blood®
Regular Article

MYELOID NEOPLASIA

A predictive model for therapy failure in patients with chronic myeloid leukemia receiving tyrosine kinase inhibitor therapy

Xiaoshuai Zhang,[1,*] Bingcheng Liu,[2,*] Jian Huang,[3-5,*] Yanli Zhang,[6,*] Na Xu,[7,*] Robert Peter Gale,[8] Weiming Li,[9] Xiaoli Liu,[7] Huanling Zhu,[10] Ling Pan,[10] Yunfan Yang,[10] Hai Lin,[11] Xin Du,[12] Rong Liang,[13] Chunyan Chen,[14] Xiaodong Wang,[15] Guohui Li,[16] Zhuogang Liu,[17] Yanqing Zhang,[18] Zhenfang Liu,[19] Jianda Hu,[20] Chunshui Liu,[11] Fei Li,[21] Wei Yang,[17] Li Meng,[22] Yanqiu Han,[23] Li'e Lin,[24] Zhenyu Zhao,[24] Chuanqing Tu,[25] Caifeng Zheng,[25] Yanliang Bai,[26] Zeping Zhou,[27] Suning Chen,[28] Huiying Qiu,[28] Lijie Yang,[16] Xiuli Sun,[29] Hui Sun,[30] Li Zhou,[31] Zelin Liu,[32] Danyu Wang,[32] Jianxin Guo,[33] Liping Pang,[34] Qingshu Zeng,[35] Xiaohui Suo,[36] Weihua Zhang,[37] Yuanjun Zheng,[37] Xiaojun Huang,[1,38,39] and Qian Jiang[1,40]

[1]Peking University People's Hospital, Peking University Institute of Hematology, National Clinical Research Center for Hematologic Disease, Beijing Key Laboratory of Hematopoietic Stem Cell Transplantation, Beijing, China; [2]National Clinical Research Center for Blood Diseases, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjing, China; [3]Department of Hematology, The First Affiliated Hospital of Zhejiang University, College of Medicine, Zhejiang University, Zhejiang, China; [4]Department of Hematology, The Fourth Affiliated Hospital of Zhejiang University, College of Medicine, Zhejiang University, Zhejiang, China; [5]Zhejiang Provincial Clinical Research Center for Hematological Disorders, Zhejiang, China; [6]Department of Hematology, Henan Cancer Hospital, Affiliated Cancer Hospital of Zhengzhou University, Henan, China; [7]Department of Hematology, Nanfang Hospital, Southern Medical University, Guangzhou, China; [8]Department of Immunology and Inflammation, Centre for Hematology, Imperial College London, London, United Kingdom; [9]Department of Hematology, Union hospital, Tongji Medical college, Huazhong University of Science and Technology, Wuhan, China; [10]Department of Hematology, Institute of Hematology, West China Hospital, Sichuan University, Sichuan, China; [11]Department of Hematology, The First Hospital of Jilin University, Jilin, China; [12]Department of Hematology, The Second People's Hospital of Shenzhen, The First Affiliated Hospital of Shenzhen University, Shenzhen, China; [13]Department of Hematology, Xijing Hospital, Airforce Military Medical University, Xi'an, China; [14]Department of Hematology, Qilu Hospital of Shandong University, Jinan, China; [15]Department of Hematology, Sichuan Academy of Medical Sciences Sichuan Provincial People's Hospital, Sichuan, China; [16]Department of Hematology, Xi'an international medical center hospital, Xi'an, China; [17]Department of Hematology, Shengjing Hospital of China Medical University, Shenyang, China; [18]Shenzhen Hospital of Southern Medical University, Shenzhen, China; [19]Department of Hematology, The First Affiliated Hospital of Guangxi Medical University, Guangxi, China; [20]Department of Hematology, The Second Affiliated Hospital of Fujian Medical University, Fujian, China; [21]Center of Hematology, The First Affiliated Hospital of Nanchang University, Nanchang, China; [22]Department of Hematology, Tongji Hospital of Tongji Medical College, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, China; [23]Department of Hematology, The Affiliated Hospital of Inner Mongolia Medical University, Inner Mongolia, China; [24]Department of Hematology, Hainan General Hospital, Hainan, China; [25]Department of Hematology, Shenzhen Baoan Hospital, Shenzhen University Second Affiliated Hospital, Shenzhen, China; [26]Department of Hematology, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan, China; [27]Department of Hematology, The Second Hospital Affiliated to Kunming Medical University, Kunming, China; [28]National Clinical Research Center for Hematologic Diseases, Jiangsu Institute of Hematology, The First Affiliated Hospital of Soochow University, Institute of Blood and Marrow Transplantation of Soochow University, Soochow, China; [29]Department of Hematology, The First Affiliated Hospital of Dalian Medical University, Dalian, China; [30]Department of Hematology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; [31]Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China; [32]Department of Hematology, Huazhong University of Science and Technology Union Shenzhen Hospital, Nanshan Hospital, Shenzhen, China; [33]Department of Hematology, The Second Affiliated Hospital of Fujian Medical University, Fujian, China; [34]Department of Hematology, Peking University Shenzhen Hospital, Shenzhen, China; [35]Department of Hematology, The First Affiliated Hospital of Anhui Medical University, Anhui, China; [36]Department of Hematology, Handan Central Hospital, Handan, China; [37]Department of Hematology, First Hospital of Shanxi Medical University, Shanxi, China; [38]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China; [39]State Key Laboratory of Natural and Biomimetic Drugs, Peking University, Beijing, China; and [40]Peking University People's Hospital, Qingdao, China

KEY POINTS

- **The predictive model stratified CML patients into 3 risk subgroups with significantly different cumulative incidences of TKI-therapy failure.**

Although tyrosine kinase inhibitor (TKI) therapy has markedly improved the survival of people with chronic-phase chronic myeloid leukemia (CML), 20% to 30% of people still experienced therapy failure. Data from 1955 consecutive patients with chronic-phase CML diagnosed by the European LeukemiaNet recommendations from 1 center receiving initial imatinib or a second-generation (2G) TKI therapy were interrogated to develop a clinical prediction model for TKI-therapy failure. This model was subsequently validated in 3454 patients from 76 other centers. Using the predictive clinical covariates associated with TKI-therapy failure, we developed a model that stratified patients into low-, intermediate- and high-risk subgroups with significantly different cumulative incidences of therapy failure ($P < .001$). There was good discrimination and calibration in the external validation data set, and the performance was consistent with that of the training data set. Our model had the better prediction discrimination than the Sokal and European Treatment and Outcome Study long-term survival scores, with the greater

time-dependent area under the receiver-operator characteristic curve values and a better ability to redefine the risk of therapy failure. Our model could help physicians estimate the likelihood of initial imatinib or 2G TKI–therapy failure in people with chronic-phase CML.

Introduction

Tyrosine kinase inhibitor (TKI) therapy has markedly improved the survival of persons with chronic-phase chronic myeloid leukemia (CML), with 85% to 90% of persons surviving for >10 years.[1-6] However, 20% to 30% of people still experienced therapy failure.[6-13] Patients who met therapy-failure milestones had a high risk of disease progression and even death during TKI therapy.[14-19] Guidance from the World Health Organization, the European LeukemiaNet (ELN), and the National Comprehensive Cancer Network suggests that switching TKI therapy should be considered for patients who failed to meet certain response milestones.[20-25] Consequently, it is important to accurately predict the likelihood of therapy failure in persons with chronic-phase CML when choosing which TKI to begin with.

Recently, we developed a predictive model of therapy failure in persons with chronic-phase CML initially treated with imatinib.[11,13,26] However, with the increasing number of persons receiving initial second-generation (2G) TKI therapy, existing predictive scoring systems seem inadequate to meet the clinical need for identifying high-risk patients for TKI-therapy failure. To achieve this aim, we interrogated data from 1955 consecutive patients with chronic-phase CML according to the ELN recommendations from the center to develop the predictive model and then externally validated it in 3454 patients from 76 other centers.[20-23] The model stratifies patients into low-, intermediate- and high-risk subgroups with significantly different cumulative incidences of therapy failure. Our model could help physicians estimate the likelihood of initial imatinib- or 2G-TKI therapy failure in people with chronic-phase CML.

Methods

Patients

Data from consecutive patients aged ≥18 years with newly diagnosed chronic-phase CML with the *e14a2* and/or *e13a2* *BCR*::*ABL* transcripts receiving initial imatinib or a 2G-TKI including nilotinib, dasatinib, or flumatinib therapy at Peking University People's Hospital from January 2006 to September 2023 were interrogated to develop the model (training data set). Data from patients from January 2006 to January 2023 with the same inclusion criteria from 76 other centers were used to validate the model (validation data set). The patients included in the study had regular follow-up during TKI therapy. Demographic and clinical covariates including complete blood count, percentages of blood blasts, basophils and eosinophils, spleen size below the costal margin, comorbidities, and initial TKI therapy, and results of hematological, cytogenetic, and molecular analyses were extracted from the medical records. Patient data were obtained before therapy. Starting doses were imatinib, 400 mg per day; nilotinib, 300 mg twice daily; dasatinib, 100 mg per day; and flumatinib, 600 mg per day. These doses were subsequently adjusted according to responses and/or adverse events on the basis of ELN recommendations.[20-23] The study was approved by the ethics committee of Peking University People's Hospital (no. 2022PHB103-001), and all patients provided written informed consent, which was consistent with the precepts of the Declaration of Helsinki.

Diagnosis, monitoring, and definitions

Diagnosis and monitoring were performed according to the ELN recommendations.[20-23] Sokal and European Treatment and Outcome Study long-term survival (ELTS) scores at diagnosis were calculated as previously described.[4,27] Bone marrow cytogenetic analyses were performed using the G-banding method. Additional cytogenetic abnormalities (ACAs) in Philadelphia chromosome–positive (Ph^+) cells were identified as described previously,[23,28] and high-risk ACAs were defined by the 2020 ELN recommendations.[23] All ACAs were independently reviewed by 2 cytogeneticists. When the interpretations were inconsistent, they were referred to a senior cytogeneticist for reevaluation. Blood samples were used to analyze *BCR*::*ABL* transcripts at diagnosis. During therapy, *BCR*::*ABL* transcript levels were analyzed by quantitative real-time polymerase chain reaction with an *ABL1* control and converted to the international scale (IS; $BCR::ABL^{IS}$) using laboratory-specific conversion factors validated at the Institute of Medical and Veterinary Science International Reference Laboratory when the value (IS) was <10%.[29]

Hematologic response was monitored every 1 to 2 weeks until a complete hematologic response (CHR) was achieved, and every 3 to 6 months thereafter. Cytogenetic response was assessed at baseline and every 3 to 6 months thereafter until a complete cytogenetic response (CCyR) was achieved, and repeated when patients failed treatment. Quantitative real-time polymerase chain reaction monitoring was performed at diagnosis and every 3 months thereafter until a major molecular response (MMR) was achieved, and every 3 to 6 months thereafter. *ABL1* mutation screening was performed in patients with a suboptimal, warning, or failure to TKI therapy according to the ELN recommendations.[20-23]

Response definitions were as follows: (1) CHR: white blood cell < 10E+9/L, platelets <450 × 10E+9/L, no blood blasts or promyelocytes, <5% blood myelocytes and metamyelocytes, <5% blood basophils and no extramedullary leukemia, and duration of ≥4 weeks; (2) CCyR: no Ph^+ cells in ≥20 bone marrow metaphases; (3) MMR: $BCR::ABL^{IS}$ ≤0.1%; (4) molecular response 4 (MR^4): $BCR::ABL^{IS}$ ≤0.01%; and (5) molecular response 4.5 ($MR^{4.5}$): $BCR::ABL^{IS}$ ≤0.0032%.[20-23] TKI-therapy failure was defined as meeting "failure" milestones in the 2020 ELN recommendations[23]; loss of responses including CHR, CCyR, or MMR; or transformation to advanced phase as defined by the ELN recommendations.[20-23] Failure-free survival (FFS) was calculated from the start of TKI therapy to the first therapy failure or censored at a transplant, death, or the last

follow-up. We analyzed only the first episode in patients with >1 therapy failure. Transformation was defined as blood or bone marrow blasts of ≥15%, increased lymphoblasts, or extramedullary leukemia during TKI therapy.[24,30] Transformation-free survival (TFS) was calculated from the start of TKI therapy to transformation or censored at a transplant, death, or the last follow-up. Survival was calculated from the start of TKI therapy to death or censored at transplant or the last follow-up. The last follow-up was 31 January 2024.

Statistical analyses

Descriptive statistics were used to summarize covariates. Categorical covariates were reported as percentages and counts. Continuous variables were reported as medians and ranges or interquartile ranges (IQRs). The Pearson χ^2 test was used to analyze categorical covariates. The Student *t* test (normal distribution) or the Mann-Whitney *U* test (nonnormal distribution) was used to analyze continuous covariates. Cumulative incidences of therapy failure were calculated using the competing risk model and compared by the Fine-Gray test considering competing events, which were defined as transplant or death. FFS was calculated by the Kaplan-Meier method and compared by the log-rank test.

Cox and Fine-Gray regression models were used for univariable and multivariable analyses to identify covariates associated with TKI-therapy failure. Candidate covariates included sex, age, spleen size below the costal margin, white blood cell, hemoglobin and platelet concentrations, percentages of blood blasts, basophils and eosinophils, comorbidity(ies), and Ph^+ ACAs at diagnosis. Fractional polynomial transformations were used for the candidate continuous prognostic covariates in Cox regression models.[31,32] The most statistically suitable polynomial transformation for each covariate was identified when modeling its influence on TKI-therapy failure. These transformations were also used for the Fine-Gray models. Interactions among candidate covariates were tested in regression models. Akaike information criterion was used to select prognostic covariates and build the best model.[33,34] Significant prognostic covariates from multivariable analyses of the training data set were used to develop the predictive model. We used the final best multivariable regression model to develop a predictive model for therapy failure. A total of 1000 bootstrap samples in the training data set were used to identify candidate cutoffs.[35,36] The Fine-Gray test, the minimal *P* value approach, and Bonferroni correction were used to select optimal cutoffs for significantly different cumulative incidences of therapy failure.[37,38] Kernel density estimator was used to fit the smooth function to visualize the distribution of cutoffs.[39] Subsequently, patients were classified into different therapy-failure risk subgroups based on the predictive model.

The time-dependent area under the receiver-operator characteristic curve (AUROC) was used to estimate the accuracy of the predictive model, and calibration plots were generated to determine how closely the predicted and observed cumulative incidences of therapy failure were concordant.[40,41] Decision curve analysis (DCA) was used to calculate the net benefit using the model.[42]

Propensity score matching (PSM) was used to adjust for differences in baseline covariates between patients receiving imatinib or a 2G-TKI as initial therapy, and balance was evaluated using the standardized absolute mean difference for which a score of <0.02 was considered balanced.[43,44]

A 2-sided $P < .05$ was considered significant. SPSS 22.0 (SPSS, Chicago, IL), R version 4.0.2 (R Core Team, Vienna, Austria), and GraphPad Prism 8 (GraphPad Software Inc, La Jolla, CA) were used for analyses and graphing.

Results

Training data set

Data from 2587 consecutive patients receiving initial imatinib or a 2G-TKI therapy in the training data set are displayed in Figure 1. A total of 632 patients were excluded because of the interval from diagnosis to starting the TKI therapy of ≥6 months (n = 41), advanced phase at diagnosis (n = 236), missing important baseline covariates (n = 194), irregular response monitoring and/or loss to follow-up (n = 112), and non-*e14a2* and/or *e13a2* transcripts (n = 49) that could not be evaluated for molecular responses based on the IS using ELN recommendations.[20-23] The remaining 1955 patients received initial imatinib (n = 1539, 79%) or a 2G-TKI (n = 416, 21%) including nilotinib (n = 280, 14%), dasatinib (n = 72, 4%), or flumatinib (n = 64, 3%).

Patient covariates are displayed in Table 1; 1195 (61%) patients were male; median age was 40 years (IQR, 30-52); and 68 (3%) patients had ACAs in Ph^+ cells at diagnosis, 41 (2%) of whom had high-risk ACAs (supplemental Table 1, available on the *Blood* website). Median follow-up was 56 months (IQR, 30-91). Overall, 1584 (81%) patients continued receiving their initial TKI therapy including imatinib (n = 1216, 79%), nilotinib (n = 249, 89%), dasatinib (n = 61, 85%), or flumatinib (n = 58, 91%). A total of 495 patients (25%) failed initial TKI therapy at a median of 9 months (IQR, 4-14), because of not meeting the ELN therapy milestones at 3 (n = 112), 6 (n = 88), or 12 (n = 82) months; loss of responses (n = 50), emergence of TKI-resistant *ABL1* mutation(s) (n = 56) or high-risk ACAs in Ph^+ cells (n = 4); and transformation to advanced phase (n = 72) alone or combined (n = 31). Only 1 patient died of hepatocellular carcinoma before therapy failure. With a median follow-up of 65 months (IQR, 38-98) in the imatinib cohort, 8-year probabilities of FFS, TFS, and survival were 71% (95% confidence interval [CI], 68-74), 93% (95% CI, 91-95), and 96% (95% CI, 95-97). With a median follow-up of 36 months (IQR, 19-69) in the 2G-TKI cohort, 6-year probabilities of FFS, TFS, and survival were 71% (95% CI, 66-76), 90% (95% CI, 87-93), and 92% (95% CI, 88-96). Although the follow-up duration in the 2G-TKI cohort was significantly shorter than that in the imatinib cohort ($P < .001$), patients receiving initial 2G-TKI therapy had comparable outcomes including FFS ($P = .69$), TFS ($P = .17$), and survival ($P = .39$) compared with those receiving imatinib therapy. PSM analyses further confirmed these results (supplemental Table 2; supplemental Figure 1).

Validation data set

For the validation data set we interrogated data from 4309 patients with chronic-phase CML from 76 other centers. A total of 855 patients were excluded because of the interval from diagnosis to starting TKI therapy of ≥6 months (n = 66), advanced phase at diagnosis (n = 125), missing several

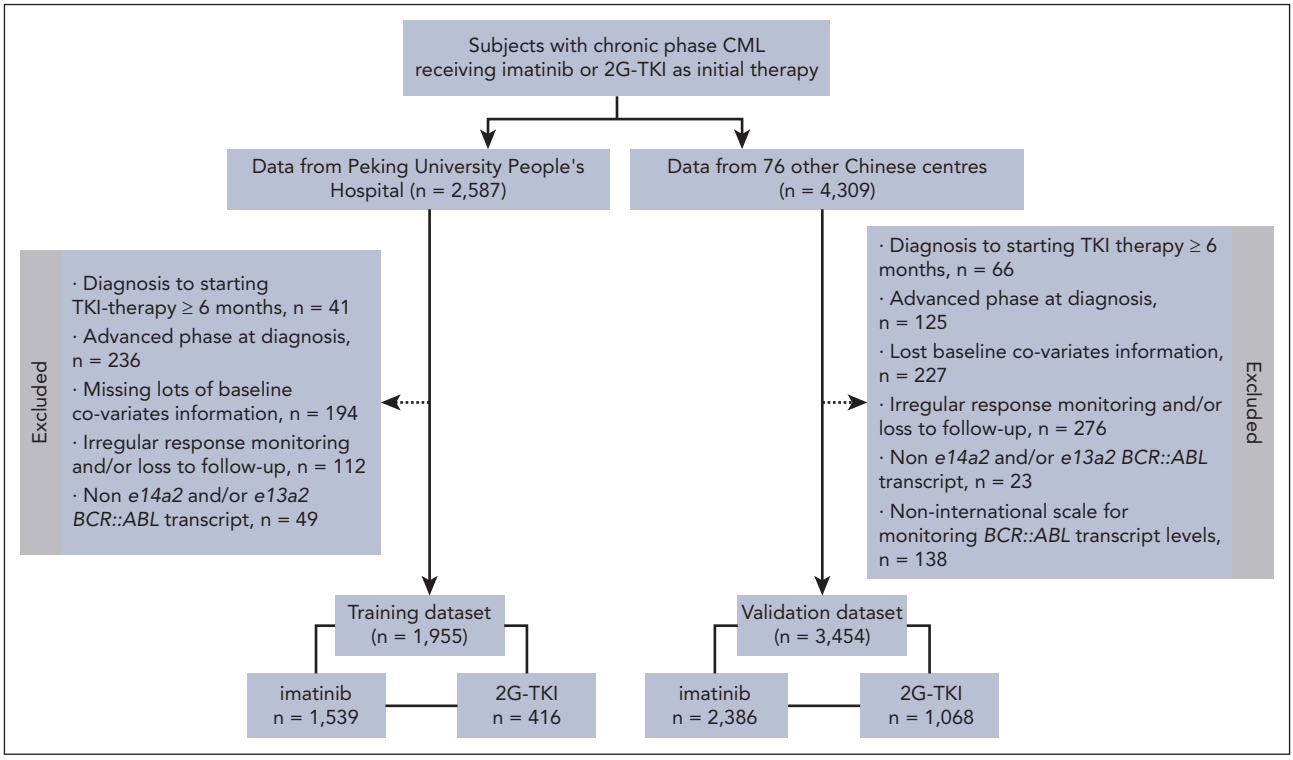


Figure 1. Study flowchart.

important baseline covariates (n = 227), irregular response monitoring and/or loss to follow-up (n = 276), non-*e14a2* and/or *e13a2* transcripts (n = 23), and non-IS monitoring (n = 138). The remaining 3454 patients received initial imatinib (n = 2386, 69%) and a 2G-TKI (n = 1068, 31%) including nilotinib (n = 620; 18%), dasatinib (n = 154; 4%), or flumatinib (n = 294, 9%). In the validation data set, 665 patients (19%) failed initial TKI therapy at a median of 9 months (IQR, 5-17) because of meeting therapy-failure criteria at 3 (n = 149), 6 (n = 107), or 12 (n = 87) months, loss of responses (n = 74), emergence of TKI-resistant *ABL* mutation(s) (n = 12) or high-risk ACAs in Ph^+ cells (n = 23), and transformation to advanced phase (n = 109) alone or combined (n = 104). A total of 60 (2%) patients died during TKI therapy, including 13 who died before therapy failure and 47 who died from disease transformation. With a median follow-up of 45 months (IQR, 26-76), 6-year probabilities of FFS, TFS, and survival were 75% (95% CI, 73-77), 96% (95% CI, 95-97), and 97% (95% CI, 95-99).

Comparison of the training and validation data sets

The distribution of baseline covariates in the validation data set differed from those in the training data set (Table 1). Patients in the validation data set were older ($P < .001$), and had a smaller spleen size below the costal margin ($P < .001$), lower percentages of blood blasts and basophils ($P < .001$), lower proportions of Sokal and ELTS high-risk patients ($P < .001$) and fewer comorbidities ($P < .001$) compared with patients in the training data set. Additionally, the median follow-up of 45 months (IQR, 26-76 months) in the validation data set was significantly-shorter than that in the training data set (median [IQR], 56 [30-91] months; $P < .001$). Probability of FFS in the validation data set was significantly-higher than that in the training data set ($P = .01$). TFS and survival were similar ($P = .25$ and .54; supplemental Figure 2).

Predictive model

Results of univariable analyses of predictive covariates in the training data set are displayed in supplemental Table 3. Multivariable Cox and Fine-Gray analyses of the 1701 patients with complete data for the 12 candidate covariates from the univariable analyses indicated that male sex, increasing age, lower hemoglobin concentration, higher percentage of blood blasts, larger spleen size below costal margin, and high-risk ACAs in Ph^+ cells at diagnosis were significantly associated with TKI-therapy failure with no significant interactions (supplemental Table 4).

After polynomial transformation, hemoglobin concentration was transformed to "$(\text{hemoglobin}/100)^{-2}$" in the regression model. Using the 6 covariates correlated with TKI-therapy failure in the 1762 evaluable patients with complete data, we repeated these analyses using Cox and Fine-Gray regression models with similar results (Table 2). Although the results of the Cox and Fine-Gray models had high concordance for therapy failure when there were competing risks, we used the Fine-Gray model that is more parsimonious and precise.[45-47] The Fine-Gray model, which included the 6 covariates, was used to develop the predictive model: initial TKI-therapy failure risk score = 0.1919 × sex (male = 1, female = 0) + 1.6160 × $(\text{age}/100) + 0.3105 \times (\text{hemoglobin concentration}/100)^{-2}$ + 0.1087 × blood blasts + 0.0671 × spleen size below costal margin + 0.5461 × high-risk ACA in Ph^+ cells (Y = 1, N = 0).

Table 1. Patient covariates at diagnosis

	Training data set (n = 1955)	Validation data set (n = 3454)	*P* value
Age, years, median (IQR)	40 (30-52)	42 (32- 54)	<.001
Male, n (%)	1195 (61)	2075 (60)	.45
Spleen size, cm below costal margin, median (IQR)	3 (0-10)	2 (0-7)	<.001
WBC, ×10E+9/L, median (IQR)	122 (47-235)	123 (51-225)	.59
Hemoglobin, ×10E+9/L, median (IQR)	115 (97-132)	111 (94-127)	<.001
Platelets, ×10E+9/L, median (IQR)	410 (270-635)	412 (263-625)	.99
Blood blasts, %, median (IQR)	1 (0-3)	1 (0-2)	<.001
Blood basophils, %, median (IQR)	5 (2-8)	4 (2-7)	<.001
Blood eosinophils, %, median (IQR)	2 (1-4)	2 (1-4)	.73
Sokal risk, n (%)			<.001
Low	819 (42)	1526 (44)	
Intermediate	555 (28)	1157 (33)	
High	394 (20)	543 (16)	
Unknown	187 (10)	228 (7)	
ELTS risk, n (%)			<.001
Low	1115 (57)	2207 (64)	
Intermediate	471 (24)	795 (23)	
High	182 (9)	224 (6)	
Unknown	187 (10)	228 (7)	
Ph$^+$ ACAs, n (%)	68 (3)	118 (3)	.90
High-risk ACAs, n (%)	41 (2)	51 (1)	.09
Comorbidity(ies), n (%)	700 (36)	694 (20)	<.001
Initial TKI therapy, n (%)			<.001
Imatinib	1539 (79)	2386 (69)	
2G TKI	416 (21)	1068 (31)	
Nilotinib	280 (14)	620 (18)	
Dasatinib	72 (4)	154 (4)	
Flumatinib	64 (3)	294 (9)	
Follow-up*, mo, median (IQR)	56 (30-91)	45 (26-76)	<.001

WBC, white blood cell.

*Censored at a transplant, death, or the last follow-up.

Identifying cutoffs for the predictive model

The TKI-therapy failure risk score was rounded to 4 decimal places. One-thousand bootstrap resamplings were performed for the 1762 patients with complete data. After confirmation of cutoffs with significance and data rationality by the Fine-Gray test, the minimal *P* value approach, and Bonferroni correction, 892 cutoffs remained. Smoothing functions of these cutoffs were determined by the kernel density plot (supplemental Figure 3). The kernel density plot for the cutoffs at the 2 highest peaks was then used to classify the patients into low- (score ≤ 1.3115; n = 716; 41%), intermediate- (1.3115 < score ≤ 2.4266; n = 812; 46%), and high-risk (score > 2.4266; n = 234; 13%) subgroups.

Cumulative incidence of therapy failure using the predictive model

In the training data set using the predictive model, 8-year cumulative incidences of therapy failure for the 3 risk subgroups were 10% (95% CI, 5-15), 34% (95% CI, 29-39), and 69% (95% CI, 63-75; *P* for trend < .001; Figure 2A), respectively. 8-year probabilities of FFS were 90% (95% CI, 87-93), 66% (95% CI, 62-70), and 32% (95% CI, 23-41; *P* for trend < .001; Figure 2B). Hazard ratios with the low-risk subgroup as reference were 3.8 (2.9, 5.0; *P* < .001) and 10.4 (7.7, 14.0; *P* < .001).

In the validation data set, 3218 patients with complete data for the 6 prognostic covariates in the model were included; 1179

Table 2. Multivariable analyses in the 1762 patients with complete data for the 6 prognostic covariates in the training data set

Covariates	Regression coefficient	HR (95% CI)	*P* value
Multivariable Cox analyses			
Male	0.2026	1.2 (1.0-1.5)	.02
Age/100, y	1.5668	4.8 (2.3-9.8)	<.001
(Hemoglobin/100)$^{-2}$, g/L	0.3108	1.4 (1.3-1.5)	<.001
Blood blasts, %	0.1085	1.1 (1.1-1.2)	<.001
Spleen size, cm below costal margin	0.0675	1.1 (1.0-1.1)	<.001
High-risk ACAs in Ph$^+$ cells*	0.5458	1.7 (1.1-2.8)	.03
Multivariable Fine-Gray analyses			
Male	0.1919	1.2 (1.0-1.5)	.01
Age/100, y	1.6160	5.0 (2.5-10.3)	<.001
(Hemoglobin/100)$^{-2}$, g/L	0.3105	1.4 (1.2-1.5)	<.001
Blood blasts, %	0.1087	1.1 (1.1-1.2)	<.001
Spleen size, cm below costal margin	0.0671	1.1 (1.0-1.1)	<.001
High-risk ACAs in Ph$^+$ cells*	0.5461	1.7 (1.1-2.8)	.02

HR, hazard ratio; N, no; Y, yes.

*High-risk ACAs in Ph$^+$ cells were scored as categorical covariates ("Y/N") in the regression models.

(37%), 1765 (55%), and 274 (8%) patients were classified as low-, intermediate-, and high-risk, respectively, using the predictive model. Six-year cumulative incidences of therapy failure for the 3 risk subgroups were 7% (95% CI, 2-12), 28% (95% CI, 24-32), and 54% (95% CI, 48-60; *P* for trend <.001; Figure 2C), respectively. Six-year probabilities of FFS were 93% (95% CI, 91-95), 70% (95% CI, 67-73), and 42% (95% CI, 35-49; *P* value for trend <.001; Figure 2D) respectively. Hazard ratios with the low-risk subgroup as reference were 5.4 (4.1, 7.2; *P* < .001) and 12.1 (8.8, 16.6; *P* < .001).

Predictive model performance

To evaluate model accuracy, we plotted time-dependent AUROCs for therapy failure at 1, 3, and 5 years in the training and validation data sets (supplemental Figure 4A,D). In the training data set, we found good prediction sensitivity and specificity with 1-, 3- and 5-year AUROCs of 0.83 (95% CI, 0.81-0.85), 0.84 (95% CI, 0.82-0.86), and 0.84 (95% CI, 0.82-0.87), respectively. Comparable AUROC values in the validation data set were 0.77 (95% CI, 0.74-0.79), 0.79 (95% CI, 0.77-0.82), and 0.80 (95% CI, 0.78-0.83), respectively. Calibration plots for the 1-, 3- and 5-year cumulative incidences of therapy failure indicated good concordance between the predicted and observed cumulative outcome incidences (supplemental Figure 4B,E). DCA curves indicated a net benefit from using the model (supplemental Figure 4C,F).

Predictive model application in the imatinib or 2G-TKI cohort

Patients with complete data from the training (n = 1762) and validation (n = 3218) data sets were integrated into the entire data set (n = 4980), in which 3621 patients receiving initial imatinib therapy were identified as the low- (n = 1389, 38%), intermediate- (n = 1907, 53%) and high-risk (n = 325, 9%) subgroups using the predictive model; 1359 patients receiving 2G-TKI therapy were identified as low- (n = 506, 37%), intermediate- (n = 670, 49%), and high-risk (n = 183, 14%) subgroups. With a median follow-up of 59 months (IQR, 37-92) in the imatinib cohort, 8-year cumulative incidences of therapy failure in the 3 risk subgroups were 11% (95% CI, 8-14%), 36% (95% CI, 32-40%), and 71% (95% CI, 67-75%; *P* < .001; Figure 3A), respectively. With a median follow-up of 28 months (IQR, 19-52) in the 2G-TKI cohort, 4-year cumulative incidences of therapy failure in the 3 risk subgroups were 5% (95% CI, 1-9%), 26% (95% CI, 20-32%), and 53% (95% CI, 46-60%; *P* < .001; Figure 3B), respectively.

Predictive model performance by age

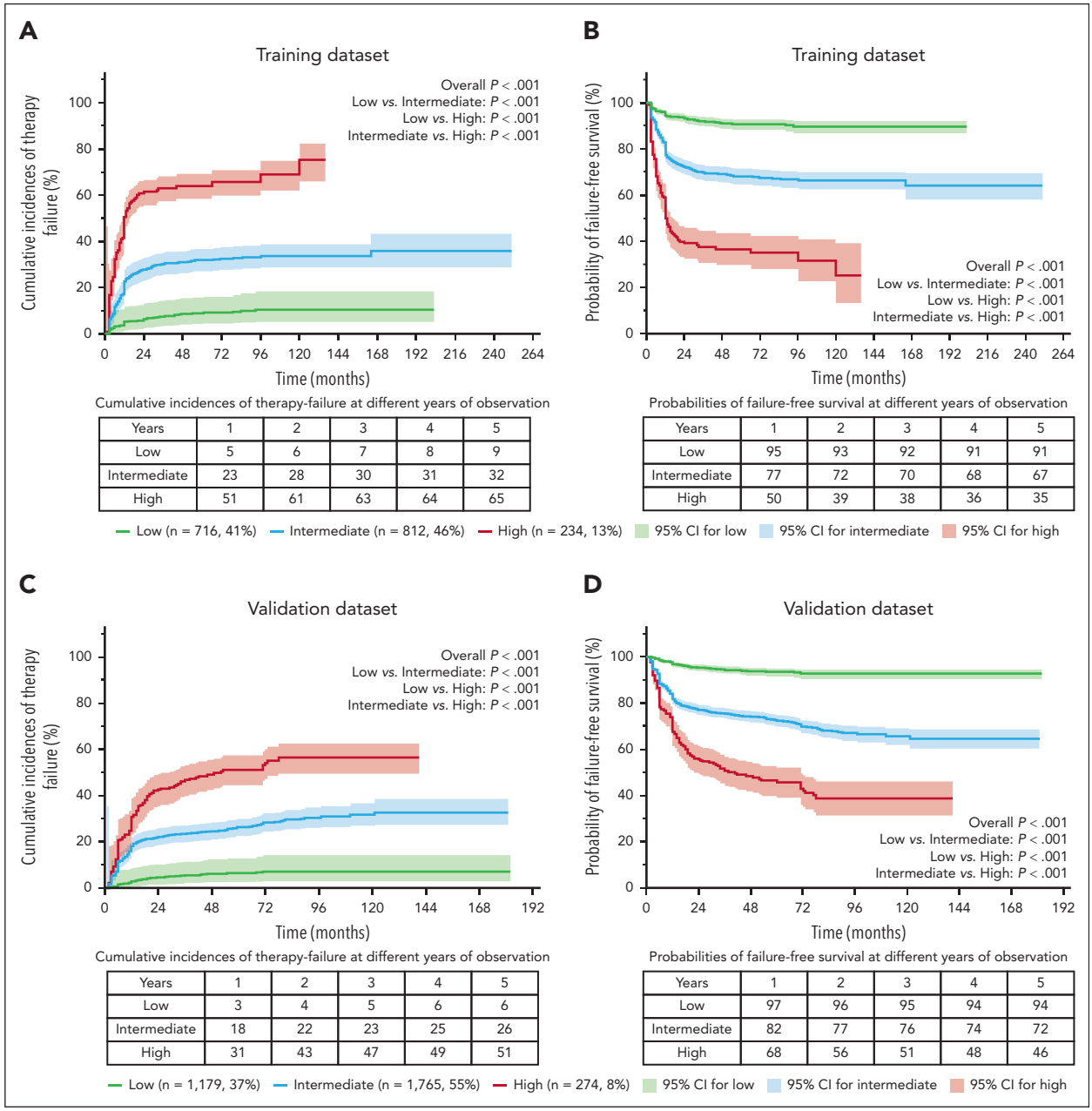
We evaluated the performance of the predictive model by age: <40 years (n = 2268, 46%), 40-60 years (n = 2022, 41%), and >60 years (n = 690, 13%). Regardless of the age subgroup, the predictive model demonstrated the excellent ability to predict TKI-therapy failure (supplemental Figure 5).

Predictive model for molecular responses, TFS, and CML-related survival

In the entire data set, 1895 (38%), 2577 (52%), and 508 (10%) patients were identified as the low-, intermediate- and high-risk, respectively, by the predictive model. There were significant differences in the cumulative incidences of MMR (*P* < .001), MR^4 (*P* < .001), and MR$^{4.5}$ (*P* = .001), as well as the probabilities of TFS (*P* < .001) and CML-related survival (*P* = .01) among the 3 subgroups (supplemental Figure 6).

Comparison of the predictive model with the Sokal and ELTS scores

A total of 4975 patients were simultaneously classified by the Sokal, ELTS, and the predictive models (Figure 4A-B). Although both the Sokal and ELTS scores could extensively



Years	1	2	3	4	5
Low	5	6	7	8	9
Intermediate	23	28	30	31	32
High	51	61	63	64	65

Years	1	2	3	4	5
Low	95	93	92	91	91
Intermediate	77	72	70	68	67
High	50	39	38	36	35

Years	1	2	3	4	5
Low	3	4	5	6	6
Intermediate	18	22	23	25	26
High	31	43	47	49	51

Years	1	2	3	4	5
Low	97	96	95	94	94
Intermediate	82	77	76	74	72
High	68	56	51	48	46

Figure 2. Cumulative incidences of the TKI-therapy failure and probabilities of FFS using the predictive model.

predict TKI-therapy failure (supplemental Figure 7), our model had better discriminatory ability than the Sokal and ELTS scores, with significantly greater AUROC values (predictive model: 0.71-0.77; Sokal: 0.63-0.68; and ELTS: 0.65-0.70; supplemental Figure 8A). This was especially the case in the 2G-TKI cohort (supplemental Figure 8C). DCA also indicated that the net benefit of using the predictive model was greater than that of using the Sokal or ELTS scores (supplemental Figure 9). Most importantly, 2340 and 3316 patients in the Sokal and ELTS low-risk subgroup, respectively, were reclassified into the low-, intermediate-, or high-risk subgroups using our model with the significantly different cumulative incidences of therapy failure (all *P* values < .001; Figure 4C,F). Results were similar in the Sokal or ELTS intermediate- and high-risk subgroups, indicating better predictive ability of the TKI-therapy failure model than the Sokal and ELTS scores (Figure 4D-H).

Intermediate- and high-risk patients identified by the predictive model receiving initial 2G-TKI therapy had lower therapy-failure rates compared with those receiving imatinib therapy

We compared the cumulative incidences of TKI-therapy failure in patients receiving initial imatinib or a 2G-TKI therapy in each model risk subgroup by the predictive model using PSM analyses to adjust for the differences in the baseline covariates (supplemental Tables 5-7). Intermediate- or high-risk patients identified by the predictive model receiving initial 2G-TKI

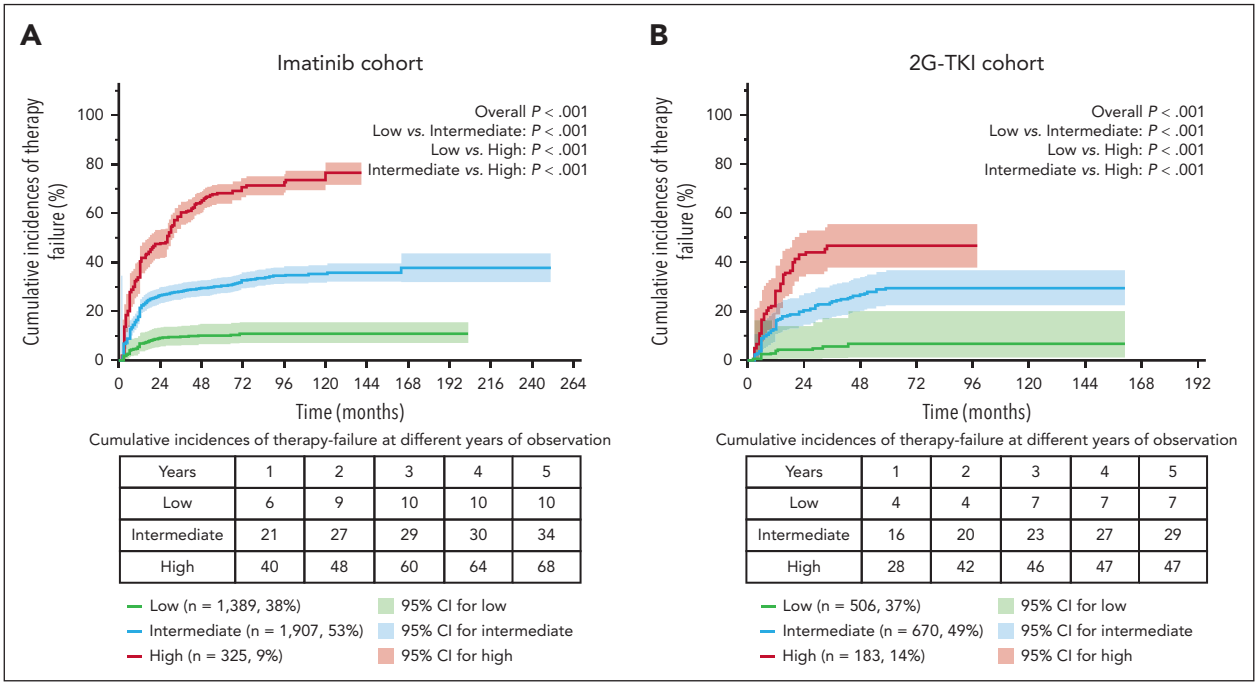


Figure 3. Cumulative incidences of the TKI-therapy failure in the imatinib and 2G-TKI cohorts using the predictive model.

therapy had significantly lower cumulative incidences of therapy failure than patients receiving initial imatinib therapy (all *P* values <.001; supplemental Figure 10). Low-risk patients treated with imatinib or a 2G-TKI had comparable therapy-failure rates (*P* = .79; supplemental Figure 10).

Discussion

Although >80% of patients with CML exhibit a long-term survival on TKI therapy, 20% to 30% of patients still experienced therapy failure. Patients with therapy failure generally have a poor prognosis and are often switched to an alternative TKI or even transplantation as salvage therapy.[14-19] Therefore, accurate prediction at diagnosis of which persons with chronic-phase CML will fail initial TKI therapy is important. We used data from 1955 patients to develop a TKI-therapy failure predictive model. The model identified 3 risk subgroups and had good predictive accuracy with the high time-dependent AUROC values. We validated our model in an independent data set of 3454 patients with similarly high time-dependent AUROC values.

Currently, the Sokal and ELTS scores, which were originally introduced to predict the survival or CML-related survival in the context of chemotherapy or TKI therapy, are the most widely used to guide initial TKI therapy in chronic-phase CML.[4,6,23,24,27] Recent studies reported that they can also be expanded to predict therapy failure and disease progression.[6,48,49] Consequently, we compared our model with these scores and unsurprisingly found that our model has better prediction accuracy. This finding does not imply that the predictive model can replace the Sokal and ELTS scoring systems. Survival remains the most crucial outcome for patients with CML, and the Sokal and ELTS scores were specifically developed for predicting survival. The more appropriate use of the predictive model is to further stratify persons who are identified as low or intermediate risk by the Sokal or ELTS score, making the risk assessment more precise. Importantly, regardless of which model identified a "high-risk" person, more attention should be given to therapeutic strategies and disease monitoring.

Despite the relatively low incidence of ACAs in Ph^+ cells at diagnosis in our study, which is consistent with previous studies,[28,50-52] high-risk ACAs rather than non–high-risk ACAs were strongly associated with TKI-therapy failure.

We found that patients in the 2G-TKI cohort had fewer late failure events than did those in the imatinib cohort. Possible reasons include the following: (1) 2G-TKIs are more effective resulting in a reduced progression risk, especially in intermediate- and high-risk patients as reported in previous studies.[3,53,54] We also found that intermediate- or high-risk patients identified by our model receiving 2G-TKI therapy had a lower therapy-failure rate than those receiving imatinib therapy. (2) Follow-up in the 2G-TKI cohort was shorter than in the imatinib cohort with fewer patients.

Cumulative incidence of therapy failure in the validation data set was lower than that in the training data set. However, patients in the validation data set were older, and had smaller spleen size below costal margin, lower percentages of blood blasts and basophils, and lower proportions of Sokal and ELTS high-risk patients.[6,13,55] Additionally, follow-up of the validation data set was shorter than that of the training data set. Regardless, our model had good prediction accuracy in both data sets.

Our study has important limitations. First, it was retrospective. However, the number of patients in the data sets was large. Second, our patients were relatively young compared with those with CML of predominantly European descent; therefore, the external validation from different races and countries is

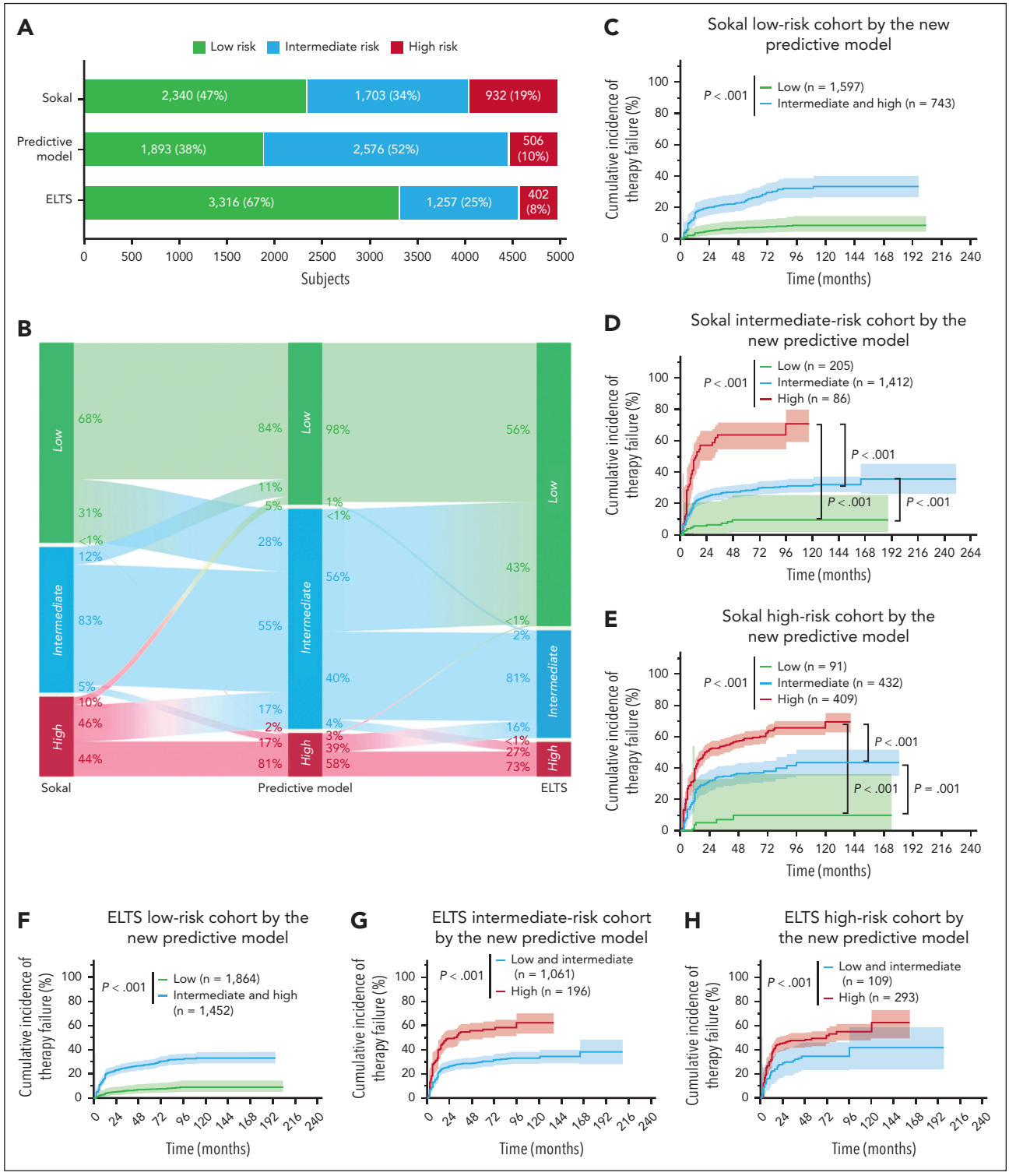


Figure 4. Comparison of the predictive discrimination for therapy failure among Sokal, ELTS, and the therapy-failure predictive models.

needed. Third, we were not able to strictly monitor the therapy compliance, especially in patients in the validation data set because of the large number of contributing centers. However, this heterogeneity increases the generalizability of our conclusions. Finally, PSM analyses can only adjust for known prognostic covariates and, although useful, are not a substitute for randomized controlled trials.

In conclusion, we developed and externally validated a TKI-therapy failure predictive model in persons with chronic-phase CML receiving initial TKI therapy with the high accuracy and precision. Intermediate- and high-risk patients identified by our model receiving 2G-TKI therapy had lower therapy-failure rate than those receiving imatinib. This conclusion needs further confirmation in a randomized controlled trial. Our model could

help physicians estimate the likelihood of therapy failure of initial TKI therapy and help physicians choose the best initial TKI.

Acknowledgments

The authors thank the following medical staff who provided patient data: Ru Feng, Junxia Meng, Xiaonan Zhang, Jijun Wang, Lan Ma, Rongxia Wei, Zhiqiang Sun, Yun Zeng, Xiaoyi Lv, Zhen Xiao, Xingxia Zhang, Yanping Ma, Xiaoyan Ge, Jianmin Luo, Yujuan Yang, Congmeng Lin, Pengliang Xin, Hai Yi, Yilan Liu, Liling Zheng, Hebing Zhou, Luoming Hua, Wangxiang Huang, Yonghuai Feng, Yu Jing, Lijun Wang, Linhua Yang, Hongguo Zhao, Hairong Fei, Qin Li, Jianhui Qiao, Hua Fan, Shasha Zhao, Ronghua Hu, Jianli Wang, Yin Sai, Wenqian Li, Lianyu Shen, Li Ding, Fang Ye, Min Ouyang, Shunjie Wu, and Liru Wang.

R.P.G. acknowledges support from the UK National Institute for Health Research Biomedical Research Centre. Q.J. acknowledges support from the National Natural Science Foundation of China (grants 81970140 and 82370161). This study was funded by the National Nature Science Foundation of China (grants 81970140 and 82370161).

Authorship

Contribution: Q.J. and X.H. designed the study; Q.J., X.Z., B.L., J. Huang, Y. Zhang, N.X., and R.P.G. analyzed the data and prepared the manuscripts; Q.J., W.L., X.L., Y.Y., H.L., B.L., X.D., R.L., C.C., J. Huang, H.Z., L. Pan, X.W., G.L., Zhuogang Liu, Y.Z., Zhenfang Liu, J. Hu, C.L., F.L., W.Y., L.M., Y.H., L.L., Z. Zhao, C.T., C.Z., Y.B., Z. Zhou, S.C., H.Q., L.Y., X. Sun, H.S., L.Z., Z.L., D.W., J.G., L. Pang, Q.Z., X. Suo, W.Z., and Y. Zheng provided most of the patient data in this study and revised the manuscript; and all the authors approved the final manuscript, were responsible for the content, and agreed to its submission for publication.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: J.H., 0000-0003-4200-8636; R.P.G., 0000-0002-9156-1676; Y.Y., 0000-0002-2924-7002; Y.Z., 0000-0003-0156-699X; Y.H., 0000-0002-2658-0803; S.C., 0000-0001-6294-972X; Q.Z., 0000-0002-6887-5255; Q.J., 0000-0001-7131-0522.

Correspondence: Qian Jiang, Peking University People's Hospital, Peking University Institute of Hematology National Clinical Research Center for Hematologic Disease, Beijing Key Laboratory of Hematopoietic Stem Cell Transplantation, No. 11 Xizhimen South St, Beijing 100044, People's Republic of China; email: jiangqian@medmail.com.cn; and Xiaojun Huang, Peking University People's Hospital, Peking University Institute of Hematology, National Clinical Research Center for Hematologic Disease, Beijing Key Laboratory of Hematopoietic Stem Cell Transplantation, Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, State Key Laboratory of Natural and Biomimetic Drugs, Peking University, No. 11 Xizhimen South St, Beijing 100044, People's Republic of China; email: huangxiaojun@bjmu.edu.cn.

Footnotes

Submitted 22 March 2024; accepted 5 July 2024; prepublished online on *Blood* First Edition 24 July 2024. https://doi.org/10.1182/blood.2024024761.

*X.Z., B.L., J. Huang, Y. Zhang, and N.X. contributed equally to this study.

For original data, please contact corresponding author Qian Jiang (jiangqian@medmail.com.cn).

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

REFERENCES

1. Hehlmann R, Lauseker M, Saußele S, et al. Assessment of imatinib as first-line treatment of chronic myeloid leukemia: 10-year survival results of the randomized CML study IV and impact of non-CML determinants. *Leukemia*. 2017;31(11):2398-2406.
2. Hochhaus A, Larson RA, Guilhot F, et al. Long-term outcomes of imatinib treatment for chronic myeloid leukemia. *N Engl J Med*. 2017;376(10):917-927.
3. Kantarjian HM, Hughes TP, Larson RA, et al. Long-term outcomes with frontline nilotinib versus imatinib in newly diagnosed chronic myeloid leukemia in chronic phase: ENESTnd 10-year analysis. *Leukemia*. 2021;35(2):440-453.
4. Pfirrmann M, Baccarani M, Saussele S, et al. Prognosis of long-term survival considering disease-specific death in patients with chronic myeloid leukemia. *Leukemia*. 2016;30(1):48-56.
5. Pfirrmann M, Lauseker M, Hoffmann VS, Hasford J. Prognostic scores for patients with chronic myeloid leukemia under particular consideration of competing causes of death. *Ann Hematol*. 2015;94(suppl 2):S209-S218.
6. Zhang XS, Gale RP, Huang XJ, Jiang Q. Is the Sokal or EUTOS long-term survival (ELTS) score a better predictor of responses and outcomes in persons with chronic myeloid leukemia receiving tyrosine-kinase inhibitors? *Leukemia*. 2022;36(2):482-491.
7. Boquimpani C, Schaffel R, Biasoli I, Bendit I, Spector N. Molecular responses at 3 and 6 months after switching to a second-generation tyrosine kinase inhibitor are complementary and predictive of long-term outcomes in patients with chronic myeloid leukemia who fail imatinib. *Leuk Lymphoma*. 2015;56(6):1787-1792.
8. Dou X, Zheng F, Zhang L, et al. Adolescents experienced more treatment failure than children with chronic myeloid leukemia receiving imatinib as frontline therapy: a retrospective multicenter study. *Ann Hematol*. 2021;100(9):2215-2228.
9. Ohanian M, Kantarjian HM, Shoukier M, et al. The clinical impact of time to response in de novo accelerated-phase chronic myeloid leukemia. *Am J Hematol*. 2020;95(10):1127-1134.
10. Yu L, Wang H, Gale RP, et al. Impact of socio-demographic co-variates on prognosis, tyrosine kinase-inhibitor use and outcomes in persons with newly-diagnosed chronic myeloid leukaemia. *J Cancer Res Clin Oncol*. 2022;148(2):449-459.
11. Zhang X, Gale RP, Liu B, et al. Validation of the imatinib-therapy failure model. *Leukemia*. 2023;37(5):1166-1169.
12. Zhang X, Xu N, Yang Y, et al. Comparison of the efficacy among nilotinib, dasatinib, flumatinib and imatinib in newly diagnosed chronic-phase chronic myeloid leukemia patients: a real-world multi-center retrospective study. *Clin Lymphoma Myeloma Leuk*. 2024;24(6):e257-e266.
13. Zhang XS, Gale RP, Zhang MJ, Huang XJ, Jiang Q. A predictive scoring system for therapy-failure in persons with chronic myeloid leukemia receiving initial imatinib therapy. *Leukemia*. 2022;36(5):1336-1342.
14. Hanfstein B, Müller MC, Hehlmann R, et al. Early molecular and cytogenetic response is predictive for long-term progression-free and overall survival in chronic myeloid leukemia (CML). *Leukemia*. 2012;26(9):2096-2102.
15. Hughes TP, Hochhaus A, Branford S, et al. Long-term prognostic significance of early molecular response to imatinib in newly diagnosed chronic myeloid leukemia: an analysis from the International Randomized Study of Interferon and STI571 (IRIS). *Blood*. 2010;116(19):3758-3765.
16. Marin D, Ibrahim AR, Lucas C, et al. Assessment of BCR-ABL1 transcript levels at 3 months is the only requirement for predicting outcome for patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors. *J Clin Oncol*. 2012;30(3):232-238.

17. Marin D, Milojkovic D, Olavarria E, et al. European LeukemiaNet criteria for failure or suboptimal response reliably identify patients with CML in early chronic phase treated with imatinib whose eventual outcome is poor. *Blood*. 2008;112(12):4437-4444.

18. Milojkovic D, Nicholson E, Apperley JF, et al. Early prediction of success or failure of treatment with second-generation tyrosine kinase inhibitors in patients with chronic myeloid leukemia. *Haematologica*. 2010; 95(2):224-231.

19. Wang L, Pearson K, Ferguson JE, Clark RE. The early molecular response to imatinib predicts cytogenetic and clinical outcome in chronic myeloid leukaemia. *Br J Haematol*. 2003;120(6):990-999.

20. Baccarani M, Cortes J, Pane F, et al. Chronic myeloid leukemia: an update of concepts and management recommendations of European LeukemiaNet. *J Clin Oncol*. 2009;27(35): 6041-6051.

21. Baccarani M, Deininger MW, Rosti G, et al. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. *Blood*. 2013;122(6):872-884.

22. Baccarani M, Saglio G, Goldman J, et al. Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood*. 2006;108(6): 1809-1820.

23. Hochhaus A, Baccarani M, Silver RT, et al. European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia*. 2020;34(4): 966-984.

24. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*. 2022;36(7):1703-1719.

25. Shah NP, Bhatia R, Altman JK, et al. Chronic myeloid leukemia, version 2.2024, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2024;22(1):43-69.

26. Ielo C, Scalzulli E, Carmosino I, et al. Validation of imatinib therapy failure score (IMTF) in chronic phase chronic myeloid leukemia in real life practice. *Leuk Lymphoma*. 2023;64(14):2324-2326.

27. Sokal JE, Cox EB, Baccarani M, et al. Prognostic discrimination in "good-risk" chronic granulocytic leukemia. *Blood*. 1984; 63(4):789-799.

28. Wang W, Cortes JE, Tang G, et al. Risk stratification of chromosomal abnormalities in chronic myelogenous leukemia in the era of tyrosine kinase inhibitor therapy. *Blood*. 2016;127(22):2742-2750.

29. Qin YZ, Jiang Q, Jiang H, et al. Which method better evaluates the molecular response in newly diagnosed chronic phase chronic myeloid leukemia patients with imatinib treatment, BCR-ABL(IS) or log reduction from the baseline level? *Leuk Res*. 2013;37(9):1035-1040.

30. Guilhot J, Baccarani M, Clark RE, et al. Definitions, methodological and statistical issues for phase 3 clinical trials in chronic myeloid leukemia: a proposal by the European LeukemiaNet. *Blood*. 2012; 119(25):5963-5971.

31. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology–with an emphasis on fractional polynomials. *Methods Inf Med*. 2005;44(4): 561-571.

32. Sauerbrei W, Meier-Hirmer C, Benner A, et al. Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. *Comput Stat Data Anal*. 2006;50(12):3464-3485.

33. Kuk D, Varadhan R. Model selection in competing risks regression. *Stat Med*. 2013; 32(18):3077-3088.

34. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods*. 2012;17(2):228-243.

35. Holl?Nder N, Schumacher MJO. On the problem of using 'optimal' cutpoints in the assessment of quantitative prognostic factors. *Onkologie*. 2001;24(2):194-199.

36. Kulesa A, Krzywinski M, Blainey P, Altman N. Sampling distributions and the bootstrap. *Nat Methods*. 2015;12(6):477-478.

37. Curtin F, Schulz P. Multiple correlations and Bonferroni's correction. *Biol Psychiatry*. 1998; 44(8):775-777, 775.

38. Polley MYC, Dignam JJ. Statistical considerations in the evaluation of continuous biomarkers. *J Nucl Med*. 2021; 62(5):605-611.

39. Wang S, Wang J, Chung FL. Kernel density estimation, kernel methods, and fast learning in large data sets. *IEEE Trans Cybern*. 2014; 44(1, 20).

40. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167-176.

41. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol*. 2017; 17(1):53.

42. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur Urol*. 2018;74(6):796-804.

43. Badhiwala JH, Karmur BS, Wilson JR. Propensity score matching: a powerful tool for analyzing observational nonrandomized data. *Clin Spine Surg*. 2021;34(1):22-24.

44. Kane LT, Fang T, Galetta MS, et al. Propensity score matching: a statistical method. *Clin Spine Surg*. 2020;33(3): 120-122.

45. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391-4400.

46. Nolan EK, Chen HY. A comparison of the Cox model to the Fine-Gray model for survival analyses of re-fracture rates. *Arch Osteoporos*. 2020;15(1):86.

47. Scrucca L, Santucci A, Aversa F. Regression modeling of competing risk using R: an in depth guide for clinicians. *Bone Marrow Transplant*. 2010;45(9):1388-1395.

48. Sato E, Iriyama N, Tokuhira M, et al. The EUTOS long-term survival score predicts disease-specific mortality and molecular responses among patients with chronic myeloid leukemia in a practice-based cohort. *Cancer Med*. 2020;9(23):8931-8939.

49. Yang X, Bai Y, Shi M, et al. Validation of the EUTOS long-term survival score in chinese chronic myeloid leukemia patients treated with imatinib: a multicenter real-world study. *Cancer Manag Res*. 2020;12:1293-1301.

50. Clark RE, Apperley JF, Copland M, Cicconi S. Additional chromosomal abnormalities at chronic myeloid leukemia diagnosis predict an increased risk of progression. *Blood Adv*. 2021;5(4):1102-1109.

51. Gong Z, Medeiros LJ, Cortes JE, et al. Cytogenetics-based risk prediction of blastic transformation of chronic myeloid leukemia in the era of TKI therapy. *Blood Adv*. 2017;1(26): 2541-2552.

52. Hehlmann R, Voskanyan A, Lauseker M, et al. High-risk additional chromosomal abnormalities at low blast counts herald death by CML. *Leukemia*. 2020;34(8): 2074-2086.

53. Cortes JE, Saglio G, Kantarjian HM, et al. Final 5-year study results of DASISION: The Dasatinib Versus Imatinib Study in Treatment-Naïve Chronic Myeloid Leukemia Patients trial. *J Clin Oncol*. 2016;34(20):2333-2340.

54. Zhang L, Meng L, Liu B, et al. Flumatinib versus imatinib for newly diagnosed chronic phase chronic myeloid leukemia: a phase iii, randomized, open-label, multi-center FESTnd Study. *Clin Cancer Res*. 2021;27(1):70-77.

55. Hoffmann VS, Baccarani M, Lindoerfer D, et al. The EUTOS prognostic score: review and validation in 1288 patients with CML treated frontline with imatinib. *Leukemia*. 2013;27(10):2016-2022.