EBioMedicine 62 (2020) 103126

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Development and validation of a prognostic model for adult patients with acute myeloid leukaemia

Ting-Ting Ma[1], Xiao-Jing Lin[1], Wen-Yan Cheng[1], Qing Xue, Shi-Yang Wang, Fu-Jia Liu, Han Yan, Yong-Mei Zhu, Yang Shen*

Shanghai Institute of Haematology, Department of Haematology, State Key Laboratory of Medical Genomics, National Research Centre for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, No.197 Ruijin Er Road, Shanghai, China

ARTICLE INFO

Article History:
Received 29 July 2020
Revised 28 October 2020
Accepted 28 October 2020
Available online xxx

Keywords:
Acute myeloid leukaemia
Prognostic model
Nomogram
Risk stratification

ABSTRACT

Background: The high heterogeneity of acute myeloid leukaemia (AML) reflected in the patient- and disease-related factors accounts for the unsatisfactory prognosis despite the introduction of novel therapeutic approaches and drugs in recent years.
Methods: In the development set (n = 412), parameters including age, hematopoietic cell transplantation-comorbidity index, white blood cell count, hemoglobin, biallelic *CEBPA* mutations, *DNMT3A* mutations, *FLT3*-ITD/*NPM1* status, and ELN cytogenetic risk status were identified as independent prognostic factors for overall survival (OS) in the multivariable Cox regression analysis. A nomogram combining these predictors for individual risk estimation was established thereby.
Findings: The prognostic model demonstrated promising performance in the development cohort. The calibration plot, C-index (0.74), along with the 1-, 2- and 3-year area under the receiver operating characteristic curve (AUC, 0.76, 0.79, and 0.74, respectively) in the validation set (n = 238) substantiated the robustness of the model. In addition to stratifying young (age $\leq$ 60 years) and elderly patients (age $>$ 60 years) into three and two risk groups with significant distinct outcomes, the prognostic model succeeded in distinguishing eligible candidates for hematopoietic stem cell transplantation.
Interpretation: The prognostic model is capable of survival prediction, risk stratification and helping with therapeutic decision-making with the use of easily acquired variables in daily clinical routine.
Funding: This work was supported in part by grants from the National Natural Science Foundation of China (81770141), the National Key R&D Program of China (2016YFE0202800), and Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant Support (20161406).

1. Introduction

Great efforts and progress have been achieved in the field of the treatment of acute myeloid leukaemia (AML) in the last decades, including the curable modality such as allogeneic hematopoietic stem cell transplantation (HSCT), several novel agents as inhibitors targeting *BCL-2, IDH1/2* and *FLT3*, antibody-drug conjugates, and hypomethylation agents, leading to a significant improvement of survival in this disease. Eight new drugs with promising effect on improving response rates and outcomes of AML patients were approved by the United States Federal Drug Administration (FDA) between 2017 and 2019 [1]. However, the prognosis of elderly AML patients remains dismal, with a long-term survival of less than 15% [2]. The high heterogeneity of AML also brings challenges to the treatment of AML, which includes not only the disease-related characteristics, such as cytogenetic and molecular abnormalities, but also individual factors including physical conditions and comorbidities. Therefore, a prognostic model incorporating these well-recognized factors is essential for precise risk stratification before treatment, and more importantly, for the implementation of therapeutic decision-making in clinical application [3]. So far, several models have been reported, as exemplified by the $PINA_{OS}$ score and $PINA_{RFS}$ score in cytogenetically normal AML (CN-AML) proposed by Pastore et al. [4], the scoring model designed for elderly AML by Djunic et al. [5], as well as the reclassification of elderly patients with intermediate-risk karyotype by Rollig et al. [6]. However, these models had certain limitations since they were only applicable to a specific subgroup of AML patients. Moreover, with the exception of the $PINA_{OS}$ and $PINA_{RFS}$ score, most of the other models reported in the literature lacked an

* Corresponding author.
E-mail address: shen_yang@126.com (Y. Shen).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ebiom.2020.103126

Research in context

Evidence before this study

We searched the PubMed on Oct. 4, 2020 according to the terms "[("*prediction*" OR "*risk prediction*" OR "*prediction model*" OR "*predictive model*" OR "*prognostic model*" OR "*model*") AND ("*acute myeloid leukaemia*" OR "*acute myeloid leukemia*" OR "AML")]" among English-language articles to identify papers aiming to propose a comprehensive model at diagnosis to predict the outcome of patients with AML. Previous studies focused on a specific age or karyotype subgroup of AML patients. None of these studies reported discrimination and calibration of the established model during the model development and validation procedure.

Added value of this study

A total of 810 patients with AML from three different clinical centres were enrolled in this study. A nomogram incorporating eight prognostic predictors was established in the development set through the multivariable Cox regression analysis with backward elimination. The area under a time-dependent receiver operating characteristic (ROC) curve (AUC), C-Index, along with the bias-corrected calibration plot of the early prediction model in the validation set showed similar promising performance as in the development set in respect to discrimination and calibration. The AML early prediction model demonstrated good performance not only in risk stratification, but also in facilitating therapeutic decisions.

Implication of all the available evidence

The predictive model incorporating the AML- and patient-related parameters was constructed and validated. The performance of the novel model was not inferior to previously reported models based on the next-generation sequencing, potentially indicating the indispensable role of the common and conveniently acquired variables including comorbidities of AML patients and routine laboratory indicators. Moreover, some easily neglected prognostic factors of AML, such as cognition, psychological state, polypharmacy, family and social support, and nutritional status, deserve to be considered in future prognostic models. We hope this study will provide a new orientation in the field of model construction of AML.

independent validation, which is indispensable for a well-established model, resulting in an insufficient predictive performance and robustness [7]. With the development of new technologies in the field of molecular biology, factors derived from high-throughput sequencing, including gene expression profiling based on microarray or RNA sequencing [8-11], and somatic mutation profiling detected by whole genome or exome sequencing have been incorporated into prognostic systems [12, 13]. Unfortunately, the C-index of these models, which reflects the discrimination ability, was usually less than 0.65 in an independent validation cohort, representing a "poor" discrimination [14]. In addition, most of these models encompassed a large number of genes, usually from dozens to thousands, that were associated with prognostic significance, which makes them difficult to be widely applied in the clinical setting. In this regard, an ideal prognostic model that might be well recognized should combine the value of accuracy, applicability, and generality.

Thus, we integrated different types of prognostic parameters, including the well-established demographic and baseline clinical characteristics, cytogenetic and molecular variables, and other recently reported prognostic factors, to develop and validate a new model predicting the overall survival (OS) and helping with the therapeutic decision-making in adult AML patients of all ages.

2. Methods

2.1. Participants and source of data

This study enrolled a total of 801 newly diagnosed non-M3 adult AML patients who were consecutively treated in three different clinical centres, including 687 patients from Ruijin Hospital (RJH) between June 2011 and August 2018, 64 patients from Ruijin Hospital North (RJHN) between January 2017 and December 2018, and 50 patients from Beizhan Hospital (BZH) between November 2013 and November 2017.

After 2015, most of the patients enrolled in this study from RJH participated in one of the three phase Ⅱ and III clinical trials, which were registered at the Chinese Clinical Trial Registry (www.chictr.org.cn Identifier: ChiCTR-OPC-15006085; ChiCTR-OIC-16007764; ChiCTR-OIN-16008955). The diagnosis and subtype classification of AML were based on the 2016 World Health Organization criteria. Patients who refused chemotherapy, received palliative treatment only, or dead before or during initial induction therapy were excluded. The follow-up of all patients ended in December 2019. This study was approved by the ethics committee of the three participating hospitals (KY-2016-2). Informed consent was obtained from all patients for treatment and cryopreservation of bone marrow and peripheral blood samples according to the Declaration of Helsinki.

2.2. Treatment protocols

Generally, young patients (age $\leq$ 60 years) were treated with the standard first-line "3 + 7″ IA/DA like induction regimens, which consisted of idarubicin/daunorubicin (10/45 mg/m^2, D1−3) and cytarabine (100 mg/m^2, D1−7). In patients who entered clinical trials, additional intervention with homoharringtonine was tried in those with D5 peripheral blast clearance rate (D5-PBCR) of less than 99.55% (see Supplementary Material). After achieving CR, four cycles of high-dose cytarabine (2 g/m^2) were given as consolidation.

Elderly patients (age $>$ 60 years) were evaluated by the treating physician and classified as "fit" or "unfit" in consideration of their physical conditions and disease risks. Fit patients received the reduced "3 + 7″-based induction regimens (idarubicin 6−10 mg/m^2 D1−3; cytarabine 100 mg/m^2, D1−7), and reduced cycles of consolidation therapy to 2 cycles of high-dose cytarabine (2 g/m^2). While the unfit patients were assigned to other less intensive therapies such as demethylation agents at the discretion of the physician. More details concerning the treatment protocols are provided in the Supplementary Materials.

2.3. Outcome

OS was the primary outcome of interest, which was measured from the date of diagnosis to death from any cause. Patients who were still alive were censored for OS at the time of last follow-up, and patients who received HSCT were censored at the beginning of transplantation. As the secondary endpoint, disease-free survival (DFS) was measured from the date of diagnosis to relapse or death from any cause, whichever came first.

2.4. Predictors

Well-established prognostic parameters and some novel factors based on recent research, which were independent predictors of OS and convenient to acquire in clinical practice, were included in this study. Parameters at diagnosis that we analysed in the model

development process included age, sex, hematopoietic cell transplantation-comorbidity index (HCT-CI) [15], Barthel Index [16], white blood cell count (WBC), hemoglobin (Hb), platelet (PLT), lactate dehydrogenase (LDH), fibrinogen (Fg), type of AML (de novo-AML vs secondary-AML), bone marrow blasts, ELN cytogenetic risk status (Supplementary Materials) [3], gene mutations and fusions including *FLT3*-ITD/TKD, *KMT2A*-PTD, *NPM1*, *NRAS*, *KIT*, *CEBPA*, *DNMT3A*, *RUNX1-RUNX1T1*, *CBFB-MYH11*, and *KMT2A* rearrangements, as well as CD34/CD38 expression status of leukemic cells. Bone marrow aspirate samples were obtained from each patient at diagnosis and were detected for morphology, immunophenotype, cytogenetics and molecular biology (MICM) following the standard operating procedures of our institution, as previously reported [17, 18].

2.5. Missing data

Missing data were assumed to occur at random and predicted by using multiple imputations on the basis of the correlations with other observed variables through mice (version 3.5.0) R package. For both training and validation datasets, Fg, PLT and LDH were imputed. Twenty different imputed datasets were created with identical non-missing information but different imputed values reflecting the uncertainty associated with the imputations. In the development dataset, patients were excluded if they lacked information on the following key predictors: age, cytogenetic or molecular data. While in the validation dataset, patients were not included if they had missing information on any predictor in the prognostic model. Details regarding the multiple imputations of missing variables were described in the Supplementary Materials.

2.6. Statistical analysis methods

All the continuous parameters, HCT-CI and Barthel Index were grouped in accordance with the generally accepted criteria in clinical practice or reports in prior literature. Univariable Cox analysis for OS was applied for each of the aforementioned parameters separately. Prognostic indicators with a P-value of less than 0.10 were entered into backward elimination for model selection. Based on the Akaike information criterion (AIC), we included all independent prognostic indicators in the final model to construct a nomogram, along with a free web-based tool used for precise risk calculation (http://121.199.26.137:3838), which was made by shiny (version 1.4.0.2) R package. The estimated risk was calculated as the weighted linear sum of all predictors in the model, where the weight of each factor was its regression coefficient in the multivariate analysis.

Model performance was evaluated through calibration and discrimination. Bias-corrected calibration for 1-year, 2-year, and 3-year OS rate was performed using 1000 bootstrap resamples to assess the consistency between the observed and estimated survival probabilities by rms (version 5.1-4) R package. Discrimination was assessed by Harrell's concordance index (C-index) adjusted through 1000 bootstrap resamples under boot (version 1.3-24) R package, as well as the area under the curve (AUC) of a time-dependent receiver operating characteristic (ROC) curve by pROC (version 1.13.0) R package. The AUC of different ROC curves was compared using the Venkatraman method [19].

Categorical variables were compared by Fisher's exact test, and continuous data by Wilcoxon rank sum test. The log-rank test was used to compare the difference of OS and DFS distribution. All survival analyses were performed using the survival (version 2.42-6) R package, and survival curves were visualized using survminer (version 0.4.3) R package. All statistical analyses were performed using the R 3.6.0 (The CRAN project, www.r-project.org) software package.

This study was reported in compliance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) standard guidelines for construction and validation of the prognostic model [20].

2.7. Role of funding source

The funders had no role in study design, data collecting, analysis and interpretation, decision to publish, or writing of the manuscript. YS had full access to all the data in the study and had final responsibility for the decision to submit for publication.

3. Results

3.1. Population characteristics

The participant flow diagram is depicted in Supplementary Figure S1. The baseline clinical characteristics of the training set (n = 412) and validation set (n = 238) are summarized and compared in Supplementary Table S1, and the median and range of all continuous variables and HCT-CI are shown in Supplementary Table S2, which indicate the scope of model application. Overall, patients in the two sets shared similar clinical characteristics and prognosis. The 1-year, 2-year, and 3-year OS rates were 76.1% (95% CI, 71.6–80.9), 59.5% (95% CI, 53.6–66.1), 51.4% (95% CI, 44.9–58.8) in the training set, and 76.3% (95% CI, 70.5–82.7), 62.1% (95% CI, 54.9–70.2), and 51.8% (95% CI, 43.6–61.6) in the validation set (log-rank p = 0.646, 0.617 and 0.733, respectively), with a median follow-up of 10.9 (range 0.4–77.0) months and 11.4 (range 0.5–77.0) months, respectively.

3.2. Model development and internal validation

Univariable Cox analysis was applied to evaluate the association of each covariate with OS of patients in the development cohort, as shown in Table 1. Through backward selection procedure based on the AIC (1313.45) in the multivariable modeling, eight prognostic factors including age, HCT-CI, WBC, Hb, biallelic *CEBPA* mutations, *DNMT3A* mutations, *FLT3*-ITD/*NPM1* status, and ELN cytogenetic risk status were incorporated in the final AML early prediction model, which was presented graphically as a nomogram (Fig. 1). The points in the nomogram were assigned in accordance with the rank order of the effect estimates [21]. Molecular biology and ELN cytogenetic risk status were important influential predictors of OS, in which *FLT3*-ITD/*NPM1* status was demonstrated as the most crucial one.

The calibration plot for internal validation showed an excellent agreement of 1-, 2-, and 3-year survival probabilities between the estimated outcomes by nomogram and actual observations (Fig. 2A). The C-index was 0.736 (95% CI, 0.693–0.780) in the development cohort, and the optimism-corrected C statistic with 1000 bootstrap replications was 0.76. Time-dependent ROC was used to assess the predictive performance of the model, which showed good discrimination and accurate overall survival prediction, with a 1-, 2- and 3-year AUCs in the development set of 0.81 (95% CI, 0.76–0.87), 0.78 (95% CI, 0.71–0.84), and 0.83 (95% CI, 0.77–0.90), respectively (Fig. 2B).

Based on the 25% and 75% quartiles of the risk score, young patients (age $\leq$ 60 years) were divided into 3 groups (Fig. 2C), with a 3-year estimated OS rate for low-, intermediate-, and high-risk group of 78.6% (95% CI, 68.3–90.5), 52.2% (95% CI, 42.1–64.7), and 9.0% (95% CI, 1.7–48.2), respectively (log-rank $p < 0.001$). In view of the relatively small number of elderly patients in both sets, and the therapeutic dilemma in this population which emphasizes the need to facilitate risk-adapted therapy, two risk groups were established with a median cut-off value of the estimated risk in elderly patients (Fig. 2D), and the estimated 3-year OS rates for low- and high-risk groups were 63.9% (95% CI, 48.1–84.8) and 9.5% (95% CI, 1.9–48.2), respectively (log-rank $p < 0.001$).

Table 1
Univariate and multivariable analyses of variables predicting overall survival in the validation set.

Variable	Comparison groups	Univariate analysis			Multivariable analysis		
		weight	HR (95% CI)	*P*	weight	HR (95%-CI)	*P*
Sex	Male vs. Female	0.027	1.028 (0.727−1.452)	0.877			
Age (years)	≥ 65 vs. < 65	0.572	1.771 (1.118−2.805)	0.015	0.435	1.545 (0.927−2.574)	0.095
HCT-CI (points)	≥ 1 vs. < 1	0.701	2.016 (1.404−2.894)	< 0.001	0.688	1.990 (1.364−2.904)	< 0.001
Barthel Index (points)	100 vs. < 100	−0.418	0.658 (0.432−1.003)	0.052			
WBC ($\times 10^9$/L)	≥ 50 vs. < 50	0.476	1.609 (1.077−2.404)	0.020	0.488	1.628 (1.044−2.539)	0.031
Hb (g/L)	60 to < 100 vs. < 60	−0.239	0.788 (0.479−1.295)	< 0.001	−0.304	0.738 (0.437−1.244)	0.254
	≥ 100 vs. < 60	−1.079	0.340 (0.186−0.621)		−0.913	0.401 (0.210−0.766)	0.006
PLT ($\times 10^9$/L)	20 to < 50 vs. < 20	−0.009	0.991 (0.600−1.637)	0.884			
	50 to < 100 vs. < 20	−0.111	0.895 (0.514−1.558)				
	≥ 100 vs. < 20	−0.176	0.838 (0.479−1.466)				
LDH (IU/L)	≥ 500 vs. < 500	0.465	1.592 (1.097−2.310)	0.014			
Fg (g/L)	≥ 1.5 vs. <1.5	−0.528	0.590 (0.240−1.451)	0.250			
AML type	Secondary-AML vs. De Novo-AML	0.108	1.114 (0.490−2.529)	0.797			
Bone marrow blasts	≥ 60 vs. < 60	0.166	1.180 (0.833−1.671)	0.351			
ELN cytogenetic risks status	Intermediate vs. Favorable	0.110	1.116 (0.697−1.787)	< 0.001	0.542	1.720 (0.989−2.989)	0.055
	Adverse vs. Favorable	1.042	2.834 (1.581−5.082)		1.443	4.232 (2.322−7.712)	< 0.001
FLT3-ITD/*NPM1* status	*FLT3*-ITD+/*NPM1*- vs. *FLT3*-ITD-/*NPM1*-	1.292	3.640 (2.029−6.532)	< 0.001	1.004	2.728 (1.427−5.215)	0.002
	FLT3-ITD-/*NPM1*+ vs. *FLT3*-ITD-/*NPM1*-	−0.258	0.772 (0.465−1.283)		−0.796	0.451 (0.250−0.813)	0.008
	FLT3-ITD+/*NPM1*+ vs. *FLT3*-ITD-/*NPM1*-	0.837	2.309 (1.159−4.602)		0.158	1.171 (0.547−2.503)	0.684
FLT3-TKD	Mutated vs. Wild-type	−1.281	0.278 (0.069−1.123)	0.072			
Biallelic *CEBPA*	Mutated vs. Wild-type &Monoallelic	−1.055	0.348 (0.199−0.610)	< 0.001	−0.980	0.375 (0.199−0.706)	0.002
DNMT3A	Mutated vs. Wild-type	0.558	1.746 (1.081−2.820)	0.023	0.792	2.208 (1.285−3.794)	0.004
N-RAS	Mutated vs. Wild-type	−0.008	0.992 (0.603−1.633)	0.974			
C-KIT	Mutated vs. Wild-type	0.107	1.113 (0.626−1.979)	0.716			
CD34/CD38 status	$CD34^+CD38^+$ vs. $CD34^-$	−0.018	0.982 (0.654−1.476)	0.024			
	$CD34^+CD38^-$ vs. $CD34^-$	1.255	3.507 (1.344−9.149)	0.024			

Abbreviations: HCT-CI hematopoietic cell transplantation-comorbidity index, WBC white blood cell count, Hb hemoglobin, PLT platelet, LDH lactate dehydrogenase, Fg fibrinogen, ELN European LeukemiaNet.

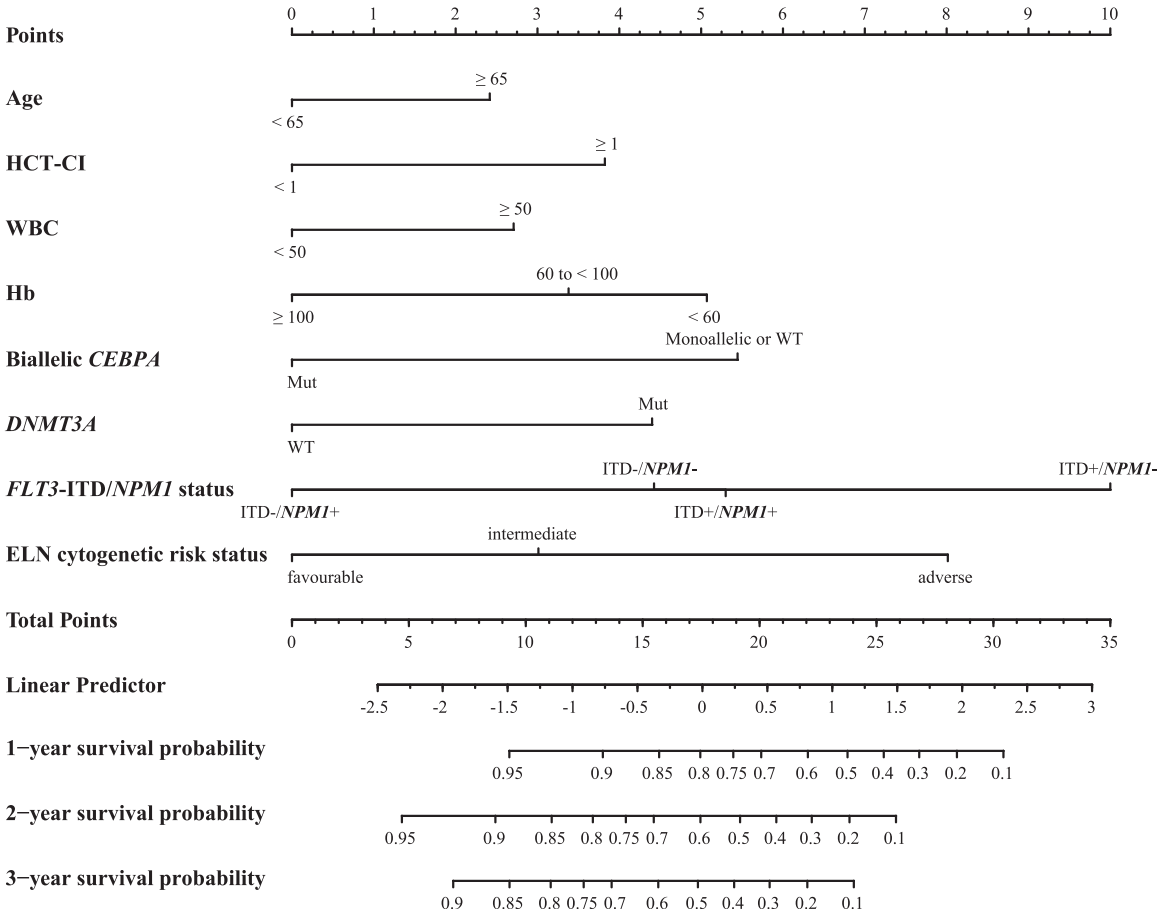

Fig. 1. A nomogram for the 1-year, 2-year, and 3-year overall survival probability prediction. HCT-CI hematopoietic cell transplantation-comorbidity index, WBC white blood cell count, Hb hemoglobin, ELN European LeukaemiaNet.

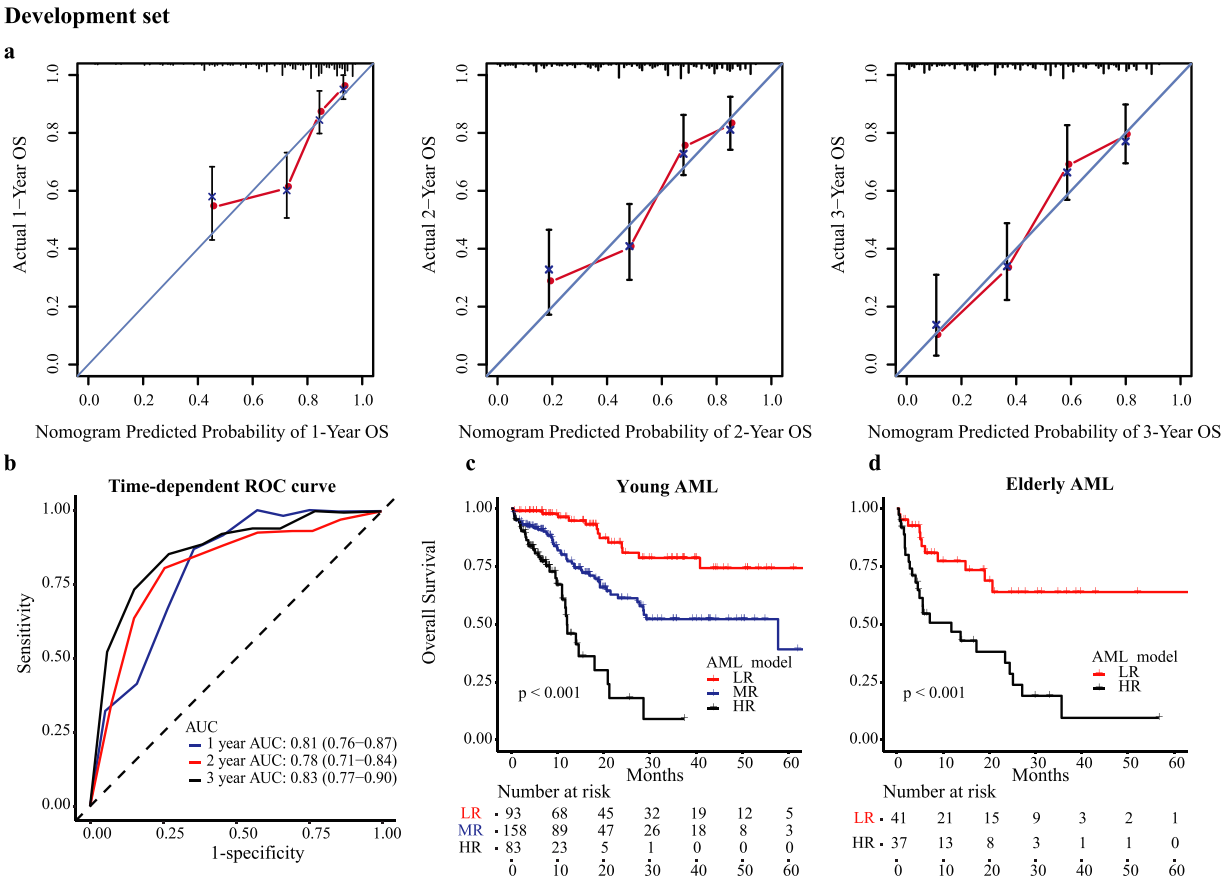

Fig. 2. Performance of the prognostic model in the development set. (a) Calibration curve showing predicted and actual 1-year, 2-year, and 3-year survival probabilities. The blue diagonal line indicates the perfect correspondence between the observation and prediction. (b) The area under the ROC curve (AUC) showing the discriminative ability of the prognostic model. The black dotted line indicates no discriminability (AUC = 0.5). Kaplan-Meier estimates the survival of young AML patients (c), and elderly AML patients (d) according to the risk groups. LR low-risk group, MR intermediate-risk group, HR high-risk group.

3.3. Model validation and performance

To evaluate the validity and performance of the prognostic model, a separate dataset comprising 238 patients was used for validation. The 1-, 2- and 3-year calibration plot of the validation set (Fig. 3A) indicated satisfactory concordance between the estimated and observed probabilities of overall survival. The C-index of the established nomogram in the validating cohort was 0.74 (95% CI, 0.70−0.80). The predictive accuracy of the model was validated by the time-dependent ROC in this cohort, with 1-, 2- and 3-year AUCs of 0.76 (95% CI, 0.67−0.84), 0.79 (95% CI, 0.71−0.87), and 0.74 (95% CI, 0.63−0.84), respectively (Fig. 3B). Collectively, these results suggested that the risk model had an equally good performance in the validation set in terms of both calibration and discrimination.

The cut-offs identified in the development set were applied directly to the validation set. For young patients (Fig. 3C), the estimated 3-year OS rate was 82.0% (95% CI, 69.3−97.2) for low-risk, 68.5% (95% CI, 58.1−80.8) for intermediate-risk, and 26.4% (95% CI, 14.9−46.7) for high-risk patients (log-rank $p < 0.001$). Of note, though there were only 51 elderly patients in the validation set (Fig. 3D), it could be seen that the high-risk patients bore a 3-year survival rate of merely 19.1% (95% CI, 8.0−45.7), which was worse than the low-risk patients with a 3-year survival rate of 39.5% (95% CI, 21.5−72.6) (log-rank $p = 0.018$).

3.4. Model comparison

To further explore and quantify the predictive performance of this novel model, we compared it in the validation cohort with the two previously reported risk stratification systems, which were Irena Djunic's model (integrating information on age, LDH, HCT-CI) in the elderly patients, and Pastore's model (incorporating age, performance status, WBC count, mutation status of *NPM1*, *CEBPA*, and *FLT3*-ITD) in CN-AML patients (Fig. 4)[4, 5]. Even though our novel risk system was not specifically designed for elderly or CN-AML patients, it provided a better discriminability than Irena Djunic's model in respect to AUC (AML early prediction model: 0.692, Irena Djunic's model: 0.630, venkatraman $p = 0.003$), and a similar discriminability compared to Pastore's model (AML early prediction model: 0.739, Pastore's model: 0.758, venkatraman $p = 0.979$).

3.5. Performance of the established model in risk stratification of the entire cohort

To appraise the value of the AML early prediction model in risk stratification for both long-term survival and risk of relapse in all patients, the OS and DFS of different risk groups without censoring for HSCT were compared after combining the two sets together. Young patients with AML were divided into three distinct groups with gradually reduced 3-year survival rate in low-risk (OS, 78.6%, 95% CI, 71.1−86.8; DFS, 62.4%, 95% CI, 54.4−71.7), intermediate-risk (OS, 63.6%, 95% CI, 57.1−70.9; DFS, 45%, 95% CI, 38.5−52.5), and high-risk (OS, 35.9%, 95% CI, 27.7−46.6; DFS, 27.4%, 95% CI, 20.2−37.1) (Fig. 5A, 5C). Besides, the model performed equally well in risk classification of elderly AML patients, with an estimated 3-year OS rate for low-, and high-risk group of 58.0% (95% CI, 38.8−72.9) and 18.6% (95% CI, 10.1−34.2) (log-rank $p < 0.001$), respectively (Fig. 5B), and

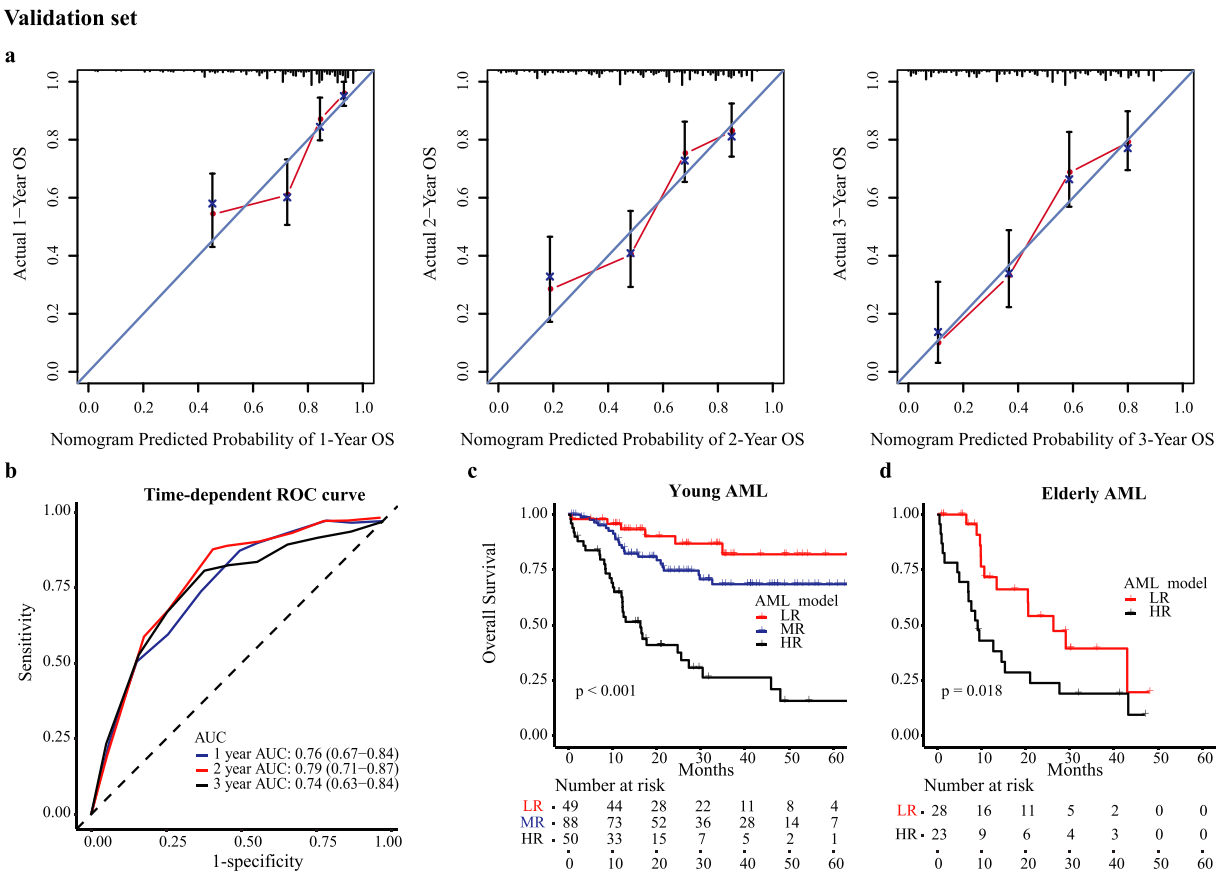

Fig. 3. Performance of the prognostic model in the validation set. (a) Calibration curve showing predicted and actual 1-year, 2-year, and 3-year survival probabilities. (b) The area under the ROC curve (AUC) showing the discriminative ability of the prognostic model. Kaplan-Meier estimates the survival of young AML patients (c), and elderly AML patients (d) according to the risk groups. LR low-risk group, MR intermediate-risk group, HR high-risk group.

an estimated 3-year DFS rate of 46.2% (95% CI, 32.6−65.5) and 8.1% (95% CI, 3.0−21.8) (log-rank $p < 0.001$), respectively (Fig. 5D).

3.6. Clinical application for treatment selection in different risk groups

Determining which young AML patients should receive HSCT and whether "3 + 7"-based or other less intensive therapy should be offered to elderly patients remain controversial in daily clinical practice. Hence, we investigated the ability of our model to assist with guiding therapeutic decision-making in both young and elderly patients. By combining 521 young patients in the training and validation cohorts without censoring for HSCT, the effects of transplantation on OS and DFS in different risk groups were compared (Fig. 6A, 6C,6E). It showed that HSCT failed to significantly improve the outcome of low-risk patients (univariable Cox regression: OS, HR = 0.666, 95% CI, 0.266−1.669, $p = 0.386$; DFS, HR = 1.118, 95% CI, 0.617−2.027, $p = 0.713$). In contrast, HSCT could considerably improve the prognosis of patients in the intermediate-risk group (OS, HR = 0.290, 95% CI, 0.164−0.513, $p < 0.001$; DFS, HR = 0.483,

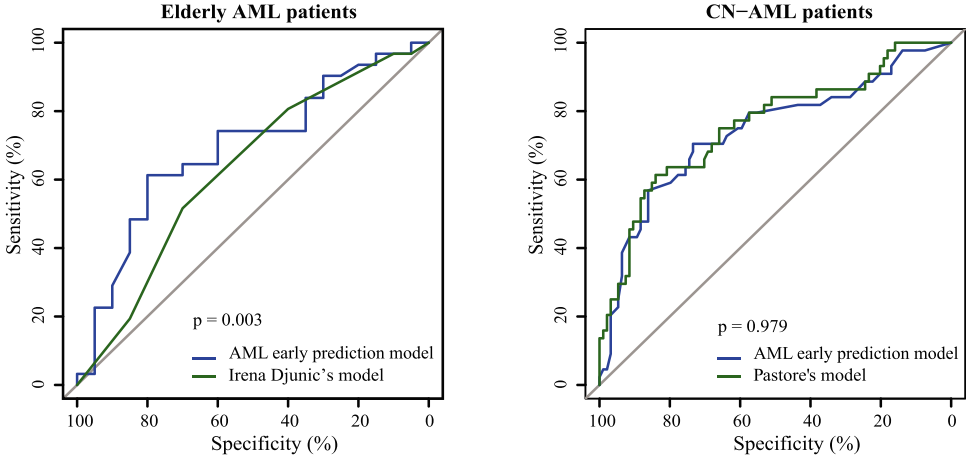

Fig. 4. Comparison between AML early prediction model and published models in the validation set. (a) ROCs of the AML early prediction model and Irena Djunic's score system in elderly AML patients (b) ROCs of the novel classification system and Pastore's $PINA_{OS}$ score in CN-AML patients. The P-value between different ROCs was calculated under the Venkatraman method.

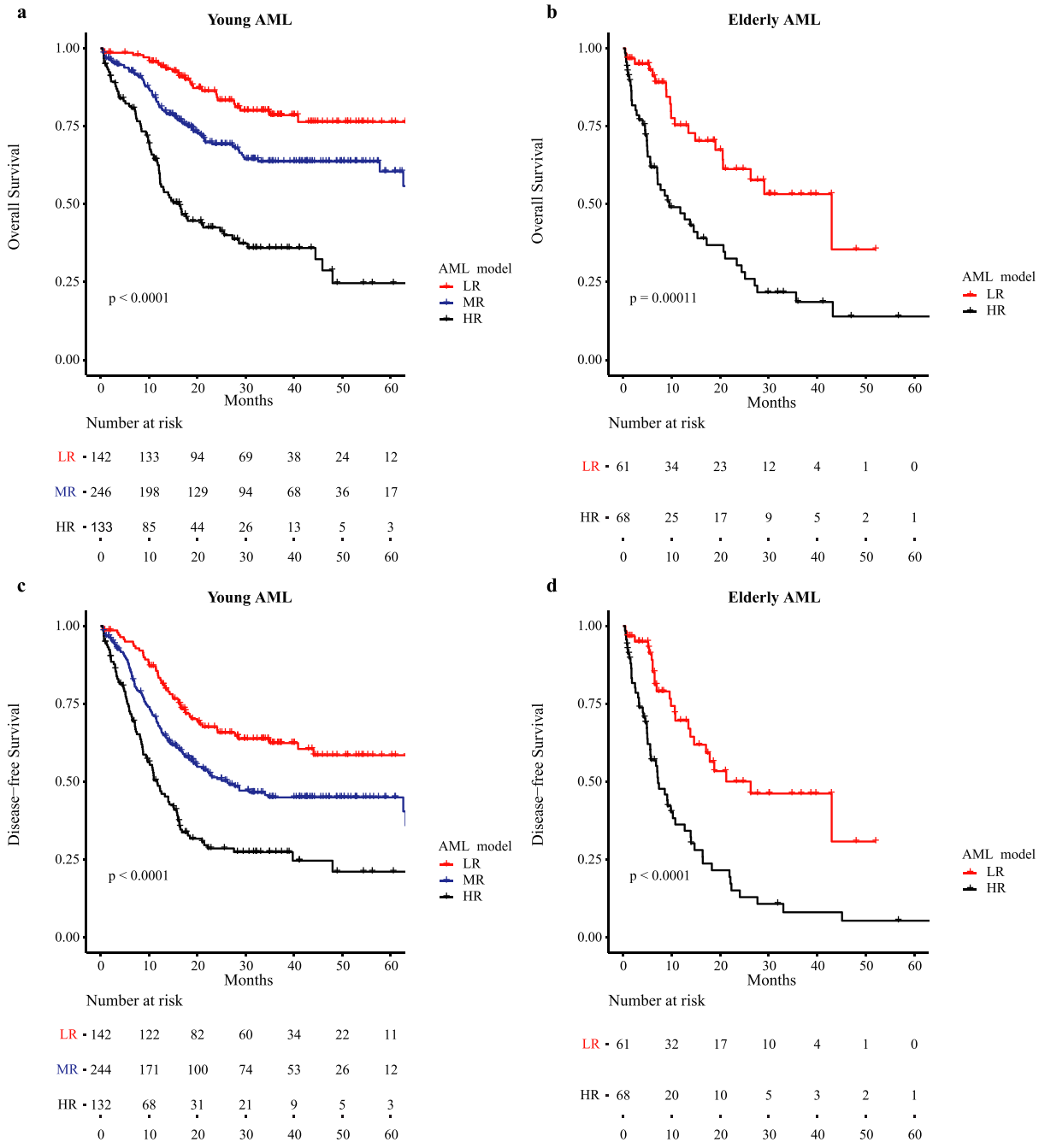

Fig. 5. Performance of the prognostic model in risk stratification of the entire cohort. Kaplan-Meier curves of OS for young AML patients (a) and elderly AML patients (b); Kaplan-Meier curves of DFS for young AML patients (c) and elderly AML patients (d) according to the AML early prediction model. LR low-risk group, MR intermediate-risk group, HR high-risk group.

95% CI, 0.325−0.720, $p < 0.001$), and to a greater extent, of those in the high-risk group (OS, HR = 0.209, 95% CI, 0.121−0.362, $p < 0.001$; DFS, HR = 0.253, 95% CI, 0.155−0.414, $p < 0.001$). Notably, patients who underwent HSCT in the intermediate-risk group achieved a similar 3-year estimated OS rate compared with transplant patients in the low-risk group (82.4 vs 83.3 months, log-rank $p = 0.751$, Supplementary Table S3).

Furthermore, we applied this new risk classification to all 129 elderly patients in both sets to identify whether elderly patients with different risks could benefit from different intensity of induction regimens (Fig. 6B, 6D, 6E). For patients in the high-risk group, the administration of reduced "3 + 7"-based induction regimens and other less intensive induction regimens conferred a similar outcome (univariable Cox regression: OS, HR = 0.872, 95% CI, 0.482−1.577, $p = 0.651$; DFS, HR = 0.922, 95% CI, 0.530−1.604, $p = 0.774$), with a 3-year estimated OS rate of 20.1% and 19.8%, and DFS rate of 5.9% and 9.9%, respectively (Supplementary Table S4). The results were similar in the low-risk group, the 3-year OS, DFS rate were 59.0%, 51.8% with reduced "3 + 7"-based therapy, and 42.7%, 37.4% with other less intensive therapy (univariable Cox regression: OS, HR = 0.945, 95% CI, 0.382−2.340, $p = 0.903$; DFS, HR = 0.835, 95% CI, 0.365−1.908, $p = 0.669$), illustrating that elderly patients in both risk groups responded similarly to different induction regimens.

4. Discussion

Precise, robust, and applicable prognostic model for predicting long-term survival, stratifying risk groups and helping with

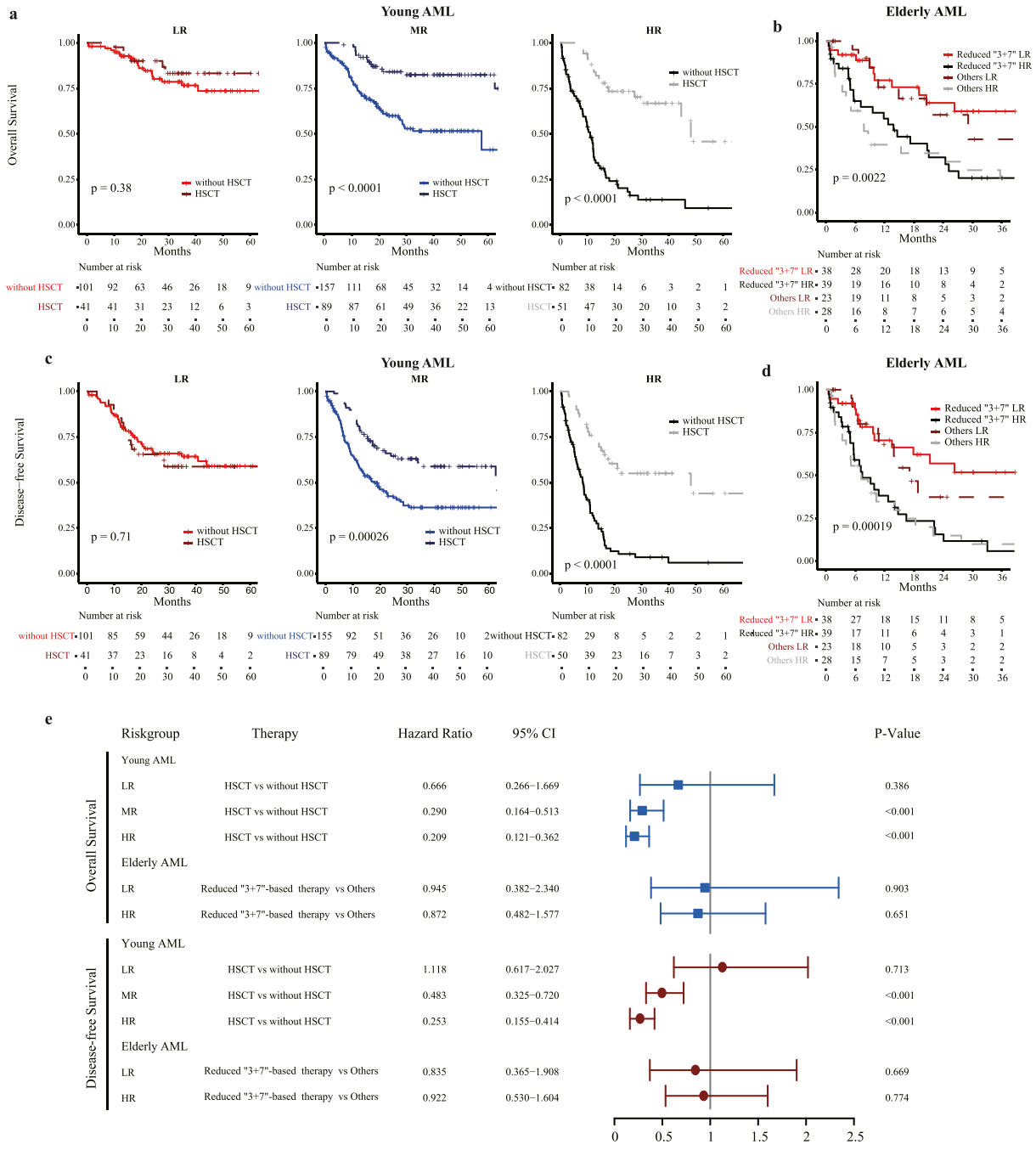

Fig. 6. Evaluation of the clinical application value in different age groups. Kaplan−Meier estimates the OS (a) and DFS (c) of young AML patients in different risk subgroups, stratified by whether or not they underwent HSCT. The benefits of transplantation for young AML patients at different risk status were calculated without censoring for HSCT. Kaplan−Meier estimates the OS (b) and DFS (d) of elderly AML patients in low- and high-risk subgroups, stratified by different induction regimens. (e) Forest plot exhibiting the P-value and Hazard ratio of OS and DFS for young AML patients who received HSCT or not within each risk subgroup, and for elderly AML patients who received reduced "3 + 7"-based induction chemotherapy or other less intensive agents within each risk subgroup. HSCT hematopoietic stem cell transplantation, LR low-risk group, MR intermediate-risk group, HR high-risk group.

therapeutic decision-making in AML is still an unmet clinical demand. In this study, by combining patient-related and AML-related predictors, we developed and validated a novel prognostic model in adult AML patients to address this issue.

Various models have been proposed to predict the prognosis of AML so far, however, the majority of which focused on specific AML cohorts, for example, models for young patients (age ≤ 60 years) [22] or elderly patients (age ≥ 60 years) only [5, 6, 23], those restricted to CN-AML [4, 24], or a combination of the above two populations [25]. In contrast, we included all age groups of adult AML patients to improve the generalizability of our prognostic model. Traditional clinical and demographic parameters such as age, performance status, physical condition, comorbidities, as well as disease-related characteristics including morphology, immunophenotype, cytogenetic risk status and molecular events were fully considered during the model construction procedure. Taking into account of multi-aspect variables allows a comprehensive view of the patient- and disease-related conditions, thereby improving the reliability and robustness of the model.

Calibration and discrimination are two key aspects when evaluating the performance of a predictive model, which was recommended and emphasized in the TRIPOD, criteria for reporting a multivariable

prediction model. Nevertheless, many of the reported prognostic models in AML lacked a comprehensive assessment of their performance, merely through stratifying patients into three or four different risk groups to evaluate the discriminative ability [6, 15, 22, 26-28]. The only exception was the nomogram predicting the leukaemia-free survival after autologous stem cell transplantation developed by shouval et al. [29]. Importantly, the AML early prediction model established in our study complied with the TRIPOD criteria and manifested promising performance concerning both calibration and discrimination in the development and validation sets in predicting the long-term survival of all adult AML patients.

It is not surprising that cytogenetic risk status, *FLT3*-ITD/*NPM1* status, and biallelic *CEBPA* mutations were powerful predictors in our model in consideration of their important roles in the ELN risk stratification and prior reports [4, 15, 28, 30]. Notably, only 12% patients carried *FLT3*-ITD mutations in our full cohort, which seemed to be considerably lower than the frequency of 20−25% in the western population [2], but similar with the reports derived from the Chinese population [17], reflecting the difference of genetic background between different races, as exemplified as the high incidence of *CEBPA* mutations in Chinese patients [31]. Even so, *FLT3*-ITD/*NPM1* status remained to be the most powerful factor for OS in our model. Interestingly, Papaemmanuil et al. also proved that the deleterious effect of *FLT3*-ITD increased in patients with concomitant *NPM1* and *DNMT3A* mutations regardless of the ratio of mutant to wild-type *FLT3*-ITD [30]. The independent negative effect of lower Hb level at diagnosis on the survival of AML patients aged 70 and older was recently revealed by Talati et al. [32], and the effect was expanded to all adult AML patients in our study. HCT-CI, a measure of comorbidities, was originally proposed by Sorror et al. in 2005 for risk assessment of AML patients before allogeneic hematopoietic cell transplantation [33], and has recently been proved to exert a significant impact on early death, 1-year mortality, and long-term survival in patients with AML [15]. Moreover, age has been widely recognized as a crucial prognostic factor, which may be related to the accumulation of molecular events and the deterioration of cognitive, psychological and physical function in the process of aging [34]. Apparently, the incorporation of these important prognostic factors into the risk score improves the robustness and validity of this prognostic model. Despite the fact that the model did not incorporate factors derived from the next-generation sequencing which has not yet been widely applied in clinical diagnosis, the well-known predictors in the model, on the contrary, are easily acquired in routine clinical practice, making it convenient to apply the novel model not only in advanced clinical centres, but also in community hospitals.

The selection of consolidation therapy for young patients after attaining CR remains challenging in clinical practice when weighing the risk of relapse and treatment-related death, especially for patients with intermediate cytogenetic and molecular risk. Several models integrating post-remission variables were developed to identify which patients could benefit from chemotherapy or HSCT for consolidation [22, 35], and more were designed to predict survival after transplantation [29, 36, 37]. Our results demonstrated that patients in both intermediate- and high-risk groups identified by our model could benefit from HSCT, and more intriguingly, patients in the intermediate-risk group who received HSCT for consolidation achieved similar prolonged survival as those in the low-risk group, indicating that our model is capable of discriminating eligible candidates for HSCT rather than simply stratifying patients, which makes it possible for physicians to select optimal consolidation therapy at the time of diagnosis.

It is of critical importance to correctly choose the treatment modality suitable for elderly AML patients, so as to achieve durable efficacy while sparing them the toxicity of chemotherapy. Elderly patients receiving reduced "3 + 7″-based induction regimens in our study were empirically evaluated as "fit" before treatment and expected to have a better prognosis than those in the "unfit" group. However, patients who experienced reduced "3 + 7″-based induction treatment and other less intensive agents had comparable survival time in both low- and high-risk groups, suggesting that the subjective criteria traditionally used might not be able to choose the proper intensity and modality of induction therapy, and the first-line "3 + 7″-based induction regimens failed to prolong the survival of elderly patients in this study. New hope relies on novel modalities, such as venetoclax, *IDH1/2,* and *FLT3* inhibitors that were not administered in these elderly patients.

Recently, several novel drugs have been approved by the FDA and introduced to patients with AML since 2017, which exerted encouraging anti-leukemic efficacy in large clinical trials, as exemplified by venetoclax, CPX-351 and *IDH1/2* inhibitors. These drugs brought new light to the treatment of AML. Furthermore, molecular prognostic profile might be changed by these new treatment strategies. We believed that it is worth noting that the AML early prediction model is suitable and beneficial for the majority of AML patients who receive the classic "3 + 7″ regimens. However, its significance in the era of mutation-directed therapy should be re-evaluated with the more extensive application of new drugs. In conclusion, we integrated comprehensive patient-related and AML-related information to develop and validate a novel score system, which could precisely predict both overall and disease-free survival of all adult AML patients. Quick evaluation and accurate calculation of individual survival probability can be made with the nomogram and a free web-based calculator, which meets the demand of the precision medicine and personally tailored cancer management. The parameters in the model are routinely evaluated and easily adopted in the clinic, making it possible for physicians to quickly complete the evaluation of survival probability, risk stratification, and therapeutic decision-making at diagnosis. Investigators and related authorities can utilize this model to select patients with specific risks for treatment research and compare performance among different clinical trials and centres. Prospective, randomized clinical trials are warranted to validate our model in the future.

Contributors

TTM, XJL, and YS designed the study. TTM, WYC, QX, SYW, FJL, HY, and YMZ collected and confirmed the data. XJL, WYC, and TTM analysed the data. TTM, XJL, WYC and YS wrote the manuscript. All authors approved the final version of the manuscript.

Declaration of Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

We would like to thank all the patients and their families participated in this study, doctors and nurses in the three hospitals who took care of the patients and documented the data in the electronic health records. This work was supported in part by grants from the National Natural Science Foundation of China (No. 81770141), the National Key R&D Program of China (No. 2016YFE0202800), and Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant Support (No. 20161406).

Data Sharing Statement

Data are available from the corresponding author on reasonable request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2020.103126.

References

[1] DiNardo CD, Perl AE. Advances in patient care through increasingly individualized therapy. Nat Rev Clin Oncol 2019 Feb;16(2):73–4.
[2] Short NJ, Rytting ME, Cortes JE. Acute myeloid leukaemia. Lancet 2018;392 (10147):593–606.
[3] Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129(4):424–47.
[4] Pastore F, Dufour A, Benthaus T, et al. Combined molecular and clinical prognostic index for relapse and survival in cytogenetically normal acute myeloid leukemia. J Clin Oncol 2014 May 20;32(15):1586–94.
[5] Djunic I, Suvajdzic-Vukovic N, Virijevic M, et al. Prognostic risk score for the survival of elderly patients with acute myeloid leukaemia comprising comorbidities. Med Oncol 2013 Mar;30(1):394.
[6] Rollig C, Thiede C, Gramatzki M, et al. A novel prognostic model in elderly patients with acute myeloid leukemia: results of 909 patients entered into the prospective AML96 trial. Blood 2010 Aug 12;116(6):971–8.
[7] Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012 May;98 (9):691–8.
[8] Bullinger L, Dohner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004 Apr 15;350(16):1605–16.
[9] Metzeler KH, Hummel M, Bloomfield CD, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood 2008 Nov 15;112(10):4193–201.
[10] Li Z, Herold T, He C, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. J Clin Oncol 2013 Mar 20;31(9):1172–81.
[11] Eppert K, Takenaka K, Lechman ER, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat Med 2011 Aug 28;17 (9):1086–93.
[12] Patel JP, Gonen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med 2012 Mar 22;366(12):1079–89.
[13] Gerstung M, Papaemmanuil E, Martincorena I, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet 2017 Mar;49 (3):332–40.
[14] Wang M, Lindberg J, Klevebring D, et al. Validation of risk stratification models in acute myeloid leukemia using sequencing-based molecular profiling. Leukemia 2017 Oct;31(10):2029–36.
[15] Sorror ML, Storer BE, Fathi AT, et al. Development and validation of a novel acute myeloid leukemia-composite model to estimate risks of mortality. JAMA Oncol 2017 Dec 1;3(12):1675–82.
[16] Lübke N, Meinck M, Von Renteln-Kruse W. [The Barthel index in geriatrics. A context analysis for the Hamburg classification manual]. Z Gerontol Geriatr 2004 Aug;37(4):316–26.
[17] Shen Y, Zhu YM, Fan X, et al. Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. Blood 2011 Nov 17;118(20):5593–603.
[18] Sui JN, Chen QS, Zhang YX, et al. Identifying leukemia-associated immunophenotype-based individualized minimal residual disease in acute myeloid leukemia and its prognostic significance. Am J Hematol 2019 May;94(5):528–38.
[19] Venkatraman ES. A permutation test to compare receiver operating characteristic curves. Biometrics 2000 Dec;56(4):1134–8.
[20] Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015 Jan 6;162(1):W1–73.
[21] Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008 Mar 10;26(8):1364–70.
[22] Pfirrmann M, Ehninger G, Thiede C, et al. Prediction of post-remission survival in acute myeloid leukaemia: a post-hoc analysis of the AML96 trial. Lancet Oncol 2012;13(2):207–14.
[23] Chen Y, Yang T, Zheng X, et al. The outcome and prognostic factors of 248 elderly patients with acute myeloid leukemia treated with standard-dose or low-intensity induction therapy. Medicine (Baltimore) 2016 Jul;95(30):e4182.
[24] Rockova V, Abbas S, Wouters BJ, et al. Risk stratification of intermediate-risk acute myeloid leukemia: integrative analysis of a multitude of gene mutation and gene expression markers. Blood 2011 Jul 28;118(4):1069–76.
[25] Damm F, Heuser M, Morgan M, et al. Integrative prognostic risk score in acute myeloid leukemia with normal karyotype. Blood 2011 Apr 28;117(17):4561–8.
[26] Marcucci G, Yan P, Maharry K, et al. Epigenetics meets genetics in acute myeloid leukemia: clinical impact of a novel seven-gene score. J Clin Oncol 2014 Feb 20;32(6):548–56.
[27] Wang M, Lindberg J, Klevebring D, et al. Development and validation of a novel RNA sequencing-based prognostic score for acute myeloid leukemia. J Natl Cancer Inst 2018 Oct 1;110(10):1094–101.
[28] Herold T, Jurinovic V, Batcha AMN, et al. A 29-gene and cytogenetic score for the prediction of resistance to induction treatment in acute myeloid leukemia. Haematologica 2018 Mar;103(3):456–65.
[29] Shouval R, Labopin M, Gorin NC, et al. Individualized prediction of leukemia-free survival after autologous stem cell transplantation in acute myeloid leukemia. Cancer 2019 Oct 15;125(20):3566–73.
[30] Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016 Jun 9;374(23):2209–21.
[31] Zhang M, Yin J, He Q, et al. Chinese and Europeans with acute myeloid leukemia have discordant mutation topographies. Leuk. Res. 2018 Jul;70::8–12.
[32] Talati C, Dhulipala VC, Mt Extermann, et al. Comparisons of commonly used frontline regimens on survival outcomes in patients aged 70 years and older with acute myeloid leukemia. Haematologica 2020;105(2):398–406.
[33] Sorror ML, Maris MB, Storb R, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. Blood 2005;106(8):2912–9.
[34] Klepin HD, Geiger AM, Tooze JA, et al. Geriatric assessment predicts survival for older adults receiving induction chemotherapy for acute myelogenous leukemia. Blood 2013 May 23;121(21):4287–94.
[35] Michelis FV, Messner HA, Atenafu EG, et al. Patient age, remission status and HCT-CI in a combined score are prognostic for patients with AML undergoing allogeneic hematopoietic cell transplantation in CR1 and CR2. Bone Marrow Transpl 2015 Nov;50(11):1405–10.
[36] Bergua JM, Montesinos P, Martinez-Cuadron D, et al. A prognostic model for survival after salvage treatment with FLAG-Ida +/- gemtuzumab-ozogamicine in adult patients with refractory/relapsed acute myeloid leukaemia. Br J Haematol 2016 Sep;174(5):700–10.
[37] Shouval R, Ruggeri A, Labopin M, et al. An integrative scoring system for survival prediction following umbilical cord blood transplantation in acute leukemia. Clin Cancer Res 2017 Nov 1;23(21):6478–86.