Check for updates

ARTICLE

ACUTE MYELOID LEUKEMIA

Comparison of scoring systems evaluating suitability for intensive chemotherapy in adults with acute myeloid leukemia—a Grand Ouest Against Leukemia (GOAL) study

Christophe Desprez[1], Jérémie Riou [2], Pierre Peterlin[3,4], Tony Marchand [4,5,6], Marie-Anne Couturier[4,7], Alban Villate[4,8], Jean-Baptiste Mear[4,5], Patrice Chevalier[3,4], Gaelle Guillerm[4,7], Emmanuel Gyan[4,8], Aline Schmidt-Tanguy[1,4,9], Roland B. Walter [10,11,12,13], Mathilde Hunault-Berger [1,4,9] and Corentin Orvain [1,4,9 ✉]

Several scoring systems have been developed to assess suitability of individual patients for intensive acute myeloid leukemia (AML) therapy. We sought to compare the performance of these scores in a cohort of 428 consecutive adults with AML who received conventional induction chemotherapy in five academic centers in France. All scoring systems identified a subset of patients with increased 28 and 56-day mortality although the prediction accuracy was overall limited with C-statistics of ranging from 0.61 to 0.71 Overall survival (OS) prediction was more limited and restricted to scoring systems that include AML-related parameters. The outcome of 104 patients (24%) considered unsuitable for intensive chemotherapy based on criteria used in recent randomized trials was similar to that of the other 324 patients (28-day mortality, odds ratio [OR] = 1.88, $P = 0.2$; 56-day mortality, OR = 1.71, $P = 0.21$; event-free survival, hazard ratio [HR] = 1.08, $P = 0.6$; OS, HR = 1.25, $P = 0.14$) with low discrimination (C-statistic: 0.57, 0.56, 0.50, and 0.52 for 28-day, 56-day mortality, EFS, and OS, respectively). Together, our findings indicate that the accuracy of currently available approaches to identify patients at increased risk of early mortality and shortened survival after intensive AML therapy is relatively limited. Caution regarding the use of available scoring systems should be warranted in clinical decision-making.

INTRODUCTION

With a median age at diagnosis of almost 70 years, acute myeloid leukemia (AML) primarily affects older adults [1, 2]. It is well recognized that advanced age and/or existing comorbid illnesses may limit the tolerance of intensive chemotherapy and resulting pancytopenia [3, 4]. Although our tools to support patients throughout this treatment period have improved over the last 20 years, there is ongoing concern that perceived risks of such therapies may not be commensurate with anticipated benefits in medically less fit individuals [5–7]. Consequently, there is interest in accurately assessing the ability of individual patients to tolerate intensive AML chemotherapy—an interest that has greatly increased with recent market approval of several new drugs that have expanded the number of lower intensity therapies for patients with AML [8–10]. With increasing treatment options available, evaluating both disease characteristics and the patient's

general medical condition is fundamental to guide informed decision-making regarding suitability for intensive therapies at the individual patient level.

Unfortunately, assessing a patient's suitability ("fitness") for intensive chemotherapy has remained a challenge, especially in older patients, those with altered general status, and/or those with comorbidities [11, 12], with no current consensus on how to approach this evaluation. Several scoring systems have been developed over the last decade to predict shorter- or longer-term outcomes in individual patients after receipt of intensive AML chemotherapy to aid this process [12–16]. Some of these scores rely primarily on previously established comorbidity scores with the incorporation of additional parameters related to AML disease characteristics (e.g., Augmented Hematopoietic Cell Transplantation Comorbidity Index [HCT-CI] or AML Composite Model [AML-CM]) while others have been specifically developed to predict

[1]Service des maladies du sang, CHU d'Angers, Angers, France. [2]Univ Angers, CHU Angers, Inserm, CNRS, MINT, SFR ICAT, Angers, France. [3]Service d'hématologie clinique, CHU de Nantes, Rennes, France. [4]Federation hospitalo-universitaire « Grand Ouest against Leukemia », Angers, France. [5]Service d'hématologie clinique, CHU de Rennes, Rennes, France. [6]INSERM U1236, Université Rennes 1, Rennes, France. [7]Service d'Hématologie Clinique, Hôpital Morvan, CHU de Brest, Rennes, France. [8]Service d'Hématologie et Thérapie Cellulaire, Hôpital Bretonneau, CHRU de Tours, CNRS EMR7001 LNOx, Université de Tours, Rennes, France. [9]Université d'Angers, Inserm UMR 1307, CNRS UMR 6075, Nantes Université, CRCI2NA, F-49000 Angers, France. [10]Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. [11]Department of Medicine, Division of Hematology/Oncology, University of Washington, Seattle, WA, USA. [12]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. [13]Department of Epidemiology, University of Washington, Seattle, WA, USA. ✉email: corentin.orvain@chu-angers.fr

early mortality (e.g., Treatment-Related Mortality [TRM] score or MD Anderson Cancer Center [MDACC] scoring system) or rely on consensus from expert panels (e.g., SIE/SIES/GITMO Consensus Criteria or Ferrara score for unfitness) [12–14, 16]. Moreover, recently conducted large controlled trials targeting adults with AML deemed unsuitable for intensive AML chemotherapy have put forth their own criteria for unfitness adapted from the Ferrara criteria [9, 10].

While these different approaches have shown some value in identifying subsets of patients at increased risk of early mortality and/or shorter survival in the cohorts in which they were derived, it is unknown how they perform relatively to each other in other patient cohorts. To address this limitation, we therefore sought to examine the predictive value of several of the most widely used approaches to estimate treatment fitness in an independent multi-centric cohort of adults with AML receiving intensive induction therapy.

PATIENTS AND METHODS
Study cohort
For our retrospective analysis, we identified all consecutive adults ≥18 years of age with newly diagnosed AML (WHO 2016) admitted to five university teaching hospitals during a 3-year period (2017–2019) provided they received induction chemotherapy with 3 + 7, idarubicin-cytarabine-lomustine (ICL), or CPX-351 [17, 18]. To ensure exhaustivity, patients were identified after consulting each hospital pharmacy's registry for induction chemotherapy prescriptions. Patients with secondary AML after antecedent myelodysplastic syndrome (MDS) or myeloproliferative neoplasm (MPN), or after exposure to a leukemogenic agent (previous chemotherapy or radiotherapy) were included. We excluded patients with acute promyelocytic leukemia (APL) or those with acute mixed-phenotypic acute leukemia, those who previously under-went allogeneic hematopoietic cell transplantation (HCT), and those who received any chemotherapy treatment directed at AML before induction chemotherapy, such as hypomethylating agent (HMA) or low dose cytarabine (LDAC). This retrospective study was approved by the Institutional Ethics Committee of the University Hospital of Angers (# 2020/63).

Data collection
We conducted a comprehensive chart review to obtain information on demographics, comorbidities, and cytogenetic/molecular abnormalities. Cytogenetic risk was classified using the refined NCRI/MRC criteria [19]. Cytogenetic/molecular risk was assigned according to the European Leukemia Net (ELN) 2017 risk classification [20]. Since many patients did not undergo next generation sequencing (NGS) testing at diagnosis, ASXL1, RUNX1, and TP53 mutational status was unknown in most patients. The augmented HCT-CI, the AML-CM (Supplementary Table 1), the TRM score, the Ferrara assessment for fitness (Supplementary Table 2), and the recent early mortality score developed from investigators from MDACC (Supplementary Table 3) were calculated as previously described with parameters collected at diagnosis [12–16, 21]. Pulmonary function testing is not routinely performed in our centers and was only available for 10 patients. Also, uricemia was not available in our database and was therefore not used in the calculation of the MDACC scoring system. The criteria used to define unfitness for intensive chemotherapy in two recent randomized clinical trials (RCTs) targeting medically less fit patients were also assessed ("Unfitness RCT eligibility criteria"; Supplementary Table 4) [9, 10].

Endpoints
Twenty-eight and 56-day mortality was used to denote early mortality. Event-free survival (EFS) was measured from the first day of induction therapy to the date of primary refractory disease, relapse from complete remission (CR)/CR with incomplete hematological recovery (CRi) [20], or death from any cause, whereas overall survival (OS) was measured from the first day of induction chemotherapy until death from any cause, with observations censored for patients last known alive [20]. Primary refractory disease was defined as absence of CR/CRi following 2 cycles of induction chemotherapy [20].

Table 1. Patient characteristics.

Characteristic	All patients (n = 428)
Age at diagnosis (IQR), years	60 (49–66)
Female gender, n (%)	197 (46%)
Performance status, n (%)	
[0]	238 (56%)
[1]	156 (36%)
[2]	26 (6%)
[3]	8 (2%)
Secondary AML, n (%)	91 (21%)
Therapy-related	28 (6%)
Antecedent myeloid disorder	63 (15%)
WBC counts at diagnosis (IQR), G/l	7 (2–39)
Cytogenetic risk (MRC), n (%)	
Favorable	34 (8%)
Intermediate	263 (62%)
Unfavorable	128 (30%)
Cytogenetic/molecular risk (ELN-2017), n (%)	
Favorable	127 (30%)
Intermediate	140 (33%)
Adverse	159 (37%)
Unfit RCT eligibility criteria, n (%)	104 (24%)
Augmented HCT-CI (IQR)	3 (2–4)
Augmented HCT-CI category, n (%)	
[0–2]	210 (49%)
[3–4]	130 (30%)
[5–6]	68 (16%)
[≥7]	20 (5%)
AML-CM (IQR)	5 (4–7)
AML-CM category, n (%)	
[1–4]	174 (41%)
[5–6]	136 (32%)
[7–9]	98 (23%)
[≥10]	20 (5%)
Ferrara criteria category, n (%)	
Fit	296 (69%)
Unfit	132 (31%)
TRM score (IQR)	1.7 (0.7–3.8)
TRM score category, n (%)	
Lower (≤ 13.1)	398 (93%)
Higher (> 13.1)	30 (7%)
MDACC scoring system category (IQR)	2 (1–4)
MDACC scoring system category, n (%)	
Low (≤ 4)	370 (86%)
High (5–8)	52 (12%)
Very High (≥ 9)	6 (2%)

AML acute myeloid leukemia, AML-CM AML composite model, ELN European LeukemiaNet, HCT-CI hematopoietic cell transplantation comorbidity index, IQR interquartile range, MDACC MD Anderson Cancer Center, MRC Medical Research Council, TRM treatment related mortality, WBC white blood cells.

Statistical analysis
Patient characteristics are reported as number and percentage for qualitative variables and as median [interquartile range (IQR)] for

Table 2. Distribution of parameters used to calculate scoring systems stratified by age categories.

Characteristic	All patients (n = 428)	[18–29] (n = 19)	[30–39] (n = 39)	[40–49] (n = 54)	[50–59] (n = 100)	[60–69] (n = 176)	[≥70] (n = 40)
Unfitness RCT eligibility criteria, n (%)							
Age	3	0	0	0	0	0	3 (8%)
Performance status	34 (8%)	1 (5%)	1 (3%)	5 (9%)	9 (9%)	17 (10%)	1 (3%)
Cardiac disease	38 (9%)	0	0	0	9 (9%)	24 (14%)	5 (12%)
Pulmonary disease	6 (1%)	0	0	1 (2%)	2 (2%)	2 (1%)	1 (3%)
Creatine levels	38 (9%)	0	0	5 (9%)	7 (7%)	22 (12%)	4 (10%)
Liver disease	9 (2%)	0	1 (3%)	1 (2%)	1 (1%)	6 (3%)	0
Augmented HCT-CI, n (%)							
Arythmia	16 (4%)	0	0	0	4 (4%)	9 (5%)	3 (8%)
Cardiovascular comorbidity	18 (4%)	0	0	0	2 (2%)	14 (8%)	2 (5%)
Inflammatory bowel disease	4 (1%)	0	0	1 (2%)	2 (2%)	1 (1%)	0
Diabetes	23 (5%)	0	0	2 (4%)	7 (7%)	10 (6%)	4 (10%)
Cerebrovascular disease	8 (2%)	0	0	0	2 (2%)	3 (2%)	3 (8%)
Psychiatric disorder	26 (6%)	0	3 (8%)	2 (4%)	6 (6%)	13 (7%)	2 (5%)
Mild hepatic comorbidity	121 (28%)	8 (42%)	13 (33%)	15 (28%)	36 (36%)	41 (23%)	8 (20%)
Moderate/severe hepatic comorbidity	38 (9%)	1 (5%)	5 (13%)	4 (7%)	3 (3%)	22 (12%)	3 (8%)
Obesity	26 (6%)	1 (5%)	3 (8%)	4 (7%)	6 (6%)	11 (6%)	1 (3%)
Infection	93 (22%)	4 (21%)	8 (21%)	10 (19%)	26 (26%)	35 (20%)	10 (25%)
Rheumatological comorbidity	4 (1%)	0	0	1 (2%)	1 (1%)	0	2 (5%)
Peptic ulcer	6 (1%)	0	0	1 (2%)	3 (3%)	2 (1%)	0
Renal comorbidity	7 (2%)	0	0	1 (2%)	1 (1%)	2 (1%)	3 (8%)
Moderate pulmonary comorbidity	9 (2%)	0	0	1 (2%)	4 (4%)	1 (1%)	3 (8%)
Severe pulmonary comorbidity	4 (1%)	0	0	1 (2%)	1 (1%)	2 (1%)	0
Prior malignancy	28 (7%)	1 (5%)	2 (5%)	4 (7%)	8 (8%)	11 (6%)	2 (5%)
Heart valve disease	6 (1%)	0	0	0	4 (4%)	2 (1%)	0
Albumine levels < 35 g/l	157 (37%)	3 (16%)	17 (44%)	20 (37%)	34 (34%)	67 (38%)	16 (40%)
Platelet counts < 20 G/l	44 (10%)	5 (26%)	4 (10%)	6 (11%)	11 (11%)	15 (9%)	3 (8%)
Moderate LDH elevation	310 (72%)	13 (68%)	30 (77%)	40 (74%)	72 (72%)	126 (72%)	29 (72%)
High LDH elevation	69 (16%)	4 (21%)	6 (15%)	10 (19%)	16 (16%)	27 (15%)	6 (15%)
Ferrara criteria, n (%)							
Age	3 (1%)	0	0	0	0	0	3 (8%)
Performance status	8 (2%)	0	1 (3%)	1 (2%)	2 (2%)	4 (2%)	0
Cardiac disease	6 (1%)	0	0	0	4 (4%)	2 (1%)	0
Pulmonary disease	7 (2%)	0	0	1 (2%)	2 (2%)	3 (2%)	1 (3%)
Renal disease	0	0	0	0	0	0	0
Liver disease	0	0	0	0	0	0	0
Infection	93 (22%)	4 (21%)	8 (21%)	10 (19%)	26 (26%)	35 (20%)	10 (25%)
Mental illness	26 (6%)	0	3 (8%)	2 (4%)	6 (6%)	13 (7%)	2 (5%)
MDACC scoring system, n (%)							
Age							
[<40]	58 (14%)	19 (100%)	39 (100%)	0	0	0	0
[40–64]	215 (50%)	0	0	54 (100%)	100 (100%)	61 (35%)	0
[64–75]	152 (36%)	0	0	0	0	115 (65%)	37 (92%)
[≥75]	3 (1%)	0	0	0	0	0	3 (8%)

Table 2. continued

Characteristic	All patients (n = 428)	[18–29] (n = 19)	[30–39] (n = 39)	[40–49] (n = 54)	[50–59] (n = 100)	[60–69] (n = 176)	[≥70] (n = 40)
Performance status							
[0–1]	394 (92%)	18 (95%)	38 (97%)	49 (91%)	91 (91%)	159 (90%)	39 (98%)
[2]	26 (6%)	1 (5%)	0	4 (7%)	7 (7%)	13 (7%)	1 (3%)
[3,4]	8 (2%)	0	1 (3%)	1 (2%)	2 (2%)	4 (2%)	0
Total bilirubin ≥ 1.3 mg/dl	94 (23%)	2 (11%)	8 (21%)	13 (25%)	22 (22%)	43 (26%)	6 (16%)
Creatine ≥ 1.3 mg/dl	31 (7%)	0	0	3 (6%)	5 (5%)	18 (10%)	5 (12%)
Chromosomal abnormalities	128 (30%)	6 (32%)	11 (28%)	17 (31%)	33 (33%)	51 (29%)	10 (25%)
Pneumonia at diagnosis	34 (8%)	1 (5%)	6 (15%)	4 (7%)	6 (6%)	14 (8%)	3 (8%)

AML acute myeloid leukemia, *AML-CM* AML composite model, *HCT-CI* hematopoietic cell transplantation comorbidity index, *LDH* lactate dehydrogenase, *MDACC* MD Anderson Cancer Center.

continuous variables. Unpaired Student's *t*-test or the Mann–Whitney *U* test for the comparison of continuous variables and the chi-squared or the Fisher's exact tests for the comparison of categorical variables were performed as appropriate. Associations between variables and 28 and 56-day mortality were assessed using logistic regression. EFS and OS were estimated using the Kaplan–Meier method and compared with the Log-Rank test; associations with RFS and OS were assessed using Cox regression. C-statistics, time-dependent AUCs, and Brier scores were used to assess the predictive ability with a graphical approach for time-dependent AUCs and Brier scores over time. An AUC/C-statistic of 1.0 indicates that a model (or covariate) is perfect at prediction, whereas 0.8–0.9 is considered good, 0.7–0.8 is considered fair, 0.6–0.7 is considered poor, 0.5–0.6 is considered very poor, and 0.5 indicates no prediction. All tests were performed with a type I error set at 0.05. Statistical analyses were performed with R (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org). This study follows "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis" (TRIPOD) guidelines [22].

RESULTS

Patient characteristics and distribution of fitness assessments

We identified 439 consecutive adults with non-APL AML who received induction therapy at 5 academic institutions in France between January 1, 2017, and December 31, 2019. Of these, 11 were excluded because of prior allogeneic HCT ($n = 4$), previous treatment with non-intensive AML chemotherapy ($n = 5$), and other types of acute leukemia ($n = 2$), leaving 428 patients with a median age of 60 years (IQR, 49–66) for analysis. Demographics of this study cohort are summarized in Table 1. Performance status (PS) was 0 to 1 in most patients (92%). Ninety-one patients (21%) had secondary AML (prior solid tumor requiring chemotherapy and/or radiation therapy [$n = 28$] or prior hematologic malignancy [$n = 63$]). Most patients received $3 + 7$ ($n = 240$, 56%) and ICL ($n = 174$, 41%) induction therapy whereas 14 (3%) received CPX-351. The distribution of each parameter used to calculate unfitness RCT eligibility criteria, Ferrara unfit criteria, augmented HCT-CI, and MDACC scoring system according to age categories is summarized in Table 2. At least one criterion defining unfitness in recent RCTs was found in 104 patients (24%) whereas 132 patients (31%) met at least one Ferrara criterion for unfitness. Only 20 patients (5%) were considered at very high risk according to the augmented HCT-CI scoring system or the AML-CM; only 7% of patients had a high TRM score (operationally defined here as a TRM score ≥ 13.1). The MDACC scoring system was low for most patients; 12% and 2% had a high or very high score, respectively.

Early mortality and survival estimates

We first examined the association of individual risk scores with measures of early mortality and estimates of survival. With a

median follow-up of 527 (IQR: 276–870) days among survivors, there were 130 relapses and 139 deaths with 19 (4%) and 26 (6%) deaths occurring at 28 and 56 days, respectively. Several parameters and scores were associated with early mortality including white blood cells (WBC) count at diagnosis (odds ratio [OR] = 1.01 [1.00–1.01], $P = 0.007$, and OR = 1.01 [1.00–1.01], $P = 0.006$, for 28 and 56-day mortality, respectively), augmented HCT-CI (OR = 1.32 [1.08–1.61], $P = 0.005$, and OR = 1.36 [1.14–1.63], $P < 0.001$, respectively), AML-CM (OR = 1.34 [1.13–1.58], $P < 0.001$, and OR = 1.39 [1.19–1.62, $P < 0.001$, respectively), Ferrara criteria (OR = 2.61 [1.03–6.73], $P = 0.042$, and OR = 3.32 [1.49–7.63], $P = 0.004$, respectively), TRM score (OR = 1.06 [1.02–1.10], $P = 0.002$, and OR = 1.07 [1.03–1.11], $P < 0.001$, respectively), and the MDACC scoring system (OR = 1.36 [1.12–1.64], $P = 0.001$, and OR = 1.39 [1.18–1.65], $P < 0.001$, respectively) (Table 3). Figure 1 shows 28 and 56-day mortality in our cohort stratified by the various risk scores. A similar proportion of patients judged fit or unfit according to criteria used for eligibility in recent RCTs targeting less fit individuals experienced early mortality (28-day mortality of 4% vs. 7%, $P = 0.27$, and 56-day mortality of 5% vs. 9%, $P = 0.21$, respectively). In contrast, all the risk assessment scores used as categorical variables were associated with 28 and 56-day mortality except for patients with higher TRM scores for whom the difference was only statistically significant for 56-day mortality (Augmented HCT-CI: $P = 0.019$ and $P = 0.017$ for 28-day and 56-mortality, respectively; AML-CM: $P = 0.010$ and $P < 0.001$; Ferrara criteria: $P = 0.035$ and $P = 0.002$; TRM score: $P = 0.14$ and $P = 0.028$; and MDACC: $P = 0.019$ and $P = 0.001$) (Supplementary Table 5).

The AML-CM (hazard ratio [HR] = 1.13 [1.07–1.18], $P < 0.001$, and HR = 1.15 [1.10–1.21], $P < 0.001$, for EFS and OS, respectively) and the recent mortality score from MDACC (HR = 1.17 [1.10–1.24], $P < 0.001$, and HR = 1.22 [1.14–1.30], $P < 0.001$, respectively) were associated with EFS and OS (Fig. 2). The HCT-CI (HR = 1.05 [0.99–1.12], $P = 0.10$, and HR = 1.07 [1.01–1.14], $P = 0.03$, respectively) and the TRM score (HR = 1.02 [1.01–1.04], $P = 0.01$, and HR = 1.03 [1.02–1.05], $P < 0.001$, respectively) were associated with EFS and OS if scores were used as continuous but not as categorical variables (Table 3). The other scores were not associated with EFS/OS. Similar results were obtained when patients who underwent allogeneic HCT were censored at the time of allografting (Supplementary Fig. 1).

Predictive accuracy of individual risk scores

We next determined the accuracy with which individual risk scores predicted early mortality and survival. Discrimination for OS after induction therapy is represented in Fig. 3 with time AUC curves describing the predictive ability of each score over time. The best

Table 3. Univariable analysis of factors associated with 28-day and 56-day mortality, EFS, and OS.

Characteristic	28-day mortality			56-day mortality			Event-free survival			Overall survival		
	OR (95% CI)	P	C-statistic	OR (95% CI)	P	C-statistic	HR (95% CI)	P	C-statistic	HR (95% CI)	P	C-statistic
Age at diagnosis, years	1.04 (1.00–1.09)	0.087	0.62	1.03 (1.00–1.08)	0.094	0.61	1.02 (1.01–1.03)	<0.001	0.59	1.03 (1.01–1.04)	<0.001	0.60
Female gender (n = 197)	0.67 (0.25–1.71)	0.41	0.55	0.72 (0.31–1.60)	0.43	0.54	0.83 (0.64–1.08)	0.2	0.52	0.79 (0.61–1.02)	0.072	0.53
Performance status	1.44 (0.77–2.51)	0.23	0.55	1.64 (0.98–2.64)	0.048	0.57	1.24 (1.05–1.47)	0.013	0.54	1.36 (1.14–1.61)	<0.001	0.56
Performance status			0.59			0.57			0.54			0.56
[0] (n = 238)	Ref.			Ref.			Ref.			Ref.		
[1] (n = 156)	0.75 (0.23–2.17)	0.61		1.15 (0.46–2.79)	0.75		1.34 (1.02–1.75)	0.036		1.44 (1.09–1.89)	0.009	
[2] (n = 26)	2.97 (0.63–10.6)	0.12		2.46 (0.53–8.43)	0.19		1.16 (0.67–2.03)	0.6		1.39 (0.80–2.41)	0.2	
[3] (n = 8)	3.26 (0.17–21.0)	0.29		6.28 (0.86–30.8)	0.034		2.40 (1.17–4.91)	0.017		3.25 (1.58–6.66)	0.001	
Secondary AML (n = 91)	1.00 (0.54–1.59)	>0.99	0.53	1.06 (0.65–1.57)	0.81	0.55	1.55 (1.16–2.08)	0.003	0.55	1.51 (1.13–2.02)	0.006	0.54
WBC counts at diagnosis, G/l	1.01 (1.00–1.01)	0.007	0.52	1.01 (1.00–1.01)	0.006	0.51	1.00 (1.00–1.00)	0.6	0.50	1.00 (1.00–1.00)	0.5	0.50
ELN-2017 classification			0.64			0.64			0.62			0.64
Favorable (n = 127)	Ref.			Ref.			Ref.			Ref.		
Intermediate (n = 140)	8.66 (1.59–161)	0.042		11.8 (2.28–217)	0.019		1.66 (1.17–2.36)	0.004		1.75 (1.23–2.49)	0.002	
Adverse (n = 159)	7.56 (1.39–140)	0.057		11.2 (2.19–205)	0.021		2.59 (1.86–3.60)	<0.001		2.70 (1.93–3.77)	<0.001	
Unfitness RCT eligibility criteria (n = 104)	1.88 (0.68–4.80)	0.20	0.57	1.71 (0.71–3.88)	0.21	0.56	1.08 (0.81–1.45)	0.6	0.50	1.25 (0.93–1.67)	0.14	0.52
Augmented HCT-CI	1.32 (1.08–1.61)	0.005	0.66	1.36 (1.14–1.63)	<0.001	0.67	1.05 (0.99–1.12)	0.10	0.54	1.07 (1.01–1.14)	0.03	0.56
Augmented HCT-CI category			0.65			0.64			0.54			0.55
[0–2] (n = 210)	Ref.			Ref.			Ref.			Ref.		
[3,4] (n = 130)	1.08 (0.27–3.85)	0.91		1.44 (0.49–4.10)	0.49		1.24 (0.93–1.67)	0.15		1.33 (0.99–1.79)	0.059	
[5,6] (n = 68)	3.29 (1.00–10.9)	0.045		2.90 (0.98–8.39)	0.048		1.07 (0.74–1.55)	0.7		1.17 (0.81–1.69)	0.4	
[≥7] (n = 20)	6.00 (1.19–25.0)	0.017		6.31 (1.55–22.5)	0.006		1.57 (0.88–2.79)	0.13		1.70 (0.95–3.02)	0.072	
AML-CM	1.34 (1.13–1.58)	<0.001	0.70	1.39 (1.19–1.62)	<0.001	0.71	1.13 (1.07–1.18)	<0.001	0.61	1.15 (1.10–1.21)	<0.001	0.63
AML-CM category			0.67			0.68			0.59			0.62
[1–4] (n = 174)	Ref.			Ref.			Ref.			Ref.		
[5–6] (n = 136)		0.5			0.91			0.004			<0.001	

Table 3. continued

Characteristic	28-day mortality			56-day mortality			Event-free survival			Overall survival		
	OR (95% CI)	P	C-statistic	OR (95% CI)	P	C-statistic	HR (95% CI)	P	C-statistic	HR (95% CI)	P	C-statistic
	1.62 (0.42–6.67)			1.07 (0.30–3.62)			1.58 (1.16–2.16)	<0.001		1.72 (1.26–2.35)	<0.001	
[7–9] (n = 98)	2.77 (0.77–11.1)	0.12		2.83 (0.99–8.69)	0.055		1.80 (1.29–2.51)	<0.001		2.09 (1.49–2.92)	<0.001	
[≥10] (n = 20)	10.6 (2.32–49.0)	0.002		12.0 (3.36–43.4)	<0.001		2.60 (1.50–4.51)	<0.001		2.95 (1.70–5.12)	<0.001	
Ferrara criteria category			0.61			0.64			0.51			0.52
Fit (n = 296)	Ref.			Ref.			Ref.			Ref.		
Unfit (n = 132)	2.61 (1.03–6.73)	0.042		3.32 (1.49–7.63)	0.004		1.05 (0.79–1.38)	0.7		1.11 (0.84–1.46)	0.5	
TRM score	1.06 (1.02–1.10)	0.002	0.64	1.07 (1.03–1.11)	<0.001	0.66	1.02 (1.01–1.04)	0.01	0.57	1.03 (1.02–1.05)	<0.001	0.60
TRM score category			0.55			0.57			0.50			0.51
Lower (≤ 13.1) (n = 398)	Ref.			Ref.			Ref.			Ref.		
Higher (> 13.1) (n = 30)	2.65 (0.59–8.59)	0.14		3.59 (1.13–9.71)	0.018		1.08 (0.65–1.79)	0.8		1.37 (0.82–2.28)	0.2	
MDACC scoring system	1.36 (1.12–1.64)	0.001	0.66	1.39 (1.18–1.65)	<0.001	0.68	1.17 (1.10–1.24)	<0.001	0.60	1.22 (1.14–1.30)	<0.001	0.62
MDACC scoring system category			0.57			0.60			0.53			0.54
Low (≤ 4) (n = 370)	Ref.			Ref.			Ref.			Ref.		
High (5–8) (n = 52)	1.56 (0.35–4.98)	0.50		2.08 (0.66–5.50)	0.17		1.56 (1.08–2.24)	0.018		1.74 (1.12–2.51)	0.003	
Very High (≥ 9) (n = 6)	12.7 (1.66–71.2)	0.005		19.6 (3.42–112)	<0.001		2.77 (1.14–6.74)	0.025		4.78 (1.95–11.7)	<0.001	

AML acute myeloid leukemia, AML-CM AML composite model, CI confidence interval, HCT-CI hematopoietic cell transplantation comorbidity index, ELN European LeukemiaNet, HR hazard ratio, IQR interquartile range, MDACC MD Anderson Cancer Center, MRC Medical Research Council, OR odds ratio, RCT randomized clinical trials, TRM treatment related mortality, WBC white blood cells.

(A)

Unfit RCT criteria

(B)

Augmented HCT-CI

(C)

AML-CM

(D)

Ferrara criteria

(E)

TRM score category
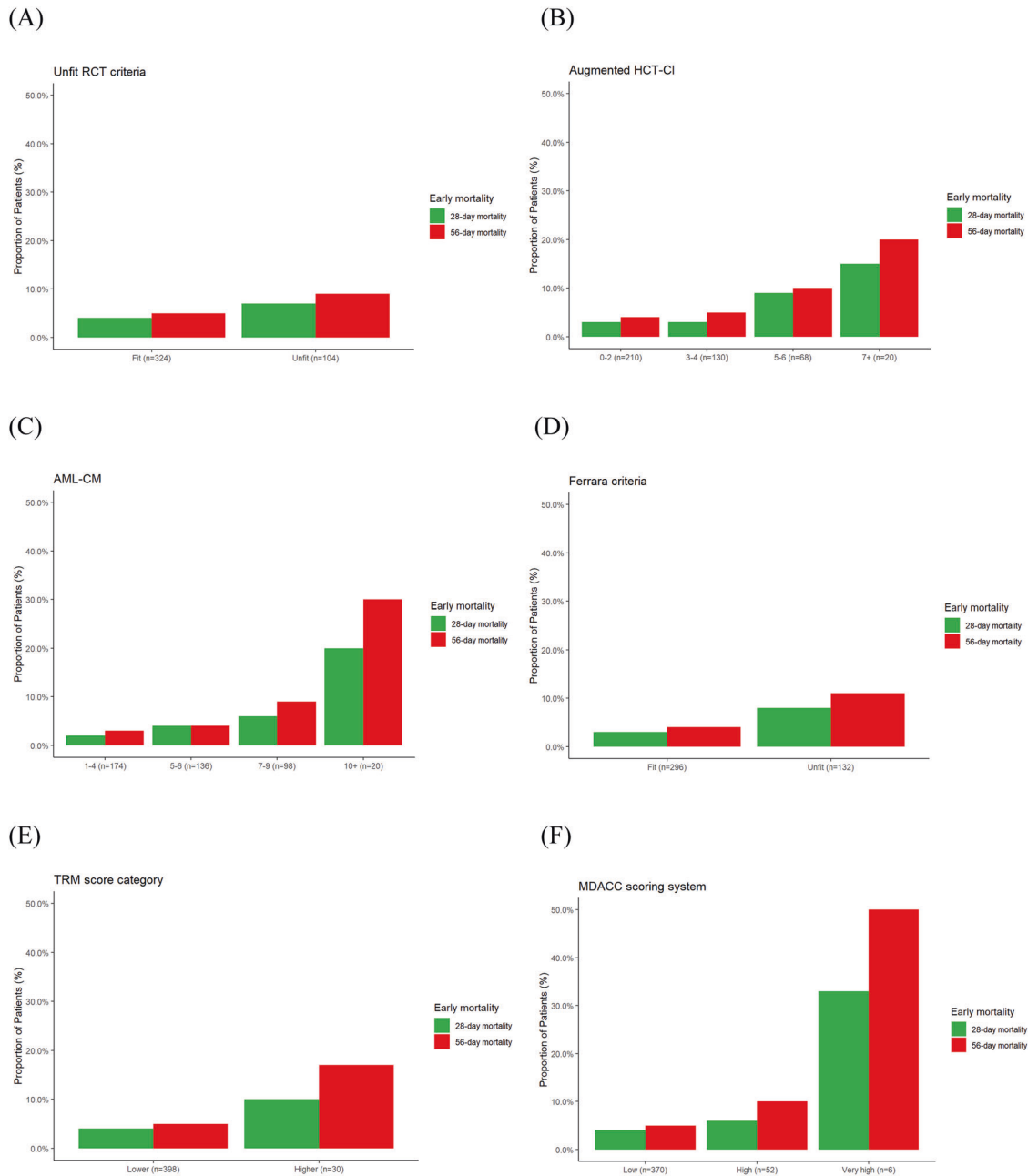
(F)

MDACC scoring system

Fig. 1 Distribution of predictive scoring systems by early (28-day and 56-day) mortality. The proportion of patients with early mortality is represented for individual patient categories, as stratified by scoring systems. A RCT criteria for unfitness, **B** Augmented HCT-CI, **C** AML-CM, **D** Ferrara criteria, **E** TRM score, and **F** MDACC scoring system. AML acute myeloid leukemia, AML-CM AML composite model, HCT-CI hematopoietic cell transplantation comorbidity index, MDACC MD Anderson Cancer Center, RCT randomized clinical trial, TRM treatment related mortality.

predictions of 28 and 56-day mortality were observed with the AML-CM (C-statistic, 0.70 and 0.71, respectively) followed by the MDACC scoring system (C-statistic, 0.66 and 0.68, respectively), the augmented HCT-CI (C-statistic, 0.65 and 0.67, respectively), the TRM score (C-statistic, 0.64 and 0.66, respectively), and the Ferrara criteria (C-statistic, 0.61 and 0.64, respectively) (Table 3). The discriminative ability of the criteria used to evaluate fitness in RCT enrolling less fit patients was very limited at all time points (maximum C-statistic of 0.57 at 28 days). The majority of the risk scores performed better for the prediction of 28 and 56-day mortality than for the prediction of EFS and OS, but the predictability remained similar for the two scores incorporating cytogenetic and molecular variables, namely the AML-CM and the

MDACC scoring system (Fig. 3). The accuracy (Brier scores) of all risk scores was somewhat similar although slightly better for the AML-CM, the MDACC scoring system, and the TRM score.

Association between individual risk scores, early mortality, and long-term survival in patients older than 60 years

As the ability to predict early mortality may be particularly relevant in older patients, we performed subset analyses restricting our dataset to the 216 individuals ≥60 years of age. This sensitivity analysis was also driven by the fact that patients judged less fit in RCTs enrolling ineligible patients for intensive induction therapy had to be 60 years or older. These 216 patients were more likely to have a PS of 1, rather than 0, in comparison to younger patients (PS 0,
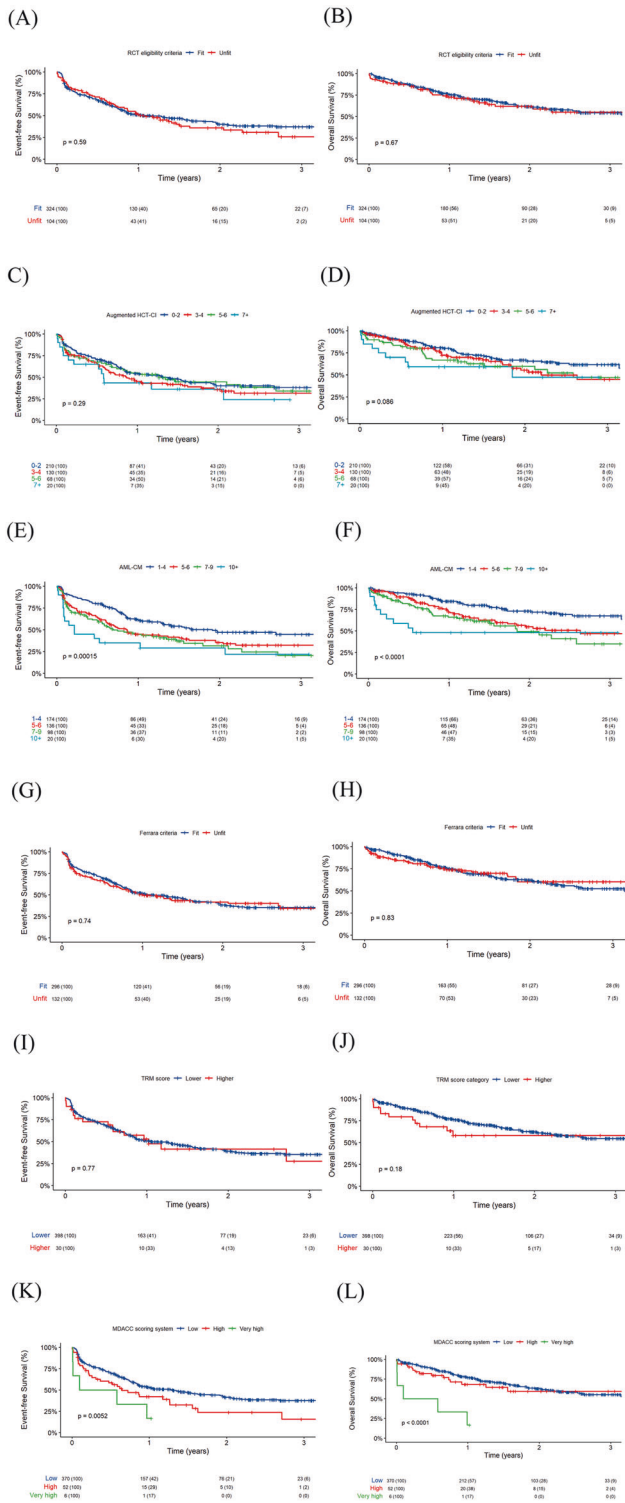
Fig. 2 Outcome of patients stratified by each scoring systems.
Kaplan-Meier plots of event-free and overall survival according to
A, **B** RCT criteria for unfitness, **C**, **D** Augmented HCT-CI, **E**, **F** AML-CM,
G, **H** Ferrara criteria, **I**, **J** TRM score, and **K**, **L** MDACC scoring system.
AML acute myeloid leukemia, AML-CM AML composite model, HCT-
CI hematopoietic cell transplantation comorbidity index, MDACC
MD Anderson Cancer Center, RCT randomized clinical trial, TRM
treatment related mortality.

64% vs. 47%; PS 1, 28% vs. 44%, for patients older than 60 years,
respectively, $P = 0.003$) and to have less favorable cytogenetic/
molecular disease risk (37% vs. 23% for patients older 60 years,
$P = 0.001$) at the expense of an increase in intermediate risk (26% vs.
39%, $P = 0.006$) whereas the proportion of patients with adverse-risk
disease was similar between younger and older individuals (37% vs.
38%; $P = 0.88$) (Supplementary Table 6). As all scoring systems
include age as a covariate, risk scores were generally higher in older
vs. younger. Despite higher risk scores observed in older patients,
early mortality was low, with 12 (6%) and 17 (8%) deaths occurring
at 28 and 56 days, respectively. The AML-CM and the MDACC
scoring systems were associated with 28-day mortality (OR = 1.26
[0.99–1.58], $P = 0.049$; and OR = 1.30 [0.99–1.66, $P = 0.047$, respec-
tively) whereas the augmented HCT-CI, the AML-CM, the TRM score,
and the MDACC scoring system were associated with 56-day
mortality (OR = 1.42 [1.13–1.79, $P = 0.002$; OR = 1.42 [1.16–1.75],
$P < 0.001$; OR = 1.07 [1.02–1.12], $P = 0.003$; and OR = 1.39
[1.10–1.74], $P = 0.004$, respectively) (Supplementary Table 7). As in
the whole cohort, EFS and OS were only associated with the AML-
CM (HR = 1.10 [1.02–1.19], $P = 0.010$; and HR = 1.17 [1.07–1.28],
$P < 0.001$, respectively) and the MDACC scoring system (HR = 1.13
[1.02–1.24], $P = 0.015$; and HR = 1.16 [1.02–1.32], $P = 0.024$, respec-
tively) (Supplementary Table 7, Supplementary Fig. 2).

DISCUSSION

Several risk scores have been developed to assess the suitability of
individual adults for intensive AML chemotherapy to help in the
decision-making process. In our study, all risk scores identified a
small subgroup of patients, with size number varying from one
score to another, with increased early mortality. However, overall,
the prediction accuracy of all scoring systems was limited with the
best performance having a C-statistic close to 0.70 for early
mortality. Only scoring systems including AML characteristics
could help in predicting long-term survival. Criteria used to define
unfitness in recent randomized clinical trials enrolling patients
judged unsuitable for intensive chemotherapy did not perform
well to stratify patients or to predict outcome. Despite the
association with outcomes seen with several risk scores, the
predictive accuracy overall was limited, highlighting important
limitations of our ability to estimate fitness.

The use of previously elaborated scores identified a subset of
"high-risk" patients with increased early mortality although early
mortality rates in our cohort were low (28 and 56-day mortality
rates of 4% and 6%, respectively), while 28-day mortality varied
from 5% to 11% in the cohorts used to develop or validate theses
risk scores [12, 16]. One may speculate that improvement in
supportive care and the use of less toxic induction therapy
regimens, such as CPX-351, although seldomly used in our
institutions, may explain this lower early mortality rate. It is
important to point out that the proportion of patients identified as
high risk for poor outcome varied greatly between scores, from
2% for the MDACC scoring system to 30% for the TRM score,
which contrasts to the training and validation cohorts used to
develop these scores were the subset of high-risk patients was
usually higher (22% for both the augmented HCT-CI and the AML-
CM, 11 to 20% for the TRM score, and 9% for the MDACC scoring
system) [12, 14, 16]. On the other hand, the Ferrara unfit criteria
was observed in 31% of patients, near identical to the rate
observed in the study by Palmieri et al., but with an impact on
early mortality but not overall survival [12].

We examined the comparative predictive accuracy of scoring
systems which was lower than in previous training and validation
cohorts. Whereas the prediction value for 56-day mortality were
0.72 and 0.78 in the validation cohorts for the augmented HCT-CI
and the AML-CM, respectively, these values were lower in our
cohort (0.67 and 0.71, respectively) [14]. This was also observed for
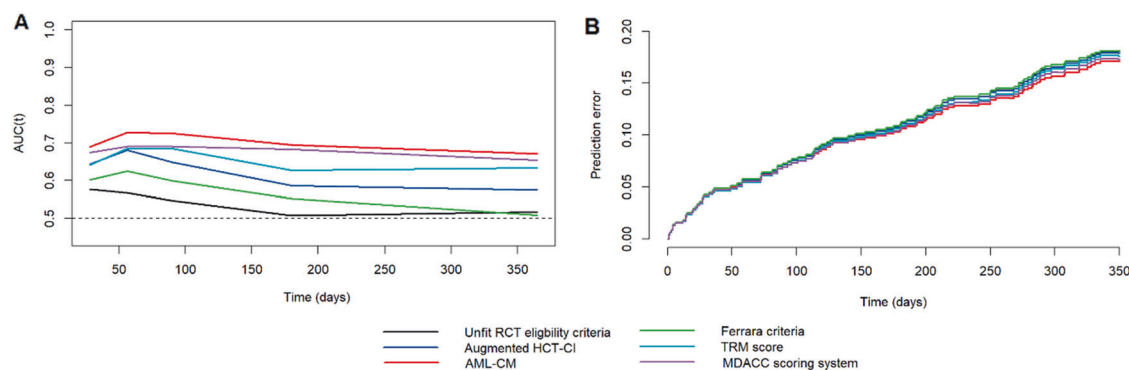the Ferrara criteria and the TRM score which had C-statistics of

Fig. 3   Time AUC and Brier score for prediction of mortality over time for different scoring systems. AUCs (**A**) and Brier scores (**B**) for overall survival according to time are depicted.

0.76 and 0.72 for 28-day mortality, respectively (0.61 and 0.64 in our cohort) [12]. As in previous studies, we observe that scoring systems including AML-specific characteristics, i.e., ELN-2017 classification for the AML-CM, secondary-type AML for the TRM score, and cytogenetics for the MDACC scoring system, were slightly more accurate in predicting OS [12]. We can hope that recent advancements in molecular characterization of AML by next-generation sequencing can further improve the prediction ability of these scoring systems [23, 24].

Assessing a patient's suitability for intensive chemotherapy remains a challenge. Whereas most RCTs in adults with AML enrolled patients eligible for intensive chemotherapy in the last decade, 13 (17%) specifically enrolled patients judged unfit for intensive chemotherapy [25]. In earlier trials, this was determined by the local investigator's own evaluation but in most recent years, eligibility criteria for "non-eligible" patients for intensive AML chemotherapy have been proposed. Although 24% of patients in our cohort would have judged not suitable and thus eligible for participation in trials evaluating less intense induction therapy regimens, these patients had a similar outcome than those judged fit. Despite the absence of formal validation, these criteria will probably be more frequently used in the future, to compare therapies in medically less-fit patients. Our results suggest that it would be more appropriate to use scores that have been specifically developed to determine which patients have a greater risk for early treatment mortality and might therefore not be eligible for intensive chemotherapy [12, 14, 16].

Although these scores are not routinely used in our institutions, there is undeniable patient selection in our cohort of patients who received intensive chemotherapy. This might explain why the proportion of patients classified as high-risk by some scores was low. Since patients in our cohort were treated more recently than patients treated in the training cohorts used to elaborate scoring systems, we cannot exclude that some high-risk patients may have rather received less-intense chemotherapy in our institutions. Similar selection processes are present in most institutions, perhaps not to the extent that is observed in our cohort, which may limit the generalizability of scoring systems. This further emphasizes the importance of external validation when developing such tools. The predictive ability of some scores may have been further limited in our cohort due to missing data used to calculate them, including uricemia levels at diagnostic and pulmonary function testing. Since all scoring systems, including the augmented HCT-CI, were evaluated retrospectively, some relevant comorbidities may have been under reported. As some criteria used to calculate these scores were occasionally observed, it was not possible to assess how individual criteria contributed to the accuracy of these scoring systems.

Previously developed risk scores could identify a small subgroup of patients with increased early mortality after intensive AML therapy, but accurate individualized prediction was relatively limited. Criteria used to define suitability for intensive AML chemotherapy in recent RCTs targeting patients judged less fit for intensive chemotherapy did not perform well to stratify patients or to predict outcome.

REFERENCES

1. Sánchez-Vizcaíno F, Tamayo C, Ramos F, Láinez-González D, Serrano-López J, Barba R et al. Identification of seasonal variation in the diagnosis of acute myeloid leukaemia: a population-based study. Br J Haematol. 2022;198:545–55.
2. Juliusson G, Antunovic P, Derolf Å, Lehmann S, Möllgård L, Stockelberg D, et al. Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. Blood. 2009;113:4179–87.
3. Appelbaum FR, Gundacker H, Head DR, Slovak ML, Willman CL, Godwin JE, et al. Age and acute myeloid leukemia. Blood. 2006;107:3481–5.
4. Etienne A, Esterni B, Charbonnier A, Mozziconacci M-J, Arnoulet C, Coso D, et al. Comorbidity is an independent predictor of complete remission in elderly patients receiving induction chemotherapy for acute myeloid leukemia. Cancer. 2007;109:1376–83.
5. Hjort Jakobsen L, Stidsholt Roug A, Kiesbye Øvlisen A, Werenberg Marcher C, Beier Ommen H, Theilgaard-Mönch K, et al. Temporal changes in survival among adult patients with acute myeloid leukaemia in the period 2000–16: a Danish population-based study. Br J Haematol. 2021;193:482–7.
6. Juliusson G, Hagberg O, Lazarevic VL, Ölander E, Antunovic P, Cammenga J, et al. Improved survival of men 50 to 75 years old with acute myeloid leukemia over a 20-year period. Blood. 2019;134:1558–61.
7. Kaplan ZLR, van Leeuwen N, Posthuma EFM, Visser O, Huls G, van de Loosdrecht AA, et al. Improved relative survival in older patients with acute myeloid leukemia over a 30-year period in the Netherlands: a long haul is needed to change nothing into something. Leukemia. 2022;36:596–8.
8. Dombret H, Seymour JF, Butrym A, Wierzbowska A, Selleslag D, Jang JH, et al. International phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with >30% blasts. Blood. 2015;126:291–9.
9. Wei AH, Montesinos P, Ivanov V, DiNardo CD, Novak J, Laribi K, et al. Venetoclax plus LDAC for newly diagnosed AML ineligible for intensive chemotherapy: a phase 3 randomized placebo-controlled trial. Blood. 2020;135:2137–45.
10. DiNardo CD, Jonas BA, Pullarkat V, Thirman MJ, Garcia JS, Wei AH, et al. Azacitidine and venetoclax in previously untreated acute myeloid leukemia. N Engl J Med. 2020;383:617–29.
11. Sorror ML, Storer BE, Fathi AT, Brunner A, Gerds AT, Sekeres MA, et al. Multisite 11-year experience of less-intensive vs intensive therapies in acute myeloid leukemia. Blood. 2021;138:387–400.
12. Palmieri R, Othus M, Halpern AB, Percival M-EM, Godwin CD, Becker PS, et al. Accuracy of SIE/SIES/GITMO consensus criteria for unfitness to predict early mortality after intensive chemotherapy in adults with AML or other high-grade myeloid neoplasm. JCO. 2020;38:4163–74.
13. Walter RB, Othus M, Borthakur G, Ravandi F, Cortes JE, Pierce SA, et al. Prediction of early death after induction therapy for newly diagnosed acute myeloid leukemia with pretreatment risk scores: a novel paradigm for treatment assignment. JCO. 2011;29:4417–24.
14. Sorror ML, Storer BE, Fathi AT, Gerds AT, Medeiros BC, Shami P, et al. Development and validation of a novel acute myeloid leukemia–composite model to estimate risks of mortality. JAMA Oncol. 2017;3:1675.

15. Ferrara F, Barosi G, Venditti A, Angelucci E, Gobbi M, Pane F, et al. Consensus-based definition of unfitness to intensive and non-intensive chemotherapy in acute myeloid leukemia: a project of SIE, SIES and GITMO group on a new tool for therapy decision making. Leukemia. 2013;27:997–9.

16. Sasaki K, Kadia T, Begna K, DiNardo CD, Borthakur G, Short NJ, et al. Prediction of early (4-week) mortality in acute myeloid leukemia with intensive chemotherapy. Am J Hematol. 2022;97:68–78.

17. Pigneux A, Béné MC, Salmi L-R, Dumas P-Y, Delaunay J, Bonmati C, et al. Improved survival by adding lomustine to conventional chemotherapy for elderly patients with AML without unfavorable cytogenetics: results of the LAM-SA 2007 FILO Trial JCO. 2018;36:3203–10.

18. Lancet JE, Uy GL, Newell LF, Lin TL, Ritchie EK, Stuart RK, et al. CPX-351 versus 7+3 cytarabine and daunorubicin chemotherapy in older adults with newly diagnosed high-risk or secondary acute myeloid leukaemia: 5-year results of a randomised, open-label, multicentre, phase 3 trial. Lancet Haematol. 2021;8:e481–e491.

19. Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood. 2010;116:354–65.

20. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129:424–47.

21. Sorror ML, Maris MB, Storb R, Baron F, Sandmaier BM, Maloney DG, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. Blood. 2005;106:2912–9.

22. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162:W1–73.

23. Duchmann M, Micol J-B, Duployez N, Raffoux E, Thomas X, Marolleau J-P, et al. Prognostic significance of concurrent gene mutations in intensively treated patients with IDH -mutated AML: an ALFA study. Blood. 2021;137:2827–37.

24. Fenwarth L, Thomas X, de Botton S, Duployez N, Bourhis J-H, Lesieur A, et al. A personalized approach to guide allogeneic stem cell transplantation in younger adults with acute myeloid leukemia. Blood. 2021;137:524–32.

25. Orvain C, Othus M, Johal G, Hunault-Berger M, Appelbaum FR, Walter RB. Evolution of eligibility criteria for non-transplant randomized controlled trials in adults with acute myeloid leukemia. Leukemia. 2022;36:2002–8.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41375-022-01677-z.

Correspondence and requests for materials should be addressed to Corentin Orvain.

Reprints and permission information is available at http://www.nature.com/reprints