CT-Based Versus FDG-PET/CT–Based NCCN International Prognostic Index Risk Stratification in DLBCL

Hugo J.A. Adams, MD; John M.H. de Klerk, MD, PhD; Rob Fijnheer, MD, PhD; Stefan V. Dubois, MD; Rutger A.J. Nievelstein, MD, PhD; and Thomas C. Kwee, MD, PhD

Abstract

Background: This study compared CT-based and 18F-fluoro-2-deoxy-D-glucose PET/CT (FDG-PET/CT)–based NCCN International Prognostic Index (NCCN-IPI) risk stratification in newly diagnosed diffuse large B-cell lymphoma (DLBCL). Materials and Methods: This retrospective study included 57 patients with newly diagnosed DLBCL who had undergone both (oral and intravenous contrast-enhanced full-dose) diagnostic CT and FDG-PET/CT. Diagnostic CT only and FDG-PET/CT were evaluated separately, and corresponding NCCN-IPI scores for the 2 datasets (NCCN-IPI\textsubscript{CT} and NCCN-IPI\textsubscript{PET/CT}) were calculated. Percentages of agreement and weighted $k$ statistic between NCCN-IPI\textsubscript{CT} and NCCN-IPI\textsubscript{PET/CT} scoring with regard to the formation of low-, low-intermediate–, high-intermediate–, and high-risk groups were calculated. Results: In 47 of 57 patients (82.5%; 95% CI, 70.4–90.4), diagnostic CT alone was in agreement with FDG-PET/CT with regard to the formation of low-, low-intermediate–, high-intermediate–, and high-risk NCCN-IPI groups, but not in the remaining 10 patients (17.5%; 95% CI, 9.6%–29.6%). All NCCN-IPI disagreements between diagnostic CT and FDG-PET/CT were from the detection of additional lesions by the latter, most of them being bone marrow lesions. Agreement between NCCN-IPI\textsubscript{CT} and NCCN-IPI\textsubscript{PET/CT} with regard to the formation of low-, low-intermediate–, high-intermediate–, and high-risk groups was considered good ($k$=0.771). Conclusions: Although agreement between NCCN-IPI\textsubscript{CT} and NCCN-IPI\textsubscript{PET/CT} risk stratification is generally good, FDG-PET/CT results in higher NCCN-IPI risk stratifications in a non-negligible proportion of patients. Future studies should investigate the prognostic implications of these imaging-based differences in NCCN-IPI scoring. (J Natl Compr Canc Netw 2015;13:171–176)

Background

Non-Hodgkin’s lymphomas (NHLs) account for 5% of all malignancies and are the sixth most common cancer in developed countries.\textsuperscript{1} Diffuse large B-cell lymphoma (DLBCL) represents 30% to 35% of all NHL subtypes.\textsuperscript{2} The International Prognostic Index (IPI) has been the most widely used risk stratification index for aggressive NHL since it was published in 1993.\textsuperscript{3} Recently, however, a new NCCN International Prognostic Index (NCCN-IPI) was developed, which is exclusively applicable to patients with DLBCL treated with R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone) chemotherapy.\textsuperscript{4} The NCCN-IPI incorporates the following parameters: age, lactate dehydrogenase (LDH) ratio, extranodal disease in major organs, Ann Arbor stage, and ECOG performance status. Unfortunately, the NCCN-IPI does not specify which imaging modality should be used to detect extranodal disease in major organs and determine Ann Arbor stage.\textsuperscript{4} CT and 18F-fluoro-2-deoxy-D-glucose PET/CT (FDG-PET/CT) are the most commonly used imaging modalities for this purpose.\textsuperscript{5} According to the 2007 guidelines published by the Imaging Subcommittee of the International Harmonization Project in Lymphoma, pretreatment FDG-PET is recommended but not obligatory for the assessment of response after treatment in DLBCL.\textsuperscript{6} Consequently, variability exists among different institutions with regard to the use of either CT alone or (more expensive) FDG-PET/CT for...
pretreatment evaluation of DLBCL. CT and FDG-PET/CT have different diagnostic performances,5 which may cause variability in the final NCCN-IPI score.7 Whether the differences between CT and FDG-PET/CT affect risk stratification is still unknown. This knowledge is crucial to validate and standardize the method of pretreatment imaging in DLBCL. Therefore, the purpose of this study was to compare CT-based versus FDG-PET/CT–based NCCN-IPI risk stratification in newly diagnosed DLBCL.

Materials and Methods

Study Design and Patients

The local Institutional Review Board approved this retrospective study and granted a waiver for informed consent. All patients with newly diagnosed DLBCL routinely undergo pretreatment FDG-PET/CT at our institution, and these patients were identified from the hospital’s database spanning the period from September 2007 to December 2013. Inclusion criteria were patients with newly diagnosed, histologically proven DLBCL who had undergone both (oral and intravenous contrast-enhanced full-dose) diagnostic CT and low-dose FDG-PET/CT, from mid-femur to base of skull, with use of the same PET/CT system, on the same day, and who had also undergone bone marrow biopsy (BMB) and serum LDH measurement. Exclusion criteria were primary mediastinal DLBCL, previously treated/relapsed lymphoma, transformed lymphoma, coexistence of another lymphoma subtype in the diagnostic biopsy, another malignancy within the past 5 years, start of anticancer therapy before diagnostic CT/low-dose FDG-PET/CT, nonavailability of either diagnostic CT or low-dose FDG-PET/CT performed with use of the same PET/CT system on the same day, suboptimal or incomplete diagnostic CT examination (either from nonadministration of oral or intravenous CT contrast agents or from incomplete body coverage [ie, less than mid-femur to base of skull]), nonavailability of BMB, and nonavailability of serum LDH measurement.

Diagnostic CT and Low-Dose FDG-PET/CT

Diagnostic CT and low-dose PET/CT images were acquired with the use of a 40-detector row PET/CT scanner (Biograph 40 TruePoint PET/CT, Siemens Healthcare, Malvern, PA). Patients ingested an oral CT contrast agent and fasted for 6 hours before 3 MBq per kilogram of body weight of FDG was injected into an antecubital vein. Blood glucose levels were checked to be less than 11 mmol/L before FDG administration. Sixty minutes after FDG injection, low-dose FDG-PET/CT and diagnostic CT imaging were performed in 3 steps. First, low-dose CT images were acquired with the following settings: 120 kV, 26 to 30 mAs (automatic dose modulation), 0.8-second tube rotation time, pitch of 1.2, and 1.5-mm slice width (reconstructed to contiguous 5-mm axial slices to match the slice thickness of the PET images). Second, PET scanning was performed from mid-femur to base of skull in 5 or 6 bed positions, with 3 minutes per bed position. Low-dose CT data were used for PET attenuation correction. PET images were reconstructed with an ordered-subset expectation maximization algorithm for 14 subsets and 4 iterations. The image reconstruction matrix was 128 × 128. Third, a nonionic iodinated contrast agent (Xenetix 300, Guerbet, Villepinte, France; 3 mL/s with bolus tracking) was administered intravenously, and a full-dose CT of neck, chest, abdomen, and pelvis (portal venous phase) was performed with the following settings: 120 kV, 60 to 160 mAs (automatic dose modulation), 0.8-second tube rotation time, pitch of 1.2, and 1.5-mm slice width.

Image Analysis

An experienced reader (T.C.K.) evaluated 2 separate datasets in each patient: (1) diagnostic CT only and (2) the combination of diagnostic CT and low-dose FDG-PET/CT images (simply referred to as “FDG-PET/CT” in this article). The interval between the reading sessions of these 2 datasets was at least 4 months in order to avoid recognition bias. During the evaluation of diagnostic CT images only, the reader was blinded to FDG-PET/CT findings. In addition, during both reading sessions, the observer was blinded to clinical and laboratory findings, other imaging modalities, BMB findings, and patient outcome. Standardized scoring forms were used for the assessment of nodal and extranodal disease. For the evaluation of diagnostic CT only, lymph nodes with a short-axis diameter greater than 10 mm in the axial plane were considered positive for lymphoma. Extranodal disease was considered present in case of any area of abnormal attenuation in the spleen, bone or bone marrow, and liver; nodule
or infiltration in the lung; and mass of soft tissue attenuation (relative to the surrounding tissue) in other extranodal sites. Splenomegaly (ie, splenic index >725 cm³ [splenic index = length × height × thickness at the hilum]) was considered positive for lymphoma, whereas hepatomegaly (without any focal liver lesions) was not.

For the evaluation of FDG-PET/CT, FDG uptake exceeding background FDG uptake in any location incompatible with normal anatomic and physiologic conditions was considered positive for lymphoma, regardless of diagnostic CT findings. Lymph nodes with a short-axis diameter greater than 10 mm in the axial plane at diagnostic CT were only considered positive for lymphoma if increased FDG uptake was also observed in these lymph nodes, otherwise they were classified as negative for lymphoma. Extranodal sites classified as positive at diagnostic CT but without any increased FDG uptake were considered negative for lymphoma, if the size of those extranodal sites exceeded the spatial resolution of PET (ie, greater than ≈10 mm). Extranodal sites classified as positive at diagnostic CT but without any increased FDG uptake were considered positive for lymphoma, if the size of those extranodal sites was below the spatial resolution of PET (ie, smaller than ≈10 mm).

Presence of extranodal disease in major organs (bone marrow, central nervous system, liver/gastrointestinal tract or lung) and Ann Arbor stage (stage I–IV)⁹ according to both diagnostic CT only and FDG-PET/CT was recorded.

NCCN-IPI Score
The NCCN-IPI uses a maximum of 8 scoring points for categorized age older than 40 to 60 years (1 point), older than 60 to 75 years (2 points), and older than 75 years (3 points); LDH ratio greater than 1 (1 point) and greater than 3 (2 points) times the upper limit of normal, extranodal disease in major organs (either bone marrow, central nervous system, liver/gastrointestinal tract, or lung; 1 point), Ann Arbor stage III/IV (1 point), and ECOG performance status (≥2; 1 point).³ Subsequently, 4 risk groups can be formed: low (0–1), low-intermediate (2–3), high-intermediate (4–5), and high (6–8).³ Age, ECOG performance status, and LDH levels of each patient were recorded at the time of diagnosis. Subsequently, 2 NCCN-IPI scores were calculated in each patient: 1 with use of diagnostic CT findings and 1 with use of FDG-PET/CT findings for extranodal disease in major organs (bone marrow, central nervous system, liver/gastrointestinal tract, or lung) and Ann Arbor stage III/IV disease (NCCN-IPI_C and NCCN-IPI_PETCT respectively). The location of the diagnostic biopsy and BMB results were also used to determine the final NCCN-IPI scores.

Statistical Analysis
Percentages of agreement and disagreement between diagnostic CT alone versus FDG-PET/CT with regard to the detection of extranodal disease in major organs (bone marrow, central nervous system, liver/gastrointestinal tract, or lung), the detection of Ann Arbor stage III/IV disease, and the formation of low, low-intermediate–, high-intermediate–, and high-risk NCCN-IPI groups were calculated, along with binomial exact 95% CIs. In addition, agreement between NCCN-IPI_C and NCCN-IPI_PETCT scoring with regard to the formation of low-, low-intermediate–, high-intermediate–, and high-risk groups was analyzed using the weighted k statistic, defined as poor (<0.2), fair (>0.2–0.4), moderate (>0.4–0.6), good (>0.6–0.8), and very good (>0.8–1.0) agreement. Statistical analyses were executed using MedCalc statistical software version 12.6.0 (Ostend, Belgium).

Results
Patients
Between September 2007 and December 2013, a total of 121 patients with newly diagnosed DLBCL were potentially eligible for inclusion. Of these 121 patients, 2 were excluded because of primary mediastinal DLBCL, 4 because of previously treated/relapsed lymphoma, 14 because of a transformed lymphoma, 6 because of coexistence of another lymphoma subtype in the diagnostic biopsy, 1 because of another malignancy within the past 5 years, 25 because of a suboptimal or incomplete diagnostic CT examination (nonadministration of intravenous CT contrast agents in 5 cases and coverage of only neck and chest/abdomen in 20 cases), and 12 because of absence of BMB. Thus, the final study population included 57 patients (34 men and 23 women; mean age, 67.4 years [range, 28–88 years]). Further patient characteristics are shown in Table 1.

Diagnostic CT Versus FDG-PET/CT
In 41 of 57 patients (71.9%; 95% CI, 59.1%–82.0%), diagnostic CT alone was in agreement with FDG-
PET/CT with regard to the detection of extranodal disease in major organs (bone marrow, central nervous system, liver/gastrointestinal tract, or lung). Disagreements between diagnostic CT alone and FDG-PET/CT with regard to the detection of extranodal disease in the remaining 16 patients (28.1%; 95% CI, 18.0%–40.9%) were from the detection of additional lesions by FDG-PET/CT in the bone marrow (n=9), gastrointestinal tract (n=5), and lung (n=2) (Figures 1 and 2).

In 45 patients (79.0%; 95% CI, 66.6%–87.7%), diagnostic CT alone was in agreement with FDG-PET/CT with regard to the detection of Ann Arbor stage III/IV disease. Disagreements between diagnostic CT alone and FDG-PET/CT with regard to the detection of Ann Arbor stage III/IV disease in the remaining 12 patients (21.1%; 95% CI, 12.3%–33.4%) were from the detection of additional lesions by FDG-PET/CT in the bone marrow (n=4), gastrointestinal tract (n=3), lung (n=1), testes (n=1), perineum (n=1), muscles (n=1), and cervical lymph nodes (n=1) (Figures 1 and 2).

In 47 of 57 patients (82.5%; 95% CI, 70.4–90.4), diagnostic CT alone was in agreement with FDG-PET/CT regarding the formation of low-, low-intermediate–, high-intermediate–, and high-risk NCCN-IPI groups (Table 2). The disagreements between CT and FDG-PET/CT in the remaining 10 patients (17.5%; 95% CI, 9.6%–29.6%) were always from the classification of higher FDG-PET/CT–based NCCN-IPI risk groups (Table 2). Agreement between NCCN-IPI$_{CT}$ and NCCN-IPI$_{PET/CT}$ with regard to the formation of low-, low-intermediate–, high-intermediate–, and high-risk groups was considered good ($k=0.771$).

### Discussion

The results of this study show that diagnostic CT and FDG-PET/CT are in agreement with regard to NCCN-IPI risk stratification in most cases. However, disagreements between the 2 imaging modalities occur in approximately 17.5% of patients and are always from the detection of additional lesions and subsequent higher risk stratification by FDG-PET/CT. Most of these additionally detected lesions by FDG-PET/CT are located in the bone marrow, followed by the gastrointestinal tract. The fact that FDG-PET/CT detects bone marrow involvement more often is not surprising, given the insensitivity of CT for the evaluation of disease involvement in this organ.\(^{10}\) However, although FDG-PET/CT detected bone marrow involvement in several cases that were not detected by CT, whether this will result in a better risk stratification remains unknown, because it has been reported in the literature that FDG-PET/CT–based bone marrow assessment is of inferior prognostic value compared with BMB.\(^{11–13}\) Unfortunately, the number of patients in this study was too low to assess the prognostic implications of the differences in CT- and FDG-PET/CT–based NCCN-IPI scoring. Future studies with larger sample sizes are required to investigate the effect of differences in imaging-based NCCN-IPI risk stratification on prognosis.

Risk stratification indices for aggressive NHL, including DLBCL, have existed for more than 2 decades.\(^{3,4,14}\) Unfortunately, similar to its predecessors (ie, the original 1993 IPI study and the 2007 R-IPI study), the 2013 NCCN-IPI study did not define which imaging modalities should be used to detect major extranodal disease and to determine Ann Arbor stage.\(^{3,4,7,14}\) In addition, there is a lack of previous

### Table 1 Patients Characteristics (N=57)

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ± SD (y)</td>
<td>67.4 ± 11.7</td>
<td></td>
</tr>
<tr>
<td>Median (y)</td>
<td>68.0</td>
<td></td>
</tr>
<tr>
<td>Range (y)</td>
<td>28–88</td>
<td></td>
</tr>
</tbody>
</table>

| Male/Female | 34/23 |

| NCCN-IPI Factors |         |         |
| Age              |         |         |
| ≤40 y            | 1       |
| >40 to ≤60 y     | 13      |
| >60 to ≤75 y     | 28      |
| >75 y            | 15      |
| ECOG PS >1       | 10      |
| LDH >ULN         |         |         |
| ≤1               | 16      |
| >1 to ≤3         | 34      |
| >3               | 7       |

Abbreviations: BMB+, positive bone marrow biopsy; LDH, lactate dehydrogenase; NCCN-IPI, NCCN International Prognostic Index; PS, performance status; ULN, upper limit of normal (according to reference value at our institution).
studies investigating the influence of different imaging modalities on risk stratification and prognostic assessment. The findings of the present study indicate a nonnegligible influence of diagnostic differences between CT and FDG-PET/CT on risk stratification. The optimal method for pretreatment imaging of DLBCL (whether diagnostic CT alone, FDG-PET/CT, MRI, or FDG-PET/MRI) is under continuing development, and surely goes beyond mere integration into the existing NCCN-IPI. Factors such as cost-effectiveness and the ability of pretreatment imaging to facilitate interim and posttreatment evaluation should also be taken into account.

This study had several limitations. First, several patients had to be excluded, most because no intravenous contrast-enhanced whole-body CT was performed. The reason for performing diagnostic CT before low-dose FDG-PET/CT was unawareness of the diagnosis of DLBCL at the time of diagnostic CT. If the diagnosis DLBCL was known before imaging, the 2 examinations would have been performed using the same PET/CT system, on the same day. To ensure scan protocol homogeneity and avoid the risk of disease progression bias, these patients had to be excluded. Nevertheless, it is unlikely that this has introduced any relevant patient selection bias. Second, the same reader evaluated both diagnostic CT and FDG-PET/CT scans. However, the interval between the 2 reading sessions was at least 4 months, and the reader may have encountered a learning curve effect. Finally, the study did not include patients with extranodal disease, but the findings may influence patient management in this setting.

Table 2  Agreement and Disagreement Between CT-Based Versus FDG-PET/CT-Based NCCN-IPI Scoring in Patients With DLBCL (N=57)

<table>
<thead>
<tr>
<th>Risk</th>
<th>FDG-PET/CT Low</th>
<th>FDG-PET/CT Low-Intermediate</th>
<th>FDG-PET/CT High-Intermediate</th>
<th>FDG-PET/CT High</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT low</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CT low-intermediate</td>
<td>0</td>
<td>8</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>CT high-intermediate</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>CT high</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

Figure 1  Example of disagreement between diagnostic CT and FDG-PET/CT with regard to the detection of extranodal disease in major organs and Ann Arbor stage III/IV disease in a 74-year-old man with newly diagnosed diffuse large B-cell lymphoma. (A, B) Diagnostic CT (B with bone window settings) shows an enlarged right internal iliac lymph node (arrowhead), but was otherwise unremarkable. Bone marrow biopsy showed no lymphomatous marrow involvement. Consequently, CT-based Ann Arbor stage was I, without any major extranodal disease. Corresponding fused FDG-PET/CT images (C,D) confirm right internal iliac lymph node involvement (C, arrowhead), but also shows bone marrow lesions in the sacrum and right iliac wing (dashed arrows; invisible at CT). Consequently, FDG-PET/CT–based Ann Arbor stage was IV, with major extranodal disease.

Figure 2  Example of disagreement between diagnostic CT and FDG-PET/CT with regard to the detection of extranodal disease in major organs and Ann Arbor stage III/IV disease in a 71-year-old woman with newly diagnosed diffuse large B-cell lymphoma. Diagnostic CT (A) shows an enlarged spleen with central necrosis (continuous arrow), indicating lymphomatous involvement. The stomach appears normal (dashed arrow). Diagnostic CT did not detect any other lesions in this patient, as a result of which CT-based Ann Arbor stage was I, without any major extranodal disease. Corresponding low-dose CT (B), FDG-PET (C), and fused FDG-PET/CT (D) confirm lymphomatous involvement of the spleen (continuous arrows), but also show pathologic FDG uptake of the stomach (dashed arrows), which appeared to involve the entire stomach (not shown). Consequently, FDG-PET/CT–based Ann Arbor stage was IV, with major extranodal disease.
which eliminated the risk of recognition bias. In addition, this approach avoided the influence of interobserver variability. Third, although FDG-PET imaging seems to have a definite advantage in detecting bone marrow disease not seen on CT, the differences in identifying gastrointestinal disease or splenic disease could also partly be related to reader interpretation and the quality of the CT scan obtained. Fourth, the FDG-PET/CT evaluation included assessment of diagnostic CT and low-dose FDG-PET/CT images. Although the prognostic performance of low-dose FDG-PET/CT–based NCCN-IPI scoring was not investigated, previous work has already shown no significant differences between unenhanced low-dose FDG-PET/CT and contrast-enhanced full-dose FDG-PET/CT with regard to both nodal assessment and organ evaluation. Fifth, as mentioned previously, the prognostic implications of the differences in NCCN-IPI risk stratification between CT and FDG-PET/CT could not be assessed because of the limited number of patients in this study.

Conclusions

Although agreement between NCCN-IPI$_{CT}$ and NCCN-IPI$_{PET/CT}$ risk stratification is generally good, FDG-PET/CT results in higher NCCN-IPI risk stratifications in a nonnegligible proportion of patients. Future studies should investigate the prognostic implications of these imaging-based differences in NCCN-IPI scoring.

References