Published in final edited form as:

Leuk Lymphoma. 2016 ; 57(4): 842–851. doi:10.3109/10428194.2015.1083098.

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 firstline 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 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 [23] to evaluate risk stratification in each population. To compare the log rank test statistic χ^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 χ^2 value for the general population. A higher χ^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).

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

1. Flowers CR, Sinha R, Vose JM. Improving outcomes for patients with diffuse large B-cell lymphoma. *CA Cancer J Clin.* 2010; 60:393–408. [PubMed: 21030533]
2. Coiffier B, Thieblemont C, Van Den Neste E, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood.* 2010; 116:2040–2045. [PubMed: 20548096]
3. Feugier P, Van Hoof A, Sebban C, et al. Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse large B-cell lymphoma: a study by the Groupe d'Etude des Lymphomes de l'Adulte. *J Clin Oncol.* 2005; 23:4117–4126. [PubMed: 15867204]
4. Pfreundschuh M, Kuhnt E, Trumper L, et al. CHOP-like chemotherapy with or without rituximab in young patients with good-prognosis diffuse large-B-cell lymphoma: 6-year results of an open-label randomised study of the MabThera International Trial (MInT) Group. *Lancet Oncol.* 2011; 12:1013–1022. [PubMed: 21940214]
5. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007; 25:579–586. [PubMed: 17242396]
6. Habermann TM, Weller EA, Morrison VA, et al. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. *J Clin Oncol.* 2006; 24:3121–3127. [PubMed: 16754935]
7. Flowers CR, Nastoupil LJ. Socioeconomic disparities in lymphoma. *Blood.* 2014; 123:3530–3531. [PubMed: 24904097]
8. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 2000; 403:503–511. [PubMed: 10676951]
9. Shipp M. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993; 329:987–994. [PubMed: 8141877]
10. Han X, Jemal A, Flowers CR, Sineshaw H, Nastoupil LJ, Ward E. Insurance status is related to diffuse large B-cell lymphoma survival. *Cancer.* 2014; 120:1220–1227. [PubMed: 24474436]
11. Shenoy P, Maggioncalda A, Malik N, Flowers CR. Incidence patterns and outcomes for hodgkin lymphoma patients in the United States. *Adv Hematol.* 2011; 2011:725219. [PubMed: 21197477]
12. Shenoy PJ, Malik N, Nooka A, et al. Racial differences in the presentation and outcomes of diffuse large B-cell lymphoma in the United States. *Cancer.* 2011; 117:2530–2540. [PubMed: 24048801]
13. Flowers CR, Shenoy PJ, Borate U, et al. Examining racial differences in diffuse large B-cell lymphoma presentation and survival. *Leuk Lymphoma.* 2013; 54:268–276. [PubMed: 22800091]

14. Ghafoor A, Jemal A, Cokkinides V, et al. Cancer statistics for African Americans. CA: A Cancer Journal for Clinicians. 2002; 52:326–341. [PubMed: 12469762]
15. Wang M, Burau KD, Fang S, Wang H, Du XL. Ethnic variations in diagnosis, treatment, socioeconomic status, and survival in a large population-based cohort of elderly patients with non-Hodgkin lymphoma. Cancer. 2008; 113:3231–3241. [PubMed: 18937267]
16. Flowers CR, Glover R, Lonial S, Brawley OW. Racial Differences in the Incidence and Outcomes for Patients with Hematological Malignancies. Current Problems in Cancer. 2007; 31:182–201. [PubMed: 17543947]
17. Sehn LH, Berry B, Chhanabhai M, et al. 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. 2007; 109:1857–1861. [PubMed: 17105812]
18. Flowers CR, Fedewa SA, Chen AY, et al. Disparities in the early adoption of chemoimmunotherapy for diffuse large B-cell lymphoma in the United States. Cancer Epidemiology Biomarkers & Prevention. 2012; 21:1520–1530.
19. Coiffier B, Lepage E, Briere J, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. N Engl J Med. 2002; 346:235–242. [PubMed: 11807147]
20. Williams JN, Rai A, Lipscomb J, Koff JL, Nastoupil LJ, Flowers CR. Disease characteristics, patterns of care, and survival in very elderly patients with diffuse large B - cell lymphoma. Cancer. 2015
21. Race Recode Changes. May http://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/. Accessed 2015 May
22. [Internet]. 2015 - [cited Date Cited Year Cited] Available from: URL
23. Klein, JP.; Moeschberger, ML. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media; 2003.
24. Bishop, CM. Pattern recognition and machine learning. springer; New York: 2006.
25. Hosmer, DW., Jr; Lemeshow, S. Applied logistic regression. John Wiley & Sons; 2004.
26. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology. 1996; 49:1225–1231. [PubMed: 8892489]
27. Eftekhari B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Medical Informatics and Decision Making. 2005; 5:3. [PubMed: 15713231]
28. Ayer T, Chhatwal J, Alagoz O, Kahn CE, Woods RW, Burnside ES. Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. Radiographics. 2010; 30:13–22. [PubMed: 19901087]
29. Shi H-Y, Lee K-T, Lee H-H, et al. Comparison of artificial neural network and logistic regression models for predicting in-hospital mortality after primary liver cancer surgery. PloS one. 2012; 7:e35781. [PubMed: 22563399]
30. Ayer T, Alagoz O, Chhatwal J, Shavlik JW, Kahn CE Jr. Burnside ES. Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration. Cancer. 2010; 116:3310–3321. [PubMed: 20564067]
31. Kröse B, Krose B, van der Smagt P, Smagt P. An introduction to neural networks. 1993
32. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837–845. [PubMed: 3203132]
34. Saba NF, Goodman M, Ward K, et al. Gender and ethnic disparities in incidence and survival of squamous cell carcinoma of the oral tongue, base of tongue, and tonsils: a surveillance, epidemiology and end results program-based analysis. Oncology. 2011; 81:12–20. [PubMed: 21912193]
35. Berry J, Bumpers K, Ogunlade V, et al. Examining Racial Disparities in Colorectal Cancer Care. Journal of Psychosocial Oncology. 2009; 27:59–83. [PubMed: 19197679]

36. Berry J, Caplan L, Davis S, et al. A black-white comparison of the quality of stage-specific colon cancer treatment. *Cancer*. 2010; 116:713–722. [PubMed: 19950126]
37. Morris AM, Billingsley KG, Hayanga AJ, Matthews B, Baldwin LM, Birkmeyer JD. Residual treatment disparities after oncology referral for rectal cancer. *J Natl Cancer Inst*. 2008; 100:738–744. [PubMed: 18477800]
38. Thornton JG, Morris AM, Thornton JD, Flowers CR, McCashland TM. Racial variation in colorectal polyp and tumor location. *J Natl Med Assoc*. 2007; 99:723–728. [PubMed: 17668638]
39. Newman LA. Breast cancer in African-American women. *The Oncologist*. 2005; 10:1–14.
40. Ries, L.; Eisner, M.; Kosary, C., et al. SEER Cancer Statistics Review, 1973-1999, National Cancer Institute. Bethesda, MD: Table VI-1. Available from URL: http://seer.cancer.gov/csr/1973_1999 2002
41. Mayberry RM, Stoddard-Wright C. Breast cancer risk factors among black women and white women: similarities and differences. *American journal of epidemiology*. 1992; 136:1445–1456. [PubMed: 1288274]
42. Mayberry RM. Age-specific patterns of association between breast cancer and risk factors in black women, ages 20 to 39 and 40 to 54. *Annals of epidemiology*. 1994; 4:205–213. [PubMed: 8055121]
43. Bernstein L, Teal CR, Joslyn S, Wilson J. Ethnicity-related variation in breast cancer risk factors. *Cancer*. 2003; 97:222–229. [PubMed: 12491485]
44. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*. 1989; 81:1879–1886. [PubMed: 2593165]
45. Newman L. Proposed revision of the Gail breast cancer risk assessment model for African American women. 2003:52.
46. Shenoy PJ, Malik N, Sinha R, et al. Racial differences in the presentation and outcomes of chronic lymphocytic leukemia and variants in the United States. *Clinical Lymphoma Myeloma and Leukemia*. 2011; 11:498–506.
47. Abouyabis AN, Shenoy PJ, Lechowicz MJ, Flowers CR. Incidence and outcomes of the peripheral T-cell lymphoma subtypes in the United States. *Leuk Lymphoma*. 2008; 49:2099–2107. [PubMed: 19021052]
48. Muringampurath-John D, Flowers CR, Toscano M, et al. Rituximab-hyperfractionated cyclophosphamide, vincristine, adriamycin and dexamethasone alternating with high-dose cytarabine and methotrexate for aggressive non-Hodgkin lymphoma. *Leuk Lymphoma*. 2012; 53:725–727. [PubMed: 21888615]
49. Nabhan C, Byrtek M, Taylor MD, et al. Racial differences in presentation and management of follicular non-Hodgkin lymphoma in the United States: Report from the National LymphoCare Study. *Cancer*. 2012
50. Flowers CR, Nastoupil L, Borate U, et al. Racial Disparities in Cell of Origin Among DLBCL Patients. *Blood*. 2012:120.
51. Chastain EC, Fisher KE, Bumpers K, et al. Racial Differences in Prognostic Biomarkers of Diffuse Large B-Cell Lymphoma. *United States and Canadian Academy of Pathology*. 2012 Abstract#1377.
52. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype. *Cancer*. 2007; 109:1721–1728. [PubMed: 17387718]
53. Choi WW, Weisenburger DD, Greiner TC, et al. A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin Cancer Res*. 2009; 15:5494–5502. [PubMed: 19706817]
54. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004; 103:275–282. [PubMed: 14504078]
55. Lenz G, Wright G, Dave S, et al. Gene expression signatures predict survival in diffuse large B cell lymphoma following rituximab and CHOP-like chemotherapy. *Annals of Oncology*. 2008; 19:93–93.

56. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002; 346:1937–1947. [PubMed: 12075054]
57. Read JA, Koff JL, Nastoupil LJ, Williams JN, Cohen JB, Flowers CR. Evaluating Cell-of-Origin Subtype Methods for Predicting Diffuse Large B-Cell Lymphoma Survival: A Meta-Analysis of Gene Expression Profiling and Immunohistochemistry Algorithms. *Clin Lymphoma Myeloma Leuk*. 2014
58. Müller C, Murawski N, Wiesen MH, et al. The role of sex and weight on rituximab clearance and serum elimination half-life in elderly patients with DLBCL. *Blood*. 2012; 119:3276–3284. [PubMed: 22337718]
59. Pfreundschuh M, Müller C, Zeynalova S, et al. Suboptimal dosing of rituximab in male and female patients with DLBCL. *Blood*. 2014; 123:640–646. [PubMed: 24297867]
60. Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*. 2007; 26:2389–2430. [PubMed: 17031868]
61. Kong CY, McMahon PM, Gazelle GS. Calibration of disease simulation model using an engineering approach. *Value in Health*. 2009; 12:521–529. [PubMed: 19900254]
62. Clarke LD, Plevritis SK, Boer R, Cronin KA, Feuer EJ. A comparative review of CISNET breast models used to analyze US breast cancer incidence and mortality trends. *JNCI Monographs*. 2006; 2006:96–105.

Figure 1.
Selection of study cohort.

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.

Variables		Values	
Risk stratification model		Survival prognostic model	
AA population		General population	AA population
Input			
Age at diagnosis	Age as a continuous variable	Age categories [‡] : 54, 55-68, 69-78, 79	Age categories [‡] : 44, 45-55, 56-68, 69
Sex	Male, female	Male, female	Male, female
Stage	Stage I/II, III/IV, unknown	Stage I/II, III/IV, unknown	Stage I/II, III/IV, unknown
B symptoms	Present, absent, unknown	Present, absent, unknown	Present, absent, unknown
Adjusted IPI scores	0, 1, 2, 3	-	-
Output	Survival time	Survival status: 5 years (=1), <5 years (=0)	

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.

Total SEER DLBCL population				Risk stratification analysis				Compare prognostic models				
Characteristics	White		AA		White		AA		White		AA	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
N	25,447		2,171		1,820		127		17,583		1,514	
Age (year)												
Median (IQR)	68	(54-78)	55	(44-68)	65	(53-77)	53	(43-65)	71	(57-80)	56	(44-69)
60	17,390	68.34%	879	40.49%	1,139	62.58%	41	32.28%	12,575	71.52%	638	42.14%
Male	14,007	55.04%	1,248	57.49%	1,012	55.60%	69	54.33%	7,920	45.04%	650	42.93%
Stage												
I/II	11,953	46.97%	894	41.18%	772	42.42%	45	35.43%	7,995	45.47%	587	38.77%
III/IV	12,284	48.27%	1,200	55.27%	1,039	57.09%	80	62.99%	8,732	49.66%	874	57.73%
Unknown	1,210	4.75%	77	3.55%	9	0.49%	2	1.57%	856	4.87%	53	3.50%
B symptoms												
No	12,567	49.38%	992	45.69%	1,029	56.54%	66	51.97%	8,003	45.52%	633	41.81%
Yes	6,215	24.42%	720	33.16%	636	34.95%	56	44.09%	4,423	25.15%	513	33.88%
Unknown	6,665	26.19%	459	21.14%	155	8.52%	5	3.94%	5,157	29.33%	368	24.31%
Adjusted-IPI												
0	527	2.07%	36	1.66%	527	28.96%	36	28.35%				
1	669	2.63%	41	1.89%	669	36.76%	41	32.28%				
2	449	1.76%	33	1.52%	449	24.67%	33	25.98%				

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Total SEER DLBCL population				Risk stratification analysis		Compare prognostic models	
	White	AA		White	AA	White	AA
3	175	0.69%	17	0.78%	17	13.39%	
Unknown IPI	23,429	92.07%	2,013	92.72%			

Table III

Performance of survival prognostic model developed using the general population (GM) and tested in the general and African American populations

Model	Testing population	H-L test		AUC
		X ² statistics	p-value	
GM-LR	General	5.684	0.683	0.736
	AA	73.279	<0.001	0.679 (p=0.003)
GM-ANN(3)†	GN	9.109	0.333	0.750
	AA	78.195	<0.001	0.681 (p<0.001)
GM-ANN(10)	General	11.509	0.175	0.740
	AA	76.506	<0.001	0.684 (p=0.003)
GM-ANN(15)	General	6.656	0.574	0.740
	AA	75.878	<0.001	0.684 (p=0.003)

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

Model	Training set	H-L test χ^2 statistics	p-value
GM2-LR	GN	98.272	<0.001
AM-LR	AA	18.664	0.017
GM2-ANN(2)	GN	89.646	<0.001
AM-ANN(2)	AA	12.355	0.136
GM2-ANN(3)	GN	119.288	<0.001
AM-ANN(3)	AA	24.317	0.002
GM2-ANN(10)	GN	118.605	<0.001
AM-ANN(10)	AA	57.550	<0.001

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