Annals of Hematology (2019) 98:1937–1946
https://doi.org/10.1007/s00277-019-03676-0

ORIGINAL ARTICLE

Clinical risk scores do not accurately identify a very high risk population with diffuse large B cell lymphoma—an analysis of 386 Portuguese patients

Rita Coutinho[1] · J. Lobato[2] · S. Esteves[3] · J. Cabeçadas[4] · M. Gomes da Silva[2]

Received: 17 July 2018 / Accepted: 17 March 2019 / Published online: 4 April 2019

Abstract

The identification of high-risk patients deserving alternative first-line treatments to R-CHOP is a research priority in diffuse large B cell lymphoma (DLBCL). Despite the increasing recognition of biological features underlying aggressive behavior, clinical scores remain the basis for prognostic evaluation and treatment stratification in DLBCL. We performed a retrospective analysis of patients with DLBCL uniformly treated with immunochemotherapy with the aim of assessing the discriminative power of the NCCN international prognostic index (IPI) and the GELTAMO-IPI scores in risk group stratification and compared them with the IPI. Additionally, we investigated if bulky disease, gender, beta-2 microglobulin (β2m), body mass index, and B-symptoms have independent prognostic impact. We confirmed the discriminative ability of the three prognostic scores in terms of progression-free survival and overall survival and found that the NCCN-IPI performs better in the identification of a high-risk population compared to the IPI and the GELTAMO scores. In an attempt to improve the prognostic power of the NCCN-IPI we analyzed additional clinical variables. Bulky disease and elevated β2m were found to be independent predictors of prognosis when controlling for the NCCN-IPI risk groups. However, they seem to bring no incremental power to the latter in the identification of poor outcome patients. We support the use of the NCCN-IPI for the clinical identification of high-risk patients in DLBCL. Future studies to unravel the biological heterogeneity within NCCN-IPI groups are needed to improve risk prediction and design targeted therapies for poor prognosis patients.

Keywords Diffuse large B cell lymphoma · GELTAMO-IPI · NCCN-IPI · Outcome · High risk

Introduction

Diffuse large B cell Lymphoma (DLBCL) accounts for approximately 30-40% of new non-Hodgkin lymphoma (NHL) cases and is morphologically, clinically, and biologically heterogeneous [1–5].

Rituximab associated with anthracyclin-containing regimens are the mainstay of treatment and lead to cure in over 60% of patients. However, the 30–40% of DLBCL patients exhibiting primary refractory or relapsed disease are hardly salvaged, even with aggressive chemotherapy regimens and autologous stem-cell transplant [6, 7]. In contrast, patients free of disease 24 months after first-line treatment have an excellent long time outcome [8]. It is critical to better identify at diagnosis patients in whom R-CHOP will be insufficient, so that alternative first-line treatments can be considered. This prognostic divergence is a reflection of the genetic heterogeneity of DLBCL. It is acknowledged that poor-risk patients harbor particular genetic changes, including a non-germinal center cell-of-origin molecular profile [2], double MYC and BCL2 breaks [9–11], double MYC and BCL2 protein expression [12, 13], and specific mutational changes [5]. The use of therapies targeted against some of these

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00277-019-03676-0) contains supplementary material, which is available to authorized users.

✉ Rita Coutinho
ritarcoutinho@hotmail.com

[1] Department of Hematology, Centro Hospitalar Universitário do Porto, Porto, Portugal

[2] Department of Hematology, Instituto Português de Oncologia de Lisboa Francisco Gentil, E.P.E., Lisbon, Portugal

[3] Clinical Research Unit, Instituto Português de Oncologia de Lisboa Francisco Gentil, E.P.E., Lisbon, Portugal

[4] Department of Pathology, Instituto Português de Oncologia de Lisboa Francisco Gentil, E.P.E., Lisbon, Portugal

alterations might generate an opportunity for improvement of the outcome of high-risk DLBCL.

However, clinical scores are still the basis for prognostic evaluation and treatment stratification in 2018. The International Prognostic Index (IPI) has been the basis for predicting outcomes in DLBCL [14] and has been largely validated in population-based studies and clinical trials despite treatment evolution over time. With the improvement of patients' outcomes with immunochemotherapy, the predictive power of the IPI has been challenged. More specifically a very high risk group of patients, with a uniformly poor outcome after R-CHOP, cannot be identified. In this context, modifications in the analysis of the IPI factors [15] and a number of patient- and disease-specific factors have been proposed to improve outcome prediction in this disease [16–20].

Newly developed prognostic scores based on clinical characteristics of large, retrospectively reviewed series of patients, aimed to better distinguish DLBCL risk groups. The enhanced NCCN-IPI [21] differs from the IPI by stratifying age and lactate dehydrogenase (LDH) levels and recognizing the risk associated with particular extranodal sites of involvement. The more recent GELTAMO-IPI [22] incorporates β2-microglobulin (β2m) in the prognostic algorithm. Both indices seem to further discriminate clinical courses among risk groups and span a larger range of outcomes. The NCCN-IPI has been applied to independent retrospective series with reproducible results [17, 23]. However, none of the scores accurately identifies very high risk DLBCL patients [24].

The relative prognostic value and clinical impact of the NCCN-IPI and GELTAMO-IPI, as well as the additional information provided by other clinical characteristics, remains to be widely validated.

In this retrospective analysis of patients with DLBCL uniformly treated with R-CHOP, we compared the discriminative power of the NCCN-IPI, GELTAMO-IPI, and IPI scores in risk group stratification. Additionally, we investigated if bulky disease, gender, β2m, body mass index (BMI), and presence of B-symptoms have independent prognostic impact and contribute to prognostic discrimination improvement in a series of 321 Portuguese patients treated with immunochemotherapy.

Patients and methods

We conducted a retrospective analysis of consecutive adult patients diagnosed with de novo DLBCL treated at the Portuguese Institute of Oncology Lisbon (IPOLFG) between January 2002 and December 2013. Diagnosis was performed by an expert hematopathologist at the center; diagnostic specimens were not specifically reviewed for this study. All patients received R-CHOP or R-CHOP-like chemotherapy with curative intent. Patients who were never treated because of frailty, or who were treated with palliative intent only and without rituximab, were excluded from the analysis. Other reasons for patient exclusion were HIV infection, evidence of transformation from an indolent lymphoma, or histologies other than DLBCL not otherwise specified. R-CHOP-21 was administered to all patients, with dose reductions in older patients with cardiac dysfunction or other significant comorbidities and according to observed toxicities. In cases of bulky disease, defined as any mass with a maximal diameter ≥ 7 cm, radiotherapy was recommended. Central nervous system (CNS) prophylaxis was prescribed in patients considered to be at significant risk due to high IPI and/or with involvement of specific extranodal locations.

Score distribution for each patient was re-calculated for the study. Treatment response and outcome were assessed using the standardized response criteria for malignant lymphoma (1999) or the 2007 revised criteria depending on the availability of an end-of-treatment PET/CT [25, 26].

Progression-free survival (PFS) was calculated from the date of diagnosis until first documented disease progression/relapse or death of any cause. Overall survival (OS) was calculated from the date of diagnosis until death of any cause. Survival distributions for each risk group within each prognostic index (IPI, NCCN-IPI, and GELTAMO-IPI) were estimated by the Kaplan-Meier method. Pairwise comparisons between risk groups within each index were performed using log rank test with p value adjustment for multiple comparison using Hochberg method. All tests were two-tailed and p values less than 0.05 were considered statistically significant.

The discrimination of the different prognostic scores was assessed using the summary measures Harrel's C-index [27] and the Uno's C-statistic [28].

We also conducted an exploratory analysis of the incremental predictive value of the factors gender, β2m (below or above the upper limit normal), categorized BMI (normal or underweight, i.e., BMI < 25, vs overweight or obese patients, i.e., BMI > 24.9), presence of B-symptoms, and bulky disease controlling for the risk score showing better discrimination in the previous comparative analysis. The independent association of each of these variables with OS and PFS was first assessed by multivariable analysis using Cox regression model by means of a backward variable selection procedure using likelihood ratio test and a critical alpha of 0.20 for variable exclusion. Harrel's C and Uno's C-statistics were then calculated to assess the incremental discrimination value of these covariates compared to the risk score.

Data was analyzed using the Statistical Software R [29]. Ethical approval for this study was obtained from Local Regional Ethics Boards.

Results

Patient characteristics

Five-hundred and twenty-one patients were diagnosed with DLBCL from 2002 to 2013 in our center. Of those, 416 fulfilled the eligibility criteria (Fig. 1). Thirty patients were further excluded due to missing data concerning the NCCN-IPI and GELTAMO-IPI scores, leading to an analysis set of 386 patients (dataset A), which was used for assessment and comparison of the performance of the IPI, NCCN-IPI, and GELTAMO-IPI prognostic indexes.

A second analysis set (dataset B; Fig. 1) comprising 321 patients with complete information concerning all variables of interest was used to explore the incremental predictive value of the factors gender, β2m, BMI, B-symptoms, and bulky disease. Clinical and demographic characteristics of patients excluded due to missing data are depicted in Fig. 1.

Table 1 summarizes the patient characteristics in the IPOLFG (dataset A), BCCA series [21], and GELTAMO series [22].

The mean age and the proportion of patients above 60 years old were similar in the three cohorts. Some baseline differences were found between the Portuguese, Spanish, and Canadian series. IPOLFG patients had better performance status than GELTAMO and BCCA (ECOG performance status ≥ 2 in 15% vs 30% in GELTAMO and 37% in BCCA, $p < 0.001$ in both comparisons), less advanced disease (stage III/IV in 53% vs 63%, $p = 0.003$), and a lower proportion of cases involving bone marrow, CNS, liver/gastrointestinal tract, or lung than GELTAMO cases (31% vs 41%, $p = 0.011$). Conversely, they presented with higher LDH than other series (67% vs 55% in GELTAMO vs 49% in BCCA, $p < 0.001$ in both comparisons) and more frequent bulky disease than the Spanish patients (39% vs 29%, $p = 0.003$).

Despite these differences, the overall IPI risk group distribution was identical between IPOLFG and GELTAMO cohorts (Table 1). Compared to BCCA, our cohort had a higher proportion of low-risk (36% vs 12%) and a lower proportion of high-intermediate risk patients (19% vs 37%). The proportion of patients with high risk IPI was similar in the three series, ranging between 14 and 19% (Table 1).

The median number of R-CHOP cycles administered to dataset A was eight, with 84% receiving 6–8 cycles. The overall response rate was 88% (with 78% complete responses); 7% had progressive disease and in the remaining 5% the response was not evaluable.

The median follow-up of the censored patients was 59 months. One-hundred and twenty-three patients died (57% due to lymphoma, 16% due to toxicity, 9% due to other cancer, 14% due to other causes, and 4% of unknown cause) and 94 patients relapsed. From the 263 censored patients, 134

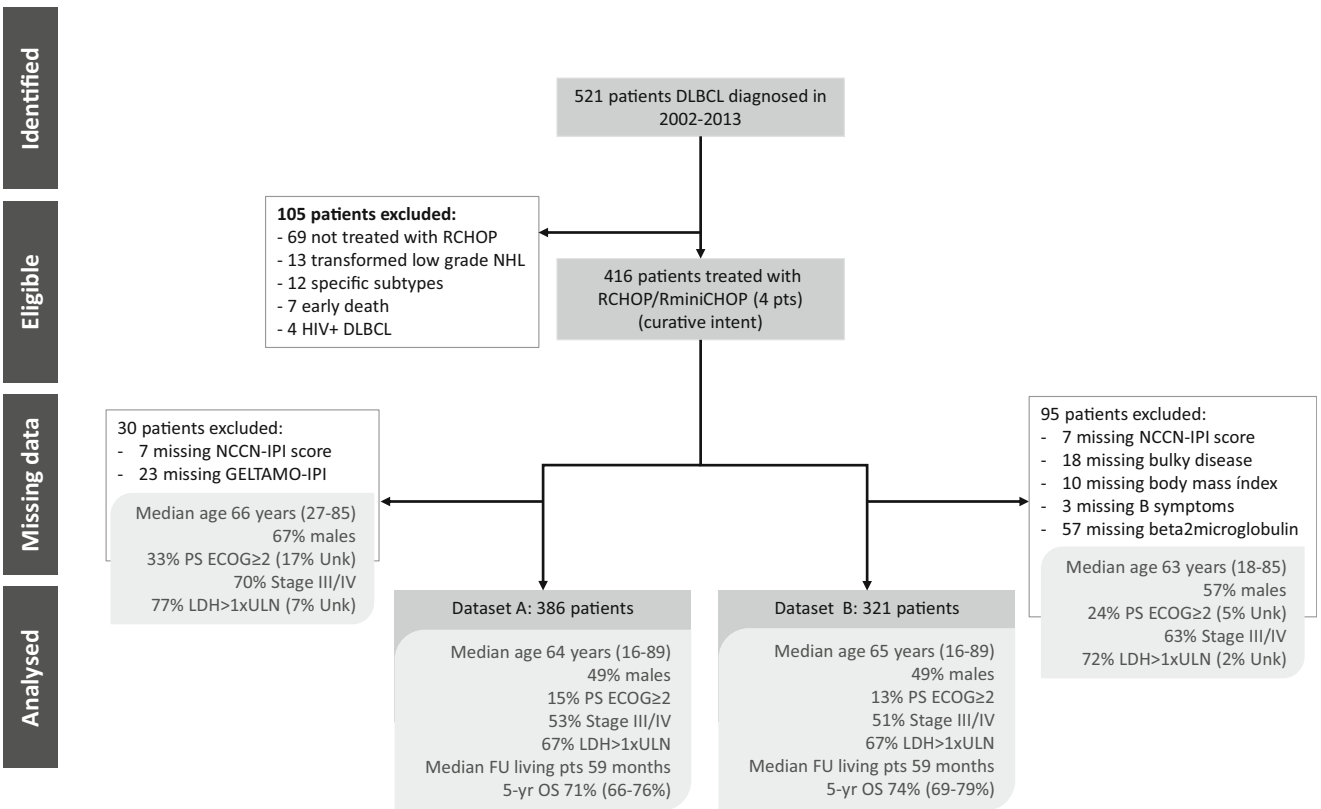


Fig. 1 Patient flow of the study cohort

Table 1 Patient's characteristics from IPOLFG, GELTAMO, and BCCA cohorts

Cohort characteristics	IPOLFG	GELTAMO	BCCA	*P* value[a] IPOLFG vs GELTAMO	*P* value[a] IPOLFG vs BCCA
Number of patients	386	1848	1138	–	–
Male sex (%)	49	51	60	1	0.004
Age (years), mean (SD)	61 (18)	60 (16)	63 (15)	–	–
Age > 60 years (%)	57	54	60	1	1
ECOG performance status ≥ 2 (%)	15	30	37	< 0.001	< 0.001
Ann Arbor stage III–IV (%)	53	63	55	0.003	1
Normalized LDH > 1 ULN (%)	67	55	49	< 0.001	< 0.001
≥ 2 extranodal sites (%)	23	18	NA	0.322	–
BM, CNS, liver/GI, or lung involvement (%)	31	41	25	0.013	0.065
Bulky disease (%)	39[b]	29	NA	0.003	–
Normalized β2-microglobulin > 1 ULN (%)	45[c]	46	NA	1	–
IPI score (%)					
Low	36	34	12	0.322	< 0.001
Low-intermediate	28	23	37		
High-intermediate	19	25	37		
High	17	19	14		
Median follow-up (months)	59	57	38	–	–

IPOLFG Instituto Português de Oncologia de Lisboa Francisco Gentil, GELTAMO Grupo Español de Linfomas y Transplantes de Médula óssea, BCCA British Columbia Cancer Agency, SD standard deviation, ECOG Eastern Cooperative Oncology Group, LDH lactate dehydrogenase, BM bone marrow, CNS central nervous system, GI gastrointestinal tract, ULN upper limit normal, IPI international prognostic index, NA Not reported in the original paper

[a] Pearson's chi-squared test with Yates' continuity correction and *p* value adjustment for multiple comparison using the Hochberg method

[b] Bulky disease unknown in 15 patients (4%)

[c] β2-microglobulin unknown in 44 patients (11%)

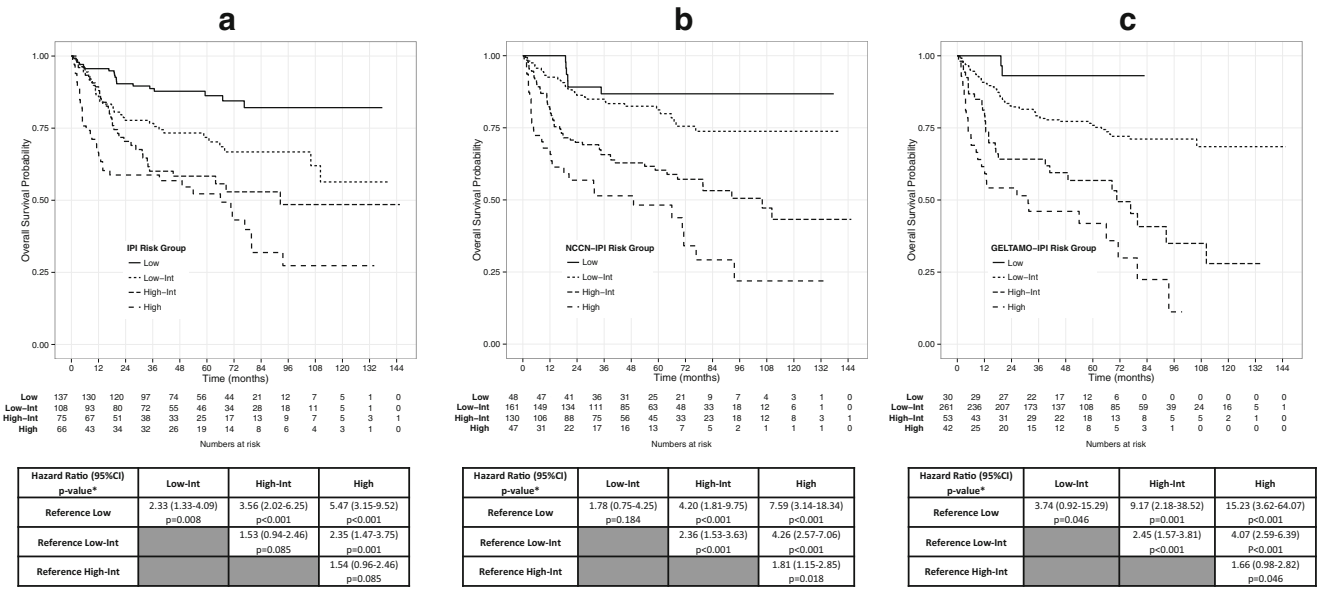
(51%) had a follow-up time inferior to 5 years. The 5-year OS and PFS in dataset A were 71% (95% CI 66–76%) and 67% (95% CI 62–72%), respectively.

Comparative performance of IPI, NCCN-IP, and GELTAMO-IPI risk scores

Figure 2 shows the OS of IPOLFG patients according to risk groups within the IPI (A), NCCN (B), and GELTAMO (C) prognostic indexes. Altogether, the estimated hazard ratios (HR) and the adjusted *p* values for pairwise comparisons indicate a significant difference in OS between risk groups for all index scores. Nevertheless, the IPI and GELTAMO-IPI performed better in the differentiation between low and low-intermediate risk groups (IPI: HR 2.33, $p = 0.008$; GELTAMO-IPI: HR 3.74, $p = 0.046$; NCCN-IPI: HR 1.78, $p = 0.184$), whereas NCCN-IPI better distinguished high from high-intermediate patients (NCCN-IPI: HR 1.81, $p = 0.018$; GELTAMO-IPI: HR 1.66, $p = 0.046$; IPI: HR 1.54, $p = 0.085$).

When comparing the patient distribution according to the three risk scores (Table 2), it becomes evident that, whereas IPI classifies the majority of patients as low and low-intermediate risk (36% and 28%, respectively), NCCN-IPI tends to classify more patients in the intermediate categories (42% low-intermediate and 34% high-intermediate). The GELTAMO-IPI further concentrates the majority of the patients (68%) in the low-intermediate category. This seems to result in higher 5-year OS and PFS in the low-risk groups and lower 5-year OS and PFS in the high risk when using GELTAMO-IPI and NCCN-IPI risk stratifications as compared to IPI (Table 2). Nevertheless, the overlapping 95% confidence intervals hinder a clear interpretation of these apparent differences (Table 2). To further clarify this issue we evaluated the accuracy of the scores to predict survival status at 5 years (Fig. 3).

Figure 3a shows the risk-group distribution of the three scores in the 106 patients who died in the 5-year follow-up period (patients with poor outcome). Within this subgroup, the proportion classified as high risk by the three scores was below 30% (28% in IPI, 22% in both NCCN-IPI and GELTAMO-IPI). Fifty five of 106 poor outcome patients (52%) were reclassified by the NCCN-IPI into a risk category different from IPI. Of these, 35 were correctly moved into an upper risk while 20 were reclassified into lower risk, resulting in a relative improvement of 14% in the classification accuracy for patients with poor outcome by NCCN-IPI. Conversely, when comparing GELTAMO-IPI with NCCN-IPI, there is a relative improvement of −21%, meaning that GELTAMO-IPI is 21% worse than



Hazard Ratio (95%CI) p-value*	Low-Int	High-Int	High
Reference Low	2.33 (1.33-4.09) p=0.008	3.56 (2.02-6.25) p<0.001	5.47 (3.15-9.52) p<0.001
Reference Low-Int		1.53 (0.94-2.46) p=0.085	2.35 (1.47-3.75) p=0.001
Reference High-Int			1.54 (0.96-2.46) p=0.085

Hazard Ratio (95%CI) p-value*	Low-Int	High-Int	High
Reference Low	1.78 (0.75-4.25) p=0.184	4.20 (1.81-9.75) p<0.001	7.59 (3.14-18.34) p<0.001
Reference Low-Int		2.36 (1.53-3.63) p<0.001	4.26 (2.57-7.06) p<0.001
Reference High-Int			1.81 (1.15-2.85) p=0.018

Hazard Ratio (95%CI) p-value*	Low-Int	High-Int	High
Reference Low	3.74 (0.92-15.29) p=0.046	9.17 (2.18-38.52) p=0.001	15.23 (3.62-64.07) p<0.001
Reference Low-Int		2.45 (1.57-3.81) p<0.001	4.07 (2.59-6.39) P<0.001
Reference High-Int			1.66 (0.98-2.82) p=0.046

A: IPI; B: NCCN-IPI; C: GELTAMO-IPI
*Logrank test with p-value adjustment for multiple comparison using Hochberg method

Fig. 2 Overall survival according to risk groups within each prognostic index

NCCN-IPI in the proper classification of patients with poor outcome.

Figure 3b shows the risk-group distribution of the three scores in the 146 patients who were alive at 5 years (patients with good outcome). In the comparative performance concerning the classification of this population, both IPI and GELTAMO-IPI performed better than NCCN-IPI (relative improvement 25% and 16%, respectively) but the GELTAMO-IPI was worse than IPI (relative improvement − 12%; not shown in Fig. 3b). These results suggest that the IPI score better identifies the patients with good outcome.

To substantiate these observations, we applied complementary statistical measures (Harrel's and Uno's C-statistics) for discrimination assessment and found that the three risk-stratification models show a significant discrimination ability (Supplementary Table 1). However, the GELTAMO-IPI had the lowest C-statistics values, suggesting that for both OS and PFS, this score has the lowest discrimination ability. The pairwise comparisons (using Delta Uno's C-statistic and the reclassification calibration statistic) suggest a better risk assessment by the NCCN-IPI. However, the difference between NCCN-IPI and GELTAMO-IPI, although significant, corresponds only to a marginal improvement.

Altogether, these results suggest that NCCN-IPI better identifies poor outcome patients. However, none of the scores can accurately identify very high risk patients.

CNS relapse prediction according to the CNS-IPI and the remaining risk scores

CNS relapses are rare events that are difficult to predict but portend a very dim prognosis in DLBCL. The recently developed CNS-IPI score [30], including the IPI risk factors in addition to renal and/or adrenal

Table 2 Five-year overall and progression-free survival according to IPI, NCCN-IPI, and GELTAMO-IPI scores

	IPI			NCCN-IPI			GELTAMO-IPI		
Risk group	*N* (%)	5-year OS, % (95% CI)	5-year PFS, % (95% CI)	*N* (%)	5-year OS, % (95% CI)	5-year PFS, % (95% CI)	*N* (%)	5-year OS, % (95% CI)	5-year PFS, % (95% CI)
L	137 (36)	86 (80–93)	85 (80–92)	48 (12)	87 (78–97)	83 (73–95)	30 (8)	93 (84–100)	93 (85–100)
LI	108 (28)	72 (64–81)	68 (59–78)	161 (42)	81 (75–88)	78 (72–85)	261 (68)	76 (71–82)	72 (67–78)
HI	75 (19)	58 (48–71)	51 (41–64)	130 (34)	60 (52–70)	55 (47–65)	53 (14)	57 (45–73)	49 (37–65)
H	66 (17)	52 (41–67)	47 (36–62)	47 (12)	48 (35–66)	46 (33–64)	42 (11)	41 (29–62)	39 (26–59)

L low, LI low-intermediate, HI high-intermediate, H high

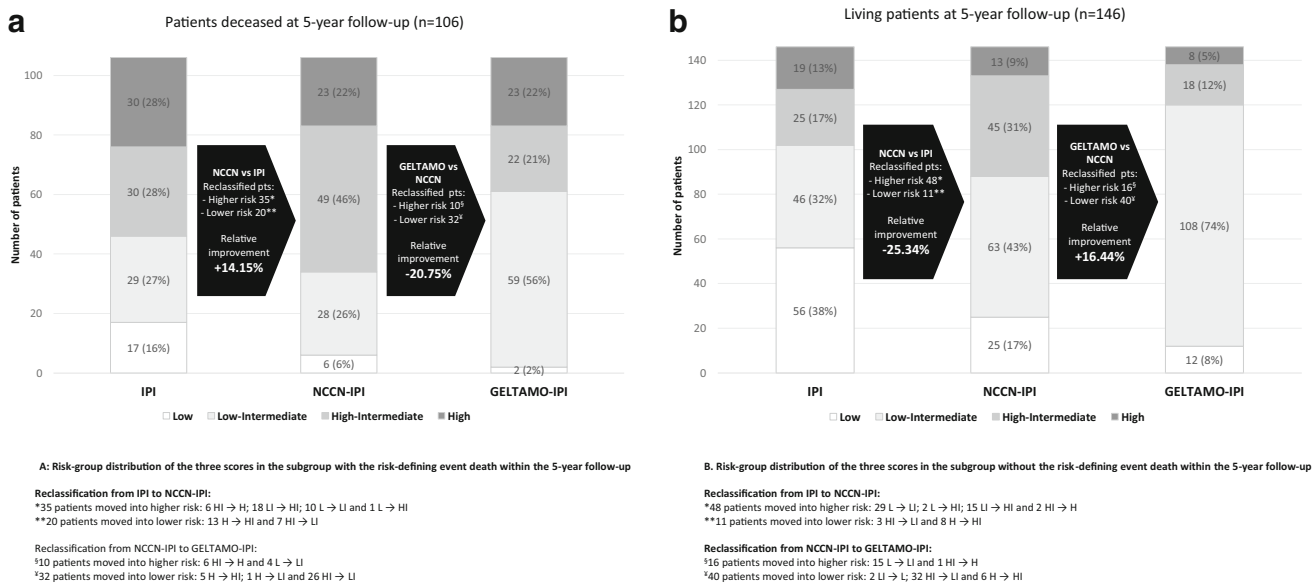


Fig. 3 Accuracy of IPI, NCCN-IPI, and GELTAMO-IPI to predict survival status at 5-year follow-up

infiltration, identifies a group of patients with an increased incidence of CNS relapse.

During the study period we identified 15 CNS relapses (3.9%, 95% CI 2.2 to 6.4%). This corresponds to an incidence rate of 0.8 per 1000 patient-years (95% CI 0.4 to 1.2), in agreement with other series.

According to the CNS-IPI score, 19% of our patients (74 of 386) were classified as high risk. Of these, only 47% (35/74) and 32% (24/74) were considered as high risk according to the NCCN-IPI and GELTAMO-IPI, respectively. Hence, although far from perfect, the NCCN-IPI seems to better identify CNS-IPI high-risk patients.

We also looked at the proportion of CNS relapse within each risk group of the GELTAMO, NCCN, and CNS-IPI scores and compared the predictive capacity of the three scores for the CNS relapse. Among the patients classified as high risk by the NCCN, GELTAMO, and CNS-IPI scores the percentage that eventually exhibits CNS recurrence is 4% (2/47), 2% (1/42), and 7% (5/74), respectively. The majority of CNS recurrence events occurred in patients classified as intermediate risk by all scores (12 events in intermediate-risk NCCN-IPI, 14 events in intermediate-risk GELTAMO-IPI, and 10 in intermediate-risk CNS-IPI). We emphasize that the incidence of CNS relapse in our high-risk group according to the CNS-IPI is around 10% and similar to published data.

Incremental predictive value of additional clinical variables

Having determined that NCCN-IPI provides the best risk stratification in our series, we aimed to investigate if additional variables could affect its discrimination ability.

We analyzed the cell of origin defined by immunohistochemistry in 199 patients with available data. According to the Hans algorithm, 59% were classified as germinal center B cell-like (GCB) and 41% as non-GCB DLBCL (data not shown). As previously demonstrated by us and others [31, 32], the Hans classification did not help segregate two populations with different outcome.

Firstly, we conducted a multivariable analysis to evaluate the independent association of gender, dichotomized β2m and BMI, B-symptoms, and bulky disease with OS and PFS, controlling for NCCN-IPI risk stratification. This analysis was undertaken in dataset B (Fig. 1). As can be appreciated in Table 3, elevated β2m (OS: HR 2.26, 95% CI 1.34–3.81, $p = 0.002$; PFS: HR 1.86, 95% CI 1.27–2.96, $p = 0.008$) and the presence of bulky disease (OS: HR 1.68, 95% CI 1.10–2.58, $p = 0.016$; PFS: HR 1.69, 95% CI 1.14–2.50, $p = 0.01$) were the only covariates independently associated with poor outcome.

We also analyzed the incremental value of these variables for discriminating poor from good outcome patients, by comparing the discrimination measures of the NCCN-IPI model [21] with risk models that additionally included elevated β2m and bulky disease. These results are presented in Supplementary Table 2. While independently associated with OS, elevated β2m and bulky disease have a small incremental effect on discrimination as evaluated by all tests used (Harrell's C, Uno's C-statistics, and estimated deltas). In conclusion, these results indicate that although these variables are independently associated with the risk of death, they present no additional value for discriminating individuals who died from those who survived at least 5 years.

Table 3 Multivariable analysis to evaluate the independent association of gender, dichotomized β2m and BMI, B-symptoms, and bulky disease controlling for the NCCN-IPI

	Overall survival			Progression-free survival		
Variable	HR	95% CI	*p* value[a]	HR	95% CI	*p* value[a]
Gender						
Female (ref)	1		0.388	1		0.558
Male	1.20	0.79–1.82		1.12	0.77–1.64	
B-symptoms						
No (ref)	1		0.681	1		0.267
Yes	1.10	0.70–1.73		1.27	0.83–1.93	
BMI						
< 25 (ref)	1		0.899	1		0.845
> 24.9	0.97	0.64–1.48		0.96	0.65–1.42	
Beta2-microglobulin						
< ULN (ref)	1		0.002	1		0.008
> 1 × ULN	2.26	1.34–3.81		1.86	1.17–2.96	
Bulky disease						
No (ref)	1		0.016	1		0.010
Yes	1.68	1.10–2.58		1.69	1.14–2.50	
NCCN-IPI risk group						
Low (ref)	1		0.014	1		0.076
Low-intermediate	2.10	0.72–6.11		1.50	0.65–3.44	
High-intermediate	3.25	1.13–9.37		2.15	0.94–4.91	
High	4.62	1.48–14.39		2.78	1.11–6.98	

[a] Likelihood ratio test comparing model with all covariates with model without the respective covariate

Discussion

This study was set out to validate and compare the prognostic efficacy of the GELTAMO-IPI, NCCN-IPI, and IPI scores in a large series of DLBCL patients uniformly treated with R-CHOP. We confirmed the discriminative ability of the three prognostic scores for PFS and OS and found that the NCCN-IPI performs better in the identification of a high-risk population compared to the IPI and the GELTAMO scores. Since patients belonging to the NCCN high-risk group still have a 5-year OS between 40 and 50% we attempted to improve the prognostic discrimination ability of this index by additionally analyzing a number of independent clinical variables. Bulky disease and elevated β2m were found to be independent prognostic factors when controlling for NCCN-IPI risk groups. However, they seem to bring no incremental discrimination ability to the latter in the identification of poor outcome patients.

The recognition of high-risk patients who will relapse or be primarily refractory to R-CHOP is a research priority in DLBCL. Clinical scores based on simple parameters have been the basis of DLBCL prognostic stratification but are only surrogate markers for the biological heterogeneity of the disease [5]. Molecular characterization in DLBCL allows the identification of genetic lesions associated with low treatment responses [2, 5, 9–13]. However, the techniques required for a detailed molecular characterization of DLBCL specimens are not widely available. Moreover, only in the last update of the WHO classification were these subgroups recognized [1], suggesting that further validation of the laboratory methods for their diagnosis is required.

Having this in mind, clinical factors remain the most useful tools available for treatment stratification in daily practice. Risk prognostication in DLBCL has been traditionally performed using the IPI score [14]. However, it is known that the IPI lost power to discriminate four distinct risk categories of patients in the rituximab era. The recently developed NCCN-IPI and GELTAMO-IPI scores allegedly increase the outcome discrimination between DLBCL risk groups compared to the IPI [21, 22].

Our main objective was to define the most robust tool at identifying high-risk patients deserving alternative first-line treatment approaches.

We firstly compared the clinical characteristics of our patients with the ones from the BCCA and the GELTAMO cohorts. Age and gender distribution were comparable while other clinical characteristics differed in the three cohorts. Our dataset appears to exhibit favorable prognostic features, such as better PS and less advanced disease. Despite these differences, the proportion of patients with high-risk IPI was similar in the three series. The majority (84%) of patients completed the planned first-line immunochemotherapy with

overall response rates within the expected range. The median follow-up of the cohort of 59 months is similar to others data and clearly above the NCCN cohort. Therefore, our series is a representative dataset of DLBCL and adequate to answer the study objective.

The results indicate that all three scores robustly segregate patients with different OS and PFS. However, we were interested in understanding the differences in patient allocation into score categories, since that could help in understanding outcome refinement. We show that the NCCN-IPI allocated more patients to the intermediate categories than the IPI, and the GELTAMO-IPI further concentrated almost 70% of the cases into the low-intermediate category. This might explain the higher 5-year OS and PFS in the low-risk groups and lower 5-year OS and PFS in the high risk when using GELTAMO-IPI and NCCN-IPI risk stratifications as compared to IPI.

The GELTAMO-IPI performed statistically better at identifying a low-risk DLBCL population with a 5-year OS and PFS close to 90%, superimposable to the original data [22]. On the other hand, the NCCN-IPI better distinguished high from high-intermediate risk patients, which has been previously suggested in similar studies [17, 23].

The Revised IPI (R-IPI) [15] performed slightly better than the IPI and the GELTAMO-IPI and slightly worse than the NCCN-IPI at the identification of high-risk patients in the current series (data not shown). Additionally, the application of the R-IPI did not significantly improve the identification of high-risk patients. Although 57% of our patients dying within 5 years belonged to the high-risk R-IPI group, 30% of those who were alive at the same time point also did. This data consolidates our conclusion that none of the clinical scores available in the literature is able to accurately identify this population.

We then hypothesized that looking at how patients with short survival were reclassified by the NCCN-IPI (compared to the IPI) would help us better perceive the outcome prediction. For this purpose we evaluated the accuracy of the scores to predict survival status at 5 years. The first finding was that only a third of the patients who died within 5 years were correctly classified as high risk by all scores, illustrating their limitation as sole prognostic tools. However, the reclassification from IPI to NCCN-IPI groups resulted in a relative improvement of 14% in the identification of patients with poor outcome. In addition, reclassification from NCCN-IPI to GELTAMO-IPI lead to a relative reduction of 21% in the adequate classification of these patients. Using adequate statistical tools, we further show that the NCCN-IPI better assesses OS and PFS compared the remaining two scores.

Given the dismal outcome of such a clinical scenario, we looked at the predictive capacity of all risk scores for CNS relapse in our dataset. Our analysis shows that, although the CNS-IPI performs better, none of the scores precisely identifies CNS relapses within a single risk group and hence shows effective clinical usefulness in the identification of patients at high risk of CNS recurrence. This would, in our opinion, be the main objective of a predictive score, allowing for the use of personalized first-line treatment approaches targeting the CNS prophylaxis.

Considering the herein illustrated limitations of clinical prognostic scores for the selection of poor outcome patients, we investigated the impact of additional variables on survival controlling for the NCCN-IPI in 321 R-CHOP-treated patients.

A variety of clinical and laboratory factors have been claimed to be prognostic independently of the IPI in the rituximab era. We tested variables previously validated in the context of large studies [16–18] or in clinical trials [20] that could be collected retrospectively. In this study, raised β2m and the presence of bulky disease were independently associated with a worse outcome, as can be appreciated in multivariate analysis. This goes along with previously published data. However, neither of these variables added significant predictive value to the NCCN-IPI for the discrimination of individuals who died from those who survived after a 5-year follow-up. Other models that use simple clinical parameters and focus on the analysis of continuous rather than dichotomized variables may be of additional interest [33].

A biological characterization in parallel with a prognostic score classification is of utmost importance in DLBCL. Despite recognizing the limitations of current immunohistochemistry algorithms for the molecular classification of DLBCL [30, 31], this remains an easily available and commonly used tool in clinical practice. We thus analyzed the cell of origin according to the Hans algorithm in 199 Portuguese patients with available data, and found no additional prognostic impact for GC and non-CG subtypes. We are pursuing the molecular characterization by other techniques and the identification of DHL and DEL in the Portuguese series.

Our work is limited by its retrospective nature. The impact of occasional missing data is however reduced by the significant size of the datasets. We did not evaluate biological characteristics known to impact on the clinical course of DLBCL patients, but this mirrors current clinical practice outside trials.

In summary, in this first retrospective study comparing the NCCN and the GELTAMO-IPI, we confirm the power of discrimination of NCCN-IPI in a large European cohort treated with immunochemotherapy and support its use for the clinical identification of high-risk patients in DLBCL. Future studies to unravel the biological heterogeneity within NCCN-IPI groups are needed to improve risk prediction. The success of targeted therapies for poor prognosis patients will require reliable and reproducible strategies that adequately segregate patients into distinct biological risk groups.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study was approved by the Ethics Committee of Instituto Português de Oncologia de Lisboa Francisco Gentil, E.P.E., Lisbon, Portugal.

References

1. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD, Jaffe ES (2016) The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood 127(20):2375–2390
2. Alizadeh AA, Eisen MB, Davis RE et al (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511
3. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, López-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM, Lymphoma/Leukemia Molecular Profiling Project (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346:1937–1947
4. Lenz G, Wright GW, Emre NC et al (2008) Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc Natl Acad Sci U S A 105:13520–13525
5. Schmitz R, Wright GW, Huang DW et al (2018) Genetics and pathogenesis of diffuse large B-cell lymphoma. N Engl J Med 378(15):1396–1407
6. Gisselbrecht C, Glass B, Mournier N et al (2010) Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. J Clin Oncol 28(27):4184–4190
7. Crump M, Neelapu SS, Farooq U, van den Neste E, Kuruvilla J, Westin J, Link BK, Hay A, Cerhan JR, Zhu L, Boussetta S, Feng L, Maurer MJ, Navale L, Wiezorek J, Go WY, Gisselbrecht C (2017) Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. Blood 130(16):1800–1808
8. Maurer M, Ghesquières H, Jais JP, Witzig TE, Haioun C, Thompson CA, Delarue R, Micallef IN, Peyrade F, Macon WR, Jo Molina T, Ketterer N, Syrbu SI, Fitoussi O, Kurtin PJ, Allmer C, Nicolas-Virelizier E, Slager SL, Habermann TM, Link BK, Salles G, Tilly H, Cerhan JR (2014) 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 32(10):1066–1073
9. Aukema SM, Siebert R, Schuuring E, van Imhoff GW, Kluin-Nelemans HC, Boerma EJ, Kluin PM (2011) Double-hit B-cell lymphomas. Blood 117(8):2319–2331
10. Ennishi D, Mottok A, Ben-Neriah S, Shulha HP, Farinha P, Chan FC, Meissner B, Boyle M, Hother C, Kridel R, Lai D, Saberi S, Bashashati A, Shah SP, Morin RD, Marra MA, Savage KJ, Sehn LH, Steidl C, Connors JM, Gascoyne RD, Scott DW (2017) Genetic profiling of MYC and BCL2 in diffuse large B-cell lymphoma determines cell-of-origin-specific clinical impact. Blood 129(20):2760–2770
11. Sesques P, Johnson NA (2017) Approach to the diagnosis and treatment of high-grade B-cell lymphomas with MYC and BCL2 and/or BCL6 rearrangements. Blood 129(3):280–288
12. Johnson NA, Slack GW, Savage KJ, Connors JM, Ben-Neriah S, Rogic S, Scott DW, Tan KL, Steidl C, Sehn LH, Chan WC, Iqbal J, Meyer PN, Lenz G, Wright G, Rimsza LM, Valentino C, Brunhoeber P, Grogan TM, Braziel RM, Cook JR, Tubbs RR, Weisenburger DD, Campo E, Rosenwald A, Ott G, Delabie J, Holcroft C, Jaffe ES, Staudt LM, Gascoyne RD (2012) Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clin Oncol 30(28):3452–3459
13. Staiger AM, Ziepert M, Horn H, Scott DW, Barth TFE, Bernd HW, Feller AC, Klapper W, Szczepanowski M, Hummel M, Stein H, Lenze D, Hansmann ML, Hartmann S, Möller P, Cogliatti S, Lenz G, Trümper L, Löffler M, Schmitz N, Pfreundschuh M, Rosenwald A, Ott G, German High-Grade Lymphoma Study Group (2017) Clinical impact of the cell-of-origin classification and the MYC/ BCL2 dual expresser status in diffuse large B-cell lymphoma treated within prospective clinical trials of the German high-grade non-Hodgkin's lymphoma study group. J Clin Oncol 35(22):2515–2526
14. The International Non-Hodgkin's Lymphoma Prognostic Factors Project (1993) A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. N Engl J Med 329:987–994
15. Sehn LH, Berry B, Chhanabhai M, Fitzgerald C, Gill K, Hoskins P, Klasa R, Savage KJ, Shenkier T, Sutherland J, Gascoyne RD, Connors JM (2007) 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 109:1857–1861
16. Seo S, Hong JY, Yoon S, Yoo C, Park JH, Lee JB, Park CS, Huh J, Lee Y, Kim KW, Ryu JS, Kim SJ, Kim WS, Yoon DH, Suh C (2016) Prognostic significance of serum beta-2 microglobulin in patients with diffuse large B-cell lymphoma in the rituximab era. Oncotarget 7(47):76934–76943
17. Melchardt T, Troppan K, Weiss L, Hufnagl C, Neureiter D, Tränkenschuh W, Hopfinger G, Magnes T, Deutsch A, Neumeister P, Hackl H, Greil R, Pichler M, Egle A (2015) A modified scoring of the NCCN-IPI is more accurate in the elderly and is improved by albumin and beta2-microglobulin. Br J Haematol 168: 239–245
18. Müller C, Murawski N, Wiesen MH et al (2012) The role of sex and weight on rituximab clearance and serum elimination half-life in elderly patients with DLBCL. Blood 119(14):3276–3284
19. Hohloch K, Altmann B, Pfreundschuh M, Loeffler M, Schmitz N, Zettl F, Ziepert M, Trümper L (2018) Obesity negatively impacts outcome in elderly female patients with aggressive B-cell lymphomas treated with R-CHOP: results from prospective trials of the German high grade non-Hodgkin's lymphoma trial group. Br J Haematol 180(2):236–245
20. Pfreundschuh M, Ho AD, Cavallin-Stahl E, Wolf M, Pettengell R, Vasova I, Belch A, Walewski J, Zinzani PL, Mingrone W, Kvaloy S, Shpilberg O, Jaeger U, Hansen M, Corrado C, Scheliga A, Loeffler M, Kuhnt E, MabThera International Trial (MInT) Group (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
21. Zhou Z, Sehn LH, Rademaker AW, Gordon LI, LaCasce AS, Crosby-Thompson A, Vanderplas A, Zelenetz AD, Abel GA, Rodriguez MA, Nademanee A, Kaminski MS, Czuczman MS, Millenson M, Niland J, Gascoyne RD, Connors JM, Friedberg JW, Winter JN (2014) An enhanced International Prognostic

Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. Blood 123:837–842

22. Montalbán C, Díaz-López A, Dlouhy I, Rovira J, Lopez-Guillermo A, Alonso S, Martín A, Sancho JM, García O, Sánchez JM, Rodríguez M, Novelli S, Salar A, Gutiérrez A, Rodríguez-Salazar MJ, Bastos M, Domínguez JF, Fernández R, Gonzalez de Villambrosia S, Queizan JA, Córdoba R, de Oña R, López-Hernandez A, Freue JM, Garrote H, López L, Martin-Moreno AM, Rodriguez J, Abraira V, García JF, the GELTAMO-IPI Project Investigators (2017) Validation of the NCCN-IPI for diffuse large B-cell lymphoma (DLBCL): the addition of β2-microglobulin yields a more accurate GELTAMO-IPI. Br J Haematol 176(6):918–928
23. Bishton MJ, Hughes S, Richardson F, James E, Bessell E, Sovani V, Ganatra R, Haynes AP, McMillan AK, Fox CP (2016) Delineating outcomes of patients with diffuse large b cell lymphoma using the national comprehensive cancer network-international prognostic index and positron emission tomography-defined remission status; a population-based analysis. Br J Haematol 172:246–254
24. El-Galaly TC, Villa D, Alzahrani M et al (2015) Outcome prediction by extranodal involvement, IPI, R-IPI, and NCCN-IPI in the PET/CT and rituximab era: a Danish-Canadian study of 443 patients with diffuse-large B-cell lymphoma. Am J Hematol 90: 1041–1046
25. Cheson BD, Horning SJ, Coiffier B et al (1999) Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. J Clin Oncol 17(4):1244
26. Cheson BD, Pfistner B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, Coiffier B, Fisher RI, Hagenbeek A, Zucca E, Rosen ST, Stroobants S, Lister TA, Hoppe RT, Dreyling M, Tobinai K, Vose JM, Connors JM, Federico M, Diehl V, International Harmonization Project on Lymphoma (2007) Revised response criteria for malignant lymphoma. J Clin Oncol 25(5):579–586
27. Harrell FE Jr, Califf RM, Pryor DB et al (1982) Evaluating the yield of medical tests. JAMA 247(18):2543–2546
28. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ (2011) On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 30(10):1105–1117
29. R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
30. Schmitz N, Zeynalova S, Nickelsen M, Kansara R, Villa D, Sehn LH, Glass B, Scott DW, Gascoyne RD, Connors JM, Ziepert M, Pfreundschuh M, Loeffler M, Savage KJ (2016) CNS international prognostic index: a risk model for CNS relapse in patients with diffuse large B-cell lymphoma treated with R-CHOP. J Clin Oncol 34:3150–3156
31. Gutiérrez-García G, Cardesa-Salzmann T, Climent F et al (2011) Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. Blood 117(18):4836–4843
32. Coutinho R, Clear AJ, Owen A, Wilson A, Matthews J, Lee A, Alvarez R, da Silva MG, Cabecadas J, Calaminici M, Gribben JG (2013) Poor concordance among nine immunohistochemistry classifiers of cell-of-origin for diffuse large B-cell lymphoma: implications for therapeutic strategies. Clin Cancer Res 19(24):6686–6695
33. Biccler J, Eloranta S, Brown N et al (2018) Simplicity at the cost of predictive accuracy in diffuse large B-cell lymphoma: a critical assessment of the R-IPI, IPI, and NCCN-IPI. Cancer Med 7(1): 114–122