

Status and prognostic nomogram of patients with Burkitt lymphoma

JIELUN LU¹, HUO TAN², BO LI¹, SHUYI CHEN², LIHUA XU² and YAWEI ZOU¹

Departments of ¹Pediatrics and ²Hematology, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong 510000, P.R. China

Received November 26, 2018; Accepted October 23, 2019

DOI: 10.3892/ol.2019.11155

Abstract. The purpose of the present study was to evaluate the newest status of patients diagnosed Burkitt lymphoma (BL), an aggressive lymphoma subset with a high cure rate. Furthermore, the study aimed to create prognostic nomograms to consider various prognostic factors and estimate patient survival, paving the way for clinical decision-making. A total of 4,600 patients diagnosed with BL between 1983 and 2015 were investigated, via data collected from the SEER database. The overall status of the patients was analyzed through several aspects, including incidence and survival analysis of the previous three decades using the log-rank test and the Kaplan-Meier method. In order to construct and validate the nomograms, the patient diagnosed during 2005-2015 were randomly assigned to the training cohort and validation cohort. Univariate and multivariate analyses were applied to identify independent factors that were further included in the nomograms, predicting 3- and 5-year overall survival (OS) and cancer-specific survival (CSS). The data of the training cohort were used for internal validation and validation cohort used to external validation. C-index and calibration plots were used to validate the nomograms, comparing predicted values with actual outcomes. The incidence of BL was gradually increased from 1984 and reached its peak in 2009, at a rate of 0.491 per 100,000 [95% confidence interval (CI), 0.412-0.581]. From 2009, the incidence slowly declined year by year and dropped to 0.280 per 100,000 (95% CI, 0.224-0.346). The OS and CSS rates of patients diagnosed between 2005 and 2015 were increased, in contrast with

those of patients diagnosed from 1983-1993 and 1994-2004. A total of five variables, including age, race, chemotherapy, primary site and stage, proved to be the prognostic factors of BL and were used to construct the nomograms predicting 3- and 5-year OS and CSS. The internal and external calibration plots for the probability of 3- and 5-year OS and CSS were consistent between nomogram prediction and observed outcomes. The slow decline in incidence and the significantly improved cure rate make BL a disease that is no longer an urgent problem. Effective nomograms were developed to predict the OS and CSS of patients with BL.

Introduction

Burkitt lymphoma (BL) is a highly aggressive B-cell non-Hodgkin lymphoma that was discovered by Denis Burkitt in 1958 in Africa (1). This was the first tumor shown to have a chromosomal translocation activating an oncogene (c-MYC) (2), as well as being associated with Epstein-Barr virus (3). Owing to the introduction of dose-intense multi-drug combination chemotherapy, prophylaxis for central nervous system disease and improvements in supportive care, BL has become the aggressive lymphoma subset that is associated with the highest cure rate in both adults and children (4).

However, compared with published single or multi-center clinical studies, the 3-year survival rate according to the 2002-2008 Surveillance, Epidemiology and End Results (SEER) US study, including 3,691 patients with BL, was notably low (56%) (5). In addition, current guidelines are still recommending uniform, aggressive treatment for all patients without considering prognosis and risk categorization. As with other malignancies, identifying the prognostic factors of BL is critical, as these can guide treatment selection. Castillo *et al* (6) reported that age, race and stage have independent prognostic roles in survival, by analyzing data from 1,428 patients recorded between 1998 and 2009 from the SEER program database. Nevertheless, systematically identifying prognostic factors to arrive at a more precise estimate of prognosis is of importance, as survival is undoubtedly multifactorial. Nomograms, a simple graphical presentation of a multivariate predictive model, show the impact of each included variable on an outcome of interest, thereby providing a numerical probability of the outcome (7). One of the advantages of nomograms is the ability to integrate multiple prognostic

Correspondence to: Dr Yawei Zou, Department of Pediatrics, The First Affiliated Hospital of Guangzhou Medical University, 151 Yanjiang West Road, Yuexiu, Guangzhou, Guangdong 510000, P.R. China
E-mail: zouyawei@gzhmu.edu.cn

Dr Lihua Xu, Department of Hematology, The First Affiliated Hospital of Guangzhou Medical University, 151 Yanjiang West Road, Yuexiu, Guangzhou, Guangdong 510000, P.R. China
E-mail: xlhua@gzhmu.edu.cn

Key words: Burkitt lymphoma, incidence, survival analysis, nomogram, prediction model

factors into the individual survival numerical estimate of a single patient, providing individualized survival predictions. Nomograms have gradually become a useful tool in the field of oncology, for predicting cancer prognosis (8-10). However, to the best of our knowledge, a nomogram for survival prediction in BL has not been reported.

The present study aimed to use data from a large US population dataset, the SEER database, to provide an update on the incidence and evaluate treatment outcomes in patients with BL over the past three decades, between 1983 and 2015. Furthermore, nomograms were developed and validated to predict the overall survival (OS) and cancer-specific survival (CSS) of patients with BL with different values of prognostic factors.

Patients and methods

Data sources and patient selection. The SEER Program that was used to obtain the data for the present study was created in 1973 to collect cancer statistics in the US, to reduce the cancer burden among the US population (<http://seer.cancer.gov>). The data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status (through linkage with the National Center for Health Statistics) were routinely collected in all 18 registries spread across the US. Currently, SEER registries cover 28% of the US population; coverage includes 25% of white Americans, 26% of African Americans, 38% of Hispanics, 44% of Native Americans and Alaska natives, 50% of Asians and 67% of Hawaiian/Pacific islanders. The SEER database was accessed via SEER*Stat software (version 8.3.5; National Cancer Institute).

The inclusion criteria of the study were as follows: i) Diagnosed with BL as the primary malignancy, according to the 3rd edition (ICD-O-3) histology code 9687 (Burkitt lymphoma) (earlier patients coded as ICD-O-1 and ICD-O-2 were previously re-coded as ICD-O-3); ii) only one malignant tumor present, in order to exclude interference from other tumors; and iii) followed up for vital status until December 31st, 2015, to ensure a minimal follow-up length of two years. The exclusion criteria were as follows: i) Diagnosed by autopsy or reported only on a death certificate; and ii) one or more statistical factors unknown. In total, 4,600 patients were included in the analysis.

Definition of variables. Patients were divided into three groups according to the time of diagnosis (1983-1993, 1994-2004 and 2005-2015), in order to evaluate the development of the diagnosis and treatment of BL. The patients were further stratified by other clinicopathological features including age, sex (male or female), race (black, white, American Indian/Alaska native, Chinese and other Asian/Pacific islander) marital status (married, single, or widowed/divorced/separated), Ann Arbor stage (I, II, or IV), primary site (lymph nodes, digestive system, oral cavity and pharynx, bone marrow, nervous system and other), percentage of families with incomes below the poverty line of the county (% families below poverty; ≤ 5 , 5-10, 10-20 or ≥ 20), median household income, chemotherapy (yes, or no/unknown), outcome at study cutoff (alive or succumbed), and survival time.

To assess the prognostic significance of age and to avoid arbitrary predetermined cutoff points, X-tile plots were constructed to assess age using four groups, and the most effective cutoff points were identified following correction for the use of minimum P statistics by Miller-Siegmund P-value correction, and dividing into four parts. The X-tile program (Yale University) was first introduced to identify the best cutoff values of several prognostic factors in cohorts of patients with breast cancer (11).

Incidence and survival analysis. To assess changes in the incidence of BL, the overall population incidence was estimated by calculating from the patient data in the SEER database. Incidence rates were expressed per 1,000,000 in the population and were age-adjusted to the 2000 US standard population (12).

As improvements in diagnosis and treatment are made over time, the present study intended to assess changes in survival rates, including OS, meaning the time from diagnosis to mortality for any reason, and CSS, defined as the time from diagnosis to mortality attributed to BL, in patients with BL grouped by different diagnostic time periods (1983-1993, 1994-2004 and 2005-2015).

Construction and validation of the nomograms. To improve predictions for the construction of the nomogram plots for BL, 2,669 patients diagnosed during 2005-2015, identified from the SEER database, were randomly assigned to the training cohort (n=1,762, for the construction and internal validation of the nomograms) and validation cohort (n=907, for external validation of the nomograms). Potential prognostic factors including age, sex, race, marital status, % families below poverty, chemotherapy, primary site and stage, that were associated with OS or CSS were included in the Kaplan-Meier analyses and log-rank tests. To eliminate the interactions between various factors, a multivariate Cox proportional hazard model was used to identify independent prognostic factors and, furthermore, hazard ratios were calculated for the significant prognostic factors. To predict whether a patient will experience the event (mortality for any reason or due to BL) at 3 or 5 years, nomograms incorporate prognostic factors and present them in an intuitive way. To validate the nomogram plots internally and externally, bootstrap validation was used with 1,000-iteration resampling. The data of the training cohort were used for internal validation and validation cohort was used for external validation. Harrell's concordance index (C-index), was used to estimate the predictive accuracy of the nomogram, for the consistency of the predicted and actual values (13). The range of the C-index is 0.5 to 1; 0.5 is completely inconsistent, indicating that the model has no predictive effect, while 1 is completely consistent, indicating that the model prediction results are completely consistent with the actual results. It is generally believed that the C-index is less accurate at 0.50-0.70, has moderate accuracy between 0.71-0.90 and has high accuracy at >0.90 . In addition, calibration plots were also used to validate the nomograms, comparing predicted values with actual outcomes.

Statistical analysis. OS was defined as the time from diagnosis to death from all possible causes. CSS was defined as the time from diagnosis to death attributed to BL. Patients who were

Table I. Demographics and clinical characteristics of patients with Burkitt lymphoma.

| Characteristics | No. (%) of patients (n=4,600) |
|--|-------------------------------|
| Age, years, mean \pm SD | 39.71 \pm 22.86 |
| Sex | |
| Male | 3,490 (75.9) |
| Female | 1,110 (24.1) |
| Ethnicity | |
| White | 3,738 (81.3) |
| Black | 469 (10.2) |
| Chinese | 65 (1.4) |
| Other Asian/Pacific islander | 299 (6.5) |
| American Indian/Alaska native | 29 (0.6) |
| Marital status | |
| Single (never married) | 2,552 (55.5) |
| Married (including common law) | 1,548 (33.7) |
| Widowed/divorced/separated | 500 (10.9) |
| % Families below poverty | |
| \leq 5 | 616 (13.4) |
| 5-10 | 1,911 (41.5) |
| 10-20 | 1,953 (42.5) |
| \geq 20 | 120 (2.6) |
| Median household income (in tens) \pm SD | 6,073.37 \pm 1,360.855 |
| Primary site | |
| Lymph nodes | 3,192 (69.4) |
| Digestive system | 707 (15.4) |
| Oral cavity and pharynx | 204 (4.4) |
| Bone marrow and nervous system | 126 (2.7) |
| Other | 371 (8.1) |
| Stage | |
| I | 939 (20.4) |
| II | 728 (15.8) |
| III | 528 (11.5) |
| IV | 2,405 (52.3) |
| Year of diagnosis | |
| 1983-1993 | 440 (9.6) |
| 1994-2004 | 1,491 (32.4) |
| 2005-2015 | 2,669 (58.0) |
| Chemotherapy | |
| Yes | 4,090 (88.9) |
| No/unknown | 510 (11.1) |
| Outcome at study cut-off | |
| Alive | 2,432 (52.9) |
| Dead | 2,168 (47.1) |
| Survival time, months, mean \pm SD | 59.74 \pm 75.241 |

SD, standard deviation.

Table II. Age grouping profile of patients with Burkitt lymphoma.

| Characteristics | No. (%) of patients (n=4,600) |
|-------------------|-------------------------------|
| Age, group, years | |
| \leq 20 | 1,183 (25.7) |
| 21-40 | 1,147 (24.9) |
| 41-55 | 1,148 (25.0) |
| \geq 56 | 1,122 (24.4) |

the patients were compared between the training and validation cohort and among the different groups of diagnostic time periods (1983-1993, 1994-2004 and 2005-2015). A Chi-square test was used for counting data (age group, sex, race, marital status, % families below poverty, chemotherapy, primary site and stage). An analysis of variance test (followed by Least Significant Difference) was used for analysis of measurement data (age, median household income and survival time). Potential prognostic factors including age, sex, race, marital status, % families below poverty, chemotherapy, primary site and stage were incorporated in the univariate analysis using Kaplan-Meier analyses and log-rank tests for OS and CSS, respectively. Significant variables were further assessed using multivariate Cox proportional hazards model hazard ratios of each variable with corresponding 95% confidence intervals (CI) were calculated regarding OS and CSS.

All statistical analysis was conducted using SPSS version 23.0 (IBM Corp.). R version 3.3.5 (<http://www.r-project.org/>) was used to construct and validate nomograms. $P < 0.05$ (two-tailed) was considered to indicate a statistically significant difference.

Results

Patient characteristics. Followed the inclusion and exclusion criteria, 4,600 patients diagnosed with BL between 1983 and 2015 were identified. All the patients were classified by their characteristics, including age, sex, race, marital status, Ann Arbor stage, primary site, % families below poverty, median household income, chemotherapy, radiotherapy, outcome at study cutoff and survival time (Table I).

Using the X-tile program, it was possible to confirm the best cutoff values of age as a prognostic factor for BL. The X-tile program divided age into four subgroups by choosing the grouping scheme that valued the highest χ^2 record. The X-tile program identified the most appropriate cutoff values of age as 20, 40 and 56 years based on OS (Fig. 1). Thus, age was divided into the following categories: \leq 20, 21-40, 41-55 and \geq 56 years. The age grouping profile of the patients with BL is described in Table II.

Incidence and survival analysis. The incidence of BL between 1983 and 2015 was calculated by updating the SEER data to explore the recent development trends of the disease (Fig. 2), although the incidence between 1973 and 2008 has been previously reported (5). Incidence gradually increased from 1983 and reached its peak in 2009, with a rate of

alive at the time of last follow-up were counted as censored observations. Baseline clinicopathological characteristics of

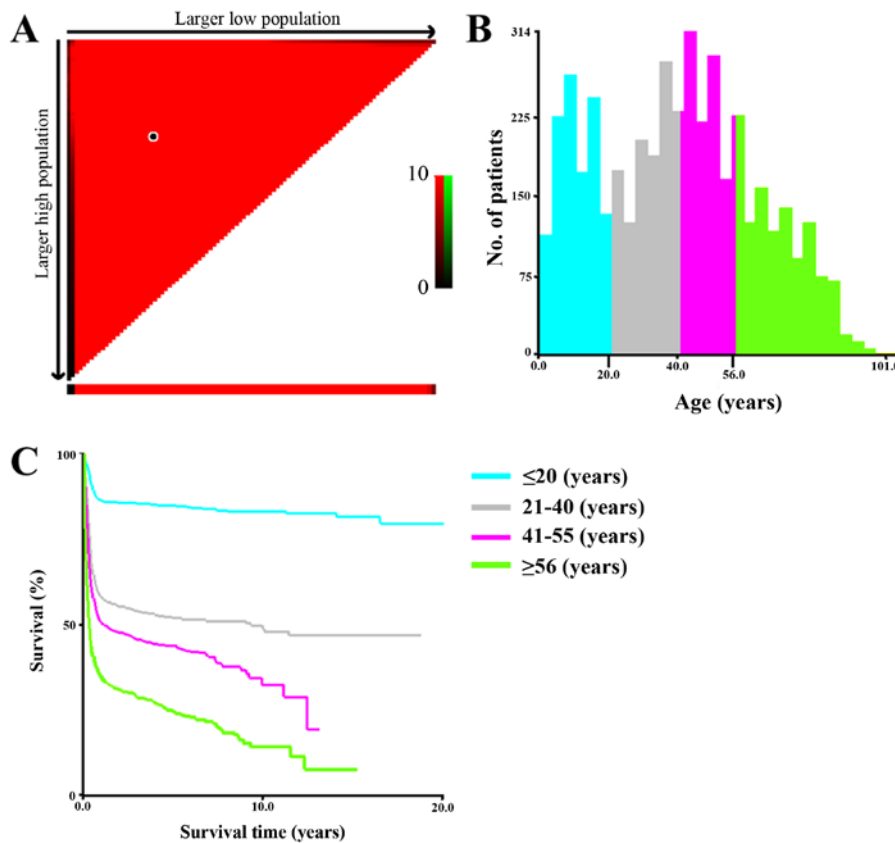


Figure 1. Optimal cutoff values of age identified by X-tile analysis of the patients with BL. (A) The optimal cutoff points calculated and selected by X-tile software evaluating the χ^2 log-rank value. The x-axis represents all potential cut-points from low to high (left to right) that define a low subset, whereas the y-axis represents cut-points from high to low (top to bottom), that define a high subset. The arrows represent the direction in which the low subset (x-axis) and the high subset (y-axis) increase in size. Red coloration of cut-points indicates an inverse correlation with survival, whereas green coloration represents direct associations. The optimal cut-point occurs at the brightest pixel (green or red), which was recognized by computer software. The optimal cutoff point occurs at the brightest pixel with red coloration, indicating an inverse correlation with survival, highlighted by the black circle. (B) Histogram based on the optimal cutoff point. (C) Kaplan-Meier analysis based on the optimal cutoff points. The optimal cutoff points for age were 20.0, 40.0 and 56.0 years based on the survival analysis ($P < 0.001$). BL, Burkitt lymphoma.

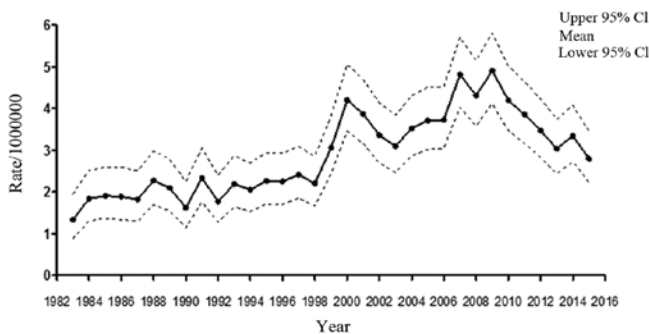


Figure 2. Estimated incidence of Burkitt lymphoma during 1983-2015. Incidence gradually increased from 1983 and reached its peak in 2009, with a rate of 0.491 per 100,000 (95% CI, 0.412-0.581). From 2009, the incidence slowly declined year by year and dropped to 0.280 per 100,000 (95% CI, 0.224-0.346). CI, confidence interval.

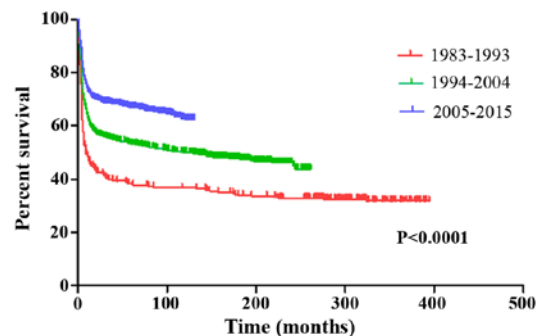


Figure 3. Kaplan-Meier overall survival of the independent validation set according to three groups as determined by diagnostic year. The results showed that Burkitt lymphoma patients in early diagnostic time group have poorer overall survival than those in late diagnostic time group (log-rank $P < 0.0001$).

0.491 per 100,000 (95% CI, 0.412-0.581). From 2009, the incidence slowly declined year by year and dropped to 0.28 per 100,000 (95% CI, 0.224-0.346).

Among the different groups of diagnostic time periods (1983-1993, 1994-2004 and 2005-2015), the aforementioned characteristics of the patients significantly differed ($P < 0.01$; Table III), revealing the differences in patient composition. The

univariate survival analysis showed that an earlier diagnostic time was associated with worse outcomes, while a more recent diagnosis time was associated with better outcomes ($P < 0.0001$; Fig. 3). The OS rates of patients diagnosed during 2005-2015 were increased, with a 3-year OS of 62.6% and a 5-year OS of 61.2%, in contrast to the 3-year OS of 35.2% and 5-year OS of 33.8% of patients diagnosed in the period 1983-1993 (Fig. 4A). Similarly, the CSS rates of patients diagnosed in 2005-2015

Table III. Demographics and clinical characteristics of Burkitt lymphoma patients with different diagnostic time.

| Characteristics | No. (%) of patients | | | P-value ^a |
|--|-------------------------|-------------------------|-------------------------|----------------------|
| | 1983-1993 | 1994-2004 | 2005-2015 | |
| Age, years, mean \pm SD | 32.02 \pm 20.74 | 39.56 \pm 23.50 | 41.07 \pm 22.58 | <0.001 |
| Age, group, years | | | | <0.001 |
| \leq 20 | 153 (34.77) | 401 (26.89) | 629 (23.57) | |
| 21-40 | 150 (34.09) | 363 (24.35) | 634 (23.75) | |
| 41-55 | 78 (17.73) | 357 (23.94) | 713 (26.71) | |
| \geq 56 | 59 (13.41) | 370 (24.82) | 693 (25.96) | |
| Sex | | | | 0.002 |
| Male | 364 (82.73) | 1,125 (75.45) | 2,001 (74.97) | |
| Female | 76 (17.27) | 366 (24.55) | 668 (25.03) | |
| Race | | | | 0.003 |
| White | 386 (87.73) | 1,216 (81.56) | 2,136 (80.03) | |
| Black | 29 (6.59) | 147 (9.86) | 293 (10.98) | |
| Chinese | 2 (0.45) | 22 (1.48) | 41 (1.54) | |
| Other Asian/Pacific islander | 21 (4.77) | 103 (6.91) | 175 (6.56) | |
| American Indian/Alaska native | 2 (0.45) | 3 (0.20) | 24 (0.90) | |
| Marital status | | | | <0.001 |
| Single (never married) | 303 (68.86) | 832 (55.80) | 1,417 (53.09) | |
| Married (including common law) | 94 (21.36) | 494 (33.13) | 960 (35.97) | |
| Widowed/divorced/separated | 43 (9.77) | 165 (11.06) | 292 (10.94) | |
| % Families below poverty | | | | <0.001 |
| \leq 5 | 79 (17.95) | 284 (19.05) | 253 (9.48) | |
| 5-10 | 261 (59.32) | 565 (37.89) | 1,085 (40.65) | |
| 10-20 | 95 (21.59) | 619 (41.52) | 1,239 (46.42) | |
| \geq 20 | 5 (1.14) | 23 (1.54) | 92 (3.45) | |
| Median household income (in tens) \pm SD | 6,485.27 \pm 1,233.39 | 6,071.02 \pm 1,324.73 | 6,006.77 \pm 1,389.06 | <0.001 |
| Primary site | | | | 0.004 |
| Lymph nodes | 323 (73.41) | 1,070 (71.76) | 1,799 (67.40) | |
| Digestive system | 67 (15.23) | 228 (15.29) | 412 (15.44) | |
| Oral cavity and pharynx | 12 (2.73) | 61 (4.09) | 131 (4.91) | |
| Bone marrow and nervous system | 13 (2.95) | 35 (2.35) | 78 (2.92) | |
| Other | 25 (5.68) | 97 (6.51) | 249 (9.33) | |
| Stage | | | | <0.001 |
| I | 92 (20.91) | 330 (22.13) | 517 (19.37) | |
| II | 43 (9.77) | 241 (16.16) | 444 (16.64) | |
| III | 34 (7.73) | 170 (11.40) | 324 (12.14) | |
| IV | 271 (61.59) | 750 (50.30) | 1,384 (51.85) | |
| Chemotherapy | | | | 0.001 |
| Yes | 376 (85.45) | 1,303 (87.39) | 2,411 (90.33) | |
| No/Unknown | 64 (14.55) | 188 (12.61) | 258 (9.67) | |
| Outcome at study cut-off | | | | <0.001 |
| Alive | 126 (28.64) | 638 (42.79) | 1,668 (62.50) | |
| Dead | 314 (71.36) | 853 (57.21) | 1,001 (37.50) | |
| Survival time, months, mean \pm SD | 100.96 \pm 140.57 | 81.80 \pm 84.29 | 40.62 \pm 40.28 | <0.001 |

^aBaseline clinicopathological characteristics of the patients were compared among the different groups of diagnostic time periods (1983-1993, 1994-2004 and 2005-2015). Chi-square test was used for counting data (age group, sex, race, marital status, % families below poverty, chemotherapy, primary site and stage). Analysis of variance test was used for analysis of measurement data (age, median household income and survival time). SD, standard deviation.

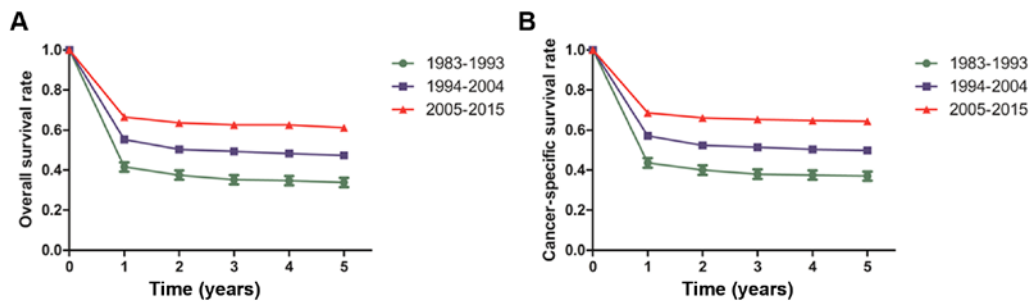


Figure 4. Trends in 5-year survival rates for patients with BL from the SEER database between 1983 and 2015. Data are shown with different groups classified by different diagnostic years (1983-1993, 1994-2004 and 2005-2015). (A) The OS rates for patients with BL. (B) The CSS rates for patients with BL. OS, overall survival; CSS, cancer-specific survival; BL, Burkitt lymphoma.

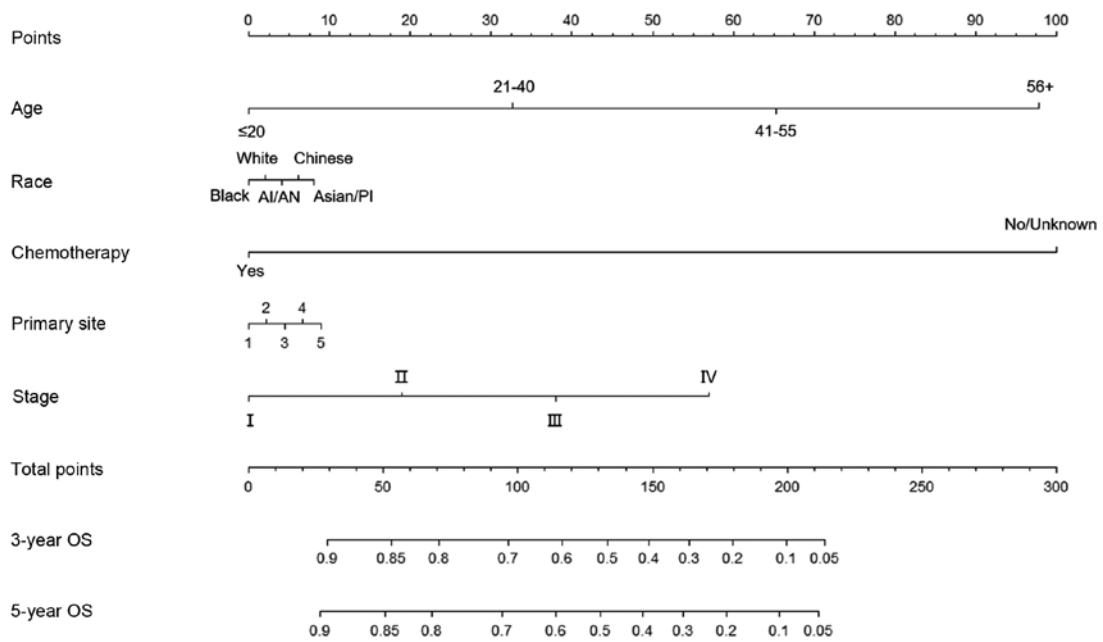


Figure 5. Nomogram predicting the 3- and 5-year OS of patients with BL. PI, Pacific islander; AI, American Indian; AN, Alaska Native; 1, lymph nodes; 2, digestive system; 3, oral cavity and pharynx; 4, other; 5, bone marrow and nervous system; OS, overall survival.

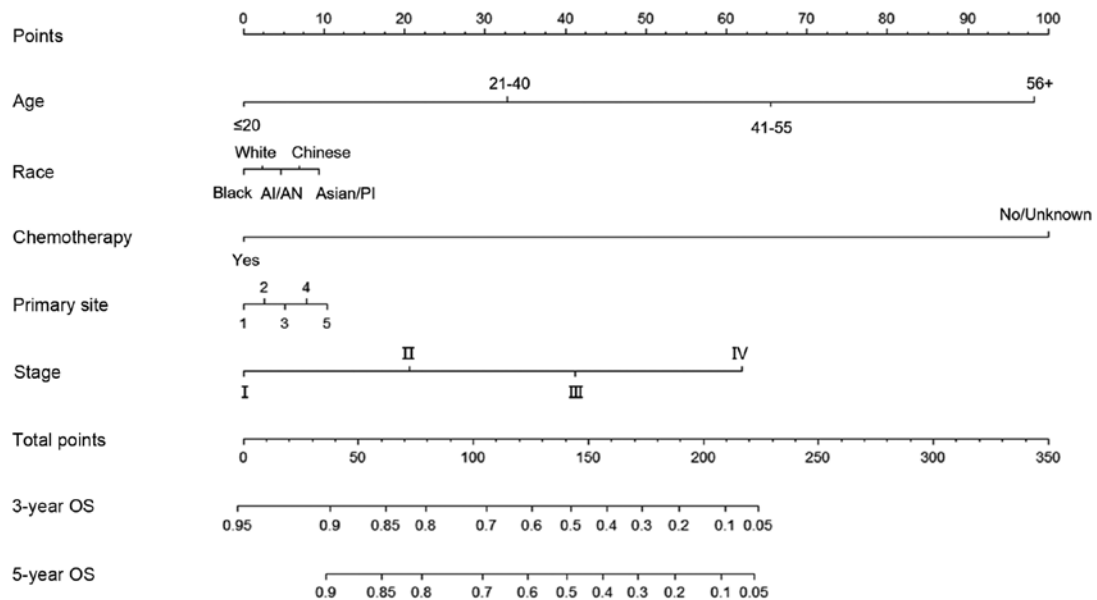


Figure 6. Nomogram predicting the 3- and 5-year CSS of patients with BL. PI, Pacific islander; AI, American Indian; CSS, cancer-specific survival; AN, Alaska native; 1, lymph nodes; 2, digestive system; 3, oral cavity and pharynx; 4, other; 5, bone marrow and nervous system.

Table IV. Demographics and clinical characteristics of training and validation cohort for BL patients.

| Characteristics | No. of patients ^b | | P-value ^a |
|--|------------------------------|---------------------------|----------------------|
| | Training cohort (n=1,762) | Validation cohort (n=907) | |
| Age, years, mean \pm SD | 41.45 \pm 22.50 | 40.33 \pm 22.73 | 0.913 |
| Age, group, years | | | 0.137 |
| \leq 20 | 398 (22.59) | 231 (25.47) | |
| 21-40 | 426 (24.18) | 208 (22.93) | |
| 41-55 | 461 (26.16) | 252 (27.78) | |
| \geq 56 | 477 (10.05) | 216 (23.81) | |
| Sex | | | 0.741 |
| Male | 1,317 (74.74) | 684 (75.41) | |
| Female | 445 (25.26) | 223 (24.59) | |
| Ethnicity | | | 0.677 |
| Black | 185 (10.50) | 108 (11.91) | |
| White | 1,418 (80.48) | 718 (79.16) | |
| American Indian/Alaska native | 18 (1.02) | 6 (0.66) | |
| Chinese | 24 (1.36) | 17 (1.87) | |
| Other Asian/Pacific islander | 117 (6.64) | 58 (6.39) | |
| Marital status | | | 0.354 |
| Single (never married) | 919 (52.16) | 498 (54.91) | |
| Married (including common law) | 650 (36.89) | 310 (34.18) | |
| Widowed/divorced/separated | 193 (10.95) | 99 (10.92) | |
| % Families below poverty | | | 0.974 |
| \leq 5 | 172 (9.76) | 81 (8.93) | |
| 5-10 | 715 (40.58) | 370 (40.79) | |
| 10-20 | 822 (46.65) | 417 (45.98) | |
| \geq 20 | 53 (3.01) | 39 (4.30) | |
| Median household income (in tens) \pm SD | 6,022.31 \pm 1,374.65 | 5,976.59 \pm 1,416.92 | 0.845 |
| Primary site | | | 0.812 |
| Lymphonodus | 1,201 (68.16) | 598 (65.93) | |
| Digestive system | 266 (15.10) | 146 (16.10) | |
| Oral cavity and pharynx | 86 (4.88) | 45 (4.96) | |
| Bone marrow and nervous system | 51 (2.89) | 27 (2.98) | |
| Other | 158 (8.97) | 91 (10.03) | |
| Stage | | | 0.140 |
| I | 320 (18.16) | 197 (21.72) | |
| II | 298 (16.40) | 146 (16.10) | |
| III | 211 (11.98) | 113 (12.46) | |
| IV | 933 (52.95) | 451 (49.72) | |
| Chemotherapy | | | 0.356 |
| Yes | 1,585 (89.95) | 826 (91.07) | |
| No/Unknown | 177 (10.05) | 81 (8.93) | |
| Outcome at study cut-off | | | 0.933 |
| Alive | 1,100 (62.43) | 568 (62.62) | |
| Dead | 662 (37.57) | 339 (37.38) | |
| Survival time, Months, mean \pm SD | 40.33 \pm 40.43 | 41.16 \pm 40.00 | 0.352 |

^aBaseline clinicopathological characteristics of the patients were compared between the training and validation cohort. Chi-square test was used for counting data (age group, sex, race, marital status, % families below poverty, chemotherapy, primary site and Stage). Analysis of variance test was used for analysis of measurement data (age, median household income and survival time). ^b2,669 patients diagnosed BL during 2005-2015 were randomly assigned to the training cohort and validation cohort by SPSS version 23.0. SD, standard deviation; BL, Burkitt lymphoma.

Table V. Univariate analysis and multivariate analysis of overall survival in the training cohort.

| Characteristics | Univariate analysis P-value | Multivariate analysis | |
|--------------------------------|-----------------------------|-----------------------|--------------------|
| | | HR (95% CI) | P-value |
| Age, group, years | 0.000 | | |
| ≤20 | | 0.106 (0.071-0.157) | <0.001<0.001<0.001 |
| 21-40 | | 0.389 (0.308-0.492) | |
| 41-55 | | 0.589 (0.487-0.711) | |
| ≥56 | | Reference | |
| Sex | 0.108 | | NI |
| Male | | NI | |
| Female | | | |
| Ethnicity | 0.000 | | |
| White | | 1.432 (0.995-2.062) | 0.054 |
| Black | | 0.911 (0.673-1.233) | 0.545 |
| American Indian/Alaska native | | 2.942 (1.546-5.597) | 0.001 |
| Chinese | | 1.272 (0.570-2.838) | 0.557 |
| Other Asian/Pacific islander | | Reference | |
| Marital status | 0.000 | | |
| Single (never married) | | 1.027 (0.803-0.312) | 0.834 |
| Married (including common law) | | 0.952 (0.761-1.190) | 0.665 |
| Widowed/divorced/separated | | Reference | |
| % Families below poverty | 0.001 | | |
| ≤5 | | Reference | |
| 5-10 | | 0.989 (0.838-1.167) | 0.895 |
| 10-20 | | 1.012 (0.859-1.192) | 0.886 |
| ≥20 | | 0.956 (0.669-1.367) | 0.806 |
| Primary site | 0.000 | | |
| Oral cavity and pharynx | | 0.82 (0.563-1.197) | 0.305 |
| Lymphonodus | | 0.744 (0.488-1.132) | 0.168 |
| Digestive system | | 0.876 (0.469-1.635) | 0.678 |
| Bone marrow and nervous system | | Reference | |
| Other | | 0.957 (0.619-1.480) | 0.845 |
| Stage | 0.000 | | |
| I | | 0.415 (0.322-0.535) | <0.001<0.0010.004 |
| II | | 0.461 (0.356-0.596) | |
| III | | 0.688 (0.532-0.889) | |
| IV | | Reference | |
| Chemotherapy | 0.000 | | |
| Yes | | 4.895 (3.982-6.017) | <0.001 |
| No/unknown | | | |

CI, confidence interval; NI, ; HR, hazard ratio.

were 65.3% for 3-year CSS and 64.4% for 5-year CSS, while the 3-year CSS was 37.0% and the 5-year CSS was 38.0% for patients diagnosed during 1983-1993 (Fig. 4B).

Construction and validation of the nomograms. Demographics and clinical characteristics of the training and validation cohorts of BL patients were collated (Table IV). Using univariate and multivariate analysis, four variables (age, race, chemotherapy

and stage) were found to be significant factors associated with OS (Table V) and CSS (Table VI). Though primary site did not produce meaningful results in the multivariate analysis, it was still included in the nomogram construction to verify its precise clinical significance. The nomograms for 3- and 5-year OS and CSS prediction were constructed by integrating the impact of five independent predictors (Figs. 5 and 6). The predicted probabilities of each covariate were mapped on a

Table VI. Univariate analysis and multivariate analysis of cancer-specific survival in the training cohort.

| Characteristics | Univariate analysis P-value | Multivariate analysis | |
|--------------------------------|-----------------------------|-----------------------|--------------------|
| | | HR (95% CI) | P-value |
| Age, group, years | 0.000 | | <0.001<0.001<0.001 |
| ≤20 | | 0.106 (0.071-0.157) | |
| 21-40 | | 0.389 (0.308-0.492) | |
| 41-55 | | 0.589 (0.487-0.711) | |
| ≥56 | | Reference | |
| Sex | 0.182 | NI | NI |
| Male | | | |
| Female | | | |
| Ethnicity | 0.000 | | |
| White | | 1.432 (0.995-2.062) | 0.054 |
| Black | | 0.911 (0.673-1.233) | 0.545 |
| American Indian/Alaska native | | 2.942 (1.546-5.597) | 0.001 |
| Chinese | | 1.272 (0.570-2.838) | 0.557 |
| Other Asian/Pacific islander | | Reference | |
| Marital status | 0.000 | | |
| Single (never married) | | 1.027 (0.803-1.312) | 0.834 |
| Married (including common law) | | 0.952 (0.761-1.190) | 0.665 |
| Widowed/divorced/separated | | Reference | |
| % Families below poverty | 0.002 | | |
| ≤5 | | 1.011 (0.847-1.207) | 0.707 |
| 5-10 | | 1.035 (0.869-1.233) | 0.879 |
| 10-20 | | 1.014 (0.698-1.474) | 0.449 |
| ≥20 | | Reference | |
| Primary site | 0.005 | | |
| Oral cavity and pharynx | | 0.821 (0.563-1.197) | 0.305 |
| Lymphonodus | | 0.744 (0.488-1.132) | 0.168 |
| Digestive system | | 0.876 (0.469-1.635) | 0.678 |
| Bone marrow and nervous system | | 0.957 (0.619-1.480) | 0.845 |
| Other | | Reference | |
| Stage | 0.000 | | <0.001<0.0010.004 |
| I | | Reference | |
| II | | 0.415 (0.322-0.535) | |
| III | | 0.461 (0.356-0.596) | |
| IV | | 0.688 (0.532-0.889) | |
| Chemotherapy | 0.000 | | |
| Yes | | 3.898 (3.358-4.525) | <0.001 |
| No/Unknown | | | |

CI, confidence interval; NI, ; HR, hazard ratio.

scale from 0 to 100, where 100 represented the highest effect and 0 the lowest. The total points accumulated were used to predict the probability for the patients. Each definition of the predictors has an accurate value displayed in the digital axis. The exact score of each point of the various covariates was calculated (Table VII). For the practical application of the nomograms, an example is provided: The total number of points of a 25-year old black patient diagnosed with stage II

BL, with a lymph node primary site, would be 60 if treated with chemotherapy and the estimated 3-year OS rate would be 85%.

The nomograms were validated both internally and externally. The C-indexes of the nomograms predicting 3- and 5-year OS in the internal and external validation were 0.777 (95% CI, 0.758-0.796) and 0.76 (95% CI, 0.735-0.786), respectively. The C-indexes of the nomograms predicting

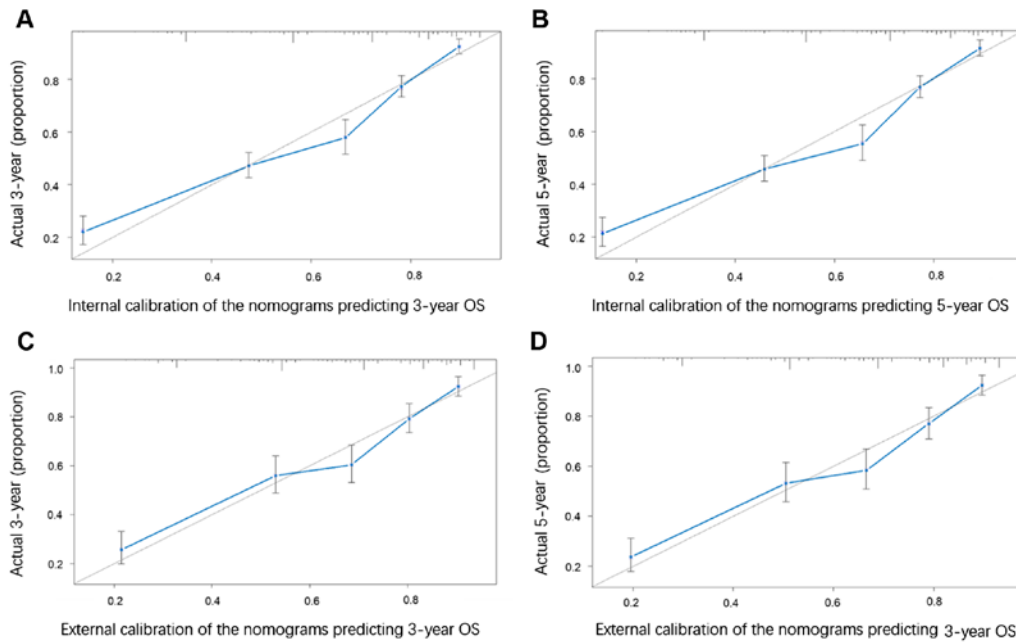


Figure 7. Internal calibration and external calibration of the nomograms predicting 3 and 5-year OS for patients with BL. (A) Internal calibration for 3-year OS. (B) Internal calibration for 5-year OS. (C) External calibration for 3-year OS. (D) External calibration for 5-year OS. The x-axis of each plot represents the nomogram prediction and the y-axis represents the observed outcomes. Each cohort equality divided into 5 and the distance between the points was calculated; the diagonal lines on the plots indicate the consistency between predictions and observed outcomes. OS, overall survival; CSS, cancer-specific survival.

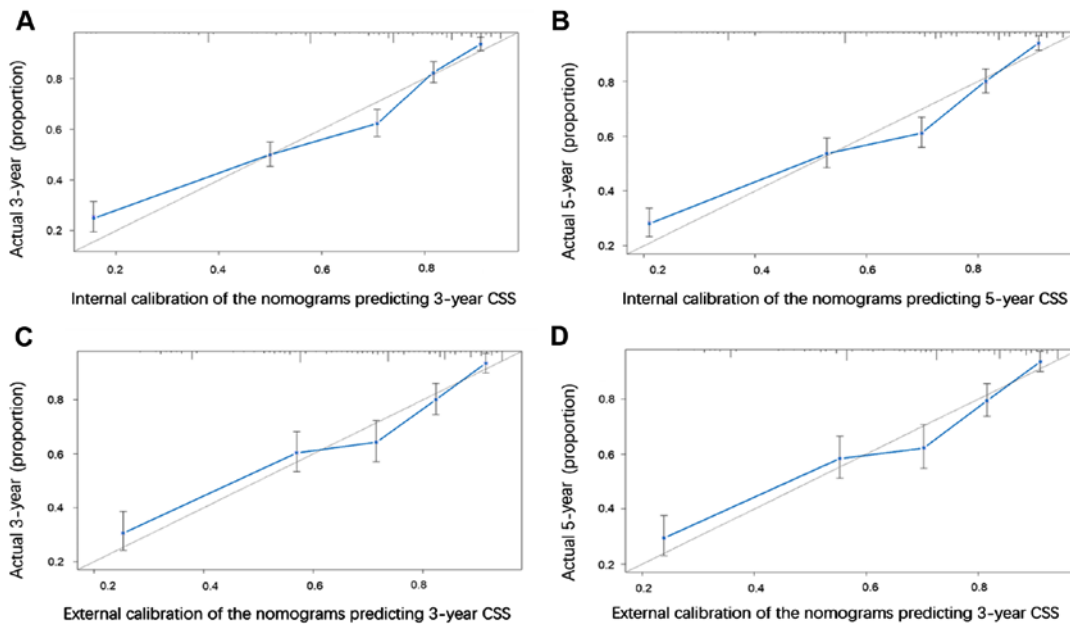


Figure 8. Internal calibration and external calibration of the nomograms predicting 3 and 5-year CSS for patients with BL patients. (A) Internal calibration for 3-year CSS. (B) the Internal calibration for 5-year CSS. (C) External calibration for 3-year CSS. (D) External calibration for 5-year CSS. The x-axis of each plot represents the nomogram prediction and the y-axis represents the observed outcomes. Each cohort equality divided into 5 and the distance between the points was calculated; the diagonal lines on the plots indicate the consistency between predictions and observed outcomes. OS, overall survival; CSS, cancer-specific survival.

3- and 5-year CSS in the internal and external validation were 0.777 (95% CI, 0.757-0.797) and 0.755 (95% CI, 0.728-0.782), respectively. In conclusion, the nomograms constructed in the present study were quite accurate. The internal and external calibration plots for 3- and 5-year OS and CSS showed good consistency between the nomogram predictions and the observed outcomes, in that the calculated points were close to the diagonal line (Figs. 7 and 8).

Discussion

As a highly malignant disease with a demonstrated propensity for dissemination, the incidence rates of BL had gradually increased and reached 4.91 per 1,000,000 (95% CI, 4.12-5.81) in 2009. Based on the relationship between the occurrence of BL and acquired immune deficiency syndrome (AIDS), it is possible that this incidence peak is related to the AIDS

Table VII. Detailed scores of each predictor in the nomograms.

| Characteristics | Overall survival nomogram | Cancer-specific survival nomogram |
|--------------------------------|---------------------------|-----------------------------------|
| Age, group, years | | |
| ≤20 | 0 | 0 |
| 20-40 | 33 | 33 |
| 41-56 | 67 | 65 |
| ≥56 | 100 | 98 |
| Ethnicity | | |
| White | 11 | 7 |
| Black | 8 | 5 |
| American Indian/Alaska native | 5 | 3 |
| Chinese | 3 | 2 |
| Other Asian/Pacific islander | 0 | 0 |
| Primary site | | |
| Lymphonodus | 0 | 0 |
| Digestive system | 2 | 1 |
| Oral cavity and pharynx | 3 | 2 |
| Bone marrow and nervous system | 6 | 3 |
| Other | 5 | 4 |
| Stage | | |
| I | 0 | 0 |
| II | 19 | 21 |
| III | 39 | 42 |
| IV | 58 | 62 |
| Chemotherapy | | |
| Yes | 0 | 0 |
| No/unknown | 90 | 100 |

epidemic (14). The incidence has slowly declined year by year and dropped to 2.80 per 1,000,000 (95% CI, 2.24-3.46) by 2015.

For different diagnostic times, the present study showed that the OS and CSS within 5 years of patients being diagnosed in 2005-2015 were increased compared with patients diagnosed in 1994-2004 and 1983-1993. This is likely to be related to improvements in the therapeutic regimens. In the 1990s, the introduction of dose-intense multi-drug combination chemotherapy and prophylaxis for central nervous system disease led to the high cure rate of BL. It has been reported that complete response and OS reached 67-95 and 54-74%, respectively (15-22). Furthermore, the addition of rituximab resulted in another prognostic improvement in BL, both in children and adults. The 5-year survival rates were reportedly >90% in children (23,24) and the 3-year survival rate was close to 80% in adults (25-27). The 5-year OS rate was improved from 60 to 70% when the Hyper-CVAD regimen with rituximab was compared with the Hyper-CVAD alone. Nevertheless, possibly due to selection bias and therapeutic regimens, the OS rate of the population was decreased.

To estimate the prognosis of diverse patient groups, numerous factors, including age, race, advanced stage, poor

performance status, central nervous system or bone marrow involvement, anemia, the presence of circulating blasts and elevated lactate dehydrogenase (LDH), were proven to be individual risk factors in previous studies (6,28-31). Nevertheless, no nomograms to comprehensively consider prognostic factors and predict the outcomes of patients with BL had yet been developed. Therefore, in the present study, nomograms for predicting the 3- and 5-year OS and CSS of patients with BL were constructed using data obtained from the SEER database. Nomograms for BL may allow clinicians to implement differential therapy for patients in different situations, similar to practices in other malignancy such as osteosarcoma, prostate cancer, gastric cancer, lung cancer and breast cancer. As mentioned above, the cure rate of Burkitt leukemia has increased considerably in the past three decades (1983-2015). In order to improve the predictive level of nomogram, the patient data diagnosed during 2005-2015 was selected for modeling. By analyzing the clinicopathological features with univariate and multivariate analysis, four of these, including age, race, chemotherapy and stage, were eventually regarded as independent prognostic factors for OS and CSS. In particular, race was identified as a prognostic factor and this may be related to genotypic milieu and gene specificity. Lymphoma at different primary sites has different clinical features and prognosis (32,33), even though primary sites is not significant in the Cox multivariate analysis, it was still imported into the nomogram prediction. The nomograms of OS and CSS were constructed and were used to predict prognosis with exact scores corresponding to different survival rates, which was validated both internally and externally via C-indexes and calibration plots. American Indian single patients of older age, whose primary sites were in the bone marrow or nervous system and who had disease of stage IV without chemotherapy, were identified to have the worst survival outcomes.

There are still several limitations that should be considered with respect to the nomograms for BL. Firstly, it is true, as the reviewer suggests, that lab data, especially that of LDH, is an important prognostic factor in lymphoma. It is regrettable that the data obtained from SEER do not cover lab data, so the present study was unable to assess the potential prognostic factors and incorporate these into the prediction models to make the nomograms more accurate. Secondly, the nomograms predicted the outcome of future patients via SEER data collected previously and these predictions will remain constant over time. This may lead to a lack of accuracy of the nomogram as time passes, due to improvements in the detection and treatment of BL. Thirdly, the data obtained from SEER do not cover specific chemotherapy regimens, as explained by an official statement and the classification is not completely accurate in that 'No' and 'Unknown' are bracketed together. The authors of the present study may possibly further investigate the effect of different chemotherapy regimens on prognosis in a single-center study at The First Affiliated Hospital of Guangzhou Medical University (Guangzho, China).

Moreover, the present study constructed and validated the nomograms using data from the same retrospective database. Another non-overlapping dataset or a new study is recommended for stricter validation of the predictive ability of the nomograms. Fourthly, even though multivariate analysis was used to control confounding variables, it is still difficult to completely avoid

influences among correlating variables, such as the relationship between primary site and stage. Finally, though the present study has already begun to collect cases of diagnosed BL; it is unfortunate that the limited number of cases is insufficient to verify the accuracy of model. In a follow-up study, the authors will solve this problem by collecting more virtual data.

The results of the present study demonstrated that the nomogram plots for BL may be a useful tool for clinicians to establish treatment options that vary according to the patient's situation and to identify different categories of patients for scientific research.

Acknowledgements

Not applicable.

Funding

The present study was supported by the National Natural Science Foundation of China (grant nos. 81672661 and 81870113), the Guangdong Province Natural Science Foundation (grant nos. 2016A030313598 and 2106A03313654), and Guangzhou Science and Technology Project (grant nos. 201804010199 and 20161A011062).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

JL and YZ conceived and designed the study. HT and LX analyzed and interpreted the data and JL, SC and BL analyzed data, were involved in drafting the manuscript and revised it critically for important intellectual content. HT, LX and YZ revised the paper. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

- Burkitt D: A sarcoma involving the jaws in African children. *Br J Surg* 46: 218-223, 1958.
- Magrath I: The pathogenesis of Burkitt's lymphoma. *Adv Cancer Res* 55: 133-270, 1990.
- Bellan C, Lazzi S, Hummel M, Palumbo N, de Santi M, Amato T, Nyagol J, Sabattini E, Lazure T, Pileri SA, *et al*: Immunoglobulin gene analysis reveals 2 distinct cells of origin for EBV-positive and EBV-negative Burkitt lymphomas. *Blood* 106: 1031-1036, 2005.
- Dozzo M, Carobolante F, Donisi PM, Scattolin A, Maino E, Sancetta R, Viero P and Bassan R: Burkitt lymphoma in adolescents and young adults: Management challenges. *Adolesc Health Med Ther* 8: 11-29, 2016.
- Costa LJ, Xavier AC, Wahlquist AE and Hill EG: Trends in survival of patients with Burkitt lymphoma/leukemia in the USA: An analysis of 3691 cases. *Blood* 121: 4861-4866, 2013.
- Castillo JJ, Winer ES and Olszewski AJ: Population-based prognostic factors for survival in patients with Burkitt lymphoma: An analysis from the surveillance, epidemiology, and end results database. *Cancer* 119: 3672-3679, 2013.
- Iasonos A, Schrag D, Raj GV and Panageas KS: How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 26: 1364-1370, 2008.
- Kattan MW, Leung DH and Brennan MF: Postoperative nomogram for 12-year sarcoma-specific death. *J Clin Oncol* 20: 791-796, 2002.
- International Bladder Cancer Nomogram Consortium, Bochner BH, Kattan MW and Vora KC: Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. *J Clin Oncol* 24: 3967-3972, 2006.
- Touijer K and Scardino PT: Nomograms for staging, prognosis, and predicting treatment outcomes. *Cancer* 115 (Suppl 13): S3107-S3111, 2009.
- Camp RL, Dolled-Filhart M and Rimm DL: X-tile: A new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 10: 7252-7259, 2004.
- Dickman PW and Adami HO: Interpreting trends in cancer patient survival. *J Intern Med* 260: 103-117, 2006.
- Pencina MJ and D'Agostino RB: Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. *Stat Med* 23: 2109-2123, 2004.
- Hesseling PB, Kouya F, Katayi E, Mbah G and Wharin P: Burkitt's lymphoma: The prevalence of HIV/AIDS and the outcome of treatment. *S Afr Med J* 108: 84-85, 2018.
- Pees HW, Radtke H, Schwamborn J and Graf N: The BFM-protocol for HIV-negative Burkitt's lymphomas and L3 ALL in adult patients: A high chance for cure. *Ann Hematol* 65: 201-205, 1992.
- Soussain C, Patte C, Ostronoff M, Delmer A, Rigal-Huguet F, Cambier N, Leprisé PY, François S, Cony-Makhoul P, Harousseau JL, *et al*: Small noncleaved cell lymphoma and leukemia in adults. A retrospective study of 65 adults treated with the LMB pediatric protocols. *Blood* 85: 664-674, 1995.
- Magrath I, Adde M, Shad A, Venzon D, Seibel N, Gootenberg J, Neely J, Arndt C, Nieder M, Jaffe E, *et al*: Adults and children with small non-cleaved-cell lymphoma have a similar excellent outcome when treated with the same chemotherapy regimen. *J Clin Oncol* 14: 925-934, 1996.
- Rizzieri DA, Johnson JL, Niedzwiecki D, Lee EJ, Vardiman JW, Powell BL, Barcos M, Bloomfield CD, Schiffer CA, Peterson BA, *et al*: Intensive chemotherapy with and without cranial radiation for Burkitt leukemia and lymphoma: Final results of cancer and leukemia group B study 9251. *Cancer* 100: 1438-1448, 2004.
- Lacasse A, Howard O, Lib S, Fisher D, Weng A, Neuberg D and Shipp M: Modified magrath regimens for adults with Burkitt and Burkitt-like lymphomas: Preserved efficacy with decreased toxicity. *Leuk Lymphoma* 45: 761-767, 2004.
- Smeland S, Blystad AK, Kvaløy SO, Ikonomou IM, Delabie J, Kvalheim G, Hammerstrøm J, Lauritzsen GF and Holte H: Treatment of Burkitt's/Burkitt-like lymphoma in adolescents and adults: A 20-year experience from the Norwegian Radium Hospital with the use of three successive regimens. *Ann Oncol* 15: 1072-1078, 2004.
- Di Nicola M, Carlo-Stella C, Mariotti J, Devizzi L, Massimino M, Cabras A, Magni M, Matteucci P, Guidetti A, Gandola L and Gianni AM: High response rate and manageable toxicity with an intensive, short-term chemotherapy programme for Burkitt's lymphoma in adults. *Br J Haematol* 126: 815-820, 2004.
- Tauro S, Cochrane L, Lauritzsen GF, Baker L, Delabie J, Roberts C, Mahendra P and Holte H: Dose-intensified treatment of Burkitt lymphoma and B-cell lymphoma unclassifiable, (with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma) in young adults (<50 years): A comparison of two adapted BFM protocols. *Am J Hematol* 85: 261-263, 2010.
- Patte C, Auperin A, Michon J, Behrendt H, Leverger G, Frappaz D, Lutz P, Coze C, Perel Y, Raphaël M, *et al*: The Société Française d'Oncologie Pédiatrique LMB89 protocol: Highly effective multiagent chemotherapy tailored to the tumor burden and initial response in 561 unselected children with B-cell lymphomas and L3 leukemia. *Blood* 97: 3370-3379, 2001.

24. Gerrard M, Cairo MS, Weston C, Auperin A, Pinkerton R, Lambilliotte A, Sposto R, McCarthy K, Lacombe MJ, Perkins SL, *et al*: Excellent survival following two courses of COPAD chemotherapy in children and adolescents with resected localized B-cell non-Hodgkin's lymphoma: Results of the FAB/LMB 96 international study. *Br J Haematol* 141: 840-847, 2008.
25. Thomas DA, Faderl S, O'Brien S, Bueso-Ramos C, Cortes J, Garcia-Manero G, Giles FJ, Verstovsek S, Wierda WG, Pierce SA, *et al*: Chemoimmunotherapy with hyper-CVAD plus rituximab for the treatment of adult Burkitt and Burkitt-type lymphoma or acute lymphoblastic leukemia. *Cancer* 106: 1569-1580, 2006.
26. Corazzelli G, Frigeri F, Russo F, Frairia C, Arcamone M, Esposito G, De Chiara A, Morelli E, Capobianco G, Becchimanzi C, *et al*: RD-CODOX-M/IVAC with rituximab and intrathecal liposomal cytarabine in adult Burkitt lymphoma and 'unclassifiable' highly aggressive B-cell lymphoma. *Br J Haematol* 156: 234-244, 2012.
27. Todeschini G, Bonifacio M, Tecchio C, Balter R, Carli G, Stefani PM, Adami F, Zamò A, Dei Tos AP, Marino F, *et al*: Intensive short-term chemotherapy regimen induces high remission rate (over 90%) and event-free survival both in children and adult patients with advanced sporadic Burkitt lymphoma/leukemia. *Am J Hematol* 87: 22-25, 2012.
28. Mead GM, Sydes MR, Walewski J, Grigg A, Hatton CS, Pescosta N, Guarnaccia C, Lewis MS, McKendrick J, Stenning SP, *et al*: An international evaluation of CODOX-M and CODOX-M alternating with IVAC in adult Burkitt's lymphoma: Results of United Kingdom Lymphoma Group LY06 study. *Ann Oncol* 13: 1264-1274, 2002.
29. Thomas DA, Cortes J, O'Brien S, Pierce S, Faderl S, Albitar M, Hagemester FB, Cabanillas FF, Murphy S, Keating MJ and Kantarjian H: Hyper-CVAD program in Burkitt's-type adult acute lymphoblastic leukemia. *J Clin Oncol* 17: 2461-2470, 1999.
30. Lopez TM, Hagemester FB, McLaughlin P, Velasquez WS, Swan F, Redman JR, Rodriguez MA, Tucker SL, Silvermintz K, Johnson J, *et al*: Small noncleaved cell lymphoma in adults: Superior results for stages I-III disease. *J Clin Oncol* 8: 615-622, 1990.
31. Diviné M, Lepage E, Brière J, Pautier P, Dupriez B, Lederlin P, Mineur P, Tilly H, Blanc M, Audhuy B, *et al*: Is the small non-cleaved-cell lymphoma histologic subtype a poor prognostic factor in adult patients? A case-controlled analysis. The Groupe d'Etude des Lymphomes de l'Adulte. *J Clin Oncol* 14: 240-248, 1996.
32. Wang G, Chang Y, Wu X, Li X, Li L, Zhang L, Fu X, Sun Z, Zhang X and Zhang M: Clinical features and prognostic factors of primary bone marrow lymphoma. *Cancer Manag Res* 11: 2553-2563, 2019.
33. Korfel A and Schlegel U: Diagnosis and treatment of primary CNS lymphoma. *Nat Rev Neurol* 9: 317-327, 2013.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.