Population-specific prognostic models are needed to stratify outcomes for African-Americans with diffuse large B-cell lymphoma

Qiushi Chen, Georgia Institute of Technology
Turgay Ayer, Georgia Institute of Technology
Loretta J. Nastoupil, University of Texas
Jean Koff, Emory University
Ashley D. Staton, Emory University
Jagpreet Chhatwal, University of Texas
Christopher Flowers, Emory University

Journal Title: Leukemia & Lymphoma
Volume: Volume 57, Number 4
Publisher: Taylor & Francis: STM, Behavioural Science and Public Health Titles | 2015-12-15, Pages 842-851
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.3109/10428194.2015.1083098
Permanent URL: https://pid.emory.edu/ark:/25593/rzsvh

Final published version: http://dx.doi.org/10.3109/10428194.2015.1083098

Copyright information:
© 2015 Taylor & Francis.

Accessed April 22, 2020 8:20 AM EDT
Population-specific prognostic models are needed to stratify outcomes for African-Americans with diffuse large B-cell lymphoma

Qiushi Chen1, Turgay Ayer1, Loretta J. Nastoupil2, Jean L. Koff3, Ashley D. Staton3, Jagpreet Chhatwal4, and Christopher R. Flowers3

1 H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA
2 Department of Lymphoma/Myeloma, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
3 Department of Hematology/Oncology, Winship Cancer Institute, Emory University, Atlanta, GA, USA
4 Department of Health Services Research, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

BACKGROUND

Diffuse large B-cell lymphoma (DLBCL) is the most common type of lymphoma in the United States (US), affecting >20,000 people/year and accounting for nearly one-third of adult non-Hodgkin lymphoma (NHL) [1]. From the perspective of cancer disparities and outcomes research, DLBCL is a disease of considerable clinical and public health interest, because it is often curable with standard therapy but is universally fatal if untreated or improperly treated. Untreated DLBCL patients have an expected survival of <1 year [1], whereas standard modern chemo-immunotherapy (i.e., rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone [R-CHOP]) produces high 5-year overall survival (OS), with a cure rate nearing 60% [2-6]. Despite these advances, patients with DLBCL experience disparate outcomes based not only on clinical prognostic factors, but also race, biological factors, and insurance status [7-16].

A landmark study of the general population of DLBCL patients identified five adverse prognostic factors (stage III/IV disease, elevated lactate dehydrogenase (LDH), age >60 years, ECOG performance status ≥2, and involvement of >1 extranodal site) and utilized these to construct the international prognostic index (IPI) score for DLBCL [9]. A revised formulation of the IPI (R-IPI) was developed in the era when R-CHOP was the most commonly used frontline therapy and demonstrated that these same factors stratified DLBCL patients into three distinct prognostic groups [17]. However, nearly all patients included in
the development and validation of the IPI model and the construction of the R-IPI were of European ancestry. Previous large population-based studies have demonstrated racial disparities in clinical presentation and outcomes for patients with DLBCL in the US. For instance, African American (AA) patients with DLBCL are diagnosed on average a decade younger than whites, are more likely to have advanced stage disease, and are less likely to reach the milestone of 5-year survival (38% for AA vs 46% for white) [12,18]. As a result, it remains unclear whether the IPI and R-IPI accurately stratify risk and predict OS for this patient population.

The aim of this study was to examine disparities in survival risk stratification and prognostication when models developed for the general population are applied to an AA population with DLBCL. Using the Surveillance, Epidemiology, and End Results (SEER) dataset, we assessed risk stratification by the IPI in the general and AA SEER DLBCL populations, and compared clinical prognostic models that were developed for general and AA DLBCL patients separately.

MATERIALS AND METHODS

We first examined risk stratification for AA DLBCL patients in SEER using IPI scores [9]. Next, we built separate prognostic models to predict 5-year survival for general and AA populations, respectively, to examine whether a general-population model provides adequate survival prediction for AA DLBCL patients, and whether an AA population-specific model could improve survival prediction for AA patients.

Data Source

We selected cases in SEER that were identified from 2002 to 2012 in all 13 registries. The SEER program has collected data on cancer cases since 1973, and includes 13 population-based registries that account for approximately 14% of the US population. We used the third edition of the International Classification of Diseases for Oncology (ICD-O-3) to identify DLBCL cases in SEER, including codes 9680 (DLBCL), 9679 (primary mediastinal large B-cell lymphoma), 9684 (immunoblastic large B-cell lymphoma), and 9678 (primary effusion lymphoma), in line with the case identification approach in prior analyses [12]. We restricted to the data after 2002, which represents the rituximab era, the standard of care of DLBCL patients transitioned to the use of first-line immunochemotherapy (R-CHOP) after that time point and improved observed outcomes [18-20]. All cases had known age at diagnosis ≥18 and race coded as white or black (used to identify AA patients for this study) in SEER. The major categories of the race attribute recorded in SEER include white, black, American Indian/Alaskan Native, and Asian or Pacific Islander [21,22]. Hispanic ethnicity was not considered as it was not a mutually exclusive race category in the SEER and was not reliably recorded [12,21]. Cases with unknown age or other/unknown race were excluded.

For each patient, we extracted data on survival months, vital status, follow-up status, and baseline clinical variables including age at diagnosis, sex, Ann Arbor stage, and presence of B symptoms. IPI scores were only available (i.e., collected as one of the collaborative stage site-specific factors in SEER) in some patients diagnosed after 2004.
In our risk stratification analyses, only patients with IPI data since 2004 were included. For developing and comparing clinical prognostic models, we did not include the IPI score such that we could utilize a larger SEER cohort (diagnosed from 2002-2012) for model training and testing based on other clinical variables. Alive cases with follow-up time less than 5 years were excluded in the analyses of 5-year survival prediction. Figure 1 displays the allocation of patients for model training and evaluation.

**Risk Stratification Analysis for AA DLBCL Population**

We selected DLBCL cases with valid IPI scores from 2004 to 2010 in SEER for stratification analysis. Kaplan-Meier OS curves for general and AA population were stratified by IPI categories (0-1, 2, 3, and 4-5 for low, intermediate-low, intermediate-high, and high risk groups, respectively). We applied the log rank test \( \chi^2 \) to evaluate risk stratification in each population. To compare the log rank test statistic \( \chi^2 \) between general and AA populations of comparable sizes, we sampled the same number of cases of the general DLBCL population as the AA population, repeated the sampling 100 times, and then calculated the average \( \chi^2 \) value for the general population. A higher \( \chi^2 \) value indicates a better separation of survival curves between each risk group given the same number of risk groups.

**Comparing Prognostic Models**

We developed prognostic models based on two different statistical learning methods, namely logistic regression (LR) and artificial neural network (ANN) [12,24]. LR has been one of the most commonly used predictive models in medicine and has intuitive interpretation in its model structure [25]. Compared with LR, ANN has a more flexible structure and is potentially able to detect more complex relationships and implicit interactions across input variables [26]. Both methods have been applied successfully in predicting and estimating clinical outcomes in various diseases, such as breast cancer, prostate cancer, and coronary heart disease [27-30].

In the prognostic models using either statistical method, input variables \( X_i \)'s included age (based on quartiles in the race-specific population), sex, Ann Arbor stage, and presence of B symptoms (Table I). The output of these prognostic models was the predicted probability \( p \) of survival at 5 years. The 5-year landmark was selected for these analyses because DLBCL patients without recurrence at 5 years can be considered cured. [1-3,6,7,13]. The two statistical methods explore different forms of the relations between model inputs and output. In particular, LR examines the linear relationship between input variables and the log-odds of the event presence probability \( p \), i.e.,

\[
\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n.
\]

A typical ANN consists of three layers: input nodes in the input layer represent each input variable \( X_i \) respectively; the single output node in the output layer represents the outcome probability \( p \), and the hidden layer (with hidden nodes) connects input and output layers, which contain the intermediate values of the network, but these values do not have physical
meaning or explicit interpretation (Figure 2). In our analysis, we used a feed-forward network structure (the most commonly utilized structure) [31] and tested ANNs with different number of hidden nodes (e.g., ANN with 5 hidden nodes was denoted as ANN(5)).

We compared the prognostic models using different training and testing populations. Our analysis aimed to exploit two questions. First, we developed a prognostic model for the general DLBCL population (i.e., trained by the general population data), denoted as GM, and hypothesized that it would performed better when tested on general than AA DLBCL population. Next, we examined that whether an AA-population specific prognostic model (AM) would outperform a general population model (GM2) when tested on the same AA population data.

To compare the performance of prognostic models on independent datasets, we used a modified 10-fold cross-validation approach to accommodate the unbalanced sizes of the general and AA population datasets. Similar to the standard 10-fold cross-validation [12], we first divided the entire SEER DLBCL dataset into 10 folds with approximately equal size. We then utilized 9 folds (i.e., 90% of all SEER DLBCL data) for model training. In the 9 folds, all data (i.e., general population) were used for training GM, all AA data for training AM, and an undersampled general population data with the same size of AA data were used for training GM2 because we need to maintain comparable sizes of training datasets for GM2 and AA for a fair comparison. The remaining one fold was used for model testing. GM was tested on two different populations: (1) AA DLBCL cases from the remaining fold, and (2) a sampled subset of general population data with the same size of AA population from the remaining fold; GM2 and AM were both evaluated on all AA population data in the remaining fold.

We iterated the above process until each fold was used once for testing. In this way, all models could be trained and tested on independent sets, and in the meanwhile general and AA DLBCL populations for both training and testing maintained the same size for fair comparisons. We then combined model predicted results for the testing set from each iteration. Finally, we used these combined results to evaluate the overall performance of each survival prognostic model.

The primary performance measure of prognostic models is model calibration, which was assessed using the Hosmer-Lemeshow (H-L) goodness-of-fit test [25]. The H-L test for survival prognostic models assesses whether or not the observed number of patients alive at 5 years matches the expected number in subgroups of the model population. If the H-L p-value is <0.01, the model is poorly calibrated, implying that a different model is needed to adequately predict survival in the given population. We also generated calibration curve plots, in which the 45-degree line represents the perfect calibration; the points to the left and right represent underestimations and overestimations of risks, respectively. We also assessed model discrimination by receiver-operator characteristics (ROC) curves [32]. In particular, we calculated the area under the curve (AUC, also known as the c-statistic), and used the 2-tailed DeLong method [33] to compare the AUC of different models.
RESULTS

Patient Characteristics and Outcome

From 2002 to 2012, 31,490 cases of DLBCL were diagnosed in the 13 SEER registries. After excluding cases with age unknown or <18 years, a population of 27,618 cases remained, with 25,447 white and 2,171 AA patients for the survival prediction analyses. For risk stratification analysis, we identified 1,820 white and 127 AA patients with DLBCL with valid IPI scores recorded in SEER (Figure 1). Table II summarizes the clinical characteristics of the study population by race. As noted in prior studies [12,18], we found that AA patients exhibited younger age at diagnosis than white patients (55 vs. 68 years; p<0.001) and more AA patients presented with advanced (III/IV) stage disease (55.3% vs. 48.3%; p<0.001).

Results of Risk Stratification Models

Figure 3 presents the Kaplan-Meier survival curves stratified by IPI risk categories for general and AA populations, respectively. To evaluate the survival stratification and compare between populations, we performed log-rank tests on the sampled general population (with the same size of the AA population) and obtained the average log-rank statistics from multiple samples, which was higher than the statistics for the AA population ($\chi^2 = 19.92$ for general vs. 8.09 for AA population, df = 3), indicating better risk stratification in the general population than in the AA population.

Results of LR and ANN Prognostic Models

In the model based on all AA DLBCL patients in SEER (N=1514) using multivariable LR, four factors significantly predicted worse 5-year OS in AA patients: age greater than the median in the AA population (>55 years; odds ratio [OR] 0.45, 95% confidence interval [CI] 0.36-0.56), male sex (OR 0.75, CI 0.60-0.93), and stage III/IV disease (OR 0.43, CI 0.34-0.54).

Next, we compared the performance of general population prognostic models on general and AA test populations. The total general population consisting of 17,583 white and 1,514 AA DLBCL patients was used to construct training and testing sets for the prognostic models following the modified 10-fold cross validation procedure. Each model was well fitted to its own training data (see the calibration plots in Appendix). We evaluated GM models’ performance on the combined testing sets for general (white = 1393, and AA = 121) and AA populations (AA=1514), respectively. GM models demonstrated good calibration for the general DLBCL population, but not for AA patients with DLBCL (p<0.001, Table III), irrespective of the model development approach used. For example, in the calibration plots for the GM-LR model (Figure 4), the calibration curve for the general test population closely approximated the perfect calibration line (i.e., the 45-degree line) with small deviations (H-L statistics 5.684, 8 df, p=0.683); whereas the curve for the AA test population showed significantly worse calibration (H-L statistics 73.279, 8 df, p<0.001). GM models also showed higher AUC, implying better model discrimination for the general population compared with that for the AA population (0.736 vs. 0.679, p=0.003 in GM-LR model; 0.740 vs. 0.684, p=0.003 in GM-ANN(10) model; also see ROC curves in Figure 4).

Leuk Lymphoma. Author manuscript; available in PMC 2017 April 01.
Additionally, we compared the performance of a general risk model (GM2) and an AA-specific risk model (AM) on the same testing dataset (N=1514) of AA patients with DLBCL (Table IV). GM2 using LR (GM2-LR) and ANN with two hidden nodes (GM2-ANN(2)) showed poor calibration (H-L statistics>89, 8 df, p<0.001) for the AA DLBCL population, whereas AM-LR and AM-ANN(2) were better calibrated for this population (H-L statistics<19, 8 df, p>0.015; see Figure 5). We used the ANN with two hidden nodes because the ANN tended to overfit small training data as the number of hidden nodes increased (i.e., worse risk calibration in testing data of AA population by ANN with increasing number of hidden nodes). In fact, the ANN with two hidden nodes had sufficient model complexity (at least more complex than logistic regression) to capture the underlying relations between input and output variables, as it had nearly perfect model fitting on the training dataset (see calibration plots in Appendix 1). However, AM models did not demonstrate superior discrimination ability with higher AUCs than GM2 for the AA population in the SEER dataset.

**DISCUSSION**

Multiple studies have identified differences in baseline characteristic and inferior OS in AA DLBCL patients [12,14], but this disparity has not yet translated into a race-specific prognostic model. In this study, we found that the most commonly used prognostic models to date, the IPI and R-IPI. Our survival prediction analysis also showed that a prognostic model trained on the general population had poor calibration for AA patients with DLBCL. A population-specific prognostic model provided better survival estimations in the AA DLBCL population.

Racial disparities have been recognized in other cancers [34-39]. For example, several studies have found that AA women with breast cancer have significantly different incidence of disease and mortality compared to their white counterparts [40-43]. The Gail model, used to predict the risk of developing breast cancer, was initially developed using white patient data [44]; it was later adapted successfully to account for racial variations, and now accurately predicts breast cancer risk for a broader population [45]. We believe comparable adjustments can be made to the IPI to better predict survival for patients with DLBCL.

Racial disparities in presentation and survival of lymphoma had not been thoroughly evaluated until recently [7,12,16,18,46-50]. Analyses from two national cancer datasets, the SEER registry and the National Cancer Data Base (NCDB), have shown that AA patients with DLBCL in the US were diagnosed on average >10 years younger than their white counterparts, were more likely to have stage III/IV disease, and had worse 5-year OS survival [10,12,18]. Moreover, in a clinic-based cohort study that showed inferior survival for AA patients given identical treatment regimens to white patients treated in the same setting, the IPI did not adequately categorize expected outcomes for AA patients with DLBCL [13]. While the relationships between race and survival in DLBCL are complex [7], racial differences in age at presentation and distribution of more aggressive DLBCL subtypes could play a role in these observed differences in outcome [50,51], perhaps analogous to those observed for triple-negative breast cancer in AA woman under the age of 50 [52]. The logistic regression model for AA population in our analysis showed that age
above the median of 55 year-old, earlier than general population, remained a significant adverse prognostic factor. Our analyses provide early insights regarding the age adjustment in AA population-specific prognostic model. Racial differences in optimal age cut-off values can be explored in a prospective analysis using more comprehensive clinical data of large cohort.

Gene expression profiling studies have identified two major “cell-of-origin” subtypes of DLBCL, germinal center B-cell-like (GCB) and activated B-cell (ABC)-like. Importantly, these subtypes are associated with significant differences in survival in patients treated with R-CHOP \[8,53-56\] (3-year OS: 87% for GCB vs. 44% for ABC [53]; HR of ABC: 1.80 (1.36-2.38) for PFS and 1.85 (1.46-2.35) for OS) [57]. Preliminary data based on the clinic-based cohort study described above showed significantly higher rate of ABC subtype in AA patients, suggesting racial differences in the prevalence of ABC DLBCL [50,51]. These findings may partially explain disparities in clinical outcome, suggesting that biological factors should be incorporated to further improve risk prediction for individual patients with DLBCL.

Our analysis had several limitations. Our analysis had several limitations. First, in the survival stratification analysis, we had limited sample size of the AA population with complete IPI data, which is insufficient to definitively disprove the use of IPI model for the AA population or to derive a new prognostic model for the AA population ready for practical use. However, the SEER dataset provides the largest available cohort to examine the racial disparities. We restricted use of these data to compare the stratification of IPI model between the general and AA populations from the SEER data. These analyses highlight the need for the development on a cohort study to address this issue. A new AA-specific prognostic model needs to be developed and extensively validated using clinical data (ideally prospectively collected) with an adequate size of AA population. A large prospective cohort study of patients with NHL, the Lymphoma Epidemiology of Outcomes (LEO), is now underway and will be enriched for AA and Hispanic patients to address these issues.

Second, several clinical variables (e.g., LDH and extranodal involvement) were not included for developing and comparing prognostic models due to data limitation in SEER. Although our model structure may not represent the optimal configuration of model parameters, we performed the analysis using available clinical variables only to evaluate the concept as to whether a population-specific model might provide a benefit, and the AA population-specific logistic regression model represents our best attempt to identify the factors and their effects on predicting the survival. These provide the motivation for thorough examination of population-specific prognostic models based on comprehensive clinical data, which has not been attempted before. Future research in this area should encourage the incorporation of lymphoma-specific prognostic variables like LDH into population-based cancer registries, and the construction of large prospective lymphoma cohort studies to address the current deficiencies of available data resources.

These findings indicate that race is a demographic factor that impacts prognostic scoring systems. The same may be true for other demographic factors. Studies have identified that
men and women treated with rituximab-based therapies may have different outcomes [58,59]. It has also been shown that elderly patients (age greater than 70) with DLBCL are not well stratified using the IPI system [20]. Large cohorts of DLBCL patients assembled through epidemiologic studies or clinical trials are needed to assess the value of the IPI and novel prognostic scoring systems that permit inclusion of these and other demographic factors in prognostic models.

Given the prognostic values of various demographic, clinical, socioeconomic, biological, and treatment factors, a comprehensive prognostic model is needed to improve survival estimates for patients. Standard statistic methods such as survival analysis and machine learning require a large single dataset consisting of all risk factors over a wide spectrum, which may not be obtainable. Novel modeling approaches, such as the multi-state Markov model [60] and simulation calibration [61,62], may represent promising alternatives to integrate clinical outcomes from multiple data sources and emerging evidences, and thus improve risk stratification and prognostic models for DLBCL patients.

REFERENCE


Leuk Lymphoma. Author manuscript; available in PMC 2017 April 01.


Figure 1.
Selection of study cohort.

Identified patients with diffuse large B-cell lymphoma diagnosed between 1992-2010 in SEER 13 registries (White = 40,828; AA = 3,331)

Excluded patients
- Unknown age at diagnosis: White = 0, AA = 0
- Age at diagnosis < 18 years: White = 311, AA = 54

Excluded patients with unknown age or <18 years (White = 40,517; AA = 3,277)

Excluded patients
- Unknown IPI scores:
  White = 38,864; AA = 3,138
- Inconsistent IPI scores:
  White = 187; AA = 30

Excluded patients
- Alive but followed < 5 years:
  White = 7,600, AA = 635

Survival stratification model
White = 1,466; black = 109

5-year survival prediction model
White = 32,917, AA = 2,642

Leuk Lymphoma. Author manuscript; available in PMC 2017 April 01.
Figure 2.
Structure of artificial neural network model.
Figure 3.
Kaplan-Meier curves for patients with IPI and revised IPI scores in SEER diagnosed 2004-2010. A, overall survival (OS) for general population stratified by IPI scores with 4 categories; B, OS for African American (AA) population stratified by IPI scores.
Figure 4.
Performance of GM-LR prognostic model for general and African American (AA) populations: A, risk calibration for general population; B, risk calibration for AA population; C, receiver operating characteristic (ROC) curve for general and AA populations.
Figure 5.
Calibration plots of GM2 and AM for African American population: A, GM2 logistic regression (GM2-LR) model; B, AM-LR model; C, GM2 ANN model with 2 hidden nodes (GM2-ANN(2)); D, AM-ANN(2) model.
Figure 6.
Calibration plots of GM2 and AM for AA populations. A, GM2 logistic regression model; B, AM logistic regression model; C, GM2 ANN model with 3 hidden nodes; AM ANN model with 3 hidden nodes.
Figure 7.
ROC curve of GM2 and AM for AA populations. A, Logistic regression model; B, ANN(3) with 3 hidden nodes.
Table I
Input factors for risk stratification and survival prognostic models.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk stratification model</td>
</tr>
<tr>
<td></td>
<td>AA population</td>
</tr>
<tr>
<td><strong>Input</strong></td>
<td></td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>Age as a continuous variable</td>
</tr>
<tr>
<td>Sex</td>
<td>Male, female</td>
</tr>
<tr>
<td>Stage</td>
<td>Stage I/II, III/IV, unknown</td>
</tr>
<tr>
<td>B symptoms</td>
<td>Present, absent, unknown</td>
</tr>
<tr>
<td>Adjusted IPI scores</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>Survival time</td>
</tr>
</tbody>
</table>

AA, African American; IPI, international prognostic index.

†Quartiles of age distribution for general population in SEER;
‡Quartiles of age distribution of AA population in SEER.
Table II

Patient characteristics of different study cohorts.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total SEER DLBCL population</th>
<th>Risk stratification analysis</th>
<th>Compare prognostic models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>AA</td>
<td>White</td>
</tr>
<tr>
<td>N</td>
<td>25,447</td>
<td>2,171</td>
<td>1,820</td>
</tr>
<tr>
<td>Age (year)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>68 (54-78)</td>
<td>55 (44-68)</td>
<td>65 (53-77)</td>
</tr>
<tr>
<td>≥60</td>
<td>17,390</td>
<td>68.34%</td>
<td>879</td>
</tr>
<tr>
<td>Male</td>
<td>14,007</td>
<td>55.04%</td>
<td>1,248</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/II</td>
<td>11,953</td>
<td>46.97%</td>
<td>894</td>
</tr>
<tr>
<td>III/IV</td>
<td>12,284</td>
<td>48.27%</td>
<td>1,200</td>
</tr>
<tr>
<td>Unknown</td>
<td>1,210</td>
<td>4.75%</td>
<td>77</td>
</tr>
<tr>
<td>B symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>12,567</td>
<td>49.38%</td>
<td>992</td>
</tr>
<tr>
<td>Yes</td>
<td>6,215</td>
<td>24.42%</td>
<td>720</td>
</tr>
<tr>
<td>Unknown</td>
<td>6,665</td>
<td>26.19%</td>
<td>459</td>
</tr>
<tr>
<td>Adjusted-IPI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>527</td>
<td>2.07%</td>
<td>36</td>
</tr>
<tr>
<td>1</td>
<td>669</td>
<td>2.63%</td>
<td>41</td>
</tr>
<tr>
<td>2</td>
<td>449</td>
<td>1.76%</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Total SEER DLBCL population</td>
<td>Risk stratification analysis</td>
<td>Compare prognostic models</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>AA</td>
<td>White</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>175</td>
<td>0.69%</td>
</tr>
<tr>
<td>Unknown IPI</td>
<td>23,429</td>
<td>2,013</td>
<td>92.07%</td>
</tr>
</tbody>
</table>
### Table III
Performance of survival prognostic model developed using the general population (GM) and tested in the general and African American populations

<table>
<thead>
<tr>
<th>Model</th>
<th>Testing population</th>
<th>H-L test</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$X^2$</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>statistics</td>
<td></td>
</tr>
<tr>
<td>GM-LR</td>
<td>General</td>
<td>5.684</td>
<td>0.683</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>73.279</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GM-ANN(3)†</td>
<td>GN</td>
<td>9.109</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>78.195</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GM-ANN(10)</td>
<td>General</td>
<td>11.509</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>76.506</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GM-ANN(15)</td>
<td>General</td>
<td>6.656</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>75.878</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

GM, general model; LR, logistic regression; ANN, artificial neural network; AA, African-American; AUC, area under curve. † The number in the parenthesis represents the number of hidden nodes in the ANN.
### Table IV

Comparisons of the risk calibration for prognostic models developed using data from the general (GM2) and African American (AM) populations and tested in a separate African American population

<table>
<thead>
<tr>
<th>Model</th>
<th>Training set</th>
<th>H-L test $\chi^2$ statistics</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM2-LR</td>
<td>GN</td>
<td>98.272</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AM-LR</td>
<td>AA</td>
<td>18.664</td>
<td>0.017</td>
</tr>
<tr>
<td>GM2-ANN(2)</td>
<td>GN</td>
<td>89.646</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AM-ANN(2)</td>
<td>AA</td>
<td>12.355</td>
<td>0.136</td>
</tr>
<tr>
<td>GM2-ANN(3)</td>
<td>GN</td>
<td>119.288</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AM-ANN(3)</td>
<td>AA</td>
<td>24.317</td>
<td>0.002</td>
</tr>
<tr>
<td>GM2-ANN(10)</td>
<td>GN</td>
<td>118.605</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AM-ANN(10)</td>
<td>AA</td>
<td>57.550</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

GM, general model; LR, logistic regression; ANN, artificial neural network; AA, African-American, area under curve.