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

Coreline N. Burggraaff[1], Jakoba J. Eertink[1], Pieternella J. Lugtenburg[2], Otto S. Hoekstra[3], Anne I.J. Arens[4], Bart de Keizer[5], Martijn W. Heymans[6], Bronno van der Holt[7], Sanne E. Wiegers[1], Simone Pieplenbosch[1], Ronald Boellaard[3], Henrica C.W. de Vet[6], and Josée M. Zijlstra[1] on behalf of the HOVON Imaging Working Group and the HOVON Lymphoma Working Group

*[1]Department of Hematology, Amsterdam UMC, Cancer Center Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [2]Department of Hematology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands; [3]Department of Radiology and Nuclear Medicine, Amsterdam UMC, Cancer Center Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [4]Department of Radiology, Nuclear Medicine, and Anatomy, Radboud University Medical Center, Nijmegen, The Netherlands; [5]Department of Radiology and Nuclear Medicine, University Medical Center Utrecht, Utrecht, The Netherlands; [6]Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; and [7]Department of Hematology, HOVON Data Center, Erasmus MC Cancer Institute, Rotterdam, The Netherlands*

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 $SUV_{max}$ [$\Delta SUV_{max}$] 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, $\Delta SUV_{max}$ (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 $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ (4.8% and 53%, respectively) than for DS (3.1% and 38%, respectively). aaIPI and $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ of more than 70% (37% of patients) yielded an NPV of 93%, and the combination of high-intermediate or high aaIPI and a $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ outperformed DS in 2-y PFS prediction.

**Key Words:** DLBCL; PET; Deauville score; $\Delta SUV_{max}$; metabolic tumor volume

Received Feb. 27, 2021; revision accepted Sep. 29, 2021.

For correspondence or reprints, contact Josée M. Zijlstra (j.zijlstra@amsterdamumc.nl).

Published online Oct. 21, 2021.



**J Nucl Med 2022; 63:1001–1007**
DOI: 10.2967/jnumed.121.262205

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 of 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 (Fig. 1) (*14*) 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 group (>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 $SUV_{max}$ exceeded hepatic $SUV_{max}$ by fewer than 3 times, and DS5 was assigned when there were new lymphoma lesions or when tumor $SUV_{max}$ was 3 or more times hepatic $SUV_{max}$ (*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 $SUV_{max}$ between baseline and I-PET4 ($\Delta SUV_{max}$). For DS1, $\Delta SUV_{max}$ was set at 100% reduction (*9*). We applied a prespecified $\Delta SUV_{max}$ 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, $\Delta SUV_{max}$) 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, 2 × 2 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 $\Delta SUV_{max}$ evaluations (Table 1).

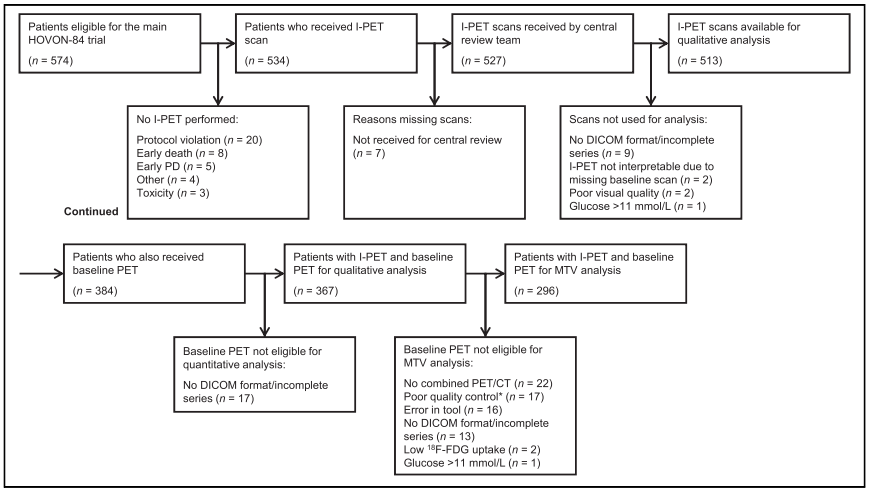


**FIGURE 1.** Flowchart of PET scans available for I-PET4, $\Delta SUV_{max}$, and baseline MTV analyses. *PET quality was acceptable when liver $SUV_{mean}$ was 1.3–3.0 and total image activity was between 50%–80% of total injected dose. PD = progressive disease.

### 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

**TABLE 1**
Baseline Patient Characteristics

| Characteristic | I-PET4 | $\Delta SUV_{max}$ | MTV |
|---|---|---|---|
| Number of patients | 513 (100) | 367 (100) | 296 (100) |
| Age at diagnosis (y) | | | |
| Median | 65 (range, 23–80) | 65 (range, 23–80) | 65 (range, 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 | 20 (IQR, 13–28) | 20 (IQR, 13–28) | 20 (IQR, 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.

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 for 474 patients (Supplemental Table 2), with an overall agreement of 87% (95% CI, 84%–90%).

### 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,

**TABLE 2**
Diagnostic and Prognostic Measures for aaIPI, Baseline MTV, Different Cutoffs of Deauville 5-Point Scale at I-PET4, and $\Delta SUV_{max}$ for 2-Year PFS

| Measure | Parameter | Patients (*n*) | Diagnostic information | | | | Prognostic information | | Discrimination (AUC) |
|---|---|---|---|---|---|---|---|---|---|
| | | | NPV | PPV | Sensitivity | Specificity | Univariate HR | *P* | |
| 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) |
| $\Delta SUV_{max}$ | >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 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.

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 (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.

$\Delta SUV_{max}$ analysis was feasible in 367 of 574 patients (64%; Fig. 1). In patients with no more than a 70% $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ was compared with the most commonly used DS cutoff, DS4–5, $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ 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 in $\Delta SUV_{max}$ were independently associated with 2-y PFS (Supplemental Table 4). A low or low-intermediate aaIPI and a $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ of 70% or less (6% of patients) resulted in a PPV of 65% (95% CI, 45%–81%; Supplemental Fig. 3).

Dichotomized baseline MTV did not add prognostic value to $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ as factors independently associated 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

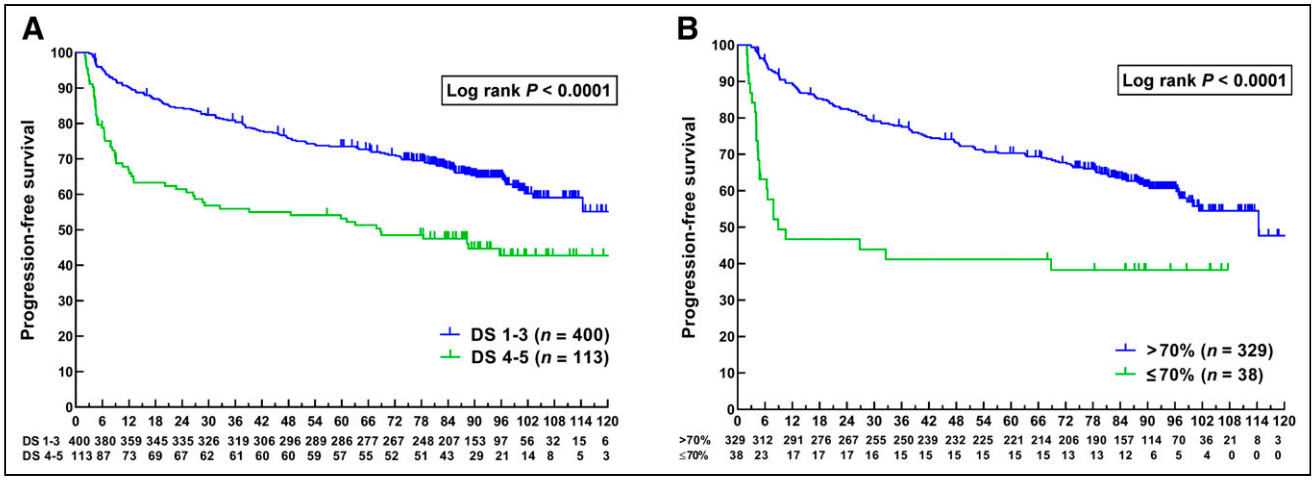


**FIGURE 2.** Kaplan–Meier curves with numbers at risk for PFS in months stratified by I-PET4 result according to DS (A) and according to $\Delta SUV_{max}$ result (B).

I-PET response criteria. However, the PPV was still relatively low. Combining clinical and PET data showed that aaIPI and $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ response criteria. As a secondary objective, we compared the most commonly used visual and semiquantitative criteria and externally validated that $\Delta SUV_{max}$ 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 a 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 $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ 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 reported a more favorable PPV for $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ (reflecting chemosensitivity) were both independent predictors of 2-y PFS. Again, the subgroup with both high-intermediate or high aaIPI and a $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$) 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 $\Delta SUV_{max}$ (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 Tables 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 (345-$cm^3$ cutoff) to $\Delta SUV_{max}$ did not improve the predictive value, but log-transformed continuous MTV added some independent predictive value when combined with $\Delta SUV_{max}$. 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 $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ 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 $\Delta SUV_{max}$ 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 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 in favor of 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 $\geq$ 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 significant difference in 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 of R-CHOP as the new standard for most patients with limited-stage disease (*32*).

## CONCLUSION

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 $\Delta SUV_{max}$ cutoffs, the optimal response criterion for I-PET4 to predict 2-y PFS was $\Delta SUV_{max}$.

## 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 $\Delta SUV_{max}$ were independent predictors for 2-y PFS in DLBCL. Six percent of patients had a high PPV of 65% resulting in poor survival outcome. $\Delta SUV_{max}$ 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 $SUV_{max}$ 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.* 2019;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, 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 $\Delta SUV_{max}$ 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.* 2014;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 [abstract]. *Blood.* 2018;132(suppl 1):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.