# Development and validation of a model for the early prediction of progression from essential thrombocythemia to post-essential thrombocythemia myelofibrosis: a multicentre retrospective study

Danhong Xiang,[a,b,o,p] Xiudi Yang,[a,o,p] Honglan Qian,[c,p] Li Zhang,[d] Yanxia Han,[e] Yongcheng Sun,[f] Ying Lu,[g] Yu Chen,[h] Dan Cao,[i] Meiwei Hu,[j] Lifeng Wang,[k] Qinli Tang,[l] Dijiong Wu,[m] Guoyan Tian,[n] Hongyan Tong,[a,o,*] Jie Jin,[a,o,**] and Jian Huang[a,o,***]

[a]Department of Haematology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[b]Department of Haematology, The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, Zhejiang, China
[c]Department of Haematology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China
[d]Department of Haematology, Taizhou Hospital of Zhejiang Province Affiliated with Wenzhou Medical University, Taizhou, Zhejiang, China
[e]Department of Haematology, The Second Hospital of Jiaxing, Jiaxing, Zhejiang, China
[f]Department of Haematology, The First Affiliated Hospital of Ningbo University, Ningbo, Zhejiang, China
[g]Department of Haematology, The Affiliated People's Hospital of Ningbo University, Ningbo, Zhejiang, China
[h]Department of Haematology, College of Medicine, Lishui Hospital, Zhejiang University, Lishui, Zhejiang, China
[i]Department of Haematology, Huzhou Central Hospital, Huzhou, Zhejiang, China
[j]Department of Haematology, The Second Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China
[k]Department of Haematology, The Affiliated Xiaoshan Hospital, Hangzhou Normal University, Hangzhou, China
[l]Department of Haematologist, The First People's Hospital of PingHu, PingHu, Zhejiang, China
[m]Department of Haematology, The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China
[n]Department of Haematology, The Affiliated Hospital of Hangzhou Normal University, Hangzhou, Zhejiang, China
[o]Zhejiang Clinical Medical Research Centre of Haematology, Hangzhou, Zhejiang, China

eClinicalMedicine
2024;67: 102378

Published Online xxx
https://doi.org/10.1016/j.eclinm.2023.102378

## Summary

**Background** Essential thrombocythemia (ET), a myeloproliferative neoplasm (MPN), has a substantial risk of evolving into post-essential thrombocythemia myelofibrosis (post-ET MF). This study aims to establish a prediction nomogram for early prediction of post-ET MF in ET patients.

**Methods** The training cohort comprised 558 patients from 8 haematology centres between January 1, 2010, and May 1, 2023, while the external validation cohort consisted of 165 patients from 6 additional haematology centres between January 1, 2010, and May 1, 2023. Univariable and multivariable Cox regression analysis was performed to identified independent risk factors and establish a nomogram to predict the post-ET MF free survival. Both bias-corrected area under the curve (AUC), calibration curves and concordance index (C-index) were employed to assess the predictive accuracy of the nomogram.

**Findings** Multivariate Cox regression demonstrated that elevated red blood cell distribution width (RDW), elevated levels of lactate dehydrogenase (LDH) and the level of haemoglobin (Hb), a history of smoking and the presence of splenomegaly were independent risk factors for post-ET MF. The C-index displayed of the training and validation cohorts were 0.877 and 0.853. The 5 years, 10 years AUC values in training and external validation cohorts were 0.948, 0.769 and 0.978, 0.804 respectively. Bias-corrected curve is close to the ideal curve and revealed a strong consistency between actual observation and prediction.

**Interpretation** We developed a nomogram capable of predicting the post-ET MF free survival probability at 5 years and 10 years in ET patients. This tool helps doctors identify patients who need close monitoring and appropriate counselling.

*Corresponding author. Department of Haematology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310003, Zhejiang, China.
**Corresponding author. Department of Haematology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310003, Zhejiang, China.
***Corresponding author. Department of Haematology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310003, Zhejiang, China.
*E-mail addresses:* tonghongyan@zju.edu.cn (H. Tong), jiej0503@zju.edu.cn (J. Jin), househuang@zju.edu.cn (J. Huang).
[p]Co-first author.

Funding This research was funded by the Key R&D Program of Zhejiang (No. 2022C03137); the Public Technology Application Research Program of Zhejiang, China (No. LGF21H080003); and the Zhejiang Medical Association Clinical Medical Research special fund project (No. 2022ZYC-D09).



Keywords: Essential thrombocythemia; Post-essential thrombocythemia myelofibrosis; Prediction nomogram

## Research in context

### Evidence before this study

We conducted a comprehensive search of PubMed and Google Scholar databases to identify relevant articles published from January 1, 2000, to May 31, 2022. The search terms employed were "post essential thrombocythemia myelofibrosis (post-ET MF)," "essential thrombocythemia (ET)," and "predictive modelling." Existing studies on ET have predominantly centred around its clinical manifestations, treatment modalities, and thromboprophylaxis. However, to the best of our knowledge, no predictive model specifically addressing the development of post-ET MF has been established thus far.

### Added value of this study

Our analysis revealed several independent predictors of ET progression to post-ET MF, including ultrasound-determined splenomegaly, smoking history, lactate dehydrogenase level, haemoglobin level, and abnormal red blood cell distribution width. To facilitate risk assessment in patients, we developed a personalized and visually intuitive nomogram model. External validation demonstrated the model's accuracy, consistency, and superior net benefit when compared to alternative approaches.

### Implications of all the available evidence

The utilization of our model equips clinicians with a straightforward and user-friendly predictive tool, thereby facilitating enhanced surveillance of high-risk populations and ultimately leading to prolonged survival among patients with ET. Furthermore, these findings serve as a foundation for further investigation into the underlying mechanisms contributing to the development of post-ET MF.

## Introduction

Essential thrombocythemia (ET) is a Ph-negative myeloproliferative neoplasm (MPN), and although it was once regarded as relatively benign, it exhibits a predisposition to evolve into secondary myelofibrosis (SMF), also called post-essential thrombocythemia myelofibrosis (post-ET MF).

The progression of ET is influenced by various factors, including genetic alterations such as allelic mutations in tumour drivers and the presence of cellular drivers that contribute to clonal haematopoiesis. Additionally, mutations in genes encoding epigenetic regulators, transcriptional regulators, splicing factors, and tumour suppressors can enhance the clonal dominance of malignant cells.[1] Besides genetic factors, the disease's progression is affected by cytokine-driven inflammation originating from malignant clones and non-hematopoietic stem cells within the bone marrow microenvironment. Inflammation-induced myelofibrosis involves excessive fibrotic tissue production by myofibroblasts, leading to impaired normal haematopoiesis and increased cytokine release, further promoting clonal expansion.[2] Recent research has identified several serum interleukins that reflect in vivo inflammation and may serve as potential predictors of ET progression to post-ET MF.[3,4]

Myelofibrosis is typified by megakaryocytic proliferation and atypia, varying degrees of bone marrow fibrosis, and extramedullary haematopoiesis accompanied by splenomegaly. The International Working Group for MPN Research and Treatment (IWG-MRT) has developed diagnostic criteria for post-ET MF. These criteria include a documented diagnosis of ET according to the World Health Organization (WHO) diagnostic criteria, the presence of grade greater than two bone marrow fibrosis, and two or more additional features such as anaemia, leucoerythroblastic, palpable splenomegaly, elevated serum lactate dehydrogenase (LDH) levels, and constitutional symptoms.

In patients with ET, it is crucial to not only focus on preventing thrombosis but also to prioritize early treatment of post-ET MF. Progression to SMF significantly impacts the lifespan and survival of individuals afflicted by ET, with a lifetime incidence rate of 13%.[5] Ensuring the improvement of overall survival and prognosis for patients necessitates timely detection of disease progression, which requires diligent and regular monitoring. ET is a chronic haematological disorder that often has minimal impact on the overall quality of life. As a result, patients may underestimate the importance of regular follow-up and may be hesitant to undergo bone marrow examinations. Performing risk assessments during the initial diagnosis helps identify patients who require close monitoring, appropriate counselling, intensified follow-up, and regular evaluation, particularly those at high risk. This approach allows

for prompt adjustments to treatment strategies, which play a crucial role in slowing disease progression and prolonging patient survival.

Reported clinical attributes indicative of SMF encompass a range of factors, including but not limited to advanced age, leucocytosis, anaemia development, elevated LDH levels, splenomegaly, ASXL1 mutations,[6–9] and JAK2V617F allele mutation.[10] While these clinical features provide some insight into the assessment of the risk of progression to post-ET MF, there is currently no comprehensive risk model to quantitatively evaluate the likelihood of progression of primary ET patients. The development of predictive models that incorporate multiple risk factors and assess their clinical utility is crucial for accurately predicting the risk of progression to post-ET MF. These models aid in identifying individuals with a high likelihood of progression at an early stage, enabling close monitoring and appropriate counselling. Early detection allows for timely adjustments in treatment strategies, ultimately leading to improved life expectancy. The objective of this study was to identify and evaluate the risk factors associated with the development of post-ET MF, construct prediction models, and assess their accuracy and effectiveness. Additionally, clinical decision curve analysis was employed to assess the clinical value of the developed models.

## Methods

### Study design and patients

This study was designed as a retrospective investigation of the data from enrolled ET patients at Zhejiang province.

We conducted a retrospective study involving patients from eight haematology centres in Zhejiang Province from January 1, 2010, to May 1, 2023. Out of the 744 patients initially included, 122 were identified as prefibrotic primary myelofibrosis (PMF) based on reassessment of bone marrow biopsies at the pathology centre. After excluding 64 patients, a total of 558 patients who met the WHO 2016 diagnostic criteria[11] were finally included in the training cohort. The external validation cohort comprised 165 patients from six haematology centres in Zhejiang Province from January 1, 2010, to May 1, 2023. The inclusion criteria were as follows: (i). meet the diagnostic criteria for ET according to WHO 2016 diagnostic criteria, (ii). complete clinical data. The exclusion criteria were: (i). bone marrow pathology indicating prefibrotic PMF, (ii). miss records at the time of diagnosis or lost follow-up visits after diagnosis. Post-ET MF was diagnosed according to the criteria of the IWG-MRT.[12] All patients in this study were Chinese.

### Ethics statement

The study protocol was approved by the Ethics Committee of the First Affiliated Hospital, School of Medicine, Zhejiang University Institutional Review Board with a waiver for informed consent (2022, No492).

### Data collection

We selected possible risk factors suggestive of progression to post-ET MF and indicators associated with poor prognosis in haematological tumours from previous studies.[13–15] So, the data we collected included the patient's sex, age, symptoms, smoking history, complete blood count (CBC), uric acid, LDH, types of genetic mutation, and ultrasonic splenomegaly (thickness exceeding 4 cm and a longitudinal diameter exceeding 12 cm). In the process of grouping continuous variables, we categorized RDW into two groups using a cut-off value of 15%. For haemoglobin values, we utilized a median value of 130 g/L as the reference point and set 120 g/L as the cut-off threshold. Consequently, the haemoglobin values were classified into three groups for analysis: <120 g/L, 120 g/L to 130 g/L, and ≥130 g/L.

### Development and assessment of the nomogram

Univariate Cox regression was used to assess the association between clinical characteristics and progression to post-ET MF. Independent predictors were assessed by multivariate Cox regression (entry criterion: $P < 0.050$, elimination criterion: $P > 0.10$) and then recruited to develop a nomogram for predicting the 5 years and 10 years post-ET MF free survival probability.

A nomogram is a valuable tool as it visually represents predicted probabilities on a scale of 0–100, making it user-friendly. The total points assigned to different covariates correspond to the predicted probability for a patient. In cases where statistical significance is not considered, the variable with the highest absolute value of beta will be assigned 100 points on the scale, while the remaining variables will be assigned a smaller number of points proportional to their effect size.

Total points were obtained by summing the points obtained for each predictor. By drawing a vertical line from the axis of the total points, the estimated post-ET MF free survival probability can be abstained. The receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC), and concordance index (C-index) were used to assess the ability of the nomogram to distinguish post-ET MF. Non-parametric bootstraps of 1000 resample were performed, and calibration curves were created to assess the agreement between the actual observations and nomogram predictions. Decision curve analysis (DCA) was conducted to assess the clinical utility of the nomogram by quantifying the net benefit at different threshold probabilities. The use of DCA represents a novel approach for evaluating the performance of predictive models. DCA combines the mathematical simplicity of accuracy measures, such as sensitivity and specificity, with the clinical applicability of decision analysis methods. DCA

curves are constructed by considering the sensitivity and specificity of the dataset. By assessing the net clinical benefit, DCA allows for a comparison of different strategies for patient intervention against a default strategy of intervention all or no patients.[16,17] The nomogram was applied to an external validation cohort to further assess its stability using C-index, AUC, calibrations and DCA.

### Statistical analysis

The sample size was estimated using PASS version 11.0.7. All data were analysed using SPSS software (version R 26.0.0.0) and the rms (version 6.7–1), timeROC (version 0.4), survival (version 3.5–5), ggplot 2 (version 3.4.4), and dcurves (version 0.4.0) packages of R software version 4.2.2. Descriptive statistics are reported as frequencies for categorical variables and medians and interquartile ranges (IQR) for continuous variables. Data comparison was performed using the Mann–Whitney U-test for continuous variables and the $\chi 2$ teat for categorical variables.

### Role of the funding source

This research was funded by the Key R&D Program of Zhejiang (No. 2022C03137); the Public Technology Application Research Program of Zhejiang, China (No. LGF21H080003); and the Zhejiang Medical Association Clinical Medical Research special fund project (No. 2022ZYC-D09). All authors had full access to all the data in the study and accept responsible to submit for publication.

## Results

### General characteristics

We collected data from a total of 622 patients after excluding 122 patients with prefibrotic PMF, 64 of whom were excluded due to ineligibility or incomplete data. Among the 64 excluded patients, 47 had only bone marrow smear results and were missing bone marrow biopsy results, 5 had missing complete blood count records at the time of diagnosis, 4 had undergone splenectomy due to trauma before diagnosis, 3 had missing complete blood count and imaging results of the spleen, and 5 had only received a diagnosis and lacked records of subsequent visits (Fig. 1). Among the eligible cases, 558 patients were included in the training cohort. 250 (44.8%) patients were male, and the median age was 57 (IQR 42–67) years. The patients in this study were followed for a median duration of 6.2 years (range 0–13.5 years). The external validation cohort comprised 165 patients from six haematology centres in Zhejiang Province from January 1, 2010, to May 1, 2023. with a media follow-up of 5.7 years (range 0–13.5 years). Furthermore. We conducted a comparison of baseline data between the training and validation cohort. Our analysis revealed that the two cohorts exhibited

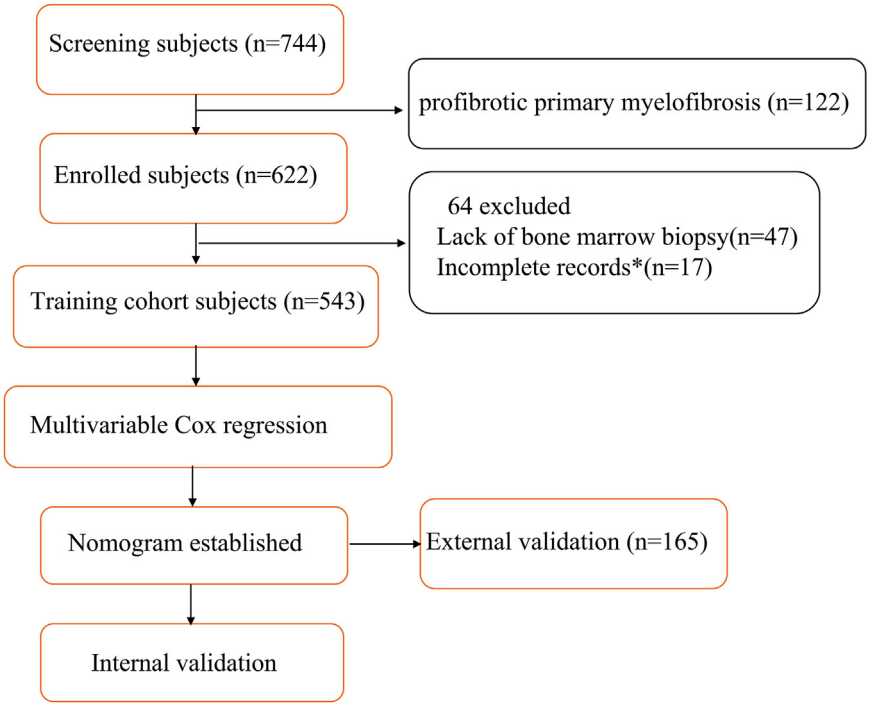


***Fig. 1:*** **The workflow of this study**. *Complete blood count (n = 5), splenectomy (n = 4), complete blood count and imaging results of the spleen (n = 3), records of subsequent visits (n = 5).

| Variables | Training cohort (n = 558) M (P25, P75)/N (%) | External validation cohort (n = 165) M (P25, P75)/N (%) | P-value |
|---|---|---|---|
| Age (years) | 57 (42–67) | 59 (46–69) | 0.11 |
| Sex | | | 0.98 |
| Male | 250 (44.8%) | 71 (44.8%) | |
| Female | 308 (55.2%) | 94 (55.2%) | |
| Splenomegaly | | | 0.022 |
| Yes | 99 (17.7%) | 17 (10.3%) | |
| No | 459 (82.3%) | 148 (89.7%) | |
| Smoking | | | 0.28 |
| Yes | 86 (15.4%) | 28 (16.9%) | |
| No | 472 (84.6%) | 137 (83.1%) | |
| Symptom | | | 0.61 |
| Yes | 141 (25.3%) | 39 (23.6%) | |
| No | 417 (74.7%) | 126 (76.4%) | |
| JAK2V617F mutation | | | 0.14 |
| Yes | 307 (57.0%) | 102 (63.0%) | |
| No | 232 (43.0%) | 60 (37.0%) | |
| CALR mutation | | | 0.56 |
| Yes | 114 (21.2%) | 30 (18.5%) | |
| No | 425 (78.8%) | 132 (81.5%) | |
| MPL mutation | | | |
| Yes | 2 (0.4%) | 2 (1.2%) | – |
| No | 537 (99.6%) | 160 (98.8%) | |
| Triple negative | 116 (21.5%) | 28 (17.3%) | 0.35 |
| Leukocyte count ( × 10⁹/L) | 8.9 (7.0–11.4) | 8.8 (6.9–11.5) | 0.49 |
| Platelet ( × 10⁹/L) | 786 (624–1016) | 737 (620–9333) | 0.16 |
| Haemoglobin (g/L) | | | 0.087 |
| <120 g/L | 133 (22.6%) | 34 (20.6%) | |
| 120 g/L-130 g/L | 154 (26.2%) | 35 (21.2%) | |
| ≥130 g/L | 271 (46.2%) | 96 (58.2%) | |
| Heamatocrit (%) | 39.7 (35.4–43.0) | 40.5 (37.2–44.5) | 0.090 |
| RDW (%) | | | 0.35 |
| ≤15% | 374 (67.0%) | 117 (70.9%) | |
| >15% | 184 (33.0%) | 48 (29.1%) | |
| Triglyceride (mmol/L) | 1.3 (0.9–1.6) | 1.2 (0.9–1.7) | 0.22 |
| Total cholesterol (mmol/L) | 3.9 (3.3–4.3) | 3.8 (3.6–4.5) | 0.38 |
| Uric Acid (umol/L) | 321 (258–376) | 305 (249–390) | 0.39 |
| LDH(U/L) | 240 (200–296) | 232 (238–319) | 0.29 |

P-values were determined using the Mann–Whitney U-test and $\chi 2$ test. Splenomegaly = ultrasonic splenomegaly; RDW = red blood cell distribution width; Hb = haemoglobin; MPV = mean platelet volume; LDH = lactate dehydrogenase; IQR = interquartile range; Symptoms included: fatigue, abdominal discomfort, early satiety, weight loss, lack of concentration, itching, night sweats, bone pain, fever, and immobility.

***Table 1:*** **Baseline characteristics of all patients in the training cohort and external validation cohort.**

differences in the proportion of splenomegaly, while no significant differences were observed in other clinical characteristics (Table 1).

## Nomogram development

In the univariable Cox regression analysis, Age (P = 0.0080), sex (P = 0.042), splenomegaly (P < 0.0001), smoking history (P < 0.0075), leukocyte count (P = 0.00024), haemoglobin (P < 0.0001), red blood cell distribution width (RDW) (P < 0.0001), LDH (P < 0.0.0001), and uric acid (P = 0.0013) exhibited significant differences between the ET and post-ET MF groups (Table 2). Based on documented factors and inflammation-related indicators of ET progression reported in the literature, we included age, sex, splenomegaly, smoking history, LDH, haemoglobin, Leukocyte count, RDW and uric acid in a multivariable Cox regression analysis, showed splenomegaly (Hazard ratio (HR) = 3.117, 95% confidence interval (CI) 1.607–6.280), smoking (HR = 3.578, 95% CI

| Variables | Post-ET MF (n = 56) M (P25, P75)/N (%) | ET (n = 502) M (P25, P75)/N (%) | P-value | HR 95% CI |
|---|---|---|---|---|
| Age (years) | 59 (50–67) | 56 (40–68) | 0.0080 | |
| Sex | | | 0.042 | |
| Male | 28 (50%) | 222 (44.2%) | | |
| Female | 28 (50%) | 280 (55.8%) | | |
| Splenomegaly | | | <0.0001 | 3.117 (1.607–6.280) |
| Yes | 40 (71.4%) | 59 (11.8%) | | |
| No | 16 (28.6%) | 443 (88.2%) | | |
| Smoking | | | <0.0001 | 3.578 (2.001–6.397) |
| Yes | 33 (58.9%) | 53 (10.6%) | | |
| No | 23 (41.1%) | 449 (89.4%) | | |
| Symptom | | | 0.25 | |
| Yes | 18 (32.1%) | 123 (24.5%) | | |
| No | 38 (67.9%) | 379 (75.5%) | | |
| JAK2V617F mutation | | | 0.72 | |
| Yes | 35 (63.6%) | 272 (56.3%) | | |
| No | 20 (36.4%) | 211 (43.7%) | | |
| CALR mutation | | | 0.96 | |
| Yes | 9 (16.4%) | 105 (22.3%) | | |
| No | 46 (83.6%) | 378 (78.2%) | | |
| MPL mutation | | | – | |
| Yes | 1 (1.8%) | 1 (0.2%) | | |
| No | 54 (98.2%) | 482 (99.8%) | | |
| Triple negative | 10 (18.2%) | 106 (21.9%) | | |
| Leukocyte count ( × $10^9$/L) | 9.5 (6.5–15.6) | 8.9 (7.0–11.3) | 0.00024 | |
| Platelet ( × $10^9$/L) | 712 (554–877) | 790 (638–1027) | 0.24 | |
| Haemoglobin (g/L) | | | <0.0001 | |
| <120 g/L | 39 (69.6%) | 94 (18.7%) | | 6.736 (1.491–30.432) |
| 120 g/L-130 g/L | 15 (26.8%) | 139 (27.7%) | | 6.441 (1.441–28.795) |
| ≥130 g/L | 2 (3.6%) | 269 (53.6%) | | |
| Heamatokrit (%) | 34.5 (31.5–40.6) | 40.0 (36.1–43.1) | 0.10 | |
| RDW | | | <0.0001 | |
| ≤15% | 11 (19.6%) | 363 (72.3%) | | |
| >15% | 45 (80.4%) | 139 (27.7%) | | 2.763 (1.346–5.672) |
| Triglyceride (mmol/L) | 1.4 (1.1–1.7) | 1.3 (0.9–1.6) | 0.24 | |
| Total cholesterol (mmol/L) | 3.8 (2.9–4.3) | 3.9 (3.3–4.3) | 0.27 | |
| Uric Acid (umol/L) | 350 (275–443) | 321 (255–368) | 0.013 | |
| LDH(U/L) | 342 (292–424) | 232 (198–280) | <0.0001 | 1.003 (1.001–1.005) |

P-values were determined using the univariate Cox regression analyses, HR 95% CI were determined using the multivariate Cox regression analyses.
Splenomegaly = ultrasonic splenomegaly; RDW = red blood cell distribution width; Hb = haemoglobin; MPV = mean platelet volume; LDH = lactate dehydrogenase; HR=Hazard Risk; CI = confident interval; IQR = interquartile range; Symptoms included: fatigue, abdominal discomfort, early satiety, weight loss, lack of concentration, itching, night sweats, bone pain, fever, and immobility; CI = confidence interval.

***Table 2:*** **General characteristics of the patients and multivariate Cox regression analyses for screening predictors.**

2.001–6.397), 120 g/L to 130 g/L level of haemoglobin (HR = 6.441, 95% CI 1.441–28.795), haemoglobin values <120 g/L (HR = 6.736, 95% CI 1.491–30.432), RDW >15% (HR = 2.763, 95% CI 1.346–5.672), LDH (HR = 1.003, 95% CI 1.001–1.005) were risk factors for post-ET MF progression. To establish a predictive model, patients in the training cohort were randomly divided into two groups: a modelling group and an internal validation group, with a ratio of 7:3. Based on the five factors on the progression to post-ET MF, each factor was assigned a specific score based on the impact of post-ET MF progression. Using the rms package of R software, we created a visual nomogram to represent the predictive model (Fig. 2).

## Nomogram validation

In the training cohort, the C-index was 0.877, while the AUC values at 5 years and 10 years were 0.948 (95% CI 0.907–0.989) and 0.769 (95% CI 0.663–0.875), respectively. In the external validation cohort, the C-index was

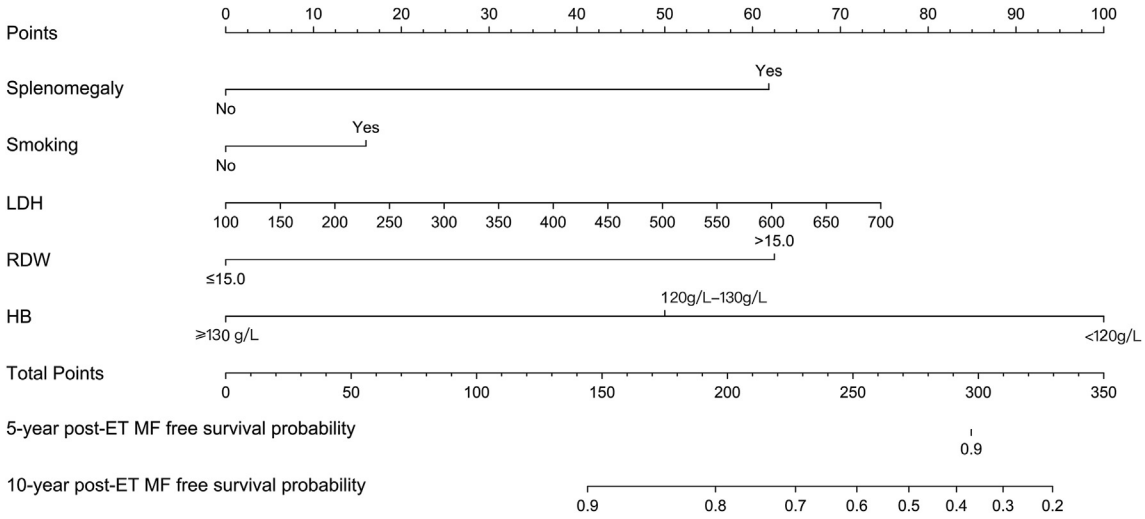


***Fig. 2:*** **Nomogram for the early prediction the post-ET MF free survival probability.** Characteristics in the nomogram to predict probability of post-ET MF free survival in patients with ET. To use the nomogram, the specific points of individual patients are located on each variable axis. Lines and dots are drawn upward to determine the points received by each variable; the sum of these points is located on the Total Points axis, and a line is drawn downward to the '5 years post-ET MF free survival probably, and 10 years post-ET MF free survival probably' axes to determine the probability of post-ET MF free survival probably. Post-ET MF = post-essential thrombocythemia myelofibrosis; LDH = lactate dehydrogenase; RDW = red blood cell distribution width; Hb = haemoglobin; Splenomegaly = ultrasonic splenomegaly.

0.853 with corresponding AUC values of 0.978 (95% CI 0.949–1.006), 0.804 (95% CI 0.641–0.968) at 5 years and 10 years, respectively. The results indicate a sustained level of accuracy (Fig. 3 A and B). The calibration curve showed that bias-corrected lines for the training and external validation cohorts closely overlapped with the ideal lines (Fig. 4A–D). DCA illustrated the clinical net benefit achievable at different risk thresholds (Fig. 5A–D). Threshold ranges for DCA were derived from the training and external cohorts based on the sensitivity and specificity of the model. Intervention was performed on patients whose assessed risk was within the threshold range. The net benefit is superior to intervening on all patients or not intervening on all patients.

## Discussion

Progression to post-ET MF is a significant determinant influencing the survival outcomes of ET patients.[18] The median survival period for post-ET MF is approximately 9.2 years. Contemporary strategies for addressing bone marrow fibrosis include JAK2 inhibitors and allogeneic stem cell transplantation.[19] In cases where physicians suspect potential disease progression, a bone marrow biopsy is recommended. Timely identification of post-ET MF can aid clinicians in implementing well-founded interventions to manage the progression of the disease.

Numerous studies have investigated the impact of risk factors on overall survival and increased thrombosis in ET. However, limited research has focused on assessing the risk of progression to post-ET MF at the time of diagnosis. The objective of our study was to develop an intuitive visual model for evaluating the post-ET MF free survival probability at 5 years and 10 years, which is both novel and significant. We enrolled a total of 558 patients from eight haematology centres in China, excluding those with prefibrotic PMF based on the WHO 2016 diagnostic criteria. Through our analysis, we identified several independent risk factors associated with post-ET MF progression, including ultrasonic splenomegaly, smoking, haemoglobin level, LDH level, and RDW level. Based on these five factors, we constructed a risk prediction model for post-ET MF progression. To validate the model, we externally tested it using data from an additional 165 patients from six different haematology centres in China. This study contributes to enhancing patients' understanding of their individual risk levels and emphasizes the importance of regular check-ups and disease management for high-risk patients. In patients suggestive of ultrasound splenomegaly, smoking, elevated RDW, elevated LDH and low haemoglobin. Regular follow-up is particularly important and provides a basis for initiating antifibrotic therapy in patients.

The progression from ET to post-ET MF displays a consistent biological continuum. Chronic inflammation is a central catalyst for clonal proliferation of tumour cells.[20] Increased levels of inflammatory markers characterize both PMF and its evolution towards SMF.[3,4] Extensive research has established a correlation

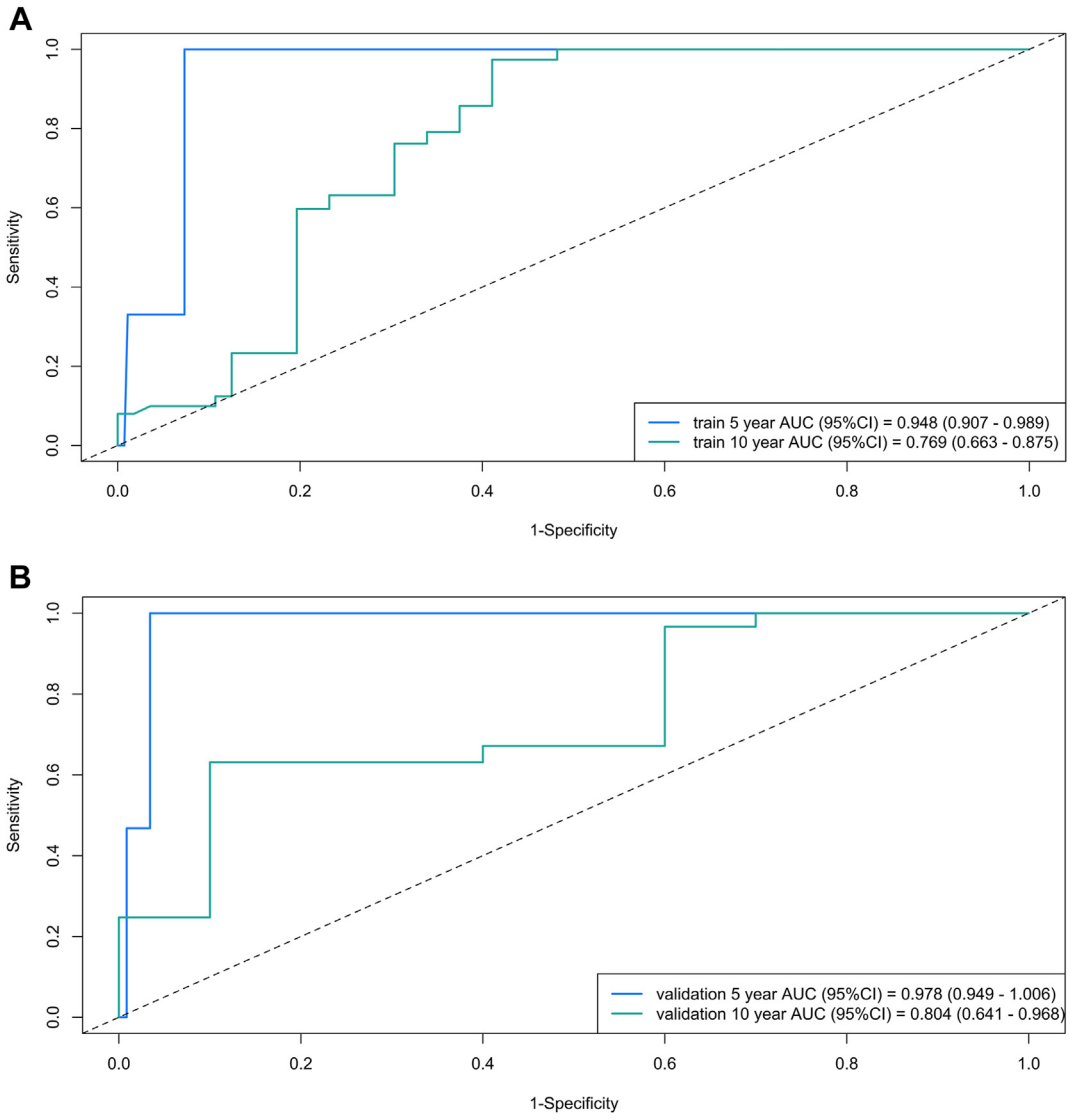


***Fig. 3:* ROC curve.** ROC curve and AUC of the nomogram. Training cohort (A) and external validation cohort (B). ROC = receiver operating characteristic; AUC = the area under ROC curve; post-ET MF = post-essential thrombocythemia myelofibrosis.

between smoking and MPNs. Smoking contributes to the release of inflammatory cytokines, thereby intensifying the inflammatory burden in individuals with MPNs. This process entails the upregulation of vital molecular signalling pathways and transcription factors that amplify the mechanisms of clonal proliferation in MPNs, creating a predisposition for their transformation into a more invasive post-ET MF state.[15] Furthermore, recent investigations have illuminated a robust and graded correlation between RDW and well-established inflammatory biomarkers, such as high-sensitivity C-reactive protein (hsCRP).[21] RDW has considerable potential as a prospective biomarker for predicting an unfavourable prognosis in haematologic malignancies, including Hodgkin's lymphoma, chronic myeloid leukaemia, and multiple myeloma.[13,22,23] In our retrospective analysis, increased RDW emerged as a noteworthy factor indicative of the progression of ET patients to post-ET MF. Moreover, splenomegaly is a distinctive manifestation of bone marrow fibrosis, often concomitant with a decline in haemoglobin levels due to marrow insufficiency. Earlier investigations have highlighted notable discrepancies in haemoglobin and LDH levels, stratified by sex, within both ET and myelofibrosis cohorts, thus establishing a foundation for assessing the progression of patients' conditions.[24]

In addition to clinical characteristic, previous studies have indicated that CALR-1 like mutations, MPL

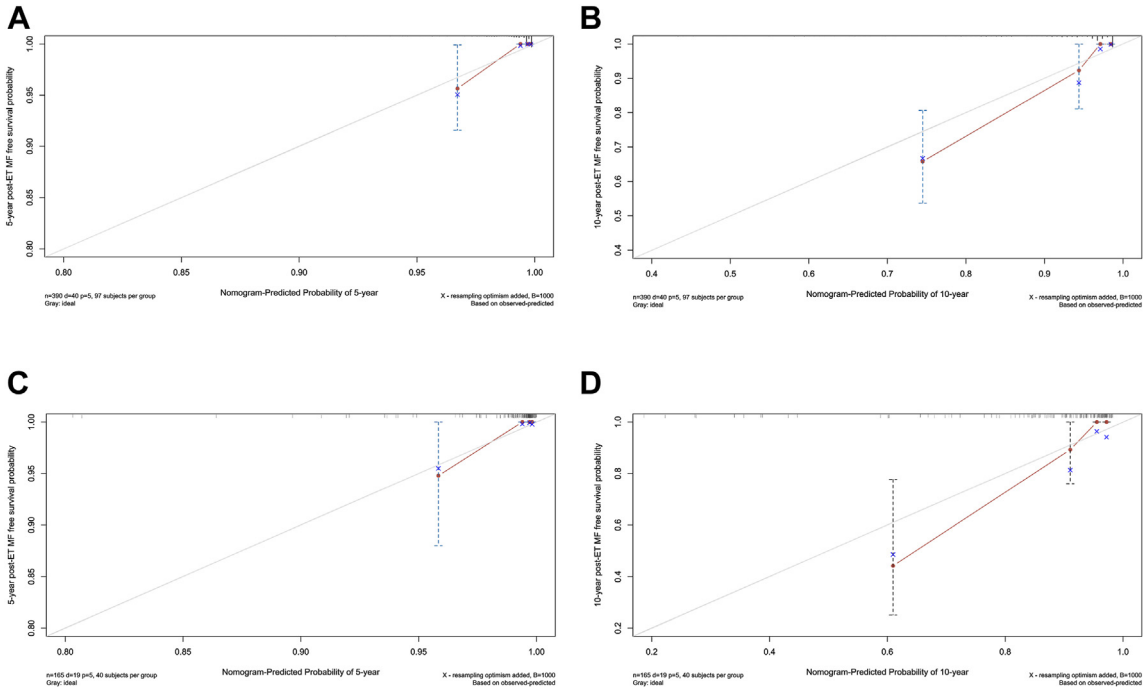


***Fig. 4:* Calibration curve.** Calibration curve of the nomogram. Calibration curves of 5 years, and 10 years post-ET MF free survival for ET in (A, B) training cohort, and (C, D) in external validation cohort. The grey line represents the ideal reference line where the predicted probability matches the observed survival rate. The blue dots are calculated by bootstrapping (resampling: 1000) and represent the performance of the nomogram. The closer the solid red line is to the grey line, the more accurate the model is in predicting overall survival. post-ET MF = post-essential thrombocythemia myelofibrosis.

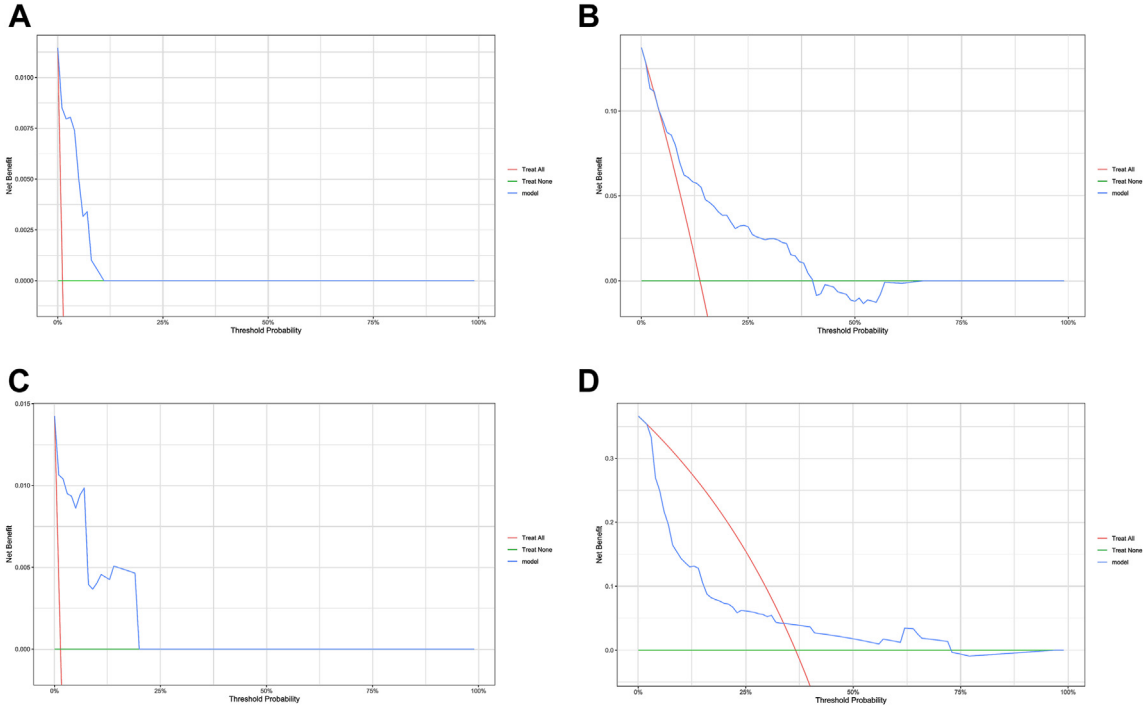


***Fig. 5:* DCA curve.** DCA curve of the nomogram. DCA curve of 5 years, and 10 years post-ET MF free survival for ET in (A, B) training cohort, and (C, D) in external validation cohort. DCA = decision curve analysis; post-ET MF = post-essential thrombocythemia myelofibrosis.

mutation are associated with an increased risk of post-ET MF progression in ET (13). Furthermore, second-generation sequencing has identified SH2B3, SF3B1, U2AF, IDH2, EZH2, and TP53 as genes linked to reduced myelofibrosis-free survival time.[25] These studies have considered the impact of gene mutations. However, it is important to acknowledge the limitations inherent in our retrospective study. Firstly, our follow-up period, in comparison to other studies, may be relatively short, resulting in potential data bias. Secondly, due to limitations in data acquisition, we were unable to differentiate between CALR gene mutation types and report the variant allele frequency of the JAK2V617F mutation. In our analysis of 190 patients with reported JAK2V617F variant allele frequency, we did not observe a significant effect of the variant allele frequency on the progression of myelofibrosis. The retrospective nature of our study led to the exclusion of several potential underlying factors, which limits our ability to assess the impact of JAK2V617F variant allele frequency and CALR mutation types on the progression of post-ET MF. Consequently, our model construction was solely based on clinical characteristics, without comprehensive inclusion of the effects of gene mutations on post-ET MF progression. Our study reveals that smoking, splenomegaly, haemoglobin, LDH, and RDW hold the potential to predict the risk of transitioning to post-ET MF. Furthermore, the nomogram was both internally and externally validated, confirming its strong predictive capacity. In future endeavours, we aim to incorporate molecular biology analyses alongside clinical features to enable a more comprehensive assessment of patients and optimize this model. However, in terms of the current study objectives, this does not impede our ability to identify patients who require close monitoring and appropriate counselling. Close monitoring enables the timely adjustment of treatment strategies, thereby enhancing life expectancy. Additionally, it serves as a fundamental framework for future model optimization in the field.

#### Contributors

DHX designed the framework of the article and drafted the manuscript. XDY, HLQ, LZ, YXH, and YC S collected and analysed the data. YL, YC and DC verified the underlying study data. MWH, LFW, QLT, DJW, and GYT generated figures and tables, and refined the study design. HYT, JJ and JH provided final modifications to the manuscript. All authors contributed to manuscript revisions and approved the final manuscript as submitted.

#### Data sharing statement

The data underlying this article will be shared on reasonable request to the corresponding authors.

#### Declaration of interests

None of the authors have conflicts of interest to disclose. None of the authors have financial relationships relevant to this article to disclose.

#### Acknowledgements

We express our gratitude for the generous support provided by the Key R&D Program of Zhejiang (No. 2022C03137); the Public Technology Application Research Program of Zhejiang, China (No. LGF21H080003); and the Zhejiang Medical Association Clinical Medical Research special fund project (No. 2022ZYC-D09) in facilitating this study. The funding sources supported the article process charge.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2023.102378.

#### References

1 Vainchenker W, Kralovics R. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood*. 2017;129(6):667–679.
2 Baumeister J, Chatain N, Sofias AM, Lammers T, Koschmieder S. Progression of myeloproliferative neoplasms (MPN): diagnostic and therapeutic perspectives. *Cells*. 2021;10(12):3551.
3 Tefferi A, Vaidya R, Caramazza D, Finke C, Lasho T, Pardanani A. Circulating interleukin (IL)-8, IL-2R, IL-12, and IL-15 levels are independently prognostic in primary myelofibrosis: a comprehensive cytokine profiling study. *J Clin Oncol*. 2011;29(10):1356–1363.
4 Panteli KE, Hatzimichael EC, Bouranta PK, et al. Serum interleukin (IL)-1, IL-2, sIL-2Ra, IL-6 and thrombopoietin levels in patients with chronic myeloproliferative diseases. *Br J Haematol*. 2005;130(5):709–715.
5 Szuber N, Mudireddy M, Nicolosi M, et al. 3023 mayo clinic patients with myeloproliferative neoplasms: risk-stratified comparison of survival and outcomes data among disease subgroups. *Mayo Clin Proc*. 2019;94(4):599–610.
6 Cerquozzi S, Tefferi A. Blast transformation and fibrotic progression in polycythemia vera and essential thrombocythemia: a literature review of incidence and risk factors. *Blood Cancer J*. 2015;5:e366.
7 Barbui T, Thiele J, Passamonti F, et al. Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: an international study. *J Clin Oncol*. 2011;29(23):3179–3184.
8 Abdulkarim K, Ridell B, Johansson P, Kutti J, Safai-Kutti S, Andreasson B. The impact of peripheral blood values and bone marrow findings on prognosis for patients with essential thrombocythemia and polycythemia vera. *Eur J Haematol*. 2011;86(2):148–155.
9 Stein BL, Williams DM, O'Keefe C, et al. Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: analysis of molecular genetics and clinical phenotypes. *Haematologica*. 2011;96(10):1462–1469.
10 Passamonti F, Rumi E, Pietra D, et al. A prospective study of 338 patients with polycythemia vera: the impact of JAK2 (V617F) allele burden and leukocytosis on fibrotic or leukemic disease transformation and vascular complications. *Leukemia*. 2010;24(9):1574–1579.
11 Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the world health organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391–2405.
12 Barosi G, Mesa RA, Thiele J, et al. Proposed criteria for the diagnosis of post-polycythemia vera and post-essential thrombocythemia myelofibrosis: a consensus statement from the international working group for myelofibrosis research and treatment. *Leukemia*. 2008;22(2):437–438.
13 Mao XL, Xi YM, Li ZJ, et al. Higher red blood cell distribution width at diagnose is a simple negative prognostic factor in chronic phase-chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: a retrospective study. *Medicine (Baltimore)*. 2021;100(10):e24003.
14 Haider M, Elala YC, Gangat N, Hanson CA, Tefferi A. MPL mutations and palpable splenomegaly are independent risk factors for fibrotic progression in essential thrombocythemia. *Blood Cancer J*. 2016;6(10):e487.
15 Lindholm Sorensen A, Hasselbalch HC. Smoking and philadelphia-negative chronic myeloproliferative neoplasms. *Eur J Haematol*. 2016;97(1):63–69.
16 Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53.

17 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18.

18 Masarova L, Bose P, Daver N, et al. Patients with post-essential thrombocythemia and post-polycythemia vera differ from patients with primary myelofibrosis. *Leuk Res*. 2017;59:110–116.

19 Mesa RAPF. Individualizing care for patients with myeloproliferative neoplasms: integrating genetics, evolving therapies, and patient-specific disease burden. *Am Soc Clin Oncol Educ Book*. 2016;35:e324–e335.

20 Hasselbalch HC. Perspectives on chronic inflammation in essential thrombocythemia, polycythemia vera, and myelofibrosis: is chronic inflammation a trigger and driver of clonal evolution and development of accelerated atherosclerosis and second cancer? *Blood*. 2012;119(14):3219–3225.

21 Lippi G, Montagnana M, Salvagno GL. Relation between red blood cell distribution width and inflammatory biomarkers in a large cohort of unselected outpatients. *Arch Pathol Lab Med*. 2009;133(4):628–632.

22 Herraez I, Bento L, Del Campo R, et al. Prognostic role of the red blood cell distribution width (RDW) in Hodgkin lymphoma. *Cancers (Basel)*. 2020;12(11):3262.

23 Lee H, Kong SY, Sohn JY, et al. Elevated red blood cell distribution width as a simple prognostic factor in patients with symptomatic multiple myeloma. *BioMed Res Int*. 2014;2014:145619.

24 Carobbio A, Finazzi G, Thiele J, et al. Blood tests may predict early primary myelofibrosis in patients presenting with essential thrombocythemia. *Am J Hematol*. 2012;87(2):203–204.

25 Tefferi A, Lasho TL, Guglielmelli P, et al. Targeted deep sequencing in polycythemia vera and essential thrombocythemia. *Blood Adv*. 2016;1(1):21–30.