepert M, et al; German Lymphoma Alliance. Four versus six cycles of CHOP chemotherapy in combination with six applications of rituximab in patients with aggressive B-cell lymphoma with favourable prognosis (FLYER): a randomised, phase 3, non-inferiority trial. *Lancet.* 2019; 394(10216):2271-2281.
26. Persky DO, Li H, Stephens DM, et al. Positron emission tomography-directed therapy for patients with limited-stage diffuse large B-cell lymphoma: Results of Intergroup National Clinical Trials Network Study S1001. *J Clin Oncol.* 2020;38(26):3003-3011.
27. Maurer MJ, Habermann TM, Shi Q, et al. Progression-free survival at 24 months (PFS24) and subsequent outcome for patients with diffuse large B-cell lymphoma (DLBCL) enrolled on randomized clinical trials. *Ann Oncol.* 2018;29(8):1822-1827.
28. 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.
29. Barrington SF, Meignan M. Time to prepare for risk adaptation in lymphoma by standardizing measurement of metabolic tumor burden. *J Nucl Med.* 2019;60(8):1096-1102.
30. 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(1):40-45.

Supplemental materials

1. Start follow-up

For 7 out of 8 studies date of baseline PET was available, whereas date of diagnosis was available for 6 out of 8 studies. Date of baseline PET was not available for all patients of the NordicUS trial (n=61) and 123 patients of the HOVON84 trial, for those patients, we used date of imaging evaluation.

2. Supplemental tables

Supplemental Table 1. Baseline characteristics and I-PET results of all evaluable patients from individual studies.

	Total (n=1692)	Bologna (n=33)[1]	GSTT15 (n=127)[2]	HOVON84 (n=497)[3]	IAEA (n=125)[4]	NCRI (n=188)[5]	Nordic US (n=61)[6]	PETAL (n=539)[7]	SAKK (n=122)[8]
Study design		Retrospective	Retrospective	Prospective	Prospective	Prospective	Prospective	Prospective	Prospective
Age (med (iqr))	62 (51-70)	54 (45-65)	57 (45-69)	65 (56-72)	57 (47-65)	61 (48-68)	65 (59-73)	61 (50-69)	59 (50-67)
Female sex (%)	778 (46.0)	15 (45.5)	66 (52)	236 (47.5)	58 (46.4)	75 (39.9)	25 (41)	246 (45.6)	57 (46.7)
Ann Arbor Stage (%)									
1	161 (9.5)	2 (6.1)	17 (13.4)		13 (10.4)	10 (5.3)		105 (19.5)	14 (11.5)
2	380 (22.5)	5 (15.2)	27 (21.3)	86 (17.3)	31 (24.8)	68 (36.2)	5 (8.2)	118 (21.9)	40 (32.8)
3	378 (22.3)	6 (18.2)	13 (10.2)	124 (24.9)	27 (21.6)	47 (25)	21 (34.4)	110 (20.4)	30 (24.6)
4	772 (45.6)	20 (60.6)	70 (55.1)	287 (57.7)	54 (43.2)	63 (33.5)	35 (57.4)	205 (38.0)	38 (31.1)
missing	1 (0.06)							1 (0.2)	
IPI (%)									
Low	513 (30.3)	4 (12.1)	44 (34.6)	87 (17.5)	52 (41.6)	70 (37.2)	5 (8.2)	192 (35.6)	59 (48.4)
Low-intermediate	412 (24.3)	14 (42.4)	15 (11.8)	128 (25.8)	27 (21.6)	40 (21.3)	19 (31.1)	142 (26.3)	27 (22.1)
High-intermediate	434 (25.7)	11 (33.3)	33 (26.0)	163 (32.8)	20 (16.0)	51 (27.1)	19 (31.1)	118 (21.9)	19 (15.6)
High	330 (19.5)	4 (12.1)	35 (27.6)	119 (23.9)	26 (20.8)	25 (13.3)	18 (29.5)	86 (16.0)	17 (13.9)
missing	3 (0.2)					2 (1.1)		1 (0.2)	
Therapy (%)									
R-CHOP14	785 (46.4)		2 (1.6)	245 (49.3)	19 (15.2)	30 (16.0)	50 (82.0)	317 (58.8)	122 (100)
R-CHOP14+ 2R	222 (13.1)							222 (41.2)	
RR-CHOP14	252 (14.9)			252 (50.7)					
R-CHOP21	425 (25.1)	33 (100)	122 (96.1)		101 (80.8)	158 (84.0)	11 (18.0)		
R-CHOP (not specified)	8 (0.5)		3 (2.4)		5 (4)				
I-PET after cycle (%)									
1	61 (3.6)						61 (100)		
2	1085 (64.1)		127 (100)	2 (0.4)	107 (85.6)	188 (100)		539 (100)	122 (100)
3	64 (3.8)	33 (100)		13 (2.6)	18 (14.4)				
4	482 (28.5)			482 (97)					

Supplemental Table 1. Continued

	Total (n=1692)	Bologna (n=33)[1]	GSTT15 (n=127)[2]	HOVON84 (n=497)[3]	IAEA (n=125)[4]	NCRI (n=188)[5]	Nordic US (n=61)[6]	PETAL (n=539)[7]	SAKK (n=122)[8]
Deauville score (%)									
1	378 (22.3)		31 (29)	174 (35)	35 (28)	34 (18.1)	12 (19.7)	87 (16.1)	5 (4.1)
2	315 (18.6)	22 (66.7)	12 (11.2)	110 (22.1)	25 (20)	39 (20.7)	5 (8.2)	74 (13.7)	28 (23)
3	386 (22.8)	6 (18.2)	28 (26.2)	107 (21.5)	34 (27.2)	46 (24.5)	6 (9.8)	126 (23.4)	33 (27)
4	496 (29.3)	1 (3)	43 (40.2)	82 (16.5)	24 (19.2)	56 (29.8)	26 (42.6)	214 (39.7)	50 (41)
5	100 (5.9)	4 (12.1)	12 (12.1)	24 (4.8)	7 (5.6)	13 (6.9)	12 (19.7)	21 (3.9)	6 (4.9)
Missing	17 (1.0)							17 (3.2)	
ΔSUVmax (%)									
≥66/70%	1336 (79.0)	26 (78.8)	107 (84.3)	324 (65.2)	110 (88)	167 (88.8)	25 (41)	470 (87.2)	107 (87.7)
<66/70%	197 (11.6)	7 (21.2)	20 (15.7)	38 (7.6)	15 (12)	20 (10.6)	15 (24.6)	68 (12.6)	14 (11.5)
missing	159 (9.4)			135 (27.2)		1 (0.5)	21 (34.4)	1 (0.2)	1 (0.8)

Abbreviations: I-PET: interim PET, IPI: International Prognostic Index, R: rituximab, RR-CHOP14: R-CHOP with rituximab intensification, ΔSUVmax: reduction in Standardized Uptake Value.

Supplemental Table 2. Risk of bias assessment using the 'Quality Assessment in Prognostic Studies' (QUIPS) tool.

	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Overall risk of bias
Bologna	Moderate	Low	Low	Low	Low
GSTT15	Low	Low	Low	Low	Low
HOVON84	Low	Low	Low	Low	Low
IAEA	Low	Low	Low	Low	Low
NCRI	Low	Low	Low	Low	Low
NordicUS	Low	Low	Low	Low	Low
PETAL	Low	Low	Low	Low	Low
SAKK	Low	Low	Low	Low	Low

Supplemental Table 3. Differences in PFS, TTP and OS of individual studies using the PETAL study as a reference.

Study ID	PFS		TTP		OS	
	HR (95% CI)	***P***	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
PETAL (reference)	1	1	1	1	1	1
Bologna	1.30 (0.70-2.42)	0.40	0.74 (0.30-1.83)	0.52	1.28 (0.59-2.77)	0.54
GSTT15	1.15 (0.80-1.67)	0.45	1.35 (0.91-2.00)	0.14	0.99 (0.62-1.58)	0.96
HOVON84	0.74 (0.57-0.97)	0.03	0.82 (0.61-1.09)	0.17	0.75 (0.55-1.04)	0.08
IAEA	1.03 (0.68-1.56)	0.88	1.32 (0.86-2.01)	0.20	0.95 (0.57-1.59)	0.84
NCRI	0.87 (0.60-1.26)	0.46	0.83 (0.54-1.27)	0.38	0.93 (0.60-1.45)	0.76
NordicUS	1.13(0.70-1.84)	0.62	1.21 (0.71-2.06)	0.48	1.45 (0.86-2.46)	0.16
SAKK	0.98 (0.64-1.52)	0.94	1.00 (0.62-1.63)	0.99	0.69 (0.38-1.27)	0.23

Abbreviations: CI: confidence interval, HR: Hazard Ratio, PFS: progression free survival, TTP: time to progression, OS: overall survival.

Supplemental Table 4. Baseline characteristics and I-PET results stratified for DS and ΔSUVmax availability.

	All patients (n=1692)	**Deauville-only (n=1675)**	**ΔSUVmax only (n=1533)**
Age (med (iqr))	62 (51-70)	62 (51-70)	62 (51-69)
Female sex (%)	778 (46.0)	776 (46.3)	706 (46.1)
Stage (%)			
1	161 (9.5)	157 (9.4)	160 (10.4)
2	380 (22.5)	378 (22.6)	351 (22.9)
3	378 (22.3)	374 (22.3)	329 (21.5)
4	772 (45.6)	765 (45.7)	692 (45.1)
missing	1 (0.1)	1 (0.1)	1 (0.1)
IPI (%)			
Low	513 (30.3)	508 (30.3)	486 (31.7)
Low-intermediate	412 (24.3)	406 (24.2)	365 (23.8)
High-intermediate	434 (25.7)	431 (25.7)	383 (25.0)
High	330 (19.5)	327 (19.5)	296 (19.3)
missing	3 (0.2)	3 (0.2)	3 (0.2)
Deauville score (%)			
1	378 (22.3)	378 (22.6)	314 (20.5)
2	315 (18.6)	315 (18.8)	288 (18.8)
3	386 (22.8)	386 (23.0)	355 (23.2)
4	496 (29.3)	496 (29.6)	471 (30.7)
5	100 (5.9)	100(6.0)	88 (5.7)
missing	17 (1.0)		17 (1.1)
ΔSUVmax (%)			
≥66/70%	1336 (79.0)	1323 (79.0)	1336 (87.1)
<66/70%	197 (11.6)	193 (11.5)	197 (12.9)
missing	159 (9.4)	159 (9.5)	
No of PFS events at 2-year (%)	378 (22.3)	373 (22.3)	342 (22.3)

Abbreviations are explained in Supplemental Table 3.

Supplemental Table 5. Number of PET positive patients for DS4-5, DS5 and ΔSUVmax PET response criteria related to the total number of events for PFS, TTP and OS outcome parameters.

Response criteria	Timing	PFS		TTP		OS	
		No true positives	No events	No true positives	No events	No true positives	No events
DS1-3 vs 4-5	I-PET1	13	19	11	16	11	17
	I-PET2	131	237	109	197	89	152
	I-PET3	6	17	3	11	4	11
	I-PET4	42	100	37	87	35	70
DS1-4 vs 5	I-PET1	6	19	6	16	5	17
	I-PET2	38	237	32	197	32	152
	I-PET3	4	17	1	11	3	11
	I-PET4	16	100	14	87	16	70
ΔSUVmax	I-PET1	5	11	4	8	4	10
	I-PET2	59	240	49	198	47	155
	I-PET3	4	16	1	10	3	10
	I-PET4	19	75	16	65	16	53

Abbreviations: DS: Deauville Score, I-PET interim PET, OS: overall survival, PFS; progression free survival, TTP: time to progression, ΔSUVmax: reduction in Standardized Uptake Value.

Progression Free Survival

Supplemental Table 6. HRs between negative I-PET timings and between positive I-PET timings compared to I-PET2 with 2-year PFS as outcome.

	DS1-3 vs 4-5 (95% CI)	DS1-4 vs 5 (95% CI)	ΔSUVmax (95% CI)
I-PET1 negative	1.31 (0.57-2.98)	1.08 (0.62-1.90)	1.06 (0.47-2.40)
I-PET2 negative	1 (reference)	1 (reference)	1 (reference)
I-PET3 negative	1.11 (0.59-2.07)	1.05 (0.60-1.84)	1.20 (0.67-2.15)
I-PET4 negative	0.70 (0.51-0.97)	0.74 (0.57-0.96)	0.72 (0.53-0.98)
I-PET1 positive	0.93 (0.52-1.64)	0.55 (0.23-1.29)	0.53 (0.21-1.32)
I-PET2 positive	1 (reference)	1 (reference)	1 (reference)
I-PET3 positive	1.90 (0.84-4.31)	1.00 (0.36-2.81)	0.92 (0.34-2.55)
I-PET4 positive	1.21 (0.85-1.71)	0.94 (0.52-1.69)	1.14 (0.68-1.92)

Abbreviations are explained in Supplemental Table 5.

Time to Progression

Supplemental Table 7. HRs for DS4-5, DS5 and ΔSUVmax PET response criteria with 2-year TTP as outcome.

Timing	DS1-3 vs 4-5	DS1-4 vs 5	ΔSUVmax
	HR (95% CI)	HR (95% CI)	HR (95% CI)
1	1.23 (0.43-3.65)	3.03 (1.10-8.36)	1.76 (0.44-7.04)
2	1.72 (1.30-2.28)	5.06 (3.45-7.41)	2.98 (2.15-4.11)
3	2.07 (0.55-7.83)	1.54 (0.20-12.08)	0.76 (0.10-6.00)
4	3.05 (1.99-4.67)	6.44 (3.62-11.44)	4.58 (2.60-8.06)

Abbreviations are explained in Table 1.

Supplemental Table 8. PPV and NPV for DS4-5, DS5 and ΔSUVmax PET response criteria at I-PET2 and I-PET4 with 2-year TTP as outcome.

	I-PET criteria	PPV (95% CI)	NPV (95% CI)
I-PET2	DS1-3 vs 4-5	26.5 (22.3-30.6)	85.5 (82.7-88.2)
	DS1-4 vs 5	63.8 (51.6-76.1)	82.9 (80.6-85.3)
	ΔSUVmax	40.8 (32.5-49.0)	83.6 (81.2-86.0)
I-PET4	DS1-3 vs 4-5	39.4 (29.9-48.8)	86.6 (83.2-90.0)
	DS1-4 vs 5	67.0 (48.2-85.8)	83.6 (80.2-87.0)
	ΔSUVmax	52.5 (36.2-68.8)	84.2 (80.1-88.2)

Abbreviations are explained in Table 2.

Supplemental Table 9. HRs between negative I-PET timings and between positive I-PET timings compared to I-PET2 with 2-year TTP as outcome.

	DS1-3 vs 4-5 (95% CI)	DS1-4 vs 5 (95% CI)	ΔSUVmax (95% CI)
I-PET1 negative	1.31 (0.53-3.26)	1.00 (0.53-1.91)	0.86 (0.32-2.32)
I-PET2 negative	1 (reference)	1 (reference)	1 (reference)
I-PET3 negative	0.97 (0.47-2.00)	0.97 (0.51-1.84)	1.09 (0.55-2.13)
I-PET4 negative	0.73 (0.51-1.03)	0.78 (0.59-1.03)	0.76 (0.55-2.13)
I-PET1 positive	0.94 (0.51-1.76)	0.60 (0.25-1.44)	0.51 (0.18-1.41)
I-PET2 positive	1 (reference)	1 (reference)	1 (reference)
I-PET3 positive	1.17 (0.37-3.68)	0.30 (0.04-2.18)	0.28 (0.04-2.00)
I-PET4 positive	1.29 (0.88-1.87)	0.99 (0.52-1.86)	1.17 (0.67-2.07)

Abbreviations are explained in Supplemental Table 5.

Overall Survival

Supplemental Table 10. HRs for DS4-5, DS5 and ΔSUVmax PET response criteria with 2-year OS as outcome.

Timing	DS1-3 vs 4-5	DS1-4 vs 5	ΔSUVmax
	HR (95% CI)	HR (95% CI)	HR (95% CI)
1	0.95 (0.35-2.57)	1.88 (0.66-5.36)	1.11 (0.31-3.92)
2	1.81 (1.31-2.50)	5.52 (3.72-8.19)	3.70 (2.62-5.21)
3	3.18 (0.93-10.90)	7.20 (1.90-27.24)	3.53 (0.91-13.66)
4	3.68 (2.29-5.89)	7.62 (4.34-13.37)	5.11 (2.84-9.19)

Abbreviations are explained in Table 1.

Supplemental Table 11. PPV and NPV for DS4-5, DS5 and ΔSUVmax PET response criteria at I-PET2 and I-PET4 with 2-year OS as outcome.

	I-PET criteria	PPV (95% CI)	NPV (95% CI)
I-PET2	DS1-3 vs 4-5	20.6 (16.8-24.4)	89.6 (87.3-92.0)
	DS1-4 vs 5	53.6 (40.9-66.3)	87.8 (85.7-89.8)
	ΔSUVmax	35.5 (27.5-43.6)	88.3 (86.3-90.4)
I-PET4	DS1-3 vs 4-5	34.2 (24.9-43.4)	90.7 (87.8-93.6)
	DS1-4 vs 5	60.7 (41.2-80.2)	88.1 (85.1-91.1)
	ΔSUVmax	45.2 (29.0-61.5)	88.2 (84.7-91.8)

Abbreviations are explained in Table 2.

Supplemental Table 12. HRs between negative I-PET timings and between positive I-PET timings compared to I-PET2 with 2-year OS as outcome.

	DS1-3 vs 4-5 (95% CI)	DS1-4 vs 5 (95% CI)	ΔSUVmax (95% CI)
I-PET1 negative	2.18 (0.94-5.06)	1.63 (0.90-2.96)	1.88 (0.83-4.29)
I-PET2 negative	1 (reference)	1 (reference)	1 (reference)
I-PET3 negative	1.18 (0.54-2.58)	1.09 (0.53-2.23)	1.18 (0.55-2.54)
I-PET4 negative	0.69 (0.45-1.04)	0.77 (0.56-1.07)	0.78 (0.54-1.14)
I-PET1 positive	1.15 (0.61-2.15)	0.56 (0.22-1.43)	0.56 (0.20-1.57)
I-PET2 positive	1 (reference)	1 (reference)	1 (reference)
I-PET3 positive	2.07 (0.76-5.66)	1.42 (0.43-4.65)	1.13 (0.35-3.63)
I-PET4 positive	1.39 (0.94-2.07)	1.07 (0.58-1.96)	1.08 (0.61-1.91)

Abbreviations are explained in Supplemental Table 5.

References

1. Zinzani PL, Gandolfi L, Broccoli A, et al: Midtreatment 18F-fluorodeoxyglucose positron-emission tomography in aggressive non-Hodgkin lymphoma. *Cancer.* 117:1010-1018, 2011.
2. Mikhaeel NG, Smith D, Dunn JT, et al: Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL. *Eur J Nucl Med Mol Imaging.* 43:1209-1219, 2016.
3. Lugtenburg PJ, de Nully Brown P, van der Holt B, et al: Rituximab-CHOP With Early Rituximab Intensification for Diffuse Large B-Cell Lymphoma: A Randomized Phase III Trial of the HOVON and the Nordic Lymphoma Group (HOVON-84). *J Clin Oncol.* JCO1903418, 2020.
4. Carr R, Fanti S, Paez D, et al: Prospective international cohort study demonstrates inability of interim PET to predict treatment failure in diffuse large B-cell lymphoma. *J Nucl Med.* 55:1936-1944, 2014.
5. Mikhaeel NG, Cunningham D, Counsell N, et al: FDG-PET/CT after two cycles of RCHOP in DLBCL predicts complete remission but has limited value in identifying patients with poor outcome - final result of a UK National Cancer Research Institute prospective study. *Br J Haematol.* 2020.
6. Mylam KJ, Kostakoglu L, Hutchings M, et al: (18)F-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.* 56:2005-2012, 2015.
7. 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.* 36:2024-2034, 2018.
8. 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.* 33:2523-2529, 2015.

CHAPTER 9

Summary, discussion and future perspectives

Summary

The aim of this thesis was to validate interim [^{18}F]FDG PET as a biomarker of response in diffuse large B-cell lymphoma (DLBCL).

Part I PET as biomarker of response in lymphoma

Chapter 2 provides a review of the accuracy of interim [^{18}F]FDG PET in DLBCL and Hodgkin lymphoma (HL) in clinical practice, including clinical trials with interim [^{18}F]FDG PET adapted therapy, published until 2016 [1]. Based on the UK RAPID and EORTC H10 trials, both investigating a strategy with de-escalation of therapy (+-radiotherapy), it was concluded that de-escalation in HL became a real option in clinical practice [2,3]. In HL patients with advanced stage disease who do not achieve CMR assessed at interim [^{18}F]FDG PET after 2 cycles, escalation from ABVD to the more intensive BEACOPP-escalated chemotherapy regimen seemed promising as shown by the RATHL trial [4]. Thus, for HL the interim [^{18}F]FDG PET became important in clinical practice both for escalation of treatment as well as de-escalation of treatment [5,6]. The Dutch guideline for treatment of HL, updated in 2019, now contains a quality indicator on the response evaluation and adaptation of first-line treatment with interim [^{18}F]FDG PET for both limited and advanced stage disease [7]. In patients with limited stage disease and a positive interim [^{18}F]FDG PET the guideline advises to escalate treatment to 2x BEACOPP-escalated followed by involved node radiotherapy. For patients with advanced stage disease and a positive interim [^{18}F]FDG PET treatment with 4x BEACOPP-escalated and additional radiotherapy on positive lesions at end of treatment. For limited stage disease patients treated with ABVD and a negative interim [^{18}F]FDG PET, treatment will be continued with either radiotherapy or AVD (without bleomycin) combined with radiotherapy, depending on presence of baseline risk factors. For advanced stage disease and a negative interim [^{18}F]FDG PET treatment is continued with either 4x AVD (in case of ABVD treatment) or 2x BEACOPP-escalated (in case of BEACOPP-escalated treatment) [7].

However, in 2016 there was no evidence for the use of interim[^{18}F]FDG PET for treatment de-escalation or escalation in clinical practice for DLBCL yet. Only preliminary published data and data presented in abstract form suggest that

current chemotherapy based escalation strategies did not overcome treatment resistance [8-10].

In **Chapter 3** a systematic review and meta-analysis of the predictive value of visual assessment of interim [^{18}F]FDG PET focusing on DLBCL patients summarizes the results of twenty published and eligible studies without treatment adaptation based on the interim [^{18}F]FDG PET result [11]. We concluded that there is indeed predictive value of interim [^{18}F]FDG PET for 2-year progression-free survival (pooled hazard ratio of 3.13). Negative predictive values exceeded 80%, but positive predictive values (ranging from 20 to 74% between studies) were too low to allow for a risk stratified treatment approach in clinical practice. Additional subgroup analyses on differences in timing of interim [^{18}F]FDG PET, patient subgroups, PET techniques and PET response criteria were limited by lack of information, heterogeneity and small sample sizes.

Based on the 2 chapters in this part we concluded that there is predictive value of interim [^{18}F]FDG PET in both HL and DLBCL. Nowadays, interim [^{18}F]FDG PET indeed plays an important role in the "state of the art" treatment of HL. For DLBCL, the role of interim [^{18}F]FDG PET as a biomarker of response was still unclear.

Part II Technical validation of PET in lymphoma

In **Chapter 4** the interobserver agreement of the Deauville 5-point scale for both interim [^{18}F]FDG PET and end-of-treatment [^{18}F]FDG PET in DLBCL patients included in the HOVON-84 study was assessed [12]. The central review of the study was performed by 2 nuclear medicine physicians (from a pool of 10). We concluded that the overall agreement was 87.7% and 91.7% for interim and end-of-treatment [^{18}F]FDG PET, respectively. The corresponding positive agreement (i.e. the probability that if one reviewer assigns a positive score, a second reviewer scores positive as well) was 73.7% and 76.3%. Based on these results, we recommended to perform dual reads for treatment evaluation in clinical practice (e.g. within multidisciplinary tumor board meetings) and for trials to perform central review procedures. Especially sites of extranodal lymphoma involvement in the gastrointestinal tract, Waldeyer's ring, skeletal system and spleen led to a

relatively large number of discrepancies between reviewers. In these tissues, the local background of [^{18}F]FDG could vary over time due to intercurrent inflammation, healing of pathologic fractures or recent administration of granulocyte colony-stimulating factor. Adjustment for these factors is not mentioned in the current Lugano classification guidelines.

Chapter 5 describes a pilot study that was performed on optimization strategies for a fast and reliable assessment of metabolic tumor volume (MTV) using twelve baseline [^{18}F]FDG PET scans [13]. An automated preselection strategy with a SUV of 4 or higher and a minimum volume of 3ml resulted in an improved interobserver reliability and ease of use compared to individual lesion selection.

Based on the 2 chapters in this part we concluded that a central review is an important prerequisite for use of visual interim [^{18}F]FDG PET criteria in future trials and for potential use in clinical practice dual reads are recommended. Training of nuclear medicine physicians should pay extra attention to the extranodal site evaluation, because these tend to have a relatively large number of discrepancies. Based on these observations, we organized several hands-on workshops on the use of the Lugano classification guidelines for radiologists, hematologists and nuclear medicine physicians in the Netherlands.

For the baseline evaluation of MTV there is still no consensus on the evaluation method that should be used. The pilot study is a first step towards a more automated MTV evaluation strategy.

Part III Clinical validation of PET in lymphoma

Chapter 6a details the interim [^{18}F]FDG PET results in the randomized clinical trial HOVON-84 in 513 newly diagnosed DLBCL patients. **Chapter 6b** describes the results of the original HOVON-84 trial aim of the randomization between standard R-CHOP14 and R-CHOP14 with rituximab intensification in the first 4 cycles. The complete remission rate, failure-free survival, progression-free and overall survival were not improved by the rituximab intensification strategy [14]. The HOVON-84 paper was added as part b of this chapter, because the complete metabolic response was determined by central review of the [^{18}F]FDG PET scans

and it gives important background information for part a. **Chapter 6a** shows that both interim [^{18}F]FDG PET and age-adjusted international prognostic index are independent response biomarkers [15]. The negative predictive value for 2-year progression-free survival of patients with a low/low-intermediate age-adjusted IPI and ΔSUVmax>70% was 93%. The positive predictive value for patients with a high-intermediate/high age-adjusted IPI and ΔSUVmax≤70% was 65%. Besides that, the external validation of semi-quantitative ΔSUVmax criterion outperformed the Deauville 5-point scale in the 2-year progression-free survival prediction (positive predictive value of 53% versus 38%, respectively).

In **Chapter 7** the results of the clinical phase II study HOVON-130 of 82 MYC positive aggressive B-cell lymphoma patients (a subgroup of DLBCL patients with worse prognosis) treated with a combination of R-CHOP and lenalidomide are described [16]. In this study both interim [^{18}F]FDG PET and end-of-treatment [^{18}F]FDG PET scans were performed. The complete metabolic response rate at end-of-treatment was 67% and was comparable (no official head to head comparison possible) to studies applying more intensive chemotherapy regimens. The positive and negative predictive value of the end-of-treatment scan for relapse within 12 months were 81% and 93%, respectively. The observational analysis of interim [^{18}F]FDG PET showed a positive and negative predictive value for the end-of-treatment [^{18}F]FDG PET result of 60% and 79%, respectively. These results do not support the use of interim [^{18}F]FDG PET in this specific patient subgroup.

Chapter 8 contains the main results of the PETRA interim [^{18}F]FDG PET project (KWF/Alpe d'Huzes grant VU 2012-5848) with individual patient data meta-analyses from 1692 DLBCL patients from 8 international studies [17]. Deauville score and ΔSUVmax criteria were compared at different timing of the interim [^{18}F]FDG PET after 2 and after 4 cycles. ΔSUVmax criteria had a higher discriminative power and predictive value for 2-year progression-free survival than the Deauville 5-point scale (with a positivity cut-off for DS 4 and 5). However, the optimal timing and response criteria may vary depending on the clinical context of the study. In general, hazard ratios increased for later timing of the interim [^{18}F]FDG PET scan, i.e. improved patient stratification. The negative predictive values were high (above 80%) for all criteria, for both interim [^{18}F]FDG PET after 2 and after 4 cycles. Generally, the positive predictive values are rather low

for all criteria, probably due to the prevalence of progression of 22% in our data. However, positive predictive values were somewhat higher at interim [^{18}F]FDG PET after 4 cycles compared to 2 cycles (ΔSUVmax 57% vs 46% and Deauville 4-5 43% vs 31%) and were higher for ΔSUVmax compared to the Deauville 4-5 positivity cut-off. A Deauville score of 5 selected the patients with the worst response (about 5% of patients) and can be identified at interim [^{18}F]FDG PET after 2 cycles, with similar positive- and negative predictive values compared to 4 cycles. In the multivariable analyses both IPI score and interim [^{18}F]FDG PET scan (all criteria) were independent predictors of outcome. We concluded that good response (defined as ΔSUVmax ≥66%) after 2 cycles of R-CHOP treatment may qualify for randomized trials evaluating de-escalation of therapy regimens. Poor response (defined as ΔSUVmax <70%) after 4 cycles of R-CHOP treatment may qualify for evaluating new therapies in a randomized trial.

The 3 chapters of this final part of the thesis comprised the clinical validation of interim [^{18}F]FDG PET in 2 HOVON studies and the individual patient data meta-analysis of studies included in the PETRA database. It can be concluded that interim [^{18}F]FDG PET indeed has predictive value for 2-year progression-free survival, 2-year overall survival and 2-year time to progression. These results can be used for future trial designs and power calculations of clinical studies with de-escalation strategies or new therapy regimens for DLBCL patients. However, this is not the case for the subgroup of DLBCL patients with a MYC translocation. A recent study demonstrated that MYC positive patients have more frequent progression during treatment after negative interim [^{18}F]FDG PET assessment and new lesion at sites that were not initially involved compared to MYC negative patients [18]. This is probably due to the aggressive disease characteristics of double (MYC with BCL2 or BCL6) and triple hit (MYC, BCL2 and BCL6) lymphoma patients. Currently, a fluorescent in situ hybridization (FISH) analysis is recommended in all DLBCL patients in the Netherlands in order to detect double or triple hit lymphoma patients and change treatment to DA-EPOCH-R with CNS prophylaxis instead of R-CHOP.

Discussion

Main findings

-For HL the interim [^{18}F]FDG PET is important in clinical practice both for escalation of treatment as well as de-escalation of treatment [5,6].

-Our systematic review showed that for DLBCL there is predictive value of interim [^{18}F]FDG PET for 2-year progression-free survival (pooled hazard ratio of 3.13). Negative predictive values exceeded 80%, but positive predictive values (ranging from 20 to 74% between studies) were too low to allow for a risk stratified treatment approach in clinical practice (**Chapter 3**).

-Interobserver agreement of Deauville scoring was 87.7% and 91.7% for interim and end-of-treatment [^{18}F]FDG PET, respectively, asking for dual reads in clinical practice and central review in research (**Chapter 4**).

-In the HOVON-84 study no added value of rituximab intensification of R-CHOP was found (**Chapter 6b**). Additional PET analyses showed that both interim [^{18}F]FDG PET and age-adjusted international prognostic index are independent response biomarkers [15]. Besides that, the external validation of semi-quantitative ΔSUVmax criterion outperformed the Deauville 5-point scale in the 2-year progression-free survival prediction (**Chapter 6a**)

-In a phase II trial with MYC positive DLBCL patients the predictive value of interim [^{18}F]FDG PET was limited (**Chapter 7**).

-In the individual patient data meta-analyses from 1692 DLBCL patients from eight international studies [17], Deauville score and ΔSUVmax criteria were compared at different timing of the interim [^{18}F]FDG PET after 2 and after 4 cycles (**Chapter 8**). ΔSUVmax criteria had a higher discriminative power and predictive value for 2-year progression-free survival than the Deauville 5-point scale (with a positivity cut-off for DS 4 and 5). The negative predictive values were high (above 80%) for all criteria, for both interim [^{18}F]FDG PET after 2 and after 4 cycles. Positive predictive values were somewhat higher at interim [^{18}F]FDG PET after 4 cycles compared to 2 cycles (ΔSUVmax 57% vs 46% and Deauville 4-5 43% vs 31%) and were higher for ΔSUVmax compared to the Deauville 4-5 positivity cut-off.

In conclusion, good response (defined as ΔSUVmax ≥66%) after 2 cycles of R-CHOP treatment may qualify for randomized trials evaluating de-escalation of therapy regimens. Poor response (defined as ΔSUVmax <70%) after 4 cycles of R-CHOP treatment may qualify for evaluating new therapies in a randomized trial.

Importance of uniform PET criteria

During this research project we performed an individual patient data meta-analysis. As we collected scans and patient data from multiple studies we had to harmonize the data first. It was soon realized that, although most studies used the Deauville criteria for their analyses, these were not applied exactly the same between studies. Mostly this was due to an unclear definition for DS5 (uptake "markedly higher than liver uptake" or "2-3 times above the maximum SUV in the liver"). We solved this by re-analyzing all scans with Deauville 5 in a semi-quantitative way by assigning a Deauville 5 if the lesional SUVmax exceeded 3 times the liver SUVmax and/or in case of new lymphoma lesions. It would be helpful for future analyses and comparisons between studies to adopt a clear and similar criteria definition integrating visual and semi-quantitative assessment.

Strengths and weakness of the current research

By building a strong consortium we were able to perform the individual patient data meta-analyses. One of the strengths of this consortium is that all patient data and scans are stored for future analyses. Future research proposals should be approved by the consortium partners before their data is used in new analyses. The consortium partners should have confidence in the correct use of their valuable data, of course. Therefore, legal policy documents were signed before entering the PETRA consortium. For 2 potential consortium partners we were not able to collaborate due to legal and logistic issues (France) and legal issues (USA). Finally, we could include 8 studies in the meta-analyses. Still we experienced a low number of patients with an interim [^{18}F]FDG PET after 1 and 3 cycles, respectively. Besides that, we did not have enough contrast in treatment strategies to investigate the effect of differences in treatment on PET prediction and -interpretation.

Current clinical guidelines and future perspectives

Role of interim [^{18}F]FDG PET in current clinical guidelines

The current thesis provides an overview of the growing evidence and validation of interim [^{18}F]FDG PET in DLBCL in the recent years. The most recent Dutch DLBCL guideline, published by HOVON in September 2021, recommends to perform an interim [^{18}F]FDG PET after 2 cycles of R-CHOP instead of a CT scan in cases where this could affect treatment de-escalation or escalation choices [19].

Interim [^{18}F]FDG PET guided de-escalation of good responders

The current Dutch guideline describes 2 patient groups where treatment could be reduced based on the interim [^{18}F]FDG PET result.

For patients with a stage I-II non bulky (<10 cm) DLBCL and a negative interim [^{18}F]FDG PET (Deauville 1-3) after 2 cycles of R-CHOP, patients can be treated with either 4 cycles R-CHOP or 3 cycles R-CHOP with additional involved node radiotherapy depending on the expected toxicity of radiotherapy versus the expected toxicity of anthracyclines [20]. For patients with a positive interim [^{18}F] FDG PET treatment with 6 cycles R-CHOP is advised.

Patients with stage II-IV disease and a negative interim [^{18}F]FDG PET after 2 cycles of R-CHOP can be treated with 6 cycles R-CHOP without 2 extra gifts of rituximab [21]. A recent study (published after the current DLBCL guideline) showed that this approach is cost-effective [22].

Escalation therapy with new (combination of) drugs?

In the overviews described in Part I we concluded that (preliminary) reports on current chemotherapy based escalation strategies do not overcome treatment resistance in DLBCL. This was also the case for a more recent large randomized trial with escalation of treatment with a "Burkitt treatment regimen" tested in patients with a poor response at interim [^{18}F]FDG PET after 2 cycles of R-CHOP immuno-chemotherapy [21]. Until very recent there were relatively few options for patients with relapsed/refractory DLBCL. Second-line therapy is not standard, as there is no generally preferred salvage regimen [23-26]. For physically fit patients, the current Dutch guideline recommends R-DHAP or R-GDP followed by BEAM and autologous stem cell transplantation in patients with responding disease to R-DHAP or R-GDP.

For older and frail patients unfit for autologous stem cell transplantation several palliative chemo-immunotherapy schemes exist (e.g. R-PECC [27], GEMOX-R [28], R-bendamustine [29] and R-lenalidomide [30]). The combination of GEMOX-R with nivolumab is currently investigated in the randomized phase I/II HOVON-153 (NIVEAU) study for this patient group.

More recently, the introduction of chimeric antigen receptor T-cells (CAR-T cell therapy) is of great interest as a new therapy strategy in DLBCL patients since this is a potential curative treatment. There are several different CAR-T therapies in development for lymphoma, most of them targeting the CD19 cell surface protein [31-33]. In short, the idea is that the ex-vivo modified re-infused CAR-T cells will expand in the patients, recognize CD19 positive cells and kill these. The CD19 CAR-T cell therapy (with axicabtagene ciloleucel) is since May 2020 a reimbursed treatment option after 2 or more lines of systemic treatment for relapsed or refractory DLBCL, transformed follicular lymphoma and primary mediastinal B-cell lymphoma in the Netherlands [31,32].

Besides the current role of CAR-T cell therapy in relapsed/refractory DLBCL, clinical trials are ongoing to investigate the role of CD19 targeting CAR-T as part of the frontline therapy in patients with high risk DLBCL. For example, the ZUMA-12 trial in patients with MYC and BCL2 and/or BCL6 (double or triple hit) or IPI>=3 enrolled patients with a positive interim [^{18}F]FDG PET phase II trial are promising, and longer follow-up results are awaited [34].

Other potential interesting "new" therapies that are being investigated are the bispecific monoclonal antibodies [35,36], tandem CAR-T (1 CAR with 2 tumor antigen targets, for example for CD19 and CD20) [37] and combination therapies with other immunomodulatory drugs.

With these new treatment strategies the role and timing of PET response assessment should again be determined for these therapies, bearing in mind that response patterns could be different with immunomodulatory agents [38] and possible criteria refinement could be needed. As these treatment strategies are generally very expensive and probably will be applied to smaller patient groups there is a clear need for defining which patients could benefit from it. For this purpose, both prognostic baseline clinical- (e.g. high IPI) and baseline PET characteristics together with a positive interim [^{18}F]FDG PET could be used to select high risk patients in order to make new clinical trials more efficient.

Future of the PETRA consortium

During the interim [^{18}F]FDG PET lymphoma project, which officially started in January 2015, the PETRA consortium and the PETRA database were established.

The project was financially supported by a KWF/Alpe d'Huzes grant (VU 2012-5848). In this consortium both the clinical individual patient data as well as imaging data from the included international studies were collected, nowadays including about 2300 DLBCL patients [39]. Studies that are currently included in the PETRA database are from Germany [21], United Kingdom [40,41], Switzerland [42], Hungary [43] (together with patients from other countries enrolled in the IAEA study), Italy [44], Nordic/US intergroup study [45] and the Netherlands/Belgium [14,16]. Not only [^{18}F]FDG PET, but also baseline [^{18}F]FDG PET and end-of-treatment [^{18}F]FDG PET data were collected. By building a sustainable database, this PETRA database can also be used to address future research questions, other technical or clinical validations, designing more efficient interim [^{18}F]FDG PET trials and extensions to e.g. other lymphoma subtypes can be made. Furthermore, the PETRA database with qualitatively good studies could also help in the design of future high quality studies. This makes the PETRA database of interest both for clinical researchers as well as for imaging or pharmaceutical companies.

After the original interim [^{18}F]FDG PET project, resulting in this thesis, The PETRA consortium continued with a project focusing on baseline [^{18}F]FDG PET. A consortium grant from KWF was obtained for the investigation of radiomics and this project has started in January 2019. Radiomics analysis of baseline [^{18}F]FDG PET provides quantitative features of tumor characteristics such as intensity, shape, volume, localization and texture and also information about intra- and intertumor heterogeneity. The aim of that project is to identify and validate those features that are robust to variability in image quality and contain prognostic information in addition to classical baseline prognostic factors of the international prognostic index (IPI) [46,47].

One of the features that is included in the radiomics analysis is metabolic tumor volume. The methods used in the pilot study from **Chapter 5** were recently also tested in a case-control study with 138 patients from the PETRA database [48]. In that study it was concluded that an automated estimation of metabolic tumor volume is feasible. Both the SUV4.0 and a majority vote strategy (MV2) were recommended to be evaluated further. The association with clinical outcome is currently being explored in a larger database within the PETRA consortium [49]. Furthermore, a project on optimization of operational characteristics of end-of-

treatment [^{18}F]FDG PET in DLBCL has recently started, aiming to optimize and/or develop criteria to improve the PPV without affecting the high NPV using the PETRA database. Recently, a new PETRA project on artificial intelligence for [^{18}F]FDG PET response prediction in DLBCL started funded by the "Hanarth Fonds". Finally, new research initiatives from other consortium members are currently proposed and investigated.

Conclusion

This thesis evaluated the role of [^{18}F]FDG PET as a biomarker in aggressive lymphoma by both technical and clinical validation. It can be concluded that the role of [^{18}F]FDG PET in treating aggressive lymphoma patients has clearly grown in the past years. Next to the role of [^{18}F]FDG PET for the baseline staging and end-of-treatment response assessment in aggressive lymphoma, the value of interim [^{18}F]FDG PET has become clearer. We validated that interim [^{18}F]FDG PET has a high negative predictive value (>80%) when using Deauville- and ΔSUVmax criteria for good response prediction in DLBCL. This has led to the inclusion of interim [^{18}F]FDG PET guided treatment de-escalation in the current Dutch guidelines for DLBCL.

References

1. Zijlstra JM, Burggraaff CN, Kersten MJ, Barrington SF; EHA Scientific Working Group on Lymphoma. FDG-PET as a biomarker for early response in diffuse large B-cell lymphoma as well as in Hodgkin lymphoma? Ready for implementation in clinical practice? *Haematologica.* 2016 Nov;101(11):1279-1283.
2. Radford J, Illidge T, Counsell N, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. *N Engl J Med.* 2015;372(17):1598-1607.
3. Raemaekers JM, André MP, Federico M, et al. Omitting radiotherapy in early positron emission tomography-negative stage I/II Hodgkin lymphoma is associated with an increased risk of early relapse: Clinical results of the preplanned interim analysis of the randomized EORTC/LYSA/FIL H10 trial. *J Clin Oncol.* 2014;32(12):1188-1194.
4. Johnson P, Federico M, Kirkwood A, et al. Adapted Treatment Guided by Interim PET-CT Scan in Advanced Hodgkin's Lymphoma. *N Engl J Med.* 2016;374(25):2419-2429.
5. André MPE, Girinsky T, Federico 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(16):1786-1794.
6. Borchmann P, Goergen H, Kobe C, et al. PET-guided treatment in patients with advanced-stage Hodgkin's lymphoma (HD18): final results of an open-label, international, randomised phase 3 trial by the German Hodgkin Study Group. *Lancet.* 2017;390(10114):2790-2802.
7. Lymfoomwerkgroep HOVON. Richtlijn Hodgkin Lymfoom bij Volwassenen. Amsterdam: Stichting Hemato-Oncologie voor Volwassenen Nederland (HOVON); 2019. Available from: https://hematologienederland.nl/wp-content/uploads/2020/01/Richtlijn-Hodgkin-lymfoom-HOVON-2019-v-2019.11.13.pdf.
8. Dührsen U, Hüttmann A, Müller S, et al. Positron Emission Tomography (PET) Guided Therapy of Aggressive Lymphomas – a Randomized Controlled Trial Comparing Different Treatment Approaches Based on Interim PET Results (PETAL Trial). *Blood.* 2014;124(21):(Abstract 391).
9. Swinnen LJ, Li H, Quon A, et al. Response-adapted therapy for aggressive non-Hodgkin's lymphomas based on early [18F] FDG-PET scanning: ECOG-ACRIN Cancer Research Group study (E3404). *Br J Haematol.* 2015;170(1):56-65.
10. Stewart DA, Kloiber R, Owen C, et al. Results of a prospective phase II trial evaluating interim positron emission tomography-guided high dose therapy for poor prognosis diffuse large B-cell lymphoma. *Leuk Lymph.* 2014;55(9):2064-2070.
11. Burggraaff CN, de Jong A, Hoekstra OS, et al. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2019 Jan;46(1):65-79.
12. Burggraaff CN, Cornelisse AC, Hoekstra OS, et al. HOVON Imaging Working Group. Interobserver Agreement of Interim and End-of-Treatment ^{18}F-FDG PET/CT in Diffuse Large B-Cell Lymphoma: Impact on Clinical Practice and Trials. *J Nucl Med.* 2018 Dec;59(12):1831-1836.
13. Burggraaff CN, Rahman F, Kaßner I, et al. PETRA Consortium. Optimizing Workflows for Fast and Reliable Metabolic Tumor Volume Measurements in Diffuse Large B Cell Lymphoma. *Mol Imaging Biol.* 2020 Aug;22(4):1102-1110.
14. Lugtenburg PJ, de Nully Brown P, van der Holt B, et al. Rituximab-CHOP With Early Rituximab Intensification for Diffuse Large B-Cell Lymphoma: A Randomized Phase III Trial of the HOVON and the Nordic Lymphoma Group (HOVON-84). *J Clin Oncol.* 2020 Oct 10;38(29):3377-3387.
15. Burggraaff CN, Eertink JJ, Lugtenburg PJ, et al. ^{18}F-FDG PET improves baseline clinical predictors of response in diffuse large B-cell lymphoma: The HOVON-84 study. *J Nucl Med.* 2022 Jul;63(7)1001-1007.
16. Chamuleau MED, Burggraaff CN, Nijland M, et al. Treatment of patients with MYC rearrangement positive large B-cell lymphoma with R-CHOP plus lenalidomide: results of a multicenter HOVON phase II trial. *Haematologica.* 2020 Dec 1;105(12):2805-2812.
17. Eertink JJ, Burggraaff CN, Heymans MW, et al. Optimal timing and criteria of interim-PET in DLBCL: a comparative study of 1692 patients. *Blood Adv.* 2021 May;5(9):2375-2384.

18. Eertink JJ, Arens AIJ, Huijbregts JE, et al. HOVON imaging workgroup. Aberrant patterns of PET response during treatment for DLBCL patients with MYC gene rearrangements. *Eur J Nucl Med Mol Imaging*. 2022 Feb; 49(3):943-952.
19. HOVON Lymfoom werkgroep. Richtlijn diffuus grootcellig B-cel non-Hodgkin lymfoom. Amsterdam: HOVON centraal bureau; 2021. Available from: https://hematologienederland.nl/wp-content/uploads/2021/10/Richtlijn-DLBCL-_NVvH.pdf.
20. Persky DO, Li H, Stephens DM, et al. Positron Emission Tomography-Directed Therapy for Patients With Limited-Stage Diffuse Large B-Cell Lymphoma: Results of Intergroup National Clinical Trials Network Study S1001. *J Clin Oncol.* 2020 Sep 10;38(26):3003-3011. Erratum in: *J Clin Oncol.* 2020 Oct 10;38(29):3459.
21. Dührsen U, Müller S, Hertenstein B, et al. PETAL Trial Investigators. Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL): A Multicenter, Randomized Phase III Trial. *J Clin Oncol.* 2018;36(20):2024-2034.
22. Greuter M, Eertink JJ, Jongeneel G, et al. PETRA consortium. Cost-Effectiveness of Shortening Treatment Duration Based on Interim PET Outcome in Patients With Diffuse Large B-cell Lymphoma. *Clin Lymphoma Myeloma Leuk*. 2022 Jun;22(6)382-392.
23. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010 Sep 20;28(27):4184-4190. Erratum in: *J Clin Oncol.* 2012 May 20;30(15):1896.
24. Crump M, Kuruvilla J, Couban S, et al. Randomized comparison of gemcitabine, dexamethasone, and cisplatin versus dexamethasone, cytarabine, and cisplatin chemotherapy before autologous stem-cell transplantation for relapsed and refractory aggressive lymphomas: NCIC-CTG LY.12. *J Clin Oncol.* 2014 Nov 1;32(31):3490-3496.
25. van Imhoff GW, McMillan A, Matasar MJ, et al. Ofatumumab Versus Rituximab Salvage Chemoimmunotherapy in Relapsed or Refractory Diffuse Large B-Cell Lymphoma: The ORCHARRD Study. *J Clin Oncol.* 2017 Feb 10;35(5):544-551.
26. Crump M, Neelapu SS, Farooq U, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. *Blood.* 2017 Oct 19;130(16):1800-1808. Erratum in: *Blood.* 2018 Feb 1;131(5):587-588.
27. Lugtenburg PJ, Zijlstra JM, Doorduijn JK, et al. Dutch HOVON group. Rituximab-PECC induction followed by ^{90} Y-ibritumomab tiuxetan consolidation in relapsed or refractory DLBCL patients who are ineligible for or have failed ASCT: results from a phase II HOVON study. *Br J Haematol.* 2019 Nov;187(3):347-355.
28. Mounier N, El Gnaoui T, Tilly H, et al. Rituximab plus gemcitabine and oxaliplatin in patients with refractory/relapsed diffuse large B-cell lymphoma who are not candidates for high-dose therapy. A phase II Lymphoma Study Association trial. *Haematologica.* 2013 Nov;98(11):1726-1731.
29. Ohmachi K, Niitsu N, Uchida T, et al. Multicenter phase II study of bendamustine plus rituximab in patients with relapsed or refractory diffuse large B-cell lymphoma. *J Clin Oncol.* 2013 Jun 10;31(17):2103-2109.
30. Zinzani PL, Pellegrini C, Argnani L, Broccoli A. Prolonged disease-free survival in elderly relapsed diffuse large B-cell lymphoma patients treated with lenalidomide plus rituximab. *Haematologica.* 2016 Sep;101(9):e385-386.
31. Neelapu SS, Locke FL, Bartlett NL, et al. Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med.* 2017 Dec 28;377(26):2531-2544.
32. Locke FL, Ghobadi A, Jacobson CA, et al. Long-term safety and activity of axicabtagene ciloleucel in refractory large B-cell lymphoma (ZUMA-1): a single-arm, multicentre, phase 1-2 trial. *Lancet Oncol.* 2019 Jan;20(1):31-42.
33. Schuster SJ, Bishop MR, Tam CS, et al. JULIET Investigators. Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *N Engl J Med.* 2019 Jan 3;380(1):45-56.
34. Neelapu SS, Dickinson M, Sci D, et al. Interim Analysis of ZUMA-12: A phase 2 Study of Axicabtagene Ciloleucel (Axi-Cel) as First-Line Therapy in Patients (Pts) With High-Risk Large B Cell Lymphoma (LBCL). *Blood.* 2020;136 (Suppl 1):49. Abstract 626.
35. Hutchings M, Morschhauser F, Iacoboni G, et al. Glofitamab, a Novel, Bivalent CD20-Targeting T-Cell-Engaging Bispecific Antibody, Induces Durable Complete Remissions in Relapsed or Refractory B-Cell Lymphoma: A Phase I Trial. *J Clin Oncol.* 2021 Jun 20;39(18):1959-1970.

36. Hutchings M, Mous R, Clausen MR, et al. Dose escalation of subcutaneous epcoritamab in patients with relapsed or refractory B-cell non-Hodgkin lymphoma: an open-label, phase 1/2 study. *Lancet.* 2021 Sep 25;398(10306):1157-1169.
37. Tong C, Zhang Y, Liu Y, et al. Optimized tandem CD19/CD20 CAR-engineered T cells in refractory/relapsed B-cell lymphoma. *Blood.* 2020 Oct 1;136(14):1632-1644.
38. Cheson BD, Ansell S, Schwartz L, et al. Refinement of the Lugano Classification lymphoma response criteria in the era of immunomodulatory therapy. *Blood.* 2016 Nov 24;128(21):2489-2496.
39. PETRA consortium. Available from: https://petralymphoma.org/.
40. 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 Jul;43(7):1209-1219.
41. Mikhaeel NG, Cunningham D, Counsell N, et al. FDG-PET/CT after two cycles of R-CHOP in DLBCL predicts complete remission but has limited value in identifying patients with poor outcome - final result of a UK National Cancer Research Institute prospective study. *Br J Haematol.* 2021 Feb;192(3):504-513.
42. 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 Aug 10;33(23):2523-2529. Erratum in: *J Clin Oncol.* 2015 Sep 20;33(27):3074.
43. Carr R, Fanti S, Paez D, et al. IAEA Lymphoma Study Group. Prospective international cohort study demonstrates inability of interim PET to predict treatment failure in diffuse large B-cell lymphoma. *J Nucl Med.* 2014 Dec;55(12):1936-1944.
44. Zinzani PL, Gandolfi L, Broccoli A, et al. Midtreatment 18F-fluorodeoxyglucose positron-emission tomography in aggressive non-Hodgkin lymphoma. *Cancer.* 2011 Mar 1;117(5):1010-1018.
45. Mylam KJ, Kostakoglu L, Hutchings M, et al. (18)F-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 Jul;56(7):2005-2012.
46. Eertink JJ, Pfaehler EAG, Wiegers SE, et al. Quantitative radiomics features in diffuse large B-cell lymphoma: does segmentation method matter? *J Nucl Med.* 2022 Mar;63(3):389-395.
47. Eertink JJ, van de Brug T, Wiegers SE, et al. ^{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 Feb;49 (3):932-942.
48. Barrington SF, Zwezerijnen BGJC, de Vet HCW, et al. Automated Segmentation of Baseline Metabolic Total Tumor Burden in Diffuse Large B-Cell Lymphoma: Which Method Is Most Successful? A Study on Behalf of the PETRA Consortium. *J Nucl Med.* 2021 Mar;62(3):332-337.
49. Mikhaeel NG, Heymans MW, Eertink JJ, et al. A proposed new dynamic prognostic index for DLBCL: International Metabolic Prognostic Index (IMPI). Accepted at *J Clin Oncol.* 2022.

Addendum

Dutch summary / Nederlandse samenvatting

Inleiding

Hoofdstuk 1 geeft een algemene introductie over lymfoom, PET en het doel van deze thesis. Lymfoom is een maligne proliferatieve ziekte van het lymfoïde weefsel, beter bekend als lymfeklierkanker. Lymfoom wordt onderverdeeld in 2 types; het Hodgkin lymfoom (HL) en het non-Hodgkin lymfoom (NHL). HL komt vooral voor bij jong-volwassenen (<40 jaar) met een grote kans op genezing (>90% overleving) na een behandeling met chemotherapie, daarentegen is bij oudere patiënten de overleving in het algemeen minder gunstig. Het NHL bestaat uit een groep van meerdere subtypes van lymfoom welke ontstaan vanuit B-cellen, T-cellen of NK-cellen. Het diffuus grootcellig B-cellymfoom is het meest voorkomende subtype (circa 40% van alle patiënten met NHL) met ongeveer 1500 nieuwe patiënten per jaar in Nederland. De standaardbehandeling bestaat uit een combinatie van immuuntherapie en chemotherapie; rituximab (een monoclonaal anti-CD20 antilichaam gericht tegen de B-cel), cyclofosfamide, doxorubicine, vincristine en prednison (R-CHOP). Patiënten die een recidief lymfoom krijgen of progressie hebben na de eerstelijnsbehandeling met R-CHOP hebben vaak een slechtere respons bij tweedelijnsbehandeling.

Met een positron-emissie tomografie (PET) scan is het mogelijk om gebieden van het lichaam met een hoog glucose metabolisme zichtbaar te maken door middel van een radioactief gelabelde glucose analoog ([^{18}F]FDG). [^{18}F]FDG PET wordt tegenwoordig standaard gebruikt voor zowel de stadiëring van het DLBCL als de responsevaluatie aan het eind van de eerstelijnsbehandeling met R-CHOP. Een PET scan tijdens de behandeling (interim [^{18}F]FDG PET) is de afgelopen jaren in opkomst als een vroege responsbiomarker. De hypothese is om patiënten die niet voldoende reageren op R-CHOP al vroeg tijdens de behandeling te kunnen identificeren en de behandeling op dat moment aan te passen naar een mogelijk meer effectieve behandeling. De beoordeling van de interim [^{18}F]FDG PET respons is mogelijk met visuele methodes (Deauville 5-puntsschaal) of semikwantitatief door middel van het meten van de gestandaardiseerde opnamewaarde van [^{18}F] FDG (SUV). Het is onbekend wat de beste responscriteria zijn en wat de optimale timing (na hoeveel cycli R-CHOP) is voor het beoordelen van de interim [^{18}F] FDG PET respons.

Het doel van deze thesis was het valideren van interim [^{18}F]FDG PET als een biomarker voor de respons op behandeling bij patiënten met DLBCL. Daarnaast was het doel om de optimale responscriteria en timing van de interim [^{18}F]FDG PET vast te stellen. Door middel van het opzetten van een internationaal PETRA (PET re-analyse) consortium konden voldoende patiëntgegevens en PET scans vanuit diverse DLBCL studies verzameld en opnieuw geanalyseerd worden om bovenstaande vragen te kunnen beantwoorden.

Deel 1 PET als een responsbiomarker bij lymfomen

Hoofdstuk 2 geeft een overzicht van wetenschappelijke artikelen gepubliceerd tot 2016 over de accuratesse van interim [^{18}F]FDG PET bij DLBCL en HL patiënten in klinische studies, waaronder studies met interim [^{18}F]FDG PET gestuurde behandeling en de rol van interim [^{18}F]FDG PET in de klinische praktijk in 2016. De Engelse RAPID en EORTC H10 studies onderzochten of een de-escalatie van behandeling (weglaten van radiotherapie) mogelijk is. Op basis van deze studies werd geconcludeerd dat de-escalatie van behandeling bij HL mogelijk is geworden in de klinische praktijk. In de veelbelovende RATHL studie, kregen HL patiënten met gevorderde ziekte die geen complete metabole remissie bereikten bij de interim [^{18}F]FDG PET na 2 kuren, escalatie van behandeling van ABVD naar het meer intensieve BEACOPP-escalated chemotherapie schema. Kortom, voor HL is interim [^{18}F]FDG PET in de klinische praktijk van belang voor zowel de-escalatie als escalatie van behandeling. De meest recente Nederlandse HL richtlijn van 2019 bevat een kwaliteitsindicator voor de responsevaluatie en aanpassing van eerstelijnsbehandeling met interim [^{18}F]FDG PET voor zowel patiënten met beperkt stadium als HL patiënten met gevorderd stadium van hun ziekte. Bij patiënten met beperkt stadium HL en een positieve interim [^{18}F]FDG PET adviseert de richtlijn om de behandeling te escaleren naar 2x BEACOPP-escalated gevolgd door radiotherapie van de initieel aangedane klieren (involved node). Bij patiënten met gevorderd stadium HL en een positieve interim [^{18}F]FDG PET is het advies om de behandeling voort te zetten met 4x BEACOPP-escalated met additionele radiotherapie op positieve laesies aan het einde van de behandeling. Bij patiënten met een beperkt stadium HL en een negatieve interim [^{18}F]FDG PET kan de behandeling worden voortgezet met alleen radiotherapie

of AVD (zonder bleomycine) in combinatie met radiotherapie (afhankelijk van de aanwezigheid van baseline risicofactoren). Bij patiënten met een gevorderd stadium HL en een negatieve interim [^{18}F]FDG PET kan de behandeling worden voortgezet met 4x AVD (in geval van behandeling met ABVD) of 2x BEACOPP-escalated (in geval van behandeling met BEACOPP-escalated). In tegenstelling tot bovenstaande was er in 2016 nog geen bewijs voor het gebruik van interim [^{18}F]FDG PET voor de-escalatie of escalatie van behandeling in de klinische praktijk bij het DLBCL. De literatuur op dat moment bestond uit preliminaire data en data gepresenteerd in abstracts en suggereerde dat de huidige chemotherapie escalatiestrategieën niet in staat waren om therapieresistentie te overwinnen.

Hoofdstuk 3 beschrijft een systematische review en meta-analyse over de predictieve waarde van visuele beoordeling van interim [^{18}F]FDG PET bij DLBCL patiënten. De studie beschrijft de resultaten van 20 gepubliceerde studies zonder aanpassing van de behandeling op basis van het interim [^{18}F]FDG PET resultaat. We concludeerden dat er predictieve waarde is van interim [^{18}F]FDG PET voor de 2 jaar progressievrije overleving met een gepoolde hazard ratio van 3.13. De negatief voorspellende waarde was boven de 80%, maar de positief voorspellende waarde (varierend van 20-74% tussen de verschillende studies) was te laag om een risicogestuurde behandeling in de klinische praktijk te adviseren. Aanvullende subgroepanalyses naar de timing van de interim [^{18}F]FDG PET, verschillende patiëntengroepen, PET technieken en het type responscriterium waren beperkt mogelijk door onvolledige informatie, heterogeniteit en de kleine steekproefomvang.

Op basis van deze 2 hoofdstukken kunnen we concluderen dat er inderdaad sprake is van predictieve waarde van interim [^{18}F]FDG PET bij zowel HL als DLBCL patiënten. Tegenwoordig speelt interim [^{18}F]FDG PET een belangrijke rol in de 'state of the art' behandeling van het HL. Voor DLBCL was de rol van interim [^{18}F]FDG PET als een responsbiomarker nog onduidelijk.

Deel 2 Technische validatie van PET bij lymfoom

Hoofdstuk 4 beschrijft de interobservervariatie van de Deauville 5-puntsschaal voor zowel interim [^{18}F]FDG PET als PET aan het eind van de eerstelijnsbehandeling bij patiënten met DLBCL geïncludeerd in de HOVON-84 studie. De centrale review van de studie werd uitgevoerd door 2 nucleair geneeskundigen (uit een groep van 10). We concludeerden dat de totale overeenkomst 87.7% en 91.7% is voor respectievelijk interim [^{18}F]FDG PET en PET aan het eind van de eerstelijnsbehandeling. De positieve overeenkomst (gedefinieerd als de kans dat als 1 reviewer de scan positief scoort, dat de andere reviewer deze scan ook positief scoort) is 73.7% en 76.3%. Op basis van deze resultaten raden we aan om een tweede beoordeling te verrichten voor de evaluatie van behandeling in de klinische praktijk (bijvoorbeeld in een multidisciplinaire patiëntenbespreking; MDO) en voor studies bevelen we een centrale reviewprocedure aan. Met name locaties van extranodale lymfoombetrokkenheid in de tractus gastrointestinalis, de ring van Waldeyer, het skelet en de milt gaven relatief veel discrepanties tussen de beoordelingen van de reviewers. In deze weefsels is de locale achtergrond van [^{18}F]FDG opname meer variabel over de tijd door de aanwezigheid van intercurrente inflammatie, genezing van pathologische fracturen of recente toediening van granulocyt koloniestimulerende factoren (G-CSF). Aanpassing op basis van deze factoren is geen onderdeel van de recente Lugano classificatie richtlijnen.

Hoofdstuk 5 beschrijft een pilotstudie waarin verschillende strategieën voor het optimaal meten van metabool tumor volume (MTV) is getest op 12 baseline [^{18}F]FDG PET scans met als doel het ontdekken van een snelle en betrouwbare meetmethode. Een geautomatiseerde preselectie strategie met een SUV van 4 of hoger en een minimaal volume van 3 ml resulteerde in een verbeterde interobserver betrouwbaarheid en meer gebruikersgemak ten opzichte van een strategie met individuele selectie van lymfoomlaesies.

Op basis van deze 2 hoofdstukken in dit deel van de thesis concluderen we dat een centrale review een belangrijke voorwaarde is voor het gebruik van visueel beoordeelde interim [^{18}F]FDG PET scans. In toekomstige studies en voor het gebruik in de klinische praktijk bevelen we aan de scans tweemaal te beoordelen. In trainingen voor nucleair geneeskundigen is extra aandacht nodig voor de evaluatie van extranodale locaties van lymfoom, aangezien deze locaties leiden

tot relatief veel discrepanties in beoordelingen. Op basis van deze observaties organiseerden we diverse hands-on workshops over het gebruik van de Lugano classificatie richtlijnen in de dagelijkse praktijk voor radiologen, hematologen en nucleair geneeskundigen in Nederland. Voor de baseline evaluatie van MTV bestaat nog geen consensus met betrekking tot de te gebruiken evaluatiemethode. De beschreven pilotstudie is een eerste stap richting een meer geautomatiseerde evaluatie van MTV.

Deel 3 Klinische validatie van PET bij lymfoom

Hoofdstuk 6a beschrijft de interim [^{18}F]FDG PET resultaten van 513 nieuw gediagnosticeerde DLBCL patiënten in de gerandomiseerde HOVON-84 studie. **Hoofdstuk 6b** bevat de resultaten van het originele studiedoel van de HOVON-84 studie met betrekking tot de randomisatie tussen standaardbehandeling met R-CHOP14 en R-CHOP14 met rituximab intensificatie tijdens de eerste 4 kuren. De complete metabole remissie, failure-free -, progressievrije - en totale overleving verbeterden niet door de intensificatie van rituximab. Het originele HOVON-84 artikel is toegevoegd als deel b van dit hoofdstuk aangezien de complete metabole respons bepaald werd door middel van de centrale review van [^{18}F]FDG PET en het artikel belangrijke achtergrondinformatie bevat voor deel a van het hoofdstuk. **Hoofdstuk 6a** toont dat zowel interim [^{18}F]FDG PET als de age-adjusted internationale prognostische index (aaIPI) onafhankelijke responsbiomarkers zijn. De negatief voorspellende waarde voor 2-jaar progressievrije overleving bij patiënten met een laag-risico of laag-intermediair risico aaIPI en een ΔSUVmax van meer dan 70% is 93%. De positief voorspellende waarde voor 2-jaar progressievrije overleving bij patiënten met een hoog-intermediair risico of hoog risico aaIPI score en ΔSUVmax ≤70% is 65%. Daarnaast laat de externe validatie zien dat de semi-kwantitatieve ΔSUVmax criteria de Deauville 5-puntsschaal overtreffen in de predictie van 2-jaar progressievrije overleving (positief voorspellende waarde van respectievelijk 53 versus 38%).

Hoofdstuk 7 beschrijft de resultaten van de fase 2 HOVON-130 studie van 82 MYC-positieve agressieve B-cellymfoom patiënten (een subgroep van DLBCL patiënten met een slechtere prognose), waarbij de behandeling bestond uit een combinatie van R-CHOP met lenalidomide. In de studie werden zowel interim

[^{18}F]FDG PET als PET scans aan het einde van de behandeling verricht. Het complete metabole respons percentage is 67% en vergelijkbaar (geen officieel vergelijk mogelijk) met eerdere studies waarin meer intensieve chemotherapie schema's werden gegeven. De positief en negatief voorspellende waarde van de PET aan het eind van de behandeling voor recidief binnen 12 maanden is respectievelijk 81% en 93%. De observationele analyse van interim [^{18}F]FDG PET toont een positief en negatief voorspellende waarde van respectievelijk 60 en 79%. Kortom, deze resultaten ondersteunen het gebruik van interim [^{18}F]FDG PET in deze specifieke patiëntsubgroep niet.

Hoofdstuk 8 beschrijft de belangrijkste resultaten van het PETRA interim [^{18}F] FDG PET project (KWF/Alpe d'Huzes grant VU 2012-5848) door middel van een individuele patiëntdata meta-analyse met 1698 DLBCL patiënten die werden geïncludeerd in 8 internationale studies. Deauville score en ΔSUVmax criteria werden vergeleken bij een interim [^{18}F]FDG PET met verschillende timing tijdens de behandeling, na respectievelijk 2 en 4 kuren R-CHOP. ΔSUVmax criteria laten een hogere discriminatie en predicitieve waarde zien voor de 2-jaars progressievrije overleving ten opzichte van de Deauville 5-puntsschaal met een afkapwaarde van positiviteit voor DS 4 en 5. Echter, de optimale timing en responscriteria kunnen afhankelijk zijn van de klinische context van de studie. In het algemeen zijn de hazard ratios hoger bij een latere timing van de interim [^{18}F]FDG PET, wat resulteert in een betere patiëntenstratificatie. De negatief voorspellende waardes zijn hoog (boven de 80%) voor alle responscriteria voor zowel de interim [^{18}F] FDG PET na 2 als na 4 kuren. In het algemeen is de positief voorspellende waarde laag voor alle responscriteria, waarschijnlijk vanwege de prevalentie van progressie van 22% in onze dataset. Echter blijkt de positief voorspellende waarde wel wat hoger bij interim [^{18}F]FDG PET na 4 kuren ten opzichte van 2 kuren (ΔSUVmax 57% versus 46%, Deauville 4-5 43% versus 31%) en tevens hoger voor ΔSUVmax ten opzichte van de Deauville cutoff 4-5 voor een positieve scan. Een Deauville score van 5 selecteert de patiënten met de meest slechte respons (ca 5% van alle patiënten), deze patiënten kunnen al na een interim [^{18}F]FDG PET na 2 kuren worden geïdentificeerd met vergelijkbare positief- en negatief voorspellende waarde als na 4 kuren. In de multivariate analyse zijn zowel IPI score en interim [^{18}F]FDG PET scan (alle responscriteria) onafhankelijke responspredictoren. We concluderen dat een goede respons (gedefinieerd als ΔSUVmax ≥66%) na 2 kuren R-CHOP behandeling in aanmerking komt voor gebruik in een gerandomiseerde

studie waarin de-escalatie van behandeling wordt onderzocht. Een slechte respons (gedefinieerd als ΔSUVmax <70%) na 4 kuren R-CHOP komt in aanmerking voor evaluaties van nieuwe behandelingen in een gerandomiseerde studie.

De 3 hoofdstukken van dit afsluitende deel van de thesis bevat de klinische validatie van interim [^{18}F]FDG PET in 2 HOVON studies en de individuele patiëntdata meta-analyse van in de PETRA database geïncludeerde studies. We kunnen concluderen dat interim [^{18}F]FDG PET inderdaad predictieve waarde heeft voor de 2-jaars progressievrije overleving, 2-jaars totale overleving en 2-jaar tijd tot progressie eindpunten. Deze resultaten kunnen gebruikt worden voor het design van toekomstige studies en powerberekeningen van klinische studies met de-escalatiestrategieën of nieuwe behandelingen voor DLBCL patiënten. Deze conclusie is echter niet van toepassing op de subgroep van DLBCL patiënten met een MYC translocatie. Een recente studie laat zien dat MYC-positieve patiënten vaker progressie hebben na een negatieve interim [^{18}F]FDG PET en ook vaker nieuwe lymfoomlaesies hebben op plaatsen die origineel niet aangedaan waren ten opzichte van MYC-negatieve patiënten. Waarschijnlijk komt dit door het meer agressieve karakter van de ziekte van de dubbel- en tripelhit (MYC, BCL2 en BCL6) lymfoompatiënten. Om die reden wordt in de huidige richtlijn een FISH analyse aanbevolen voor alle nieuw gediagnosticeerde DLBCL patiënten in Nederland om de dubbel- en tripelhit patiënten te kunnen identificeren en de R-CHOP behandeling aan te passen naar DA-EPOCH-R met centraal zenuwstelselprofylaxe.

Conclusie

Deze thesis onderzocht de rol van interim [^{18}F]FDG PET als een responsbiomarker bij agressieve lymfomen door middel van technische en klinische validatie. We kunnen concluderen dat de rol van [^{18}F]FDG PET bij lymfoom duidelijk gegroeid is in de afgelopen jaren. Naast de baseline stadiëring en de responsevaluatie aan het eind van de behandeling, is ook de rol van interim [^{18}F]FDG PET duidelijker geworden. We hebben gevalideerd dat interim [^{18}F]FDG PET een hoge negatief voorspellende waarde heeft (>80%) bij gebruik van de Deauville 5-puntsschaal en ΔSUVmax criteria voor het voorspellen van een goede respons bij DLBCL. Dit heeft geleid tot het opnemen van een interim [^{18}F]FDG PET gestuurde de-escalatie behandelstrategie in de huidige DLBCL richtlijn.

Acknowledgements / Dankwoord

Wat een avontuur is het promoveren geweest afgelopen jaren. Jaren met daarin vele “ups” en natuurlijk ook zo nu en dan de bijbehorende “downs”, schrijf ik nu dit laatste stuk voor mijn proefschrift. Bij het schrijven van dit dankwoord besef ik opnieuw hoeveel mensen ik mijn dank verschuldigd ben. Het is onmogelijk om iedereen die een bijdrage heeft geleverd aan dit proefschrift, in welke vorm dan ook, bij naam te noemen.

Allereerst natuurlijk mijn promotoren en copromotoren.
Prof.dr. Josée Zijlstra, internist-hematoloog én wetenschapper in hart en nieren; dankzij jouw altijd aanwezige enthousiasme en positivisme heb ik kennis mogen maken met de wetenschap en in het bijzonder de rol van PET bij lymfoom. Tijdens mijn semi-arts stage hematologie/oncologie bij de MDO’s in het Spaarne ziekenhuis in Hoofddorp (tegenwoordig Spaarne Gasthuis) hebben wij elkaar leren kennen. Daarna ben ik begonnen als student wetenschappelijke stage onder jouw vleugels met een retrospectief statusonderzoek naar patiënten met een systemische- en centraal zenuwstelsellokalisatie van lymfoom. Tijdens mijn keuzecoschap hematologie op 3B heb je mij benaderd voor het PETRA project. Ik had destijds mijn sollicitatiebrief voor een ANIOS plek al klaar liggen om te versturen. Hoe dit verder is gegaan is te zien aan het proefschrift dat nu voor ons ligt. Dank voor alles, veelal tussen de hectische diensten en kliniek door, wat ik van je geleerd heb en ga leren. Ik kijk uit naar onze verdere samenwerking tijdens mijn opleiding tot internist-hematoloog de komende jaren.

Prof.dr.ir. Riekie de Vet, wij kenden elkaar nog niet voordat ik begon aan dit avontuur. Ik heb je leren kennen als een heel sportief en bovenal heel fijn mens. Dank voor alle epidemiologische en methodologische kennis die ik in de afgelopen jaren heb mogen opdoen. Dankzij jouw scherpe inzichten, kunst in het vereenvoudigen van complexe begrippen en nauwkeurigheid bij het schrijven van teksten, weet ik zeker dat de artikelen naar een hoger niveau zijn gebracht. Daarnaast ben je ook altijd bereid én bereikbaar om de soms zoveelste versie van een manuscript opnieuw kritisch te bekijken of gewoon voor een gezellig gesprek. Inmiddels mag je genieten van je welverdiende pensioen, waarin er weer (meer) tijd is voor je vele vrienden en sportieve bezigheden.

A

Prof.dr. Ronald Boellaard, jij was vanaf het eerste moment zeer betrokken bij alle PET kwaliteit, kwantificaties en methodologische overwegingen in het PETRA project. Na je (gelukkig tijdelijke) vertrek naar Groningen hebben wij elkaar vooral op de vrijdagen gesproken, en tussendoor per mail. Vele versies van de tool zijn voorbij gekomen, altijd binnen no time weer een bugje gefixt, of toch mogelijk gemaakt om die ene scan uit "Lutjebroek" ook in te kunnen laden. Ik heb ontzettend veel geleerd van je enorme methodologische inzicht en praktische instelling, maar vooral ook genoten van de nuchtere en gezellige persoon die daar bij hoort.

Prof.dr. Otto Hoekstra, inmiddels alweer jaren geleden heb ik je leren kennen als "de grijze nucleair geneeskundige" tijdens de hematologische PET bespreking op 3B. Gelukkig bleek je zeker niet grijs als persoon te zijn en een ware bron van wetenschappelijke inspiratie. Jouw scherpe blik, passie voor de wetenschap en oog voor de persoonlijke ontwikkeling als academicus zijn onmisbaar geweest. Tijdens mijn promotietijd hebben wij urenlang scans bekeken op 1F voor de HOVON studies (nog net niet dag en nacht), vele goede discussies gehad over de inhoud van de artikelen en zorgde jij voor de kritische noot die het artikel altijd weer beter maakte.

Leden van de lees- en promotiecommissie.
Prof.dr. S. Stroobants, prof.dr. S. Zweegman, prof.dr. Y.M. Smulders, prof.dr. L. Hooft, heel veel dank voor de tijd en energie die jullie hebben gestoken in het lezen en het beoordelen van dit proefschrift. Dr. Martine Chamuleau, heel veel dank voor de samenwerking bij de HOVON 130 studie en natuurlijk dat je aanwezig wilt zijn bij het opponeren. Dear prof.dr. C. Kobe and prof. dr. N.G. Mikhaeel thank you for the effort and time you invested in reading and assessing my thesis.

Natuurlijk niet te vergeten zijn mijn collega's van het PETRA team, a.k.a. "de meisjes".
Nikie Hoetjes, wij hebben een heel mooi (ruim) jaar mogen samenwerken. In die tijd heb ik ontzettend veel geleerd van jouw enorme ervaring. Maar vooral ook genoten van jouw altijd mooie (bier)verhalen.

Simone Pieplenbosch, wij lopen altijd de verkeerde kant op als we samen op pad zijn. Vooral die ene keer, dat je de roltrap via de "verkeerde" kant op wilde lopen

vergeet ik nooit meer. Dankzij jou waren alle buitenlandbezoeken fantastisch. Gelukkig is dat de verkeerde kant op lopen niet merkbaar geweest met het werk zelf. Heel veel dank (waarbij dit dankwoord op papier zetten niet genoeg aangeeft hoe dankbaar ik ben) voor alle fijne jaren samenwerken.

Corinne Eertink, ondanks dat je origineel uit de neuroscience komt, bleek je een zwak te hebben voor whole body PET scans. Sinds september 2016 werd jij aangenomen als research analist, maar inmiddels ben je nu zeer succesvol PhD student op het PETRA radiomics project. Ik heb altijd heel fijn samengewerkt met je, wat zich onder andere uitte in het doen van dezelfde uitspraken op hetzelfde moment.

Sanne Wiegers, ook jij bent inmiddels alweer jaren research analist in het PETRA team, waarbij ik me inmiddels niet meer kan voorstellen hoe het zonder jou was. Wat een enorme berg werk heb jij verricht in het doen van alle PET analyses en het opzetten van de shared database. Heel veel dank daarvoor en natuurlijk ook voor alle gezelligheid tijdens lunches/koffie etcetera.

Tijdens het PETRA project heb ik ook diverse (inmiddels ex-) studenten mogen begeleiden tijdens hun wetenschappelijke stages. Alexander Cornelisse, jij was destijds de eerste student die ik mocht begeleiden. Samen hebben wij de interobserver analyses van de HOVON 84 uitgewerkt. De maandagochtenden met koffie zijn nooit meer zo geweest als met jou. Theo van den Belt, zo veel uren heb jij zitten werken aan de kwaliteitscontroles van de scans en het effect daarvan op de beoordeling van de scans, dank daarvoor. Ariënne Baijense, veel dank voor jouw onderzoek naar de diverse methodes voor het analyseren van metabool tumor volume en het effect van diverse reconstructies op de metingen.

1F collega's.
Antoinette de Jong, samen hebben wij een systematische review geschreven, niet zonder slag of stoot. Gelukkig bracht het nummer "ik heb een heel zwaar leven" van Brigitte Kaandorp verlichting op de zware momenten. Cemile Karga, mijn kamergenoot gedurende 1,5 jaar, ik heb het getroffen met jou als collega. Gem Kramer en Thomas Koopman, bedankt voor alle hilarische momenten, op 1F en tijdens het congres in Barcelona. Gem, onze zangkunsten in Barcelona vergeet ik nooit! Thomas waarschijnlijk zeker niet... Thomas, jouw spontane imitatie

van Daniël Arends in Barcelona was ronduit briljant. Dennis Heijtel, fijn dat ik je kamergenoot mocht zijn, het terugkijken van gemiste dumpert filmpjes blijf ik wel missen. Mette Stam, het maken van de cd voor je vrienden (met nogal bijzondere nummers) was toch wel een van de hoogtepunten. Alle andere 1F collega's (en aanverwanten) Marc Huisman, Maqsood Yaqub, Fiona Heeman, Matthijs Cysouw, Bernard Jansen, Sandeep Golla, Ramsha Iqbal, Hayel Tuncel, Eline Verwer, Bart de Vries, Marijke den Hollander en Esther Drees heel veel dank voor alle gezellige lunches en koffiemomenten.

Collega's van de radiologie en nucleaire geneeskunde.
Ben Zwezerijnen, Daniela Oprea, Remi Schmeits en Amanda Kroonenberg, veel dank voor al jullie hulp.
Collega's van de afdeling epidemiologie. Martijn Heijmans, Marjolein Greuter en Veerle Coupé, dank voor de samenwerking afgelopen jaren.
Rita Azevedo, samen met je team waaronder Johan en Ruud, heb jij de infrastructuur van het project helpen opzetten, waarvoor veel dank.
Leden van de HOVON imaging werkgroep, dankzij jullie heb ik me deel gevoeld van "het veld" dat HOVON imaging werkgroep heet. In het bijzonder Anne Arens, Bart de Keijzer, Julia Huijbregts en Filiz Celik voor alle reviewuren.
Medewerkers van het HOVON data centrum, in het bijzonder Ronnie van der Holt, hartelijk dank.
Lieve mensen van het CCA, dokters en verpleegkundigen van 3B, en natuurlijk al mijn huidige collega's ook jullie allemaal ontzettend bedankt.
Alle nog niet genoemde co-auteurs van de artikelen in dit proefschrift wil ik enorm bedanken voor hun bijdragen.

All members of the PETRA consortium. Thanks for your enormous efforts in making our shared database. Members from the start: Sally, George, Ulrich, Stefan, Andreas, Michel, thanks for everything. Of course also a big thanks to all "new" PETRA steering committee members and all the people involved in making the sharing of data possible. Elly, Lars, Christine, Emanuele, Luca, Tamás, Robert, Sándor, Stefano, Lale, Annika and Martin. To all, we have made a start for a database with an enormous scientific value. As the famous dutch painter Vincent van Gogh said: "*great things are done by a series of small things brought together.*"

Voor mijn vrienden en familie buiten de academische wereld kan ik nu eindelijk "ja" zeggen op de vraag of ik al klaar ben met mijn onderzoek. Heel veel dank voor jullie steun en fijne afleiding al die tijd.

Yvonne Jauw, dankzij jou is mijn promotietijd fantastisch geworden. Ik ben blij dat wij collega's zijn geworden op het "eilandje" imaging in lymfomen, een altijd wat vreemde eend in de hematologieresearch in het VUmc. Samen hebben wij ontelbare momenten gekend waarop we elkaar hebben kunnen helpen, zowel inhoudelijk als organisatorisch (hoe zorg ik dat ik tijd heb om te schrijven?) Bedankt voor alles en natuurlijk dat jij mijn paranimf wilt zijn.

Marloes Vester, jij hebt al sinds het begin van mijn geneeskundestudie in mij geloofd en ook mede gezorgd dat deze dag zou komen door me te overtuigen van het gaan doen van dit PhD onderzoek. Ik kan altijd op je bouwen en hoop dat we alle mijlpalen die nog gaan komen samen kunnen blijven vieren.

Lisanne Burggraaff, waarschijnlijk kent niemand mij beter dan jij als zus(je), inmiddels heb jij je eigen gezin met 2 kinderen. Ik ben heel trots op jou en dankbaar voor alle ondersteuning in de afgelopen jaren. Natuurlijk ook Wilbert, mijn kleine (lees: veel langere) broertje wil ik bedanken.

Lida, helaas kun jij deze dag niet meer meemaken, maar ik weet zeker dat je heel trots zou zijn geweest.

Papa en mama, zonder jullie altijd aanwezige steun had ik dit traject nooit kunnen volbrengen. Al weten jullie waarschijnlijk maar tot op zekere hoogte wat ik nu eigenlijk onderzocht heb afgelopen jaren. Ik weet dat jullie ontzettend trots zijn en ben ontzettend dankbaar dat ik deze mijlpaal in jullie aanwezigheid mag vieren.

Lieve Rick, ik ben jou zo ontzettend dankbaar voor alle liefde, steun en het vertrouwen wat je mij al die jaren hebt gegeven, ook al was dat niet altijd makkelijk. Ik hou van jou en kijk uit naar onze toekomst samen!

Curriculum Vitae

Corry Neli (Coreline) Burggraaff was born in Woerden, the Netherlands, on the 18th of February 1990. She graduated cum laude from secondary school in 2008 at the Veenlanden College in Mijdrecht. From 2008 on she studied medicine at the VU University medical center (VUmc) in Amsterdam. During her study she developed special interest in hematology and oncology. After obtaining her medical degree in 2014, she started in 2015 working as a PhD student in the PETRA consortium at the department of hematology in the Amsterdam UMC, location VUmc. The PETRA interim PET lymphoma project under the supervision of prof.dr. J.M. Zijlstra, prof.dr.ir. H.C.W. de Vet, prof.dr. R. Boellaard and prof.dr. O.S. Hoekstra resulted in this PhD thesis. In 2019 she started her residency in internal medicine in the Amsterdam UMC, location VUmc (prof.dr. Y.M. Smulders) and recently she did part of her internal medicine training at the departments of oncology and nephrology in the Spaarne Gasthuis in Hoofddorp and Haarlem (dr. W. de Ronde). Starting from January 2023 Coreline will continue her clinical training during a fellowship hematology at the department of hematology in the Amsterdam UMC.

List of Publications

Zijlstra JM, **Burggraaff CN**, Kersten MJ, Barrington SF; EHA Scientific Working Group on Lymphoma. FDG-PET as a biomarker for early response in diffuse large B-cell lymphoma as well as in Hodgkin lymphoma? Ready for implementation in clinical practice? *Haematologica.* 2016 Nov;101(11):1279-1283.

Burggraaff CN, Cornelisse AC, Hoekstra OS, Lugtenburg PJ, De Keizer B, Arens AIJ, Celik F, Huijbregts JE, De Vet HCW, Zijlstra JM; HOVON Imaging Working Group. Interobserver Agreement of Interim and End-of-Treatment ^{18}F-FDG PET/CT in Diffuse Large B-Cell Lymphoma: Impact on Clinical Practice and Trials. *J Nucl Med.* 2018 Dec;59(12):1831-1836.

Burggraaff CN, de Jong A, Hoekstra OS, Hoetjes NJ, Nievelstein RAJ, Jansma EP, Heymans MW, de Vet HCW, Zijlstra JM. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2019 Jan;46(1):65-79.

Kaalep A, **Burggraaff CN,** Pieplenbosch S, Verwer EE, Sera T, Zijlstra J, Hoekstra OS, Oprea-Lager DE, Boellaard R. Quantitative implications of the updated EARL 2019 PET-CT performance standards. *EJNMMI Phys.* 2019 Dec 26;6(1):28. doi: 10.1186/s40658-019-0257-8.

Pfaehler E, **Burggraaff CN**, Kramer G, Zijlstra J, Hoekstra OS, Jalving M, Noordzij W, Brouwers AH, Stevenson MG, de Jong J, Boellaard R. PET segmentation of bulky tumors: Strategies and workflows to improve inter-observer variability. *PLoS One.* 2020 Mar 30;15(3):e0230901. doi: 10.1371/journal.pone.0230901.

Burggraaff CN, Rahman F, Kaßner I, Pieplenbosch S, Barrington SF, Jauw YWS, Zwezerijnen GJC, Müller S, Hoekstra OS, Zijlstra JM, De Vet HCW, Boellaard R; PETRA Consortium. Optimizing Workflows for Fast and Reliable Metabolic Tumor Volume Measurements in Diffuse Large B Cell Lymphoma. *Mol Imaging Biol.* 2020 Aug;22(4):1102-1110.

Lugtenburg PJ, de Nully Brown P, van der Holt B, D'Amore FA, Koene HR, de Jongh E, Fijnheer R, van Esser JW, Böhmer LH, Pruijt JF, Verhoef

GE, Hoogendoorn M, Bilgin MY, Nijland M, van der Burg-de Graauw NC, Oosterveld M, Jie KG, Larsen TS, van der Poel MW, Leijs MB, Silbermann MH, van Marwijk Kooy M, Beeker A, Kersten MJ, Doorduijn JK, Tick LW, Brouwer RE, Lam KH, **Burggraaff CN**, de Keizer B, Arens AI, de Jong D, Hoekstra OS, Zijlstra-Baalbergen JM. Rituximab-CHOP With Early Rituximab Intensification for Diffuse Large B-Cell Lymphoma: A Randomized Phase III Trial of the HOVON and the Nordic Lymphoma Group (HOVON-84). *J Clin Oncol.* 2020 Oct 10;38(29):3377-3387.

Chamuleau MED, **Burggraaff CN**, Nijland M, Bakunina K, Mous R, Lugtenburg PJ, Dierickx D, van Imhoff GW, Vermaat JSP, Marijt EAF, Visser O, Mandigers C, Bilgin YM, Beeker A, Durian MF, van Rees B, Bohmer LH, Tick LW, Boersma RS, Snijders TJF, Schouten HC, Koene HR, de Jongh E, Hijmering N, Diepstra A, van den Berg A, Arens AIJ, Huijbregts J, Hoekstra O, Zijlstra JM, de Jong D, Kersten MJ. Treatment of patients with MYC rearrangement positive large B-cell lymphoma with R-CHOP plus lenalidomide: results of a multicenter HOVON phase II trial. *Haematologica.* 2020 Dec 1;105(12):2805-2812.

Barrington SF, Zwezerijnen BG, de Vet HC, Heymans MW, Mikhaeel NG, **Burggraaff CN**, Eertink JJ, Pike LC, Hoekstra OS, Zijlstra JM, Boellaard R. Automated Segmentation of Baseline Metabolic Total Tumor Burden in Diffuse Large B-Cell Lymphoma: Which Method Is Most Successful? A Study on Behalf of the PETRA Consortium. *J Nucl Med.* 2021 Mar;62(3):332-337.

Kersten MJ, Driessen J, Zijlstra JM, Plattel WJ, Morschhauser F, Lugtenburg PJ, Brice P, Hutchings M, Gastinne T, Liu R, **Burggraaff CN**, Nijland M, Tonino SH, Arens AIJ, Valkema R, van Tinteren H, Lopez-Yurda M, Diepstra A, De Jong D, Hagenbeek A. Combining brentuximab vedotin with dexamethasone, high-dose cytarabine and cisplatin as salvage treatment in relapsed or refractory Hodgkin lymphoma: the phase II HOVON/LLPC Transplant BRaVE study. *Haematologica.* 2021 Apr 1;106(4):1129-1137.

Eertink JJ, **Burggraaff CN**, Heymans MW, Dührsen U, Hüttmann A, Schmitz C, Müller S, Lugtenburg PJ, Barrington SF, Mikhaeel NG, Carr R, Czibor S, Györke T, Ceriani L, Zucca E, Hutchings M, Kostakoglu L, Loft A, Fanti S, Wiegers SE, Pieplenbosch S, Boellaard R, Hoekstra OS, Zijlstra JM, de Vet HCW. Optimal

timing and criteria of interim PET in DLBCL: a comparative study of 1692 patients. *Blood Adv.* 2021 May 11;5(9):2375-2384.

Zwezerijnen GJC, Eertink JJ, **Burggraaff CN**, Wiegers SE, Shaban EAIN, Pieplenbosch S, Oprea-Lager DE, Lugtenburg PJ, Hoekstra OS, de Vet HCW, Zijlstra JM, Boellaard R. Interobserver Agreement on Automated Metabolic Tumor Volume Measurements of Deauville Score 4 and 5 Lesions at Interim ^{18}F-FDG PET in Diffuse Large B-Cell Lymphoma. *J Nucl Med.* 2021 Nov;62(11):1531-1536.

Eertink JJ, Arens AIJ, Huijbregts JE, Celik F, de Keizer B, Stroobants S, de Jong D, Wiegers SE, Zwezerijnen GJC, **Burggraaff CN**, Boellaard R, de Vet HCW, Hoekstra OS, Lugtenburg PJ, Chamuleau MED, Zijlstra JM; HOVON imaging workgroup. Aberrant patterns of PET response during treatment for DLBCL patients with MYC gene rearrangements. *Eur J Nucl Med Mol Imaging.* 2022 Feb;49(3):943-952.

Greuter M, Eertink JJ, Jongeneel G, Dührsen U, Hüttmann A, Schmitz C, Lugtenburg PJ, Barrington SF, Mikhaeel NG, Ceriani L, Zucca E, Carr R, Györke T, **Burggraaff CN**, de Vet H, Hoekstra OS, Zijlstra JM, Coupé V; PETRA consortium. Cost-Effectiveness of Shortening Treatment Duration Based on Interim PET Outcome in Patients With Diffuse Large B-cell Lymphoma. *Clin Lymphoma Myeloma Leuk*. 2022 Jun;22(6):382-392.

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 Jul;63(7):1001-1007.

List of Abbreviations

[^{18}F]-FDG	^{18}F-fluoro-2-deoxy-D-glucose
95% CI	95% confidence interval
aaIPI	age-adjusted international prognostic index
ABC	activated B-cell
ABVD	doxorubicin, bleomycin, vincristine, dacarbazine
ASCT	autologous stem cell transplantation
AVD	doxorubicin, vincristine, dacarbazine
AUC	area under the curve
BEACOPP	bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, prednisone
BEAM	carmustine, etoposide, cytarabin, melphalan
CAR-T	chimeric antigen receptor T-cells
CD	cluster of differentiation
CHOP	cyclophosphamide, doxorubicin, vincristine, prednisone
CMR	complete metabolic response
CNS	central nervous system
CR	complete remission
CT	computed tomography
CoV	coefficient of variation
DFS	disease-free survival
DICOM	digital imaging and communications in medicine
DLBCL	diffuse large B-cell lymphoma
DS	Deauville 5-point scale
EANM	European Association of Nuclear Medicine
EFS	event-free survival
EoT-PET	end-of-treatment positron emission tomography
FISH	fluorescent in situ hybridization
FFS	failure-free survival
FL	follicular lymphoma
GCB	germinal center B-cell
HL	Hodgkin lymphoma
HOVON	stichting Hemato-Oncologie voor Volwassenen Nederland
HR	hazard ratio
HSROC	hierarchical summary ROC curve

ICC	intraclass correlation coefficient
IHP	international harmonization project
IPD	individual patient data
ITT	intention-to-treat
IPI	international prognostic index
I-PET	interim positron emission tomography
IQR	interquartile range
KM	Kaplan-Meier
LDH	lactate dehydrogenase
MBP	mediastinal blood pool
MTV	metabolic tumor volume
Mo	months
MV	majority vote
NA	negative agreement
NCCN-IPI	national comprehensive cancer network international prognostic index
NFT	no further treatment
NHL	non-Hodgkin lymphoma
NK	natural killer
NM	nuclear medicine
NPV	negative predictive value
OA	overall agreement
OS	overall survival
PA	positive agreement
PD	progressive disease
PET	positron emission tomography
PET/CT	positron emission tomography and computed tomography
PETRA	PET Re-Analysis
PFS	progression-free survival
PMBCL	primary mediastinal B-cell lymphoma
PMD	progressive metabolic disease
PMR	progressive metabolic response
PPV	positive predictive value
PTLD	post-transplant lymphoproliferative disease
R-ACVBP	rituximab, doxorubicin, vindesine, bleomycine, prednisone
R-CHOP	rituximab, cyclophosphamide, doxorubicin, vincristine,

	prednisone
RCT	randomized clinical trial
R-IPI	revised international prognostic index
ROC	receiver operating characteristic
RS	Reed-Sternberg
RT	radiotherapy
SD	standard deviation
SUV	standardized uptake value
TLG	total lesion glycolysis
TraIT	translational research IT
TTP	time to progression
VOI	volume of interest
WHO	world health organization