haematologica
Journal of the Ferrata Storti Foundation

Integrating information across datasets to improve prognostic models for patients with diffuse large B-cell lymphoma

by Can Xie, Dai Chihara, R. Andrew Harkins, Ruosha Li, Xuelin Huang and Christopher R. Flowers

Received: March 2, 2026.
Accepted: November 12, 2026.

Citation: Can Xie, Dai Chihara, R. Andrew Harkins, Ruosha Li, Xuelin Huang and Christopher R. Flowers. Integrating information across datasets to improve prognostic models for patients with diffuse large B-cell lymphoma.
Haematologica. 2026 Nov 19. doi: 10.3324/haematol.2026.300829 [Epub ahead of print]

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication.
E-publishing of this PDF file has been approved by the authors.
After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval, the final version of the manuscript will then appear in a regular issue of the journal.
All legal disclaimers that apply to the journal also pertain to this production process.

Integrating information across datasets to improve prognostic models for patients with diffuse large B-cell lymphoma

Short title for the running head: Integrated Prognostic Modeling in DLBCL

[*]Can Xie[1]; [*]Dai Chihara[2]; R. Andrew Harkins[2]; Ruosha Li[3]; [#]Xuelin Huang[1]; and [#]Christopher R. Flowers[2]

[1]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA;

[2]Department of Lymphoma and Myeloma, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[3]Department of Biostatistics and Data Science, The University of Texas Health Science Center at Houston, Houston, TX, USA;

* Co-first authors
# Co-corresponding authors

Co-correspondence:

Xuelin Huang, PhD, Email contact: xlhuang@mdanderson.org

Christopher R. Flowers, MD, Email contact: crflowers@mdanderson.org

Acknowledgments
The US National Institutes of Health grant R01CA272806 provided support for Huang and Xie. The research of Huang was also partially supported by the US National Institutes of Health grants U54CA096300, U01CA253911, and P50CA100632. Both Huang and Xie were partially supported by the Dr. Mien-Chie Hung and Mrs. Kinglan Hung Endowed Professorship. The research of Li was partially supported by the US National Institutes of Health grants R01DK117209 and R03NS108136. This work was supported by Cancer Prevention and

Research Institute of Texas (CPRIT) RR190079 (PI: Christopher R. Flowers, MD) Dr. Flowers is a CPRIT Scholar. The research of Flowers was also partially supported by the US National Institutes of Health grant U01CA308978, P01CA272295, and U54CA302435. Dai Chihara is a scholar of Blood Cancer United.

Authorship Contributions
XH and CRF designed the research and contributed equally as co-corresponding authors. CX, RAH and RL collected the data. CX, RL and XH analyzed the data. CX and DC wrote the draft and contributed equally as co-first authors. All authors reviewed and contributed to the final paper.

Disclosure of Conflicts of Interest
DC receives research funding from BMS, Genentech, Genmab, Treeline and ONO, and served as an advisory board member for ADC therapeutics, REGENERON, Incyte, Miltenyi Biomedicine, BMS and Genmab. The remaining authors declare no conflicts of interest. CRF has received consulting fees from: BeOne, Foresight Diagnostics, a wholly-owned subsidiary of Natera Inc, Genentech, a member of the Roche Group, Genmab US Inc, N-Power Medicine; Stock Options/Stock from: Foresight Diagnostics, a wholly-owned subsidiary of Natera Inc and N-Power Medicine; and support for contracted research from: 4D Pharma PLC, Acerta Pharma—A member of the AstraZeneca Group, Adaptimmune, Alaunos Therapeutics, Allogene Therapeutics, Amgen Inc, BostonGene, Cellectis, EMD Serono Inc, Genentech, a member of the Roche Group, Guardant Health, Iovance Biotherapeutics, Kite, A Gilead Company, MorphoSys, Nektar Therapeutics, Novartis, Pfizer Inc, Sanofi, Takeda Pharmaceuticals USA Inc, TG Therapeutics Inc, Xencor.

Data Sharing Statement
All data can be found in a data supplement available with the online version of this article.

AI Statement
No artificial intelligence was used to produce this scientific work.

Abstract

Accurate survival prediction is vital for guiding clinical decisions in diffuse large B-cell lymphoma (DLBCL), especially for advancing precision medicine within its genomically heterogeneous subgroups. However, datasets that include both detailed clinical variables and genomic profiling are usually small. In contrast, large population-based datasets often lack multi-omics data. This gap limits the development and validation of robust prognostic models in DLBCL. We developed a transfer-learning prognostic framework that leverages the external population-level information from the Surveillance, Epidemiology, and End Results program to the prognostic model for an individual-level clinical and genomic cohort. The proposed model is a Cox-model-based boosted regression with individual data for generalization and enhancement (BRIDGE-Cox) that increases predictive accuracy and reduces uncertainty. The BRIDGE-Cox model incorporates the International Prognostic Index (IPI) factors, with model training performed in the Reddy genomic dataset and external validation in an independent National Cancer Institute genomic dataset. Compared with conventional IPI-based and revised IPI-based Cox models, the BRIDGE-Cox model improved predictive performance for overall survival, achieving a higher concordance index (0.70 vs 0.66 and 0.63, respectively) and consistently higher time-dependent area under the receiver operating characteristic curves for 2-, 3-, and 5-year survival prediction by using the repeated cross-validation in the Reddy dataset. Importantly, these gains were maintained within genomic subgroups. External validation confirmed the robustness and generalizability of the BRIDGE-Cox approach. By borrowing large population-level data, this model addresses key limitations of existing tools and enables more reliable risk stratification for patients with DLBCL, particularly in genomically defined subpopulations.

Introduction

Diffuse large B-cell lymphoma (DLBCL), the most common subtype of aggressive non-Hodgkin lymphoma, is a heterogeneous disease from clinical and biological perspectives,[1,2] producing different risks of relapse and survival for patients. Accurate prognostic prediction helps guide personalized treatment and trial design. The International Prognostic Index (IPI), developed utilizing five simple clinical factors, was the first introduced prognostic model for aggressive lymphoma, most of which were DLBCL, and is still widely used in prognostication and trial eligibility criteria.[3] Starting in the era of rituximab, the revised-IPI (R-IPI)[4] and enhanced IPI using data from National Comprehensive Cancer Network (NCCN-IPI),[5] became two other commonly used prognostic models.

Besides these three common clinical prognostic models, recently proposed prognostic models have extended the IPI by incorporating more features, such as genetic and molecular biomarkers and imaging-based risk factors. Via gene-expression profiling, DLBCL is classified into three cell-of-origin (COO) subtypes, activated B-cell-like (ABC, 30-40%), germinal-center B-cell-like (GCB, 40-50%), and unclassifiable (10-15%)[6] that influence treatment response and survival.[7-10] Other factors reported that impact survival include MYC, BCL2, and BCL6 expression and rearrangements,[9-12] total metabolic tumor volume,[13] and the amount of circulating tumor DNA (ctDNA).[14]

Supervised and unsupervised machine learning methods are increasingly implemented for prognosis modeling in DLBCL to identify prognostically relevant variables.[15-17] A support vector machine (SVM) classifier based on eight gene-expression markers was proposed to classify the prognostic risk groups for DLBCL patients.[18] Stacking ensemble methods have also been applied to combine several prognostic survival curves with weights.[16,19] In addition, the least absolute shrinkage and selection operator regression, random forest survival models, and neural networks have been used to construct prognostic models using clinical, genomic, or imaging data.[20-22]

However, when developing prognostic models that incorporate extensive clinical and molecular features, assembling a large cohort of patients is time-consuming and financially challenging. A systematic review of prognostic models for DLBCL proposed between 1993 and 2018 showed that only 37% of included papers (14 of 38) had sample sizes greater than 500 patients.[23] Although institution-specific, consortium, and trial datasets often contain granular, patient-level data, they are typically limited by small to modest sample sizes, restricted

demographic diversity, and center-specific practice patterns, all of which constrain the generalizability and external validity of derived prognostic models. On the other hand, large population-based registry datasets such as the Surveillance, Epidemiology, and End Results (SEER) Program provide large and diverse patient populations but limited clinical and no biomarker data available. These limitations make both institution-specific and large population-based data challenging to use as a sole source for modeling. However, it provides an opportunity to develop a new method that borrows the strength of a large-scale registry database to improve the performance of a prognostic model built from a smaller but clinically richer dataset.

Herein, we propose a machine-learning-based prognostic model that integrates external information from a large-scale dataset while preserving the clinical features available in genomic datasets. The model incorporates the five IPI components: age, Ann Arbor stage, serum lactate dehydrogenase (LDH), Eastern Cooperative Oncology Group performance status (ECOG PS), and number of extranodal sites.

Methods

Ethics Statement

This study was conducted in the United States using publicly available, de-identified datasets and was performed in accordance with applicable ethical standards and regulations in the United States. Because the analysis was based exclusively on publicly available data without identifiable private information, formal approval by a constituted institutional review board and informed consent were not required.

Data source

We utilized three publicly available DLCBL datasets, restricting to newly diagnosed adults (age ≥18 years). The Reddy et al[9] dataset ($n = 1{,}001$) was used to develop our proposed prognostic model. For comparisons with the IPI- and R-IPI-Cox-based models, we included 744 patients with survival data and complete IPI components. SEER dataset was leveraged to enhance predictive performance. Using SEER*Stat software (version 9.0.42), we extracted adults DLBCL cases (ICD-O-3 code: 9680)[24] diagnosed in 2010—2022 from the 17 Registries database. We retained variables overlapping with the Reddy dataset (age and Ann Arbor stage), yielding 70,281 patients. The National Cancer Institute (NCI) dataset from Schmitz et al[25] ($n =$

217) was used for external validation. After excluding patients aged < 18 years, 170 patients with survival values and complete IPI components remained.

Proposed and comparator prognostic models

We proposed a Cox-model-based boosted regression with individual data for generalization and enhancement (BRIDGE-Cox) for overall survival (OS) prediction that integrates SEER and Reddy datasets. In the first step, we implemented a Trans-Cox model[26] to borrow information from SEER using shared predictors (age, Ann Arbor stage), obtaining externally adjusted coefficients and baseline cumulative hazard estimates for the Reddy dataset. Its tuning parameters were selected by the area under the receiver operating characteristics (ROC) curve (AUC). In the second step, we fit a Cox model in the Reddy dataset incorporating the remaining IPI components (LDH, ECOG PS, and number of extranodal sites) with the predicted risk scores from the first step included as offsets. For subgroup prediction (e.g., *MYD88*-mutated), we additionally included the subgroup indicator in a Cox model with step-2 risk scores as offsets. We compared the BRIDGE-Cox model with Cox models using the IPI and R-IPI risk groups as predictors (IPI- and R-IPI-Cox models, respectively) (**Figure 1**).

Predictive performance evaluation

We evaluated performance on predicted OS probabilities across follow-up using the concordance index[27] (C-index; R package survC1) and assessed 2-, 3-, and 5- year performance using the time-dependent AUC and Brier score (BS) with the R package tdROC.[28,29]

Internal validation used five-fold cross-validation (CV) repeated 100 times, summarized as the mean C-index, AUC, and BS with their interquartile ranges across 500 test evaluations (5 folds × 100 repeats). In each repeat, the Reddy dataset was randomly split into five equal folds. Each fold served as the test set once and the remaining folds as the training set. For external validation trained all three models were trained on the entire Reddy dataset and evaluated predictions in the NCI dataset with 95% confidence intervals. We also tested C-index differences between BRIDGE-Cox model and comparator models using standard errors estimated from a leaving-one-out jackknife resampling approach and *t*-based inference.[30,31]

We further examined performance within genomic subgroups (COO: GCB/ABC; mutations: *MYD88*, *ATM*, *SETD2*, and *MGA* [MAX dimerization protein]). External subgroup validation was conducted for GCB, ABC and *MYD88*-mutated patients available in NCI dataset. The

training models were extended by adding the relevant indicator (COO for GCB/ABC; *MYD88* status for *MYD88*-mutated).

We provide the R code to fit the BRIDGE-Cox model for all patients and GCB, ABC, and *MYD88* subgroups in the Reddy dataset and evaluate the OS predictive performance in the NCI dataset using the C-index, AUC, and Brier score. It is available in the online supplementary file and at https://github.com/cxie19/DLBCL_BRIDGE-Cox.

Results

Patient characteristics

The median age at diagnosis in the SEER dataset was higher than that in the Reddy dataset (68 vs 63 years), and the distribution of age categories differed significantly between two datasets ($P < 0.01$) (**Table 1**). The largest difference in the age categories was observed among patients older than 75 years (17.2% in Reddy and 28.8% in SEER), corresponding to an absolute difference of 11.6% (**Table 1**). An absolute difference of 8.8% was presented in patients age 41-60 years, who accounted for 33.3% of the Reddy cohort and 24.5% of SEER cohort (**Table 1**). The Reddy dataset included significantly more patients with stages III/IV DLBCL than the SEER dataset (60.9% vs. 55.7%; $P < 0.01$), corresponding to an absolute difference of 5.2% (**Table 1**). The NCI and Reddy datasets did not show significant difference in the five components of the IPI except the number of extranodal sites ($P = 0.046$) (*online supplementary* **Table S1**). However, these two datasets did not have significant difference in IPI or R-IPI risk groups ($P = 0.146$ and 0.147, respectively) (*online supplementary* **Table S1**).

Patients in the Reddy dataset had better Kaplan-Meier (KM) estimates of OS than those in the SEER dataset (log-rank $P < 0.01$) (**Figure 2**). The median OS duration was 10.7 years in Reddy dataset and 7.6 year in SEER, and the median follow-up durations were 4.3 and 6.1 years, respectively. The KM estimates of OS were similar in the NCI and Reddy datasets (log-rank $P = 0.76$) with a median OS of 10.5 years in the NCI dataset (*online supplementary* **Figure S1**).

Therefore, the heterogeneity of the SEER dataset enables borrowing diverse population-level information to augment the smaller Reddy dataset. The similarity between the NCI and Reddy datasets supports use of the NCI cohort as an appropriate external validation dataset for models developed in the Reddy dataset.

BRIDGE prognostic risk groups

Following Sehn and Gascoyne[32] and Mikhaeel et al[13], we defined the BRIDGE-model-based prognostic index (BPI) with three prognostic risk groups (low, intermediate, and high) using the 10th and 40th percentiles of the predicted 5-year OS in the Reddy dataset. Using the BRIDGE-Cox model trained in the Reddy dataset, cutoffs at 64.9% (40th percentile) and 37.5% (10th percentile) defined the low (5-year OS > 64.9%), intermediate (37.5%-64.9%), and high (≤ 37.5%) BPI groups.

The KM estimates of OS for BPI risk groups exhibited clear separation in both datasets (**Figure 3E**, **F**), indicating that the BPI effectively stratified patients into prognostically distinct groups (log-rank $P < 0.01$). Although the IPI and R-IPI risk groups were significantly different in the two datasets (log-rank $P < 0.01$) (**Figure 3A-D**), patients with IPI low-intermediate and high-intermediate risk groups both in the Reddy and NCI dataset showed similar OS (**Figure 3A**, **B**), and the KM curves of patients with R-IPI very good and good risk groups in the Reddy dataset crossed (**Figure 3C**).

We compared the predicted 5-year OS probabilities across risk groups defined by the BPI, IPI, and R-IPI in the Reddy and NCI datasets using models trained the Reddy dataset (*online supplementary* **Table S2**). In the Reddy dataset, the mean predicted 5-year OS from the BRIDGE-Cox model was 80.1%, 55.6%, and 28.2%, respectively, for the BPI low-, intermediate-, and high-risk groups. Under the IPI-Cox model, predicted 5-year OS probabilities were 84.2%, 68.3%, 59.6%, and 41.1% for the low-, low-intermediate-, high-intermediate-, and high-risk groups, respectively. From the R-IPI-Cox model, the very good, good, and poor risk groups had the 5-year OS of 93.2%, 73.8%, and 51.8%, respectively.

The BPI high-risk group had lower predicted 5-year OS than the highest-risk groups from the IPI and R-IPI (28.2% vs 41.1% and 51.8%) in the Reddy dataset. The same pattern was observed in the NCI dataset using the risk group classification criteria derived from the Reddy dataset. For the high-risk group, the 5-year OS in BPI was 29.6% compared with 41.1% and 51.8% in the IPI and R-IPI, respectively. Hence, the BPI was better at distinguishing the high-risk group than the IPI and R-IPI.

To establish a four-level BPI, we further divided the previous low-risk group into low-intermediate- and low-risk groups, defined by predicted 5-year OS between the 40th and 75th percentiles and greater than the 75th percentile, respectively. The BPI with four prognostic risk

groups distinctly separated the KM curves of four risk groups in both the Reddy and NCI cohorts (*online supplementary* **Figure S2**). The 5-year OS probabilities in the low and low-intermediate risk groups derived from BPI are higher than those for the corresponding IPI risk groups (*online supplementary* **Table S3**). Therefore, both the three- and four-level BPI provided clear stratification of prognostic risk groups.

We developed a Shiny-based web tool to determine a new patient's risk group and predict 5-year OS probability using the BRIDGE-Cox model (https://cxie19.shinyapps.io/DLBCL_BPI/). The source code is also available at https://github.com/cxie19/DLBCL_BRIDGE-Cox.

Predictive performance evaluation

The predictive performance results as measured by C-index and AUC in the Reddy dataset from the repeated five-fold CV. The BRIDGE-Cox model had the largest average C-index value (0.70), followed by the IPI-Cox model (0.66) and the R-IPI-Cox model (0.63). For the predicted 2-, 3-, and 5-year OS probabilities, the BRIDGE-Cox model had the best discriminative ability, with the average AUC values of 0.75, 0.77, and 0.73, respectively. The smaller average AUC values of the IPI-Cox model (0.70, 0.72, and 0.68, respectively) and the R-IPI-Cox model (0.67, 0.69, and 0.66, respectively) indicated poorer predictive performance than the BRIDGE-Cox model. The BS values by the BRIDGE-Cox model were smallest (*online supplementary* **Table S4**), indicating superior calibration. Also, the metrics values from the BRIDGE-Cox model in the 500 CV subsets tended to have smaller variability than the other two models. The best performance of the BRIDGE-Cox model revealed that the external information from the SEER dataset helped improve the predictive performance. In the BRIDGE-Cox model, we used age categorized into $\leq$40, 41-60, 61-75, and >75 years since it had better predictive performance than the model including age categorized into $\leq$ and $>$ 60 years via the repeated CV on the Reddy dataset (*online supplementary* **Table S5**). In all the patient subgroups (GCB, ABC, *MYD88*, *ATM*, *SETD2*, and *MGA*), the BRIDGE-Cox outperformed the other two models in terms of C-index, AUC, and BS from the repeated CV (*online supplementary* **Table S6**).

We validated the three Cox models trained on the entire Reddy dataset (*online supplementary* **Table S7**) using the NCI dataset. For the subgroup prediction, we incorporated the COO and *MYD88* mutation status as predictors in the three models (*online supplementary* **Tables S8-S10**)**.** The BRIDGE-Cox model had the largest C-index values on OS across all time points in the

entire NCI dataset and the GCB, ABC, and *MYD88* subgroups (**Figure 5**). In the GCB subgroup, the C-index of BRIDGE-Cox model was significantly better than the IPI-Cox model (0.70 vs 0.59, $P = 0.015$) and had the trend to significantly outperform the R-IPI-Cox model (0.70 vs 0.63, $P = 0.065$) (**Figure 5**). The C-index values of the BRIDGE-Cox model were not significantly different from the IPI- and R-IPI-Cox models in all patients and patient subgroups of ABC and *MYD88* (**Figure 5**). The BRIDGE-Cox model showed narrower distribution and 95% confidence intervals of the C-index than the comparator models, indicating more precise and more stable discrimination, particularly in the small *MYD88*-mutated subgroup (**Figure 5**).

For all patients in the NCI dataset, the BRIDGE-Cox model achieved larger time-dependent AUC values than those in the IPI- and R-IPI-Cox models (**Table 2**). The BRIDGE-Cox model's AUC values for 2-, 3-, and 5-year OS were 0.69, 0.68, and 0.69, respectively, compared with 0.68, 0.66, and 0.67 for the IPI-Cox model and 0.65, 0.65, and 0.66 for the R-IPI-Cox model. For the GCB and *MYD88* subgroups, the AUC values from the BRIDGE-Cox model were the largest. Similarly, the BRIDGE-Cox model had all BS values for OS at different time points smaller than those from the comparator models for all patients and subgroups of GCB and *MYD88* (*online supplementary* **Table S11**). These results demonstrate that the BRIDGE-Cox model outperformed the IPI- and R-IPI-Cox models in the entire NCI dataset in terms of C-index, AUC, and BS. Therefore, in small or modest sample sizes, the BRIDGE-Cox model increased the prediction accuracy and reduced the uncertainty in OS estimates.

From the perspective of clinical utility, decision curve analysis showed that BRIDGE-Cox was clinically useful, with net benefit curves above treat-all and treat-none strategies (**Figure 6**). The BRIDGE-Cox model provided higher net benefit than IPI-Cox and R-IPI-Cox at higher threshold probabilities in the overall cohort and favorable utility in the ABC and MYD88 subgroups (**Figure 6**).

Discussion

We proposed a transfer learning prognostic model for DLBCL incorporating external information from a large population-based dataset (e.g., SEER) and a genomic dataset. To our knowledge, this is the first study to link datasets under transfer-learning framework to improve the predictive performance of a prognostic model for DLBCL using the components of IPI. The proposed BRIDGE-Cox model included age and Ann Arbor stage from the Reddy and SEER

datasets and LDH, ECOG PS, and number of extranodal sites from the Reddy dataset. While borrowing the external information, it had more accurate predictive performance and outperformed the widely used IPI and R-IPI prognostic models, which was validated further with the entire NCI dataset and patient subgroups in this dataset. Leveraging a large population-based dataset enhanced the generalizability of the prognostic model and helped address the limitations due to the limited sample size of the smaller genomic dataset.

The BRIDGE-Cox model is a machine learning method that incorporates the transfer learning approach and extends the Trans-Cox model. The Trans-Cox model is a Cox-model-based approach for time-to-event data that adaptively transfers the risk estimation from an external study (i.e., the SEER dataset) into an internal study (i.e., the Reddy dataset). In its original formulation, the Trans-Cox model was developed primarily for risk estimation and inference and can be used to evaluate the statistical significance of predictors. Leveraging this inferential property, the BRIDGE-Cox model retains strong interpretability as reflected by the estimated hazard ratios and statistical significance of predictors (*online supplementary* **Tables S7** and **S10)**. We further extended the Trans-Cox model to support the prediction. To better align with the prediction objective, we changed the criterion of selecting the tuning parameters in the Trans-Cox model from the Bayesian Information Criterion[33] to the AUC.

Methodologically, the BRIDGE-Cox model is not a direct pooling approach within the transfer-learning framework. In the first step, it accommodates potential differences in covariate effects for shared predictors and the time-varying baseline hazards from the external and internal datasets. In the second step, it incorporates predictors available only in the internal dataset. Thus, the model leverages the large registry cohort to improve estimation of shared predictor effects and baseline hazards while preserving the additional clinical granularity of the smaller genomic cohort.

For the BRIDGE-Cox model, the individual-level data are assumed to be available from both the internal and external datasets, and the shared covariates must be harmonized to comparable definitions and coding schemes. In the first step, the Trans-Cox model assumes that the proportional hazards assumption holds within both cohorts. Importantly, the Trans-Cox model allows the covariate distributions, covariate effects, and baseline cumulative hazards to differ between the two datasets. These discrepancies are explicitly accounted for in the model. This flexibility makes the approach more applicable to real-world settings, where heterogeneity across

datasets commonly arises. When the two have identical risk models, the estimated discrepancy parameters for the covariate effects of the shared predictors and the baseline hazards are zero.[26]

The heterogeneity between the Reddy and SEER datasets was reflected in both baseline characteristics and survival outcomes. Compared to the Reddy dataset, the SEER cohort included a higher proportion of older patients and a slightly lower proportion of patients with advanced-stage disease. The SEER cohort had worse OS than the Reddy cohort. One plausible explanation for this heterogeneity is selection bias. The SEER dataset represented a broad population-based registry, whereas the Reddy cohort was a genomically profiled research cohort requiring available clinical and molecular information. Data from the largest prospective study of patients with DLBCL in the US also showed a difference in age distribution suggesting that there is an older population of patients with DLBCL who are not routinely represented in research studies.[34] Because comorbidity information and calendar years of diagnosis were unavailable for the Reddy dataset, and detailed treatment-regimen and comorbidity information were unavailable in the standard SEER dataset, we were limited in our ability to determine additional causes of heterogeneity.

The transfer learning method was previously applied to OS prediction in lung cancer,[35] where it pretrained a Cox deep neural network on SEER dataset and then adapted it to the internal data via fine-tuning. This method could not incorporate new predictors during fine-tuning. In addition, the original Trans-Cox model[26] uses only factors available in both the internal and external datasets. In contrast, our proposed BRIDGE-Cox model incorporates additional predictors from the genomic dataset by including the offsets from the first Cox model into the Cox model in the second step, and enables subgroup prediction by adding subgroup indicators in the third-step Cox model with the step-2 offsets.

This model-augmentation approach leverages external information while preserving all available survival information from the entire internal data when adding a genomic variable in the third step. When a genomic variable has missing values, we do not need to exclude all the patients with missing information on that variable, so this prognostic model preserves sample size and maintains statistical efficiency compared to the conventional IPI- and R-IPI-Cox models.

The second- and third-step Cox models used predicted risk scores as offsets, thereby treating prior estimates as fixed. This approach may not fully propagate uncertainty across modeling

steps for parameter inference. However, because the primary objective of this study was prognostic prediction rather than inference on model coefficients, we evaluated the BRIDGE-Cox model using repeated cross-validation and external validation. To quantify uncertainty or conduct inference, future work could conduct a perturbation scheme for uncertainty propagation or adopt joint modeling frameworks such as Bayesian dynamic-borrowing methods.[36,37]

The internal and external predictive performance evaluation demonstrated that the R-IPI-Cox model had suboptimal predictive performance. This was consistent with findings in previous studies that compared the between IPI and R-IPI.[38-40] The IPI-Cox model has comparable predictive performance with our proposed BRIDGE-Cox model since the sample size of the Reddy dataset as the training set was moderate. When the training set includes a small number of patients, the BRIDGE-Cox model borrowing external information reduces the uncertainty of the OS estimates and has better predictive performance than the IPI-Cox model.

Compared to existing clinical and molecular prognostic models, the BRIDGE-Cox model is not intended to replace established clinical indices or biologically refined molecular risk-stratification tools. Rather, it provides a flexible transfer-learning framework that can incorporate additional clinical, molecular, and imaging-related predictors when available and can be applied to other combinations of external and internal datasets. The BRIDGE-Cox model is most useful when the internal dataset is limited in sample size but contains additional prognostic variables even without molecular profiling, because external population-level data can help stabilize prediction while preserving the additional granularity of the internal dataset. This prognostic model can help physicians with clinical decision-making by predicting individual survival probabilities and identifying risk stratification, which can assist in personalized treatment planning.

This study has several limitations. First, the external validation of the prognostic models would be more informative if the sample size of the external independent validation set was larger. Second, the Reddy dataset did not contain the corresponding details of the NCCN-IPI components. We could not construct the BRIDGE-Cox model with the same predictor categories as the NCCN-IPI or compare our proposed model with a Cox model using NCCN-IPI risk groups as predictors. Next, the proposed prognostic model is Cox-model based, and the violation of the proportional hazards assumption in either the internal or external dataset could affect the estimation of the prognostic model and prediction accuracy.

The last limitation is that the validity of borrowing information from SEER is constrained by the small shared predictor space between the SEER and Reddy datasets. In this study, since only age and Ann Arbor stage were shared between the two datasets, the external information borrowed from SEER was conditional only on these two variables. Consequently, the validity of this borrowing depends on whether the age- and stage-associated survival patterns estimated from the SEER cohort are sufficiently transportable to the Reddy cohort. If important prognostic variables unavailable in SEER, such as LDH, ECOG PS, extranodal involvement, molecular subtype, treatment patterns, or comorbidity, differ between the datasets and modify or confound the association of age and stage with survival, the SEER-derived risk component may introduce bias.

When the shared age- and stage-associated risk component is reasonably transportable to the target cohort, the large SEER sample size can improve the stability of this component and reduce uncertainty in the survival prediction. However, the magnitude of the improvement from external data borrowing may vary when transportability is uncertain. Although external validation is required to evaluate the transportability and validity of the BRIDGE-Cox model, a distinct value of the BRIDGE-Cox model is that borrowing remains possible with small shared predictor space when the population dataset is large.

We present a novel machine-learning based prognostic model, BRIDGE-Cox integrating rich clinical data with molecular characteristics and large population-based data from SEER to enhance prognostic discrimination for patients with newly diagnosed DLBCL including molecular subgroups. By leveraging transfer learning, the BRIDGE-Cox approach successfully incorporates external survival information while retaining the clinical granularity of the internal dataset reducing estimate uncertainty. This hybrid strategy addresses a fundamental challenge in prognostic model development, the tradeoff between data depth and data breadth, and demonstrates that various datasets with different strengths can be utilized to improve predictive performance.

References

1. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127(20):2375-2390.
2. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. Leukemia. 2022;36(7):1720-1748.
3. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. N Engl J Med. 1993;329(14):987-994.
4. 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(5):1857-1861.
5. Zhou Z, Sehn LH, Rademaker AW, et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. Blood. 2014;123(6):837-842.
6. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403(6769):503-511.
7. 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;14(6):460-467.e2.
8. Tilly H, Morschhauser F, Sehn LH, et al. Polatuzumab vedotin in previously untreated diffuse large B-cell lymphoma. N Engl J Med. 2022;386(4):351-363.
9. Reddy A, Zhang J, Davis NS, et al. Genetic and functional drivers of diffuse large B cell lymphoma. Cell. 2017;171(2):481-494.e15.
10. Rodriguez M, Alonso-Alonso R, Fernandez-Miranda I, et al. An integrated prognostic model for diffuse large B-cell lymphoma treated with immunochemotherapy. EJHaem. 2022;3(3):722-733.
11. Hu S, Xu-Monette ZY, Tzankov A, et al. MYC/BCL2 protein coexpression contributes to the inferior survival of activated B-cell subtype of diffuse large B-cell lymphoma and demonstrates high-risk gene expression signatures: a report from The International DLBCL Rituximab-CHOP Consortium Program. Blood. 2013;121(20):4021-4031.
12. Petrich AM, Gandhi M, Jovanovic B, et al. Impact of induction regimen and stem cell transplantation on outcomes in double-hit lymphoma: a multicenter retrospective analysis. Blood. 2014;124(15):2354-2361.
13. Mikhaeel NG, Heymans MW, Eertink JJ, et al. Proposed new dynamic prognostic index for diffuse large B-cell lymphoma: international metabolic prognostic index. J Clin Oncol. 2022;40(21):2352-2360.
14. Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. J Clin Oncol. 2018;36(28):2845-2853.
15. Merdan S, Subramanian K, Ayer T, et al. Gene expression profiling-based risk prediction and profiles of immune infiltration in diffuse large B-cell lymphoma. Blood Cancer J. 2021;11(1):2.
16. Fan SL, Zhao ZQ, Yu HM, et al. Applying probability calibration to ensemble methods to predict 2-year mortality in patients with DLBCL. BMC Med Inform Decis Mak. 2021;21(1):14.
17. Mu SD, Shi DY, Ai LS, et al. International prognostic index-based immune prognostic model for diffuse large B-cell lymphoma. Front Immunol. 2021;12:732006.
18. Zhao ST, Dong XL, Shen WZ, Ye Z, Xiang R. Machine learning-based classification of diffuse large B-cell lymphoma patients by eight gene expression profiles. Cancer Med. 2016;5(5):837-852.

19. Biccler JL, Eloranta S, Brown PD, et al. Optimizing outcome prediction in diffuse large B-cell lymphoma by use of machine learning and nationwide lymphoma registries: a Nordic Lymphoma Group study. JCO Clin Cancer Inform. 2018;2:1-13.
20. Shen ZY, Zhang S, Jiao YX, et al. LASSO model better predicted the prognosis of DLBCL than random forest model: a retrospective multicenter analysis of HHLWG. J Oncol. 2022;2022:1618272.
21. Orgueira AM, Arías JAD, López MC, et al. Prognostic stratification of diffuse large B-cell lymphoma using clinico-genomic models: validation and improvement of the LymForest-25 model. Hemasphere. 2022;6(4):e706.
22. Ferrández MC, Golla SSV, Eertink JJ, et al. An artificial intelligence method using FDG PET to predict treatment outcome in diffuse large B cell lymphoma patients. Sci Rep. 2023;13(1):13111.
23. Jelicic J, Larsen TS, Maksimovic M, Trajkovic G. Available prognostic models for risk stratification of diffuse large B cell lymphoma patients: a systematic review. Crit Rev Oncol Hematol. 2019;133:1-16.
24. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program. Lymphoma Subtype Recode/WHO 2008. Available from: https://seer.cancer.gov/lymphomarecode/lymphoma-who2008.html. Accessed May 29, 2026.
25. Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-Cell lymphoma. N Engl J Med. 2018;378(15):1396-1407.
26. Li Z, Shen Y, Ning J. Accommodating time-varying heterogeneity in risk estimation under the Cox model: a transfer learning approach. J Am Stat Assoc. 2023;118(544):2276-2287.
27. Uno H. survC1: C-statistics for risk prediction models with censored survival data. CRAN: R package version 1.0-3; 2021. Available from: https://cran.r-project.org/web/packages/survC1/index.html. Accessed May 29, 2026.
28. Li L, Greene T, Hu B. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. Stat Methods Med Res. 2018;27(8):2264-2278.
29. Li X, Yin Z, Li L. tdROC: Nonparametric estimation of time-dependent ROC, Brier score, and survival difference from right censored time-to-event data with or without competing risks. CRAN: R package version 2.0; 2023. Available from: https://CRAN.R-project.org/package=tdROC. Accessed May 29, 2026.
30. Chihara D, Asano N, Ohmachi K, et al. Prognostic model for mantle cell lymphoma in the rituximab era: a nationwide study in Japan. Br J Haematol. 2015;170(5):657-668.
31. Newson RB. Comparing the predictive powers of survival models using Harrell's C or Somers' D. Stata J. 2010;10(3):339-358.
32. Sehn LH, Gascoyne RD. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. Blood. 2015;125(1):22-32.
33. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461-464.
34. Cerhan JR, Maurer MJ, Link BK, et al. The lymphoma epidemiology of outcomes cohort study: design, baseline characteristics, and early outcomes. Am J Hematol. 2024;99(3):408-421.
35. Zhu F, Zhong R, Li F, et al. Development and validation of a deep transfer learning-based multivariable survival model to predict overall survival in lung cancer. Transl Lung Cancer Res. 2023;12(3):471-482.
36. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics. 2011;67(3):1047-1056.
37. Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. Bayesian Anal. 2012;7(3):639-673.

38. Ruppert AS, Dixon JG, Salles G, et al. International prognostic indices in diffuse large B-cell lymphoma: a comparison of IPI, R-IPI, and NCCN-IPI. Blood. 2020;135(23):2041-2048.
39. Warnnissorn N, Kanitsap N, Niparuck P, et al. External validation and comparison of IPI, R-IPI, and NCCN-IPI in diffuse large B-cell lymphoma patients treated with R-CHOP to predict 2-year progression-free survival. Hematology. 2022;27(1):1237-1245.
40. Jelicic J, Juul-Jensen K, Bukumiric Z, et al. Prognostic indices in diffuse large B-cell lymphoma: a population-based comparison and validation study of multiple models. Blood Cancer J. 2023;13(1):157.

Table 1. Patient characteristics and predictors included in the Reddy (n = 744) and the Surveillance, Epidemiology, and End Results (SEER, n = 70,281) datasets.

Characteristic	Reddy (n = 744)	SEER (n = 70,281)	*P*-value*
Age at diagnosis, years			<0.01
median (IQR)	63 (52.9-71.9)	68 (57-77)	
≤40	66 (8.9%)	4792 (6.8%)	
41-60	248 (33.3%)	17209 (24.5%)	
61-75	302 (40.6%)	28029 (39.9%)	
>75	128 (17.2%)	20251 (28.8%)	
Ann Arbor stage			<0.01
I/II	291 (39.1%)	31104 (44.3%)	
III/IV	453 (60.9%)	39177 (55.7%)	
ECOG performance status			
0-1	534 (71.8%)		
2-4	210 (28.2%)		
LDH			
≤ upper limit of normal	336 (45.2%)		
> upper limit of normal	408 (54.8%)		
Number of extranodal sites			
0 or 1	566 (76.1%)		
>1	178 (23.9%)		
IPI risk group			
Low risk	237 (31.9%)		
Low-intermediate risk	175 (23.5%)		
High-intermediate risk	192 (25.8%)		
High risk	140 (18.8%)		
R-IPI risk group			
Very good risk	79 (10.6%)		
Good risk	333 (44.8%)		
Poor risk	332 (44.6%)		

* Categorical predictors were compared using the Chi-square test.
Abbreviations: IQR, interquartile range; ECOG, Eastern cooperative oncology group; LDH, serum lactate dehydrogenase; IPI, International Prognostic Index; R-IPI, revised-IPI.

Table 2. External validation of the three prognostic models using the time-dependent area under the receiver operating characteristics curve (AUC; larger is better) for 2-, 3-, and 5-year OS evaluated in all patients and in the GCB, ABC, and *MYD88* subgroups of the National Cancer Institute dataset (n = 170). The AUC values are shown with 95% confidence intervals and *p*-values for the differences relative to the BRIDGE-Cox model (compared with the IPI- and R-IPI-Cox models) for each patient subgroup.

Patient	Model	At 2 years AUC	*P*-value	At 3 years AUC	*P*-value	At 5 years AUC	*P*-value
All (n=170)	BRIDGE-Cox	**0.685** (0.575, 0.786)		**0.675** (0.568, 0.773)		**0.689** (0.593, 0.774)	
	IPI-Cox	0.683 (0.586, 0.780)	0.963	0.663 (0.566, 0.755)	0.727	0.669 (0.581, 0.756)	0.543
	R-IPI-Cox	0.653 (0.566, 0.738)	0.439	0.649 (0.561, 0.730)	0.502	0.661 (0.575, 0.739)	0.444
GCB (n = 84)	BRIDGE-Cox	**0.702** (0.505, 0.89)		**0.711** (0.541, 0.874)		**0.728** (0.583, 0.87)	
	IPI-Cox	0.640 (0.440, 0.842)	0.148	0.627 (0.448, 0.810)	0.033	0.636 (0.493, 0.801)	0.017
	R-IPI-Cox	0.613 (0.44, 0.773)	0.129	0.628 (0.476, 0.776)	0.095	0.600 (0.472, 0.742)	0.016
ABC (n = 60)	BRIDGE-Cox	0.601 (0.454, 0.763)		0.571 (0.426, 0.737)		0.589 (0.44, 0.752)	
	IPI-Cox	**0.630** (0.488, 0.781)	0.726	**0.602** (0.446, 0.749)	0.712	0.613 (0.473, 0.774)	0.766
	R-IPI-Cox	0.617 (0.497, 0.749)	0.810	**0.602** (0.478, 0.742)	0.646	**0.646** (0.504, 0.785)	0.409
MYD88 (n = 37)	BRIDGE-Cox	**0.618** (0.431, 0.83)		**0.618** (0.429, 0.829)		**0.637** (0.486, 0.876)	
	IPI-Cox	0.587 (0.394, 0.781)	0.807	0.587 (0.395, 0.782)	0.837	0.592 (0.451, 0.819)	0.782
	R-IPI-Cox	0.524 (0.364, 0.681)	0.450	0.524 (0.364, 0.681)	0.474	0.614 (0.434, 0.798)	0.876

Models: (1) IPI-Cox, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox, a Cox-model-based boosted regression with individual data for generalization and enhancement.

Figure 1. Predictors incorporated in the prognostic models across the Reddy and Surveillance, Epidemiology, and End Results (SEER) datasets. IPI, International Prognostic Index; R-IPI, revised International Prognostic Index; LDH, serum lactate dehydrogenase; ECOG PS, Eastern Cooperative Oncology Group performance status.

Figure 2. Kaplan-Meier estimates of overall survival in the Reddy dataset (n = 744) and the Surveillance, Epidemiology, and End Results (SEER) dataset (n = 70,281).

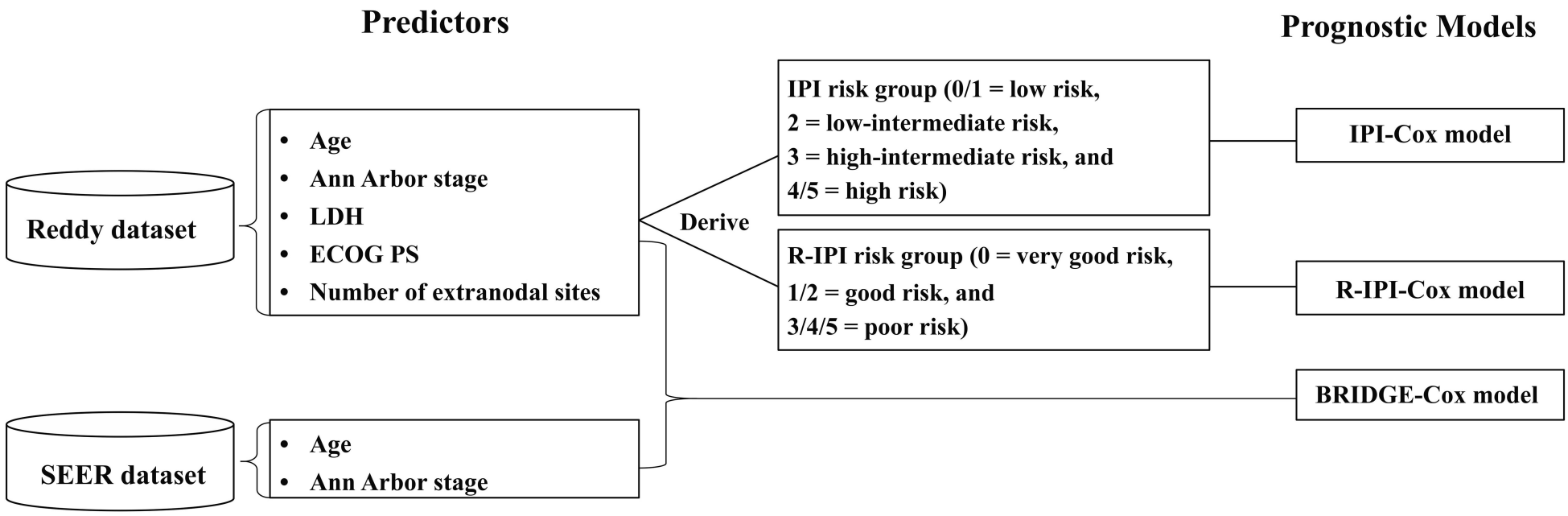
Figure 3. Kaplan-Meier estimates of overall survival by prognostic risk group in the Reddy dataset (n = 744) and National Cancer Institute (NCI) dataset (n = 170). Panels A, C, E and B, D, F show overall survival in the Reddy and NCI datasets, respectively. Patients were stratified by risk groups defined using the International Prognostic Index (IPI) in panels A and B, the revised-IPI (R-IPI) in panels C and D, and the prognostic index defined from the proposed boosted regression with individual data for generalization and enhancement model (BPI) in panels E and F.

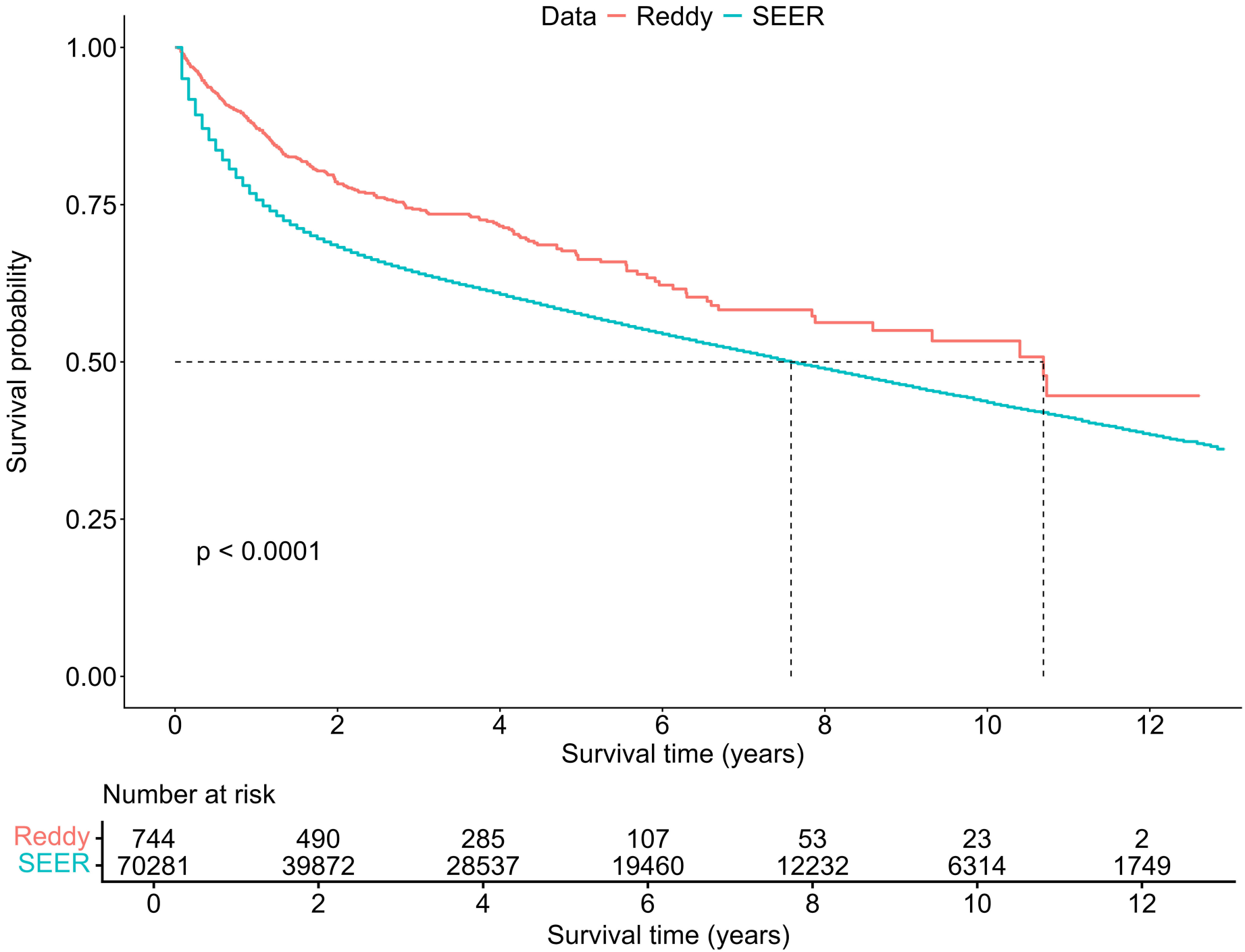
Figure 4. Internal validation performance of the three prognostic models in the Reddy dataset (n = 744), using the repeated five-fold cross-validation (CV, with 100 repetitions). The evaluation metrics include the concordance index (C-index; larger is better) for overall survival (OS) across all time points and the time-dependent area under the receiver operating characteristics curve (AUC; larger is better) for 2-, 3-, and 5-year OS. Results from the 500 CV repetitions (five test subset each from 100 repetitions) are illustrated by the violin plots with the box plots and the mean values annotated with the red asterisks. Models: (1) IPI-Cox, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox, a Cox-model-based boosted regression with individual data for generalization and enhancement.

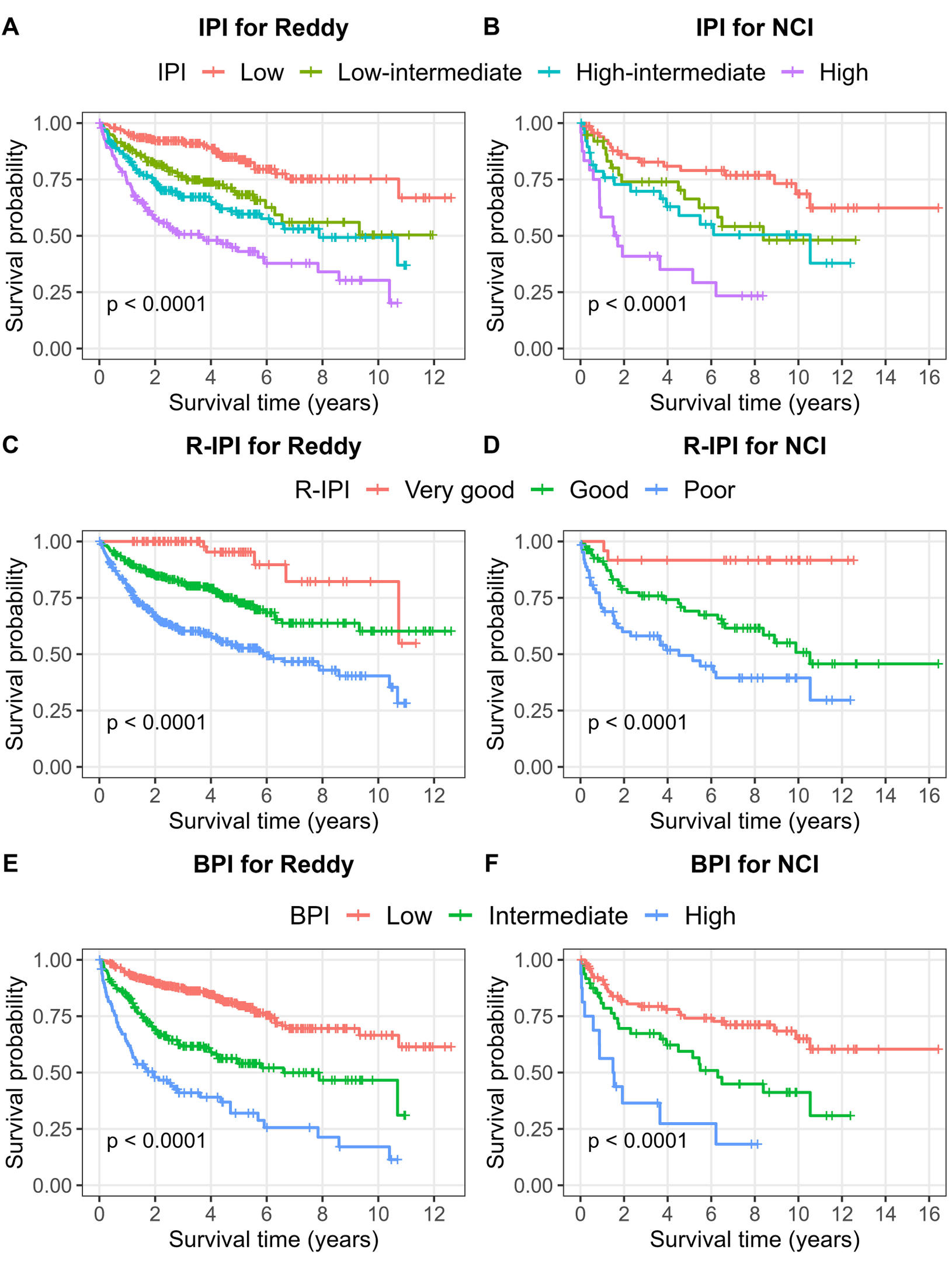
Figure 5. External validation performance of the three prognostic models in the National Cancer Institute dataset (n = 170) evaluated by the concordance index. Predictive performance was evaluated using the concordance index (C-index; larger is better) for overall survival across all time points evaluated in the overall cohort and GCB, ABC, and *MYD88* subgroups of the National Cancer Institute dataset. The C-index values are presented with 95% confidence intervals, and *p*-values for the differences relative to the BRIDGE-Cox model (compared with the IPI- and R-IPI-Cox models) are reported for each patient subgroup. Models: (1) IPI-Cox, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox, a Cox-model-based boosted regression with individual data for generalization and enhancement.

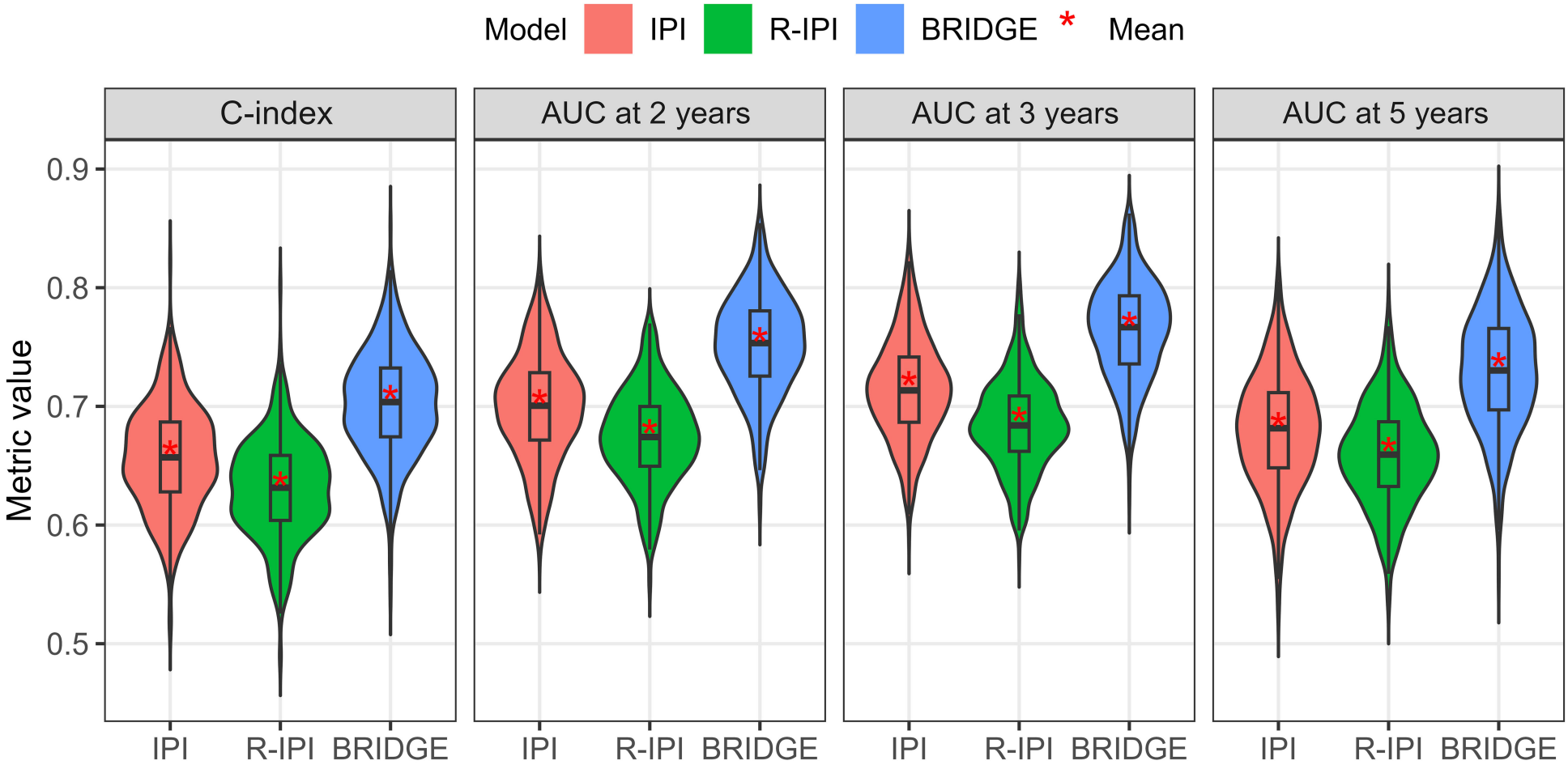
Figure 6. Decision curve analysis of the three prognostic models in the National Cancer Institute dataset (n = 170). Decision curve analysis was performed to compare the clinical

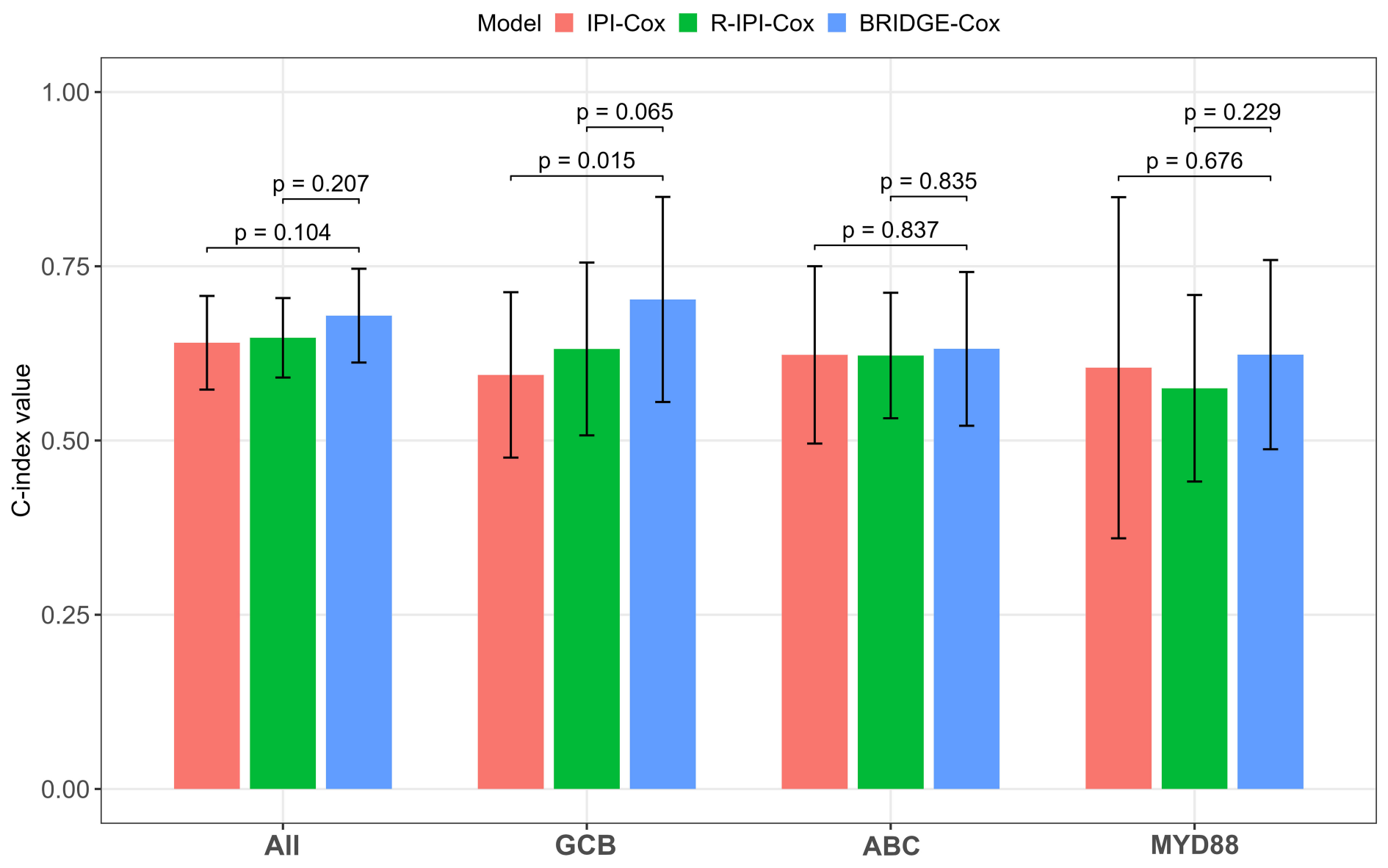
utility of the IPI-, R-IPI-, and BRIDGE-Cox models for predicting 5-year overall survival risk in the overall cohort (panel A) and GCB (panel B), ABC (panel C), and *MYD88* (panel D) subgroups. Net benefit was evaluated across threshold probabilities with treat-all and treat-none strategies shown as reference curves. Models: (1) IPI-Cox, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox, a Cox-model-based boosted regression with individual data for generalization and enhancement.

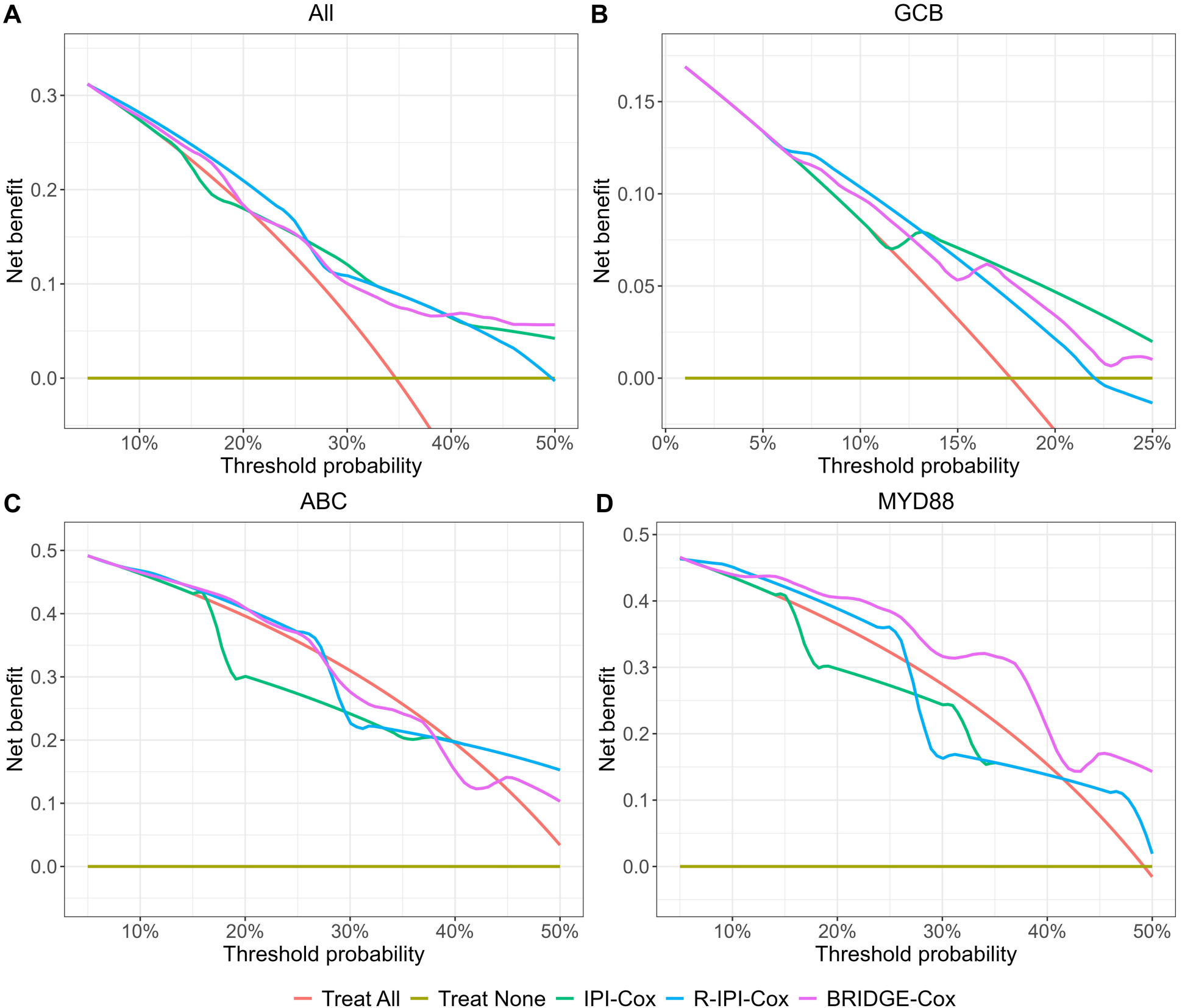

Predictors
Prognostic Models
Reddy dataset
Age
Ann Arbor stage
LDH
ECOG PS
Number of extranodal sites
SEER dataset
Age
Ann Arbor stage
Derive
IPI risk group (0/1 = low risk,
2 = low-intermediate risk,
3 = high-intermediate risk, and
4/5 = high risk)
R-IPI risk group (0 = very good risk,
1/2 = good risk, and
3/4/5 = poor risk)
IPI-Cox model
R-IPI-Cox model
BRIDGE-Cox model


Data
Reddy
SEER
1.00
0.75
0.50
0.25
0.00
Survival probability
p < 0.0001
0
2
4
6
8
10
12
Survival time (years)
Number at risk
Reddy
SEER
744
490
285
107
53
23
2
70281
39872
28537
19460
12232
6314
1749
0
2
4
6
8
10
12
Survival time (years)


A
IPI for Reddy
B
IPI for NCI
IPI Low Low-intermediate High-intermediate High
Survival probability
1.00
0.75
0.50
0.25
0.00
p < 0.0001
0 2 4 6 8 10 12
0 2 4 6 8 10 12 14 16
Survival time (years)
C
R-IPI for Reddy
D
R-IPI for NCI
R-IPI Very good Good Poor
E
BPI for Reddy
F
BPI for NCI
BPI Low Intermediate High


Model
IPI
R-IPI
BRIDGE
Mean
C-index
AUC at 2 years
AUC at 3 years
AUC at 5 years
Metric value
0.9
0.8
0.7
0.6
0.5
IPI
R-IPI
BRIDGE


Model
IPI-Cox
R-IPI-Cox
BRIDGE-Cox
C-index value
1.00
0.75
0.50
0.25
0.00
p = 0.104
p = 0.207
p = 0.015
p = 0.065
p = 0.837
p = 0.835
p = 0.676
p = 0.229
All
GCB
ABC
MYD88


A
All
B
GCB
C
ABC
D
MYD88
Net benefit
Threshold probability
0.3
0.2
0.1
0.0
0.15
0.10
0.05
0.00
0.5
0.4
10%
20%
30%
40%
50%
0%
5%
15%
25%
Treat All
Treat None
IPI-Cox
R-IPI-Cox
BRIDGE-Cox

Supplementary materials for "Integrating information across datasets to improve prognostic models for patients with diffuse large B-cell lymphoma"

Section S1. External validation data

An external independent data from the National Cancer Institute (NCI) was used for external validation for the predictive performance of the proposed and comparison models. **Table S1** presents patient characteristics and predictors included in the NCI and Reddy datasets. **Figure S1** shows the Kaplan-Meier estimates of overall survival in the NCI and Reddy datasets.

Table S1. Patient characteristics and predictors included in the National Cancer Institute (NCI) (n = 170) and Reddy (n = 744) datasets.

Characteristic	NCI (n = 170)	Reddy (n = 744)	*P*-value*
Age at diagnosis, years			
Median (IQR)	60 (49-70)	63.0 (52.9-71.9)	
≤40	19 (11.2%)	66 (8.9%)	0.181
41-60	68 (40.0%)	248 (33.3%)	
61-75	61 (35.9%)	302 (40.6%)	
>75	22 (12.9%)	128 (17.2%)	
Ann Arbor stage			0.050
I/II	81 (47.6%)	291 (39.1%)	
III/IV	89 (52.4%)	453 (60.9%)	
ECOG performance status			0.603
0-1	126 (74.1%)	534 (71.8%)	
2-4	44 (25.9%)	210 (28.2%)	
LDH			0.982
≤ upper limit of normal	76 (44.7%)	336 (45.2%)	
> upper limit of normal	94 (55.3%)	408 (54.8%)	
Number of extranodal sites			0.046
0 or 1	142 (83.5%)	566 (76.1%)	
>1	28 (16.5%)	178 (23.9%)	
IPI risk group			0.146
Low risk	69 (40.6%)	237 (31.9%)	
Low-intermediate risk	38 (22.4%)	175 (23.5%)	

Characteristic	NCI (n = 170)	Reddy (n = 744)	*P*-value*
High-intermediate risk	39 (22.9%)	192 (25.8%)	
High risk	24 (14.1%)	140 (18.8%)	
R-IPI risk group			0.147
Very good risk	24 (14.1%)	79 (10.6%)	
Good risk	83 (48.8%)	333 (44.8%)	
Poor risk	63 (37.1%)	332 (44.6%)	

* Categorical predictors were compared using the Chi-square test.

Abbreviations: IQR, interquartile range; ECOG, Eastern cooperative oncology group; LDH, serum lactate dehydrogenase; IPI, International Prognostic Index; R-IPI, Revised International Prognostic Index.

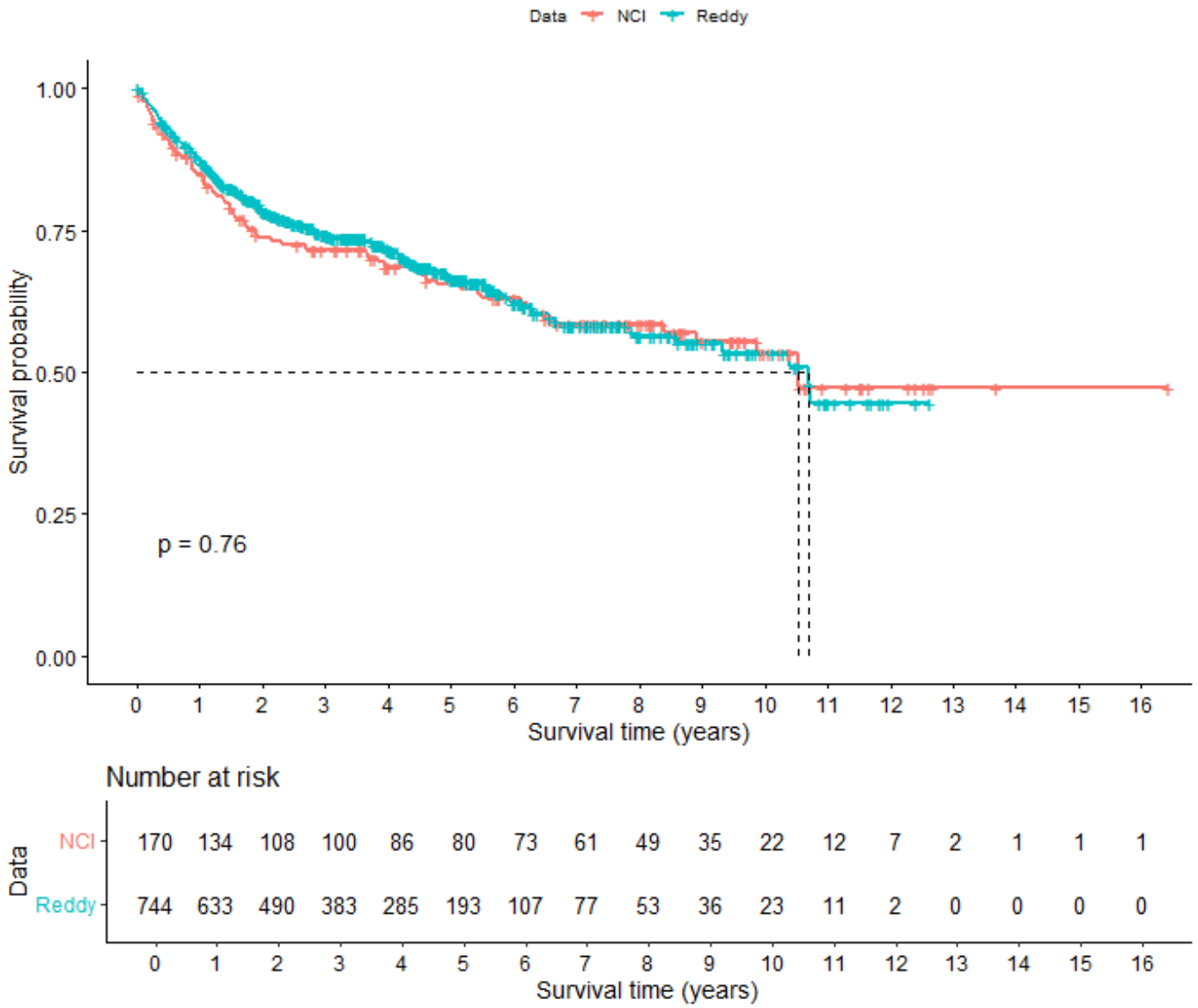


Figure S1. Kaplan-Meier estimates of overall survival in the National Cancer Institute (NCI, n = 170) and Reddy (n = 744) datasets.

Section S2. Prognostic index derived from the proposed prognostic model

The proposed prognostic model was a Cox-model-based boosted regression with individual data for generalization and enhancement, named the BRIDGE-Cox model. We defined the BRIDGE-model-based prognostic index (BPI) by categorizing patients into three prognostic risk groups, low, intermediate, and high. The three risk groups were based on the 10th and 40th percentiles of predicted 5-year OS in the Reddy dataset. Specifically, using the BRIDGE-Cox model trained in the Reddy dataset, there predicted 5-year OS cutoffs of 64.9% and 37.5%, corresponding to the 40th and 10th percentiles, respectively, were used to define the BPI groups. We compared the proposed prognostic index with the International Prognostic Index (IPI) and the revised International Prognostic Index (R-IPI). **Table S2** presents the predicted 5-year overall survival (OS) probabilities in the Reddy and NCI datasets stratified by the IPI, R-IPI, and three-level BPI risk groups.

Table S2. Five-year overall survival (OS) probabilities in the Reddy and National Cancer Institute (NCI) datasets for the risk groups from the three prognostic indices: International Prognostic Index (IPI); revised-IPI (R-IPI); the prognostic index defined from the proposed boosted regression with individual data for generalization and enhancement model (BPI) based on the corresponding trained prognostic models using the Reddy dataset with three levels.

IPI		**R-IPI**		**BPI for Reddy**		**BPI for NCI**	
Low	84.2%	Very good	93.2%	Low	80.1%	Low	81.6%
Low-intermediate	68.3%	Good	73.8%	Intermediate	55.6%	Intermediate	57.1%
High-intermediate	59.6%						
High	41.1%	Poor	51.8%	High	28.2%	High	29.6%

We further divided the previous low-risk group into low-intermediate- and low-risk groups, defined by predicted 5-year OS between the 40th and 75th percentiles and greater than the 75th percentile, respectively. Specifically, Using the BRIDGE-Cox model trained in the Reddy dataset, cutoffs at 82.0% (75th percentile), 64.9% (40th percentile) and 37.5% (10th percentile) defined the low (5-year OS > 82.0%), low-intermediate (64.9%-82.0%), high-intermediate (37.5%-64.9%), and high (≤ 37.5%) BPI groups. **Table S3** presents the predicted 5-year OS in the Reddy and NCI datasets stratified by the IPI, R-IPI, and four-level BPI risk groups. The 5-year OS probabilities in the low and low-intermediate risk groups derived from BPI are higher than the corresponding risk groups from the IPI. **Figure S2** shows that the Kaplan-Meier estimates of OS by BPI risk group in the Reddy and NCI dataset. The four-level BPI distinctly separates the risk groups of the Reddy and NCI cohorts.

Table S3. Five-year overall survival (OS) probabilities in the Reddy and National Cancer Institute (NCI) datasets for the risk groups from the three prognostic indices: International Prognostic Index (IPI); revised-IPI (R-IPI); the prognostic index defined from the proposed boosted regression with individual data for generalization and enhancement model (BPI) based on the corresponding trained prognostic models using the Reddy dataset with four levels.

IPI		R-IPI		BPI for Reddy		BPI for NCI	
Low	84.2%	Very good	93.2%	Low	88.4%	Low	88.9%
Low-intermediate	68.3%	Good	73.8%	Low-intermediate	76.1%	Low-intermediate	76.5%
High-intermediate	59.6%			High-intermediate	55.6%	High-intermediate	57.1%
High	41.1%	Poor	51.8%	High	28.2%	High	29.6%

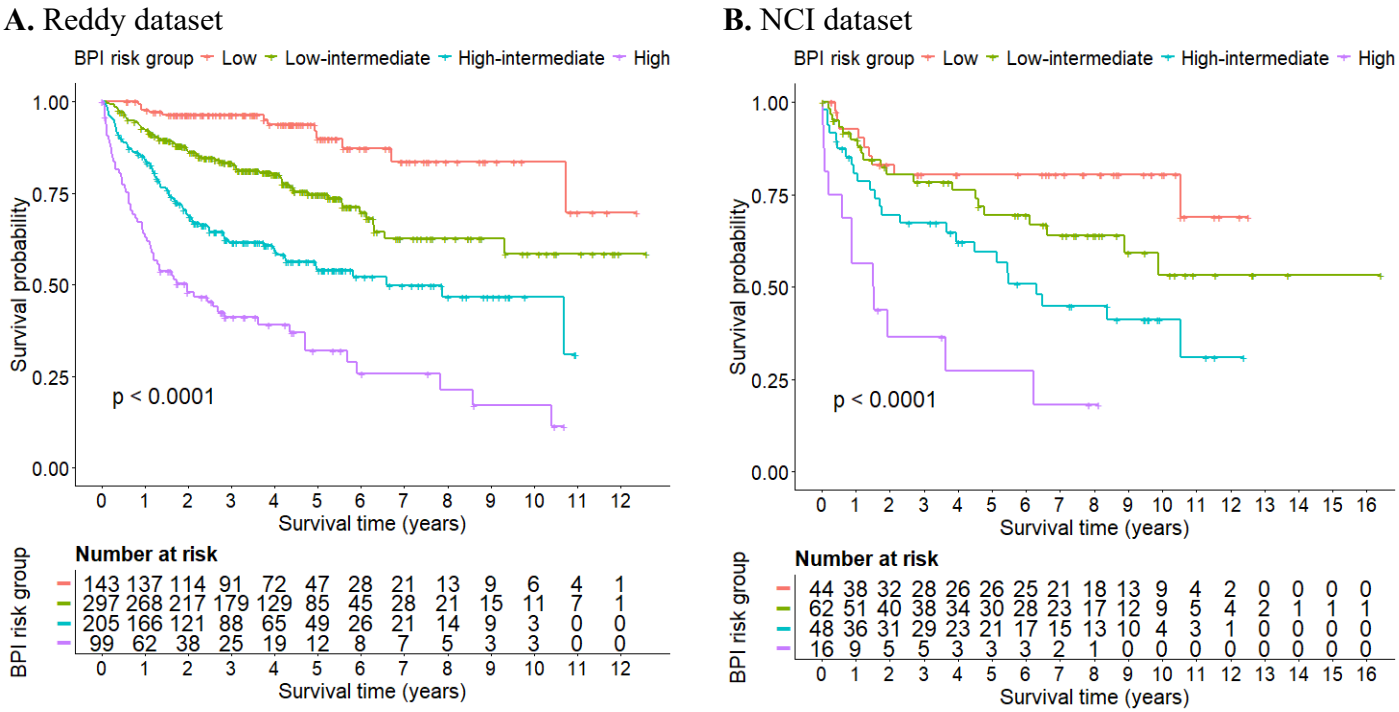


Figure S2. Kaplan-Meier estimates of overall survival by four-level BPI risk group in the Reddy dataset (n = 744) and National Cancer Institute (NCI) dataset (n = 170). BPI is the prognostic index defined from the proposed boosted regression with individual data for generalization and enhancement model.

Section S3. Additional predictive performance results in the Reddy dataset

Table S4 presents the Brier scores evaluated for 2-, 3-, and 5-year OS probabilities predicted by the three prognostic models in the Reddy dataset. The comparison models were the Cox models with the IPI and R-IPI risk groups as predictors separately, named IPI- and R-IPI-Cox models, respectively. The proposed prognostic model was the BRIDGE-Cox model.

Table S4. Average values of Brier score (smaller is better) with interquartile ranges from the repeated five-fold cross-validation (100 repetitions) evaluated for 2-, 3-, and 5-year overall survival probabilities predicted by the three prognostic models in the Reddy dataset (n = 744).

	At 2 years	At 3 years	At 5 years
IPI-Cox model	0.157 (0.144, 0.168)	0.172 (0.161, 0.182)	0.206 (0.194, 0.217)
R-IPI-Cox model	0.159 (0.145, 0.169)	0.175 (0.164, 0.185)	0.207 (0.196, 0.217)
BRIDGE-Cox model	**0.149** (0.137, 0.160)	**0.163** (0.151, 0.174)	**0.191** (0.177, 0.205)

Models: (1) IPI-Cox model, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox model, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox model, a Cox-model-based boosted regression with individual data for generalization and enhancement.

Two age variants were considered in the BRIDGE-Cox model: $\leq, >$ 60 years from the IPI and ≤40, 41-60, 61-75, >75 years from the enhanced IPI using data from National Comprehensive Cancer Network (NCCN-IPI). The corresponding BRIDGE models were named the BRIDGE-IPI- and BRIDGE-Cox models, respectively. Compared to the predictive performance of the BRIDGE-Cox model shown in **Figure 4** and **Table S4**, the BRIDGE-IPI-Cox model had worse values of C-index, AUC, and Brier score evaluated for the 2-, 3-, and 5-year OS shown in **Table S5**. Therefore, we selected the BRIDGE-Cox model with age composing of four categorical levels (≤40, 41-60, 61-75, >75 years) as our proposed model presenting in the manuscript.

Table S5. Predictive performance of the Cox-model-based boosted regression with individual data for generalization and enhancement including age as a binary predictor (≤, > 60 years), named the BRIDGE-IPI-Cox, in the Reddy dataset (n = 744). The evaluation used the repeated five-fold cross-validation (CV, 100 repetitions). The metrics included the concordance index (C-index; larger is better) for overall survival (OS) across all time points and the time-dependent area under the receiver operating characteristics curve (AUC; larger is better) and Brier score (BS; smaller is better) for 2-, 3-, and 5-year OS. Results are reported as the mean values of metrics with interquartile ranges across 500 CV repetitions.

C-index	AUC			Brier score		
Overall	At 2 years	At 3 years	At 5 years	At 2 years	At 3 years	At 5 years
0.681 (0.651, 0.711)	0.721 (0.691, 0.752)	0.725 (0.696, 0.755)	0.699 (0.666, 0.732)	0.153 (0.140, 0.165)	0.169 (0.157, 0.181)	0.200 (0.187, 0.213)

We further compared the predictive performance of the three prognostic models for genomic subgroups from the Reddy dataset shown in **Table S6**.

Table S6. Predictive performance of the three prognostic models in the germinal center B cell–like (GCB), activated B cell–like (ABC), *MYD88*, *ATM*, *SETD2*, and *MGA* subgroups in the Reddy dataset (n = 744). The evaluation used the repeated five-fold cross-validation (CV, 100 repetitions). The metrics included the concordance index (C-index; larger is better) for overall survival (OS) across all time points and the time-dependent area under the receiver operating characteristics curve (AUC; larger is better) and Brier score (BS; smaller is better) for 2-, 3-, and 5-year OS. Results are reported as the mean values of metrics with interquartile ranges across 500 CV repetitions.

Subgroup	Model	C-index	AUC			BS		
		Overall	At 2 yrs	At 3 yrs	At 5 yrs	At 2 yrs	At 3 yrs	At 5 yrs
GCB (n = 252)	IPI-Cox	0.705 (0.648, 0.765)	0.687 (0.625, 0.757)	0.704 (0.646, 0.766)	0.699 (0.633, 0.768)	0.132 (0.109, 0.152)	0.155 (0.133, 0.172)	0.191 (0.167, 0.213)
	R-IPI-Cox	0.686 (0.63, 0.738)	0.678 (0.626, 0.731)	0.686 (0.634, 0.744)	0.694 (0.641, 0.753)	0.131 (0.11, 0.15)	0.155 (0.134, 0.173)	0.189 (0.166, 0.208)
	BRIDGE-Cox	**0.753** (0.707, 0.805)	**0.726** (0.678, 0.778)	**0.743** (0.696, 0.8)	**0.729** (0.676, 0.785)	**0.13** (0.108, 0.15)	**0.148** (0.127, 0.167)	**0.169** (0.146, 0.189)
ABC (n = 254)	IPI-Cox	0.599 (0.549, 0.653)	0.659 (0.606, 0.713)	0.676 (0.625, 0.728)	0.627 (0.577, 0.682)	0.188 (0.168, 0.206)	0.197 (0.179, 0.214)	0.23 (0.211, 0.248)
	R-IPI-Cox	0.559 (0.514, 0.605)	0.62 (0.576, 0.661)	0.641 (0.599, 0.685)	0.602 (0.553, 0.657)	0.193 (0.173, 0.21)	0.202 (0.184, 0.216)	0.232 (0.213, 0.247)
	BRIDGE-Cox	**0.615** (0.561, 0.666)	**0.706** (0.657, 0.759)	**0.712** (0.663, 0.759)	**0.679** (0.634, 0.729)	**0.183** (0.162, 0.202)	**0.194** (0.174, 0.212)	**0.218** (0.197, 0.237)

Subgroup	Model	C-index	AUC			BS		
		Overall	At 2 yrs	At 3 yrs	At 5 yrs	At 2 yrs	At 3 yrs	At 5 yrs
MYD88 (n = 139)	IPI-Cox	0.608 (0.531, 0.686)	0.701 (0.634, 0.786)	0.742 (0.673, 0.814)	0.715 (0.635, 0.813)	0.167 (0.139, 0.192)	**0.17** (0.146, 0.192)	0.202 (0.171, 0.228)
	R-IPI-Cox	0.573 (0.503, 0.641)	0.663 (0.601, 0.728)	0.688 (0.627, 0.752)	0.677 (0.6, 0.749)	0.17 (0.144, 0.193)	0.178 (0.156, 0.198)	0.209 (0.185, 0.231)
	BRIDGE-Cox	**0.661** (0.589, 0.741)	**0.745** (0.688, 0.812)	**0.764** (0.706, 0.83)	**0.743** (0.682, 0.809)	**0.164** (0.137, 0.191)	**0.17** (0.145, 0.192)	**0.196** (0.17, 0.219)
ATM (n=52)	IPI-Cox	0.801 (0.719, 0.907)	0.749 (0.628, 0.9)	**0.781** (0.688, 0.909)	0.81 (0.721, 0.929)	0.115 (0.072, 0.155)	0.14 (0.093, 0.18)	0.158 (0.116, 0.192)
	R-IPI-Cox	0.771 (0.7, 0.853)	0.726 (0.667, 0.812)	0.758 (0.696, 0.833)	0.791 (0.722, 0.867)	0.116 (0.073, 0.151)	0.142 (0.096, 0.181)	0.161 (0.121, 0.195)
	BRIDGE-Cox	**0.834** (0.732, 0.965)	**0.827** (0.699, 1)	**0.781** (0.647, 1)	**0.821** (0.701, 1)	**0.09** (0.048, 0.126)	**0.119** (0.071, 0.157)	**0.127** (0.083, 0.166)
SETD2 (n=46)	IPI-Cox	0.722 (0.611, 0.875)	0.834 (0.75, 0.929)	0.863 (0.797, 0.958)	0.787 (0.693, 0.929)	0.137 (0.097, 0.169)	0.142 (0.107, 0.17)	0.173 (0.132, 0.202)
	R-IPI-Cox	0.684 (0.595, 0.813)	0.787 (0.722, 0.875)	0.805 (0.75, 0.889)	0.757 (0.667, 0.875)	0.142 (0.1, 0.174)	0.152 (0.115, 0.176)	0.178 (0.144, 0.206)
	BRIDGE-Cox	**0.821** (0.724, 0.967)	**0.958** (0.938, 1)	**0.964** (0.944, 1)	**0.903** (0.857, 1)	**0.09** (0.064, 0.111)	**0.092** (0.066, 0.112)	**0.121** (0.084, 0.152)
MGA (n=62)	IPI-Cox	0.746 (0.665, 0.828)	0.861 (0.8, 0.944)	0.861 (0.8, 0.944)	0.785 (0.681, 0.906)	0.157 (0.123, 0.19)	0.154 (0.127, 0.179)	0.19 (0.148, 0.223)
	R-IPI-Cox	0.721 (0.646, 0.803)	0.81 (0.75, 0.889)	0.81 (0.75, 0.889)	0.774 (0.7, 0.875)	0.177 (0.144, 0.206)	0.173 (0.149, 0.197)	0.193 (0.167, 0.218)
	BRIDGE-Cox	**0.809** (0.745, 0.89)	**0.902** (0.852, 0.979)	**0.902** (0.852, 0.979)	**0.872** (0.812, 0.954)	**0.153** (0.118, 0.189)	**0.148** (0.116, 0.177)	**0.164** (0.126, 0.195)

Models: (1) IPI-Cox, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox, a Cox-model-based boosted regression with individual data for generalization and enhancement.

Section S4. Estimation results of the prognostic models in the Reddy dataset

Table S7 presents the estimation results of the IPI-, R-IPI-, and BRIDGE-Cox models fitting in the Reddy dataset.

Table S7. Coefficients, hazard ratios with 95% confidence intervals (CIs), and *p*-values of predictors in the three prognostic models trained on the Reddy dataset (n = 744).

Model	Coefficient	Hazard ratio (95% CI)	*P*-value
IPI-Cox model			
IPI risk group (Low risk as ref)			
Low-intermediate risk	0.796	2.218 (1.439, 3.417)	< 0.001
High-intermediate risk	1.100	3.005 (2.002, 4.510)	< 0.001
High risk	1.642	5.165 (3.450, 7.735)	< 0.001
R-IPI-Cox model			
R-IPI risk group (Very good risk as ref)			
Good risk	1.458	4.297 (1.741, 10.610)	0.002
Poor risk	2.231	9.305 (3.814, 22.700)	< 0.001
BRIDGE-Cox model			
Step 1: Trans-Cox model			
Age at diagnosis (years, 61-75 as ref)			
≤40	-1.289	0.276 (0.220, 0.345)	< 0.001
41-60	-0.445	0.641 (0.562, 0.731)	< 0.001
>75	0.717	2.049 (1.830, 2.295)	< 0.001
Ann Arbor Stage (I/II as ref)			
III/IV	0.459	1.583 (1.476, 1.698)	< 0.001
Step 2: Cox model with offsets			
ECOG PS (0-1 as ref)			
2-4	0.819	2.269 (1.726, 2.981)	< 0.001
LDH (≤ ULN as ref)			
> ULN	0.321	1.379 (1.031, 1.845)	0.031
Number of extranodal sites (0 or 1 as ref)			
>1	0.067	1.069 (0.798, 1.432)	0.655

Models: (1) IPI-Cox model, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox model, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox model, a Cox-model-based boosted regression with individual data for generalization and enhancement.
Abbreviations: ECOG PS, Eastern Cooperative Oncology Group performance status; LDH, serum lactate dehydrogenase; ULN, upper limit of normal.

The IPI-, R-IPI-, BRIDGE-Cox model estimation results trained on the Reddy and used for the survival prediction of COO, *MYD88, ATM, SETD2,* and *MGA* subgroups in the NCI dataset are presented in **Tables S8-S10**, respectively. The IPI- and R-IPI-Cox models are extended by adding the subgroups of interest as covariates in the BRIDGE-Cox model. For example, the COO covariate was the covariate in the third-step Cox model for the patients with available COO information (n = 610) for the GCB and ABC subgroup analysis. For the *MYD88* subgroup analysis, the third-step Cox model included the *MYD88* mutation status as the covariate.

These BRIDGE-Cox models for the subgroup analyses shared the same estimation results from the first two steps. The inference of the parameter estimates was obtained from the corresponding local steps only. The covariance estimation of the parameter estimates merits future research, which could be conducted by a perturbation scheme.

Table S8. Estimation results of a Cox model with the International Prognostic Index (IPI-Cox) risk group as the predictor in the Reddy dataset (n = 744) used for the survival prediction of cell of origin (COO), *MYD88, ATM, SETD2,* and *MGA* subgroups.

IPI-Cox Model	**Coefficient**	**Hazard ratio** (95% CI)	***P*-value**
With COO as a covariate (n=610)			
IPI risk group (Low risk as ref)			
Low-intermediate risk	0.862	2.369 (1.455, 3.859)	< 0.001
High-intermediate risk	1.054	2.868 (1.805, 4.557)	< 0.001
High risk	1.658	5.248 (3.323, 8.288)	< 0.001
COO (ABC as ref)			
GCB	-0.353	0.703 (0.506, 0.977)	0.038
Unclassified	0.0007	1.001 (0.672, 1.491)	0.997
With *MYD88* as a covariate (n=744)			
IPI risk group (Low risk as ref)			
Low-intermediate risk	0.794	2.213 (1.436, 3.410)	< 0.001
High-intermediate risk	1.097	2.996 (1.995, 4.499)	< 0.001
High risk	1.640	5.153 (3.440, 7.719)	< 0.001
MYD88 (No as ref)			
Yes	0.054	1.055 (0.765, 1.455)	0.743
With *ATM* as a covariate (n=744)			
IPI risk group (Low risk as ref)			
Low-intermediate risk	0.797	2.219 (1.440, 3.420)	< 0.001
High-intermediate risk	1.101	3.007 (2.003, 4.513)	< 0.001
High risk	1.648	5.198 (3.471, 7.783)	< 0.001
ATM (No as ref)			
Yes	-0.488	0.614 (0.343, 1.098)	0.100
With *SETD2* as a covariate (n=744)			
IPI risk group (Low risk as ref)			
Low-intermediate risk	0.799	2.223 (1.443, 3.427)	< 0.001
High-intermediate risk	1.100	3.005 (2.002, 4.511)	< 0.001
High risk	1.643	5.169 (3.452, 7.740)	< 0.001
SETD2 (No as ref)			
Yes	-0.096	0.908 (0.529, 1.562)	0.728
With *MGA* as a covariate (n=744)			
IPI risk group (Low risk as ref)			
Low-intermediate risk	0.831	2.296 (1.487, 3.545)	< 0.001
High-intermediate risk	1.113	3.044 (2.027, 4.570)	< 0.001
High risk	1.677	5.347 (3.563, 8.024)	< 0.001
MGA (No as ref)			
Yes	0.410	1.507 (0.983, 2.310)	0.060

Table S9. Estimation results of a Cox model with the revised International Prognostic Index (R-IPI-Cox) risk group as the predictor in the Reddy dataset (n = 744) used for the survival prediction of cell of origin (COO), *MYD88, ATM, SETD2,* and *MGA* subgroups.

R-IPI-Cox model	**Coefficient**	**Hazard ratio** (95% CI)	***P*-value**
With COO as a covariate (n=610)			
R-IPI risk group (Very good risk as ref)			
Good risk	1.260	3.525 (1.416, 8.778)	0.002
Poor risk	1.988	7.304 (2.973, 17.943)	< 0.001
COO (ABC as ref)			
GCB	-0.343	0.710 (0.511, 0.987)	0.041
Unclassified	0.018	1.019 (0.684, 1.517)	0.928
With *MYD88* as a covariate (n=744)			
R-IPI risk group (Very good risk as ref)			
Good risk	1.458	4.296 (1.740, 10.603)	0.002
Poor risk	2.228	9.283 (3.804, 22.651)	< 0.001
MYD88 (Yes as ref)			
No	0.067	1.069 (0.775, 1.474)	0.683
With *ATM* as a covariate (n=744)			
R-IPI risk group (Very good risk as ref)			
Good risk	1.446	4.247 (1.721, 10.485)	0.002
Poor risk	2.224	9.243 (3.788, 22.553)	< 0.001
ATM (Yes as ref)			
No	-0.447	0.640 (0.358, 1.145)	0.133
With *SETD2* as a covariate (n=744)			
R-IPI risk group (Very good risk as ref)			
Good risk	1.460	4.305 (1.744, 10.628)	0.001
Poor risk	2.232	9.315 (3.818, 22.728)	< 0.001
SETD2 (Yes as ref)			
No	-0.080	0.924 (0.537, 1.587)	0.773
With *MGA* as a covariate (n=744)			
R-IPI risk group (Very good risk as ref)			
Good risk	1.448	4.241 (1.717, 10.471)	0.002
Poor risk	2.224	9.247 (3.790, 22.563)	< 0.001
MGA (Yes as ref)			
No	0.242	1.274 (0.834, 1.948)	0.263

Table S10. Estimation results of the Cox-model-based boosted regression with individual data for generalization and enhancement (BRIDGE-Cox) in the Reddy dataset (n = 744) used for the survival prediction of cell of origin (COO), *MYD88, ATM, SETD2,* and *MGA* subgroups.

BRIDGE-Cox model	**Coefficient**	**Hazard ratio** (95% CI)	***P*-value**
Step 1: Trans-Cox model (n=744)			
Age at diagnosis (years, 61-75 as ref)			
≤40	-1.289	0.276 (0.220, 0.345)	< 0.001
41-60	-0.445	0.641 (0.562, 0.731)	< 0.001
>75	0.717	2.049 (1.830, 2.295)	< 0.001
Ann Arbor Stage (I/II as ref)			
III/IV	0.459	1.583 (1.476, 1.698)	< 0.001
Step 2: Cox model with offsets (n=744)			
ECOG PS (0-1 as ref)			
2-4	0.819	2.269 (1.726, 2.981)	< 0.001
LDH (≤ ULN as ref)			
> ULN	0.321	1.379 (1.031, 1.845)	0.031
Number of extranodal sites (0 or 1 as ref)			
>1	0.067	1.069 (0.798, 1.432)	0.655
Step 3: Add a covariate with offsets			
(1) With COO as a covariate (n = 610)			
Cell of origin (ABC as ref)			
GCB	-0.277	0.758 (0.547, 1.051)	0.097
Unclassified	0.098	1.102 (0.741, 0.1641)	0.631
(2) With MYD88 as a covariate (n = 744)			
MYD88 (No as ref)			
Yes	0.031	1.032 (0.749, 1.422)	0.849
(3) With ATM as a covariate (n = 744)			
ATM (No as ref)			
Yes	-0.437	0.646 (0.361, 1.155)	0.141
(4) With SETD2 as a covariate (n = 744)			
SETD2 (No as ref)			
Yes	0.002	1.002 (0.583, 1.721)	0.996
(5) With MGA as a covariate (n = 744)			
MGA (No as ref)			
Yes	0.332	1.393 (0.912, 2.129)	0.125

Abbreviations: ECOG PS, Eastern Cooperative Oncology Group performance status; LDH, serum lactate dehydrogenase; ULN, upper limit of normal.

Section S5. Additional predictive performance results in the NCI dataset

Table S11 presents the predictive performance evaluated by the Brier score on the predicted 2-, 3-, and 5-year OS for all patients and the GCB, ABC, and *MYD88* subgroups from the NCI dataset.

Table S11. External validation of the three prognostic models using the Brier score (BS; smaller is better) for 2-, 3-, and 5-year OS evaluated in all patients and in the GCB, ABC, and *MYD88* subgroups of the National Cancer Institute dataset (n = 170). The BS values are shown with 95% confidence intervals.

Patient	Model	BS		
		At 2 years	At 3 years	At 5 years
All (n=170)	IPI-Cox	0.178 (0.139, 0.216)	0.187 (0.149, 0.225)	0.207 (0.177, 0.241)
	R-IPI-Cox	0.185 (0.148, 0.224)	0.192 (0.156, 0.228)	0.209 (0.180, 0.241)
	BRIDGE-Cox	**0.176** (0.136, 0.221)	**0.182** (0.144, 0.227)	**0.205** (0.169, 0.247)
GCB (n = 84)	IPI-Cox	0.099 (0.057, 0.146)	0.121 (0.078, 0.173)	0.146 (0.103, 0.197)
	R-IPI-Cox	0.104 (0.062, 0.154)	0.123 (0.082, 0.173)	0.153 (0.11, 0.202)
	BRIDGE-Cox	**0.093** (0.055, 0.14)	**0.108** (0.073, 0.161)	**0.127** (0.093, 0.186)
ABC (n = 60)	IPI-Cox	0.268 (0.199, 0.335)	0.267 (0.199, 0.334)	0.269 (0.213, 0.324)
	R-IPI-Cox	**0.267** (0.202, 0.33)	**0.263** (0.208, 0.321)	**0.252** (0.208, 0.3)
	BRIDGE-Cox	0.282 (0.201, 0.358)	0.284 (0.206, 0.358)	0.282 (0.215, 0.347)
MYD88 (n = 37)	IPI-Cox	0.246 (0.164, 0.333)	0.236 (0.163, 0.318)	0.262 (0.19, 0.33)
	R-IPI-Cox	0.257 (0.173, 0.346)	0.248 (0.175, 0.325)	0.261 (0.201, 0.327)
	BRIDGE-Cox	**0.236** (0.145, 0.325)	**0.227** (0.14, 0.311)	**0.248** (0.178, 0.334)

Models: (1) IPI-Cox, a Cox model with the International Prognostic Index (IPI) risk group as the predictor; (2) R-IPI-Cox, a Cox model with the revised-IPI risk group as the predictor; (3) BRIDGE-Cox, a Cox-model-based boosted regression with individual data for generalization and enhancement.