Respective prognostic values of germinal center phenotype and early $^{18}$fluorodeoxyglucose-positron emission tomography scanning in previously untreated patients with diffuse large B-cell lymphoma

Jehan Dupuis, Philippe Gaulard, Francois Hemery, Emmanuel Itti, Christian Gisselbrecht, Alain Rahmouni, Christiane Copie-Bergman, Josette Brière, Taoufik El Gnaoui, Isabelle Gaillard, Michel Meignan, Corinne Haioun

Background and Objectives
Diffuse large B-cell lymphomas (DLBCL) have a variable outcome, and powerful methods of prognostication are needed in order to choose the best treatment for each patient. Immunophenotypic classification of the tumor as germinal center (GC) or non-germinal center-like (nGC) and early response evaluation with $^{18}$fluorodeoxyglucose positron emission tomography ($^{18}$FDG-PET) scanning have been correlated with survival in DLBCL but the two methods have never been evaluated simultaneously in the same patient population. Our aim was to investigate their respective prognostic values in the same series of patients.

Design and Methods
We investigated the expression of CD10, Bcl-6, and MUM1 in 81 patients with DLBCL evaluated early with $^{18}$FDG-PET. The tumors were classified as GC or nGC using the algorithm of Hans et al. The results of both methods were correlated with the patients’ characteristics and survival.

Results
CD10 was positive in 27/76 (36%), Bcl-6 in 43/74 (58%), and MUM1 in 33/73 (45%) interpretable cases. Thirty-eight (51%) were in the GC group, and 36 (49%) in the nGC group. With a median follow-up of 33 months, estimated 3-year event-free survival (EFS) of the whole population was 67%. There was no influence of GC/nGC phenotype on survival. Three-year EFS was 46% in the early PET-positive group versus 80% in the PET-negative group ($p=0.0003$).

Interpretation and Conclusions
The prognostic value of GC/nGC phenotype is not confirmed in this heterogeneous series, whereas early PET findings are confirmed to be a powerful predictor of outcome. The impact of treatment decisions based on early PET results should be evaluated.

Key words: diffuse large B-cell lymphoma, germinal center, immunohistochemistry, PET scan, prognosis.

Haematologica 2007; 92:778-783
©2007 Ferrata Storti Foundation
Diffuse large B-cell lymphoma (DLBCL), despite being considered a single entity within the WHO classification of lymphoid tumors,\(^1\) is a heterogeneous disease in terms of clinical presentation, histopathology, and outcome. As new treatments appear, there is increasing interest in identifying the patients in whom conventional approach is likely to fail. Risk stratification currently relies mainly on the International Prognostic Index (IPI).\(^2\) This approach has proven useful in identifying high-risk patients who could benefit from consolidative high-dose therapy (HDT) after having reached a first remission.\(^3,4\) Among other parameters, expression of Bcl-2 can also predict outcome in DLBCL.\(^5,6\)

Gene-expression profiling in DLBCL has brought an insight into the biological heterogeneity of the disease. Major subgroups were identified: germinal center B cell-like (GC), activated B cell-like or non-GC (nGC), the former showing a better prognosis.\(^7,8\) Immunohistochemistry has been evaluated as a surrogate for this molecular classification.\(^9-14\) The phenotypic GC and nGC groups are defined by the expression of GC or post-GC stage markers. While using variable definitions, some\(^9,11,12\), but not all studies\(^10,13,14\), found a better prognosis for phenotypically-defined GC cases.\(^15\) Fluorodeoxyglucose positron emission tomography (\(^{18}\)FDG-PET) scanning performed after a few cycles of chemotherapy has been shown to predict treatment outcome.\(^15,20\) In our recently published series of 90 patients,\(^15\) event-free and overall survival differed significantly between patients with a negative versus positive \(^{18}\)FDG-PET after two cycles of chemotherapy, independently from the IPI score.

Both methods appear promising in order to establish optimal risk-based treatment strategies, but, to the best of our knowledge, they have never been compared within the same patient population. Our objective was to conduct such a comparison in order to determine which method performs better in the clinical setting. We retrospectively studied the expression of Bcl-2, CD10, Bcl-6 and MUM1 on biopsies from 81 consecutive patients with DLBCL who had been prospectively investigated with early \(^{18}\)FDG-PET. We aimed at evaluating the prognostic impact of immunophenotype (phenotypic classification into GC and nGC groups and Bcl-2 expression), along with that of early \(^{18}\)FDG-PET imaging.

**Design and Methods**

**Selection of patients**

We retrospectively performed immunohistochemical studies on available paraffin-embedded diagnostic material from 77 DLBCL patients who had been prospectively included between January 2000 and January 2004 in our previous \(^{18}\)FDG-PET study.\(^15\) Four additional patients were also studied, two of whom had been recruited in an extension of this study until September 2004; the two others had not been included in our previous report because they had no adverse prognostic factors of the age-adjusted IPI.

Inclusion criteria were age under 80 years, a centrally reviewed diagnosis of DLBCL, measurable disease, ECOG performance status of 0 to 2, and availability of paraffin-embedded tissue for immunohistochemical analysis. Patients with central nervous system involvement, positive human immunodeficiency serology, concomitant or previous cancer (except carcinoma \textit{in situ} of the cervix), or any serious concomitant disease contraindicating chemotherapy were not included.

According to the declaration of Helsinki, the protocol was approved by our Institutional Review Board and all patients gave written informed consent. The study was sponsored by the Délégation à la Recherche Clinique of the Assistance Publique – Hôpitaux de Paris.

**Pretreatment evaluation and follow-up**

Before treatment, all patients were evaluated by physical examination, complete blood counts, routine chemistry including measurement of lactate dehydrogenase (LDH) levels, computed tomographic scan of the thorax, abdomen and pelvis, and bone marrow biopsy. Restaging was performed after the first two and four cycles of induction, at the end of treatment, then every 6 months for 2 years, and then yearly. Responses were classified according to the International Workshop criteria.\(^21\) Additionally, all patients underwent whole-body \(^{18}\)FDG-PET examination before starting treatment and after the first two chemotherapy cycles (see section \(^{18}\)FDG-PET).

**Treatment**

Forty-four patients (54%) were treated within randomized clinical trials conducted by the Groupe d’Etude des Lymphomes de l’Adulte (GELA). Induction treatment always included an anthracycline-based regimen, which was either CHOP (doxorubicin 50 mg/m\(^2\) day 1, cyclophosphamide 750 mg/m\(^2\) day 1, vincristine 1.4 mg/m\(^2\) day 1 and prednisone 40 mg/m\(^2\) days 1-5, repeated every 21 days for 4 courses, n=23) or one of the dose-intensified ACBVP (doxorubicin 75 mg/m\(^2\) day 1, cyclophosphamide 1,200 mg/m\(^2\) day 1, vindesine 2 mg/m\(^2\) days 1 and 5, bleomycin 10 mg days 1 and 5 and prednisone 60 mg/m\(^2\) days 1-5, every 15 days for four courses, n=48) or AC/ACE (doxorubicin 75 mg/m\(^2\) day 1, cyclophosphamide 1,000 mg/m\(^2\) day 1, and prednisone 60 mg/m\(^2\) days 1-5, for one course followed by three courses repeated every 15 days of the same drugs plus etoposide 150 mg/m\(^2\), n=10) regimens. Thirty-three patients (41%) received consolidative HDT after having reached CR, the others received CHOP-based or ACBVP-type* sequential consolidation. Consolidative HDT was given to younger patients with two or three age-adjusted IPI factors at diagnosis,\(^14\) within or outside protocols (n=28), and to patients with one age-adjusted IPI factor and high Bcl-2 expression within the LNH98-2 GELA protocol.\(^22\) Thirty-seven patients (46%) received rituximab as a part of their treatment. Clinicians taking care of patients were blinded to the results of early \(^{18}\)FDG-PET imaging.
and treatment decisions were taken only on the basis of conventional staging methods.

**18FDG-PET**

Modalities of 18FDG-PET image acquisition were as previously described. Images were interpreted by a consensus of two experienced observers blinded to clinical and radiological data. All foci of abnormal FDG uptake were scored for their extent and intensity using a three-point scale (1=low, 2=moderate, 3=high) within each lymphatic area, organ, and skeletal region. Then, each post-chemotherapy scan was scored as negative or positive. Negative was defined as having no residual abnormal uptake or as having a unique residual site (with an extent score of 1 associated with an intensity score of 1), while all the other previously hyper-metabolic sites were extinguished. This approach was successfully used by Mikhaeel and co-workers in a previous study. Positive was defined as having at least one residual site (with an extent score of 1) associated with an intensity score of 2 or 3, or as having two or more residual sites with any score of extent and intensity.

**Immunohistochemical studies**

All immunohistochemical studies were performed in the same laboratory under standardized conditions. Deparaffinized tissue sections were immunostained with antibodies including CD10 (56C6, Novocastra, Newcastle, UK), Bcl-2 (clone 124), Bcl-6 (P1F6) and MUM1/IRF4 (MUM1p) (DakoCytomation, Glostrup, Denmark) using an indirect immunoperoxidase method with a manual technique (Bcl-2, CD10) or an automated immunostainer (Ventana medical systems, Tucson, AZ, USA) (Bcl-6, MUM1). Antigen retrieval involved microwave heating pretreatment with three cycles of 5 minutes in 0.01M citrate buffer, pH 7.6 for Bcl-2, CD10 and Bcl-6 or in EDTA buffer pH 9 for MUM1. Positivity was rated independently by two observers (PG, JD) and discordant cases were resolved by review on a multiheaded microscope. We used the thresholds of 50% positive cells for Bcl-2 and 30% for CD10, Bcl-6 and MUM1. Cases without any internal positive control were scored as not informative for the corresponding antibody. Patients were classified as having GC or nGC disease following the algorithm of Hans et al. (Figure 1A). Differences between the results of comparative tests were considered statistically significant at a two-sided \( p < 0.05 \). All statistical analyses were performed using Statistical Application System software (SAS, version 9, SAS Institute, Cary, NC, USA).

**Results**

Demographic and pretreatment characteristics are shown in Table 1. The median age was 52 years (26 – 79), and the male/female ratio was 1.72. Patients were mostly young (72% under 60) with a good performance status (72%). Nevertheless, 58% were in the high or intermediate-high IPI risk groups, essentially because of advanced disease stage (90% stages III – IV) and high LDH level (62%). Diagnoses had been made on nodal (n=45), extranodal (n=25) or mediastinal (n=11) specimens. Bcl-2 immunostaining gave interpretable results in 80/81 (99%), CD10 in 76/81 (94%), Bcl-6 in 74/81 (91%), and MUM1 in 73/81 (90%) cases. Thus, 91% of patients could be classified as having a GC or nGC phenotypic profile. Bcl-2 was positive in 42/80 (53%), CD10 in 76/81 (94%), Bcl-6 in 74/81 (91%), and MUM1 in 73/81 (90%) cases. Therefore, 91% of patients could be classified as having a GC or nGC phenotypic profile. Bcl-2 was positive in 42/80 (53%), CD10 in 76/81 (94%), Bcl-6 in 74/81 (91%), and MUM1 in 73/81 (90%) cases. Thus, 91% of patients could be classified as having a GC or nGC phenotypic profile. Bcl-2 was positive in 42/80 (53%), CD10 in 76/81 (94%), Bcl-6 in 74/81 (91%), and MUM1 in 73/81 (90%) cases. Thus, 91% of patients could be classified as having a GC or nGC phenotypic profile. Bcl-2 was positive in 42/80 (53%), CD10 in 76/81 (94%), Bcl-6 in 74/81 (91%), and MUM1 in 73/81 (90%) cases. Thus, 91% of patients could be classified as having a GC or nGC phenotypic profile. Bcl-2 was positive in 42/80 (53%), CD10 in 76/81 (94%), Bcl-6 in 74/81 (91%), and MUM1 in 73/81 (90%) cases. Thus, 91% of patients could be classified as having a GC or nGC phenotypic profile. Bcl-2 was positive in 42/80 (53%), CD10 in 76/81 (94%), Bcl-6 in 74/81 (91%), and MUM1 in 73/81 (90%) cases. Thus, 91% of patients could be classified as having a GC or nGC phenotypic profile.
significantly different between GC and nGC cases only in terms of bone marrow involvement (17% of GC versus 39% of nGC cases, p=0.04) and presence of more than one extranodal disease site (45% of GC versus 72% of nGC cases, p=0.02). The site of biopsy (nodal, extranodal or mediastinal – GC: 50/36/14% versus nGC: 54/31/15%, p=0.56), and the proportion of Bcl-2 immunoreactive cases (50% in GC versus 52% in nGC, p=0.92) were equally distributed between the two groups. Of note, the proportion of patients receiving rituximab and/or frontline HDT did not differ significantly between the GC and nGC groups.

On 18FDG-PET after two cycles, 49/81 (60%) patients were negative and 32/81 (40%) positive. Initial disease characteristics were not distributed in a statistically significant different manner between the early PET-positive and –negative cases. Furthermore, there was no significant difference in the proportion of GC cases or Bcl-2 immunoreactive cases between the two categories.

Three patients were not evaluated after the first four courses of chemotherapy (induction treatment) because of early death or disease progression. According to conventional staging methods, 62/78 (79%) patients were in complete remission (CR) or CR-unconfirmed (CRu), 9/78 (11%) were in partial remission (PR), and 7/78 (9%) had stable or progressive disease following induction. Treatment had to be interrupted following induction because of death or toxicity in four cases. At the end of the complete treatment procedure, 61/74 (82%) patients were in CR or CRu (including four patients who had converted from PR to CR or CRu), one patient was in persistent PR, and five patients had progressive disease after having shown an initial response.

With a median follow-up of 33 months, estimated 3-year OS and EFS rates of the entire population were 75% and 67%, respectively. Patients with a high-intermediate/high IPI score (3-5 factors) had a 3-year EFS of 52%, as opposed to 72% among the patients with a low-intermediate/low score (0-2 factors) (p=0.09). Three-year EFS was 61% and 73% in the Bcl-2-positive and negative groups, respectively (p=0.08). Survival did not differ according to the results of CD10, Bcl-6 or MUM1 immunostaining, and, most importantly, we did not observe any prognostic influence of the GC versus nGC profile (Figure 1B): 3-year EFS was 72% in the GC group and 64% in the nGC one (p=0.65). With a longer median follow-up (35 months versus 24 months), we confirmed our findings regarding the prognostic value of early 18FDG-PET scan: the 3-year EFS was 46% in the PET-positive group and 80% in the PET-negative group (p=0.0005). Three-year OS was 90% in the PET-negative group versus 52% in the PET-positive group (p=0.0001).

In a subgroup analysis, we evaluated the prognostic influence of early 18FDG-PET findings and GC versus nGC profile in the following populations: higher and lower-risk patients according to the IPI, patients who had or had not received rituximab, patients who had or had not received HDT. The predictive value of early 18FDG-PET results was observed in every analyzed subgroup, as was the absence of predictive value of the immunophenotypic GC versus nGC profile (data not shown). When excluding the group of patients in whom the diagnosis had been made on a mediastinal biopsy (which might not be only comprised of cases of primary mediastinal DLBCL), Bcl-2 positivity and GC status did not significantly predict 3-year EFS: 54% versus 74% (p=0.08) for Bcl2 and 72% versus 58% (p=0.39) for GC status. Early PET negativity was still highly predictive of outcome: 3 year EFS was 81% versus 36% in positive cases (p<0.0001).

The IPI risk groups, the results of Bcl-2 immunostaining and the results of early-PET were entered in a Cox proportional hazards model. A positive early PET was the only factor significantly associated with EFS, with a relative risk of 3.55 (p=0.0014).

Although the early PET results performed well as a prognostic predictor, 20% of the early PET-negative patients still relapse. It would be interesting to identify potential prognostic parameters within this subgroup. In this regard, Bcl-2 status and GC/nGC phenotype did not

---

**Table 1. Pretreatment characteristics (n=81).**

<table>
<thead>
<tr>
<th></th>
<th>Whole population</th>
<th>GC patients (n=38)</th>
<th>nGC patients (n=36)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>63</td>
<td>61</td>
<td>69</td>
<td>NS</td>
</tr>
<tr>
<td>Female</td>
<td>37</td>
<td>39</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 60 years</td>
<td>72</td>
<td>74</td>
<td>78</td>
<td>NS</td>
</tr>
<tr>
<td>&gt; 60 years</td>
<td>28</td>
<td>26</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I–II</td>
<td>10</td>
<td>16</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>III–IV</td>
<td>90</td>
<td>84</td>
<td>97</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Performance status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–1</td>
<td>72</td>
<td>76</td>
<td>72</td>
<td>NS</td>
</tr>
<tr>
<td>≥ 2</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td><strong>LDH level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ ULN</td>
<td>38</td>
<td>37</td>
<td>44</td>
<td>0.01</td>
</tr>
<tr>
<td>&gt; ULN</td>
<td>62</td>
<td>63</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td><strong>Number of EN disease sites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–1</td>
<td>42</td>
<td>55</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>≥ 2</td>
<td>58</td>
<td>45</td>
<td>72</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Bone marrow biopsy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involved</td>
<td>25</td>
<td>17</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Not involved</td>
<td>75</td>
<td>83</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td><strong>IPI risk group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-Low/Intermediate</td>
<td>42</td>
<td>34</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Intermediate/High-High</td>
<td>58</td>
<td>66</td>
<td>53</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthracycline-containing</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Rituximab</td>
<td>46</td>
<td>45</td>
<td>47</td>
<td>NS</td>
</tr>
<tr>
<td>Frontline HDT</td>
<td>41</td>
<td>39</td>
<td>47</td>
<td></td>
</tr>
</tbody>
</table>

LDH: lactate dehydrogenase; ULN: upper limit of normal; EN: extranodal; HDT: high-dose therapy; NS: not significant. *Bone marrow biopsy was performed in 79/81 (97%) patients.
show predictive power in our series: among early PET-negative patients with interpretable Bcl-2 status (n=47), Bcl-2 positive patients (n=23) had a 77% 3-year EFS versus 87% in negative cases (n=24) (p=0.12). In this same group, GC patients had a 3-year EFS of 86% versus 78% in nGC cases (p=0.63). Likewise, no differences were observed according to Bcl-2 or GC/nGC status in the early PET positive group.

Discussion

As more and more options are made available for the treatment of DLBCL, risk stratification becomes an increasingly important issue in patient management. The actual IPI-based stratification strategy only enables suboptimal separation of patients: the differences in long-term EFS predicted by the different IPI risk categories are of a magnitude of 10%, and the low/intermediate and intermediate/high categories identify patients with similar long-term event-free survivals. We applied two innovative risk-stratification approaches to a series of DLBCL patients with heterogeneous presentations (nodal, extranodal or primary mediastinal), and treatments, fairly closely reflecting patients seen in our everyday practice. The series was mainly composed of younger patients with high-risk disease (as determined by the IPI), who are the best candidates for innovative first-line treatment approaches, and thus for optimal risk stratification. Both methods have been recently introduced, and are thus currently undergoing active evaluation. Each method would potentially affect therapeutic decisions at different time points, as one delivers information at diagnosis, and the other only after a few cycles of therapy.

The large majority (91%) of patients could be classified as having GC or nGC disease using the algorithm of Hans, and 51% were in the GC group. We found that multiple extranodal disease sites and bone marrow involvement were more frequent in nGC cases. To the best of our knowledge, such differences in disease presentation according to phenotypic (or genotypic) profile have not been reported previously. In this series, the phenotypic profile had no impact on prognosis, although the 8% difference in EFS observed between the GC and nGC groups may suggest a trend favoring the GC subgroup. This is concordant with results observed by other investigators in the setting of first-line treatment, as well as in relapsed/refractory disease, but appears to contradict the results of others. The reasons for these discrepancies remain to be understood. We did not confirm the findings of Fields et al. who observed an impressive difference in EFS between Bcl-2 positive and negative cases within their small population of interim PET-negative patients. This might be due to differences between the two studies, in particular regarding treatment. Another explanation, among several, is that interpretation of immunostaining patterns is prone to interobserver and inter-institutional variations, as has recently been shown for MUM1. In addition, our study was based on consecutively recruited patients with heterogeneous clinical presentations, and included patients with mediastinal lymphoma who likely belong to molecularly distinct entity.

Almost half of the patients in this series received rituximab as part of their treatment. The addition of rituximab to multiagent chemotherapy has been shown to erase the predictive value of known prognostic markers, namely Bcl-2 and Bcl-6 expression. This difference from other series homogeneously treated with chemotherapy only should be taken into account. Importantly, the original observation that GC and nGC cases (as determined by gene expression profiling) had different outcomes originates from series of patients treated without rituximab. The use of first-line HDT could have further modified the predictive value of individual prognostic markers. Interestingly, in our series, early FDG-PET scanning predicted survival in both the groups treated with and without rituximab or HDT. Some patients (n=14) received their treatment within the GELA 98-B2 protocol, in which patients with one adverse factor of the age-adjusted IPI with Bcl-2 expression received consolidative HDT in order to try to overcome the poor prognosis associated with Bcl-2 expression. Such an approach might have contributed to the lack of predictive value of Bcl-2 expression in the present series, but the low number of patients (n=5) potentially represents only a minimal bias.

With a longer follow-up of our previously published series, we confirmed the prognostic value of early FDG-PET scan results: 5-year EFS was 46% in the PET-positive group and 80% in the PET-negative group (p=0.0003).

In view of the conflicting results presented here and elsewhere, we believe that GC/nGC phenotype should be further evaluated in large and homogeneously treated series of patients before it can be widely applied for stratification of patients. Definitive knowledge on its potential value will need ongoing efforts in the field of technique standardization. The strong prognostic impact of early PET, as shown again here warrants prospective trials evaluating the impact of stratification through early FDG-PET used to propose a risk-adapted treatment approach in DLBCL.

Authors’ Contributions

JD, PG, FR and CH designed research; EI, AR and MM performed and interpreted imaging studies; FH and JD analyzed data; JD, CG, TEG, IG and CH provided patient care and clinical data; PG, CC-B and JB performed the histopathological review; JD, PG, FR and CH wrote the paper; and all authors checked the final version of the manuscript.

Conflict of Interest

The authors reported no potential conflicts of interest.
References


