HEMATOLOGIC MALIGNANCY

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

Raffaele Palmieri, MD[1]; Megan Othus, PhD[2]; Anna B. Halpern, MD[1,3]; Mary-Elizabeth M. Percival, MD, MS[1,3]; Colin D. Godwin, MD, MPhil[1,3]; Pamela S. Becker, MD, PhD[1,3]; and Roland B. Walter, MD, PhD, MS[1,3,4]

PURPOSE With increasing therapeutic alternatives available, there is growing interest in tools that accurately identify patients most suitable for intensive acute myeloid leukemia (AML) chemotherapy. Nowadays, conceptual criteria proposed by an Italian panel of experts are widely used for this purpose. How accurately these Ferrara criteria predict fitness for intensive chemotherapy is unknown.

PATIENTS AND METHODS We assessed the fitness of adults undergoing intensive AML therapy based on Ferrara criteria and determined the accuracy of this assessment for early mortality and survival prediction.

RESULTS Among 655 adults who received curative-intent induction or reinduction chemotherapy with 7 days of standard-dose cytarabine and 3 days of an anthracycline ("7+3") CLAG-M (cladribine, high-dose cytarabine, granulocyte colony-stimulating factor, and mitoxantrone), or reduced-dose CLAG-M, 197 (30%) met at least one of the criteria defining unfitness for intensive chemotherapy (F-unfit). Compared with F-fit patients, the overall survival of F-unfit patients was significantly shorter (median, 4.8 months; 95% CI, 3.6 to 6.5 months *v* 36.8 months; 95% CI, 27.4 to 73.0 months; $P < .001$). When used alone, the Ferrara unfitness assessment was more accurate in predicting day 28 and day 100 mortality than the treatment-related mortality score we developed previously (used binary, ≤ 13.1 *v* > 13.1), as indicated by area under the receiver operating characteristic curve (AUC) values of 0.76 and 0.79 versus 0.66 and 0.62. The predictive accuracy of the Ferrara unfitness assessment could be significantly improved by including additional covariates such as performance status and albumin, yielding AUCs as high as 0.84-0.85 for the prediction of day 28 or day 100 mortality. Prediction of overall survival was less accurate, yielding a c-statistic value as high as 0.75 in multivariable models.

CONCLUSION Ferrara unfitness criteria provide a good prediction tool for shorter-term mortality after intensive AML chemotherapy. Our data may serve as a benchmark for expected outcomes with intensive chemotherapy in F-fit and F-unfit patients.

J Clin Oncol 38:4163-4174. © 2020 by American Society of Clinical Oncology

INTRODUCTION

Intensive chemotherapy is central to the treatment of acute myeloid leukemia (AML).[1,2] Although improvements in supportive care have substantially reduced early deaths (treatment-related mortality [TRM]),[3-5] overwhelming toxicities remain concerning, particularly for older individuals and those with comorbidities. Thus, there is ongoing interest in accurately assessing fitness for intensive AML chemotherapy.[6-11] This interest has only increased with the availability of less-intense treatment alternatives.[12-17]

Various factors are associated with early death after intensive AML chemotherapy[10,11] and can be incorporated into scoring systems for TRM prediction. Some of these, including the TRM score we developed previously,[18] attain a good (but far from perfect) predictive ability, as indicated by area under the receiver operating characteristic curve (AUC) values of $\geq$ 0.7-0.8.[11]

As an alternative to quantitative scoring systems, a panel convened by the Italian Society of Hematology (SIE), the Italian Society of Experimental Hematology (SIES), and the Italian Group for Bone Marrow Transplantation (GITMO) selected conceptual (Ferrara) criteria to classify patients as fit for intensive chemotherapy, fit for nonintensive chemotherapy, or unfit for nonintensive chemotherapy.[20] Although widely used, it is unknown how useful these criteria are for fitness evaluation. We used a large cohort of adults treated with intensive AML-like chemotherapy to assess the ability of the Ferrara criteria to predict early death and survival and compared results with those obtained with the TRM score.

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on July 23, 2020 and published at ascopubs.org/journal/jco on October 8, 2020: DOI https://doi.org/10.1200/JCO.20.01392

CONTEXT

Key Objective

Fitness evaluations based on criteria such as those proposed by Ferrara et al[20] are commonly used in AML, but how accurately they predict early mortality after intensive chemotherapy is unknown.

Knowledge Generated

Studying a large number of adults treated with intensive AML-like chemotherapy, we observed a day 28/100 mortality of 2%/5% for Ferrara-fit and 14%/42% for Ferrara-unfit patients, as well as a median survival of > 3 years for Ferrara-fit versus < 6 months for Ferrara-unfit patients. Ferrara criteria–based fitness assessments, either alone or with a small number of additional parameters, had good to very good accuracy in predicting day 28 and day 100 mortality for individual patients.

Relevance

Our findings indicate that the Ferrara criteria provide a useful tool to predict early mortality after intensive AML chemotherapy, which, in conjunction with molecular/genetic data, could serve as a basis for informed decision making, particularly in older patients and those with comorbidities.

PATIENTS AND METHODS

We identified adults ≥ 18 years of age with AML[21] or other myeloid neoplasm presenting with ≥ 10% blasts in the blood and/or marrow who received induction or reinduction chemotherapy with 7 + 3, CLAG-M (cladribine, high-dose cytarabine, granulocyte colony-stimulating factor, and mitoxantrone), or dose-reduced CLAG-M between January 2006 and January 2020 at our institution. The TRM score was computed with an online calculator [18,19] and corresponds to the predicted probability of death within 28 days of beginning intensive chemotherapy.[18] The criteria proposed by Ferrara et al[20] (Data Supplement, online only) were used to categorize patients into Ferrara-fit (F-fit) and Ferrara-unfit (F-unfit).[20] Results from cardiac and pulmonary function tests performed after administration of chemotherapy were used for classification if not defining unfitness.

Overall survival (OS) was estimated using the Kaplan-Meier method. Fisher's exact and Kruskal-Wallis tests assessed differences between categorical and quantitative variables across categories. We used multivariable logistic regression and Cox models to assess the relationship between individual covariates and outcomes of interest, and then used AUCs and c-statistics to quantify predictive ability. Two-sided *P* values are reported. Additional information regarding patient selection and classification, treatment, and the statistical methods are provided in the Data Supplement. This retrospective study was approved by the Fred Hutchinson Cancer Research Center Institutional Review Board.

RESULTS

We identified 703 adults with AML (n = 567; 81%) or other high-grade myeloid neoplasm (n = 136; 19%) who received 1 (n = 585; 83%), 2 (n = 102; 15%), or 3 (n = 16; 2%) cycles of induction chemotherapy with 7 + 3, CLAG-M, or dose-reduced CLAG-M for newly diagnosed or relapsed/refractory disease (Table 1). Because pulmonary function testing is not routinely performed for patients with AML not undergoing transplantation at our institution, 335 patients did not have results from such tests available. Among the other 368 patients, 159 had pulmonary function tests performed before chemotherapy initiation, with abnormal findings in 79; in 209 patients, baseline pulmonary function tests were not available but normal results were obtained at one or more later time points, at a median of 98 days (interquartile range, 73-149 days) after the start of induction or reinduction chemotherapy. For our overall analyses, we considered patients with missing baseline pulmonary function tests to lack pulmonary compromise for the purpose of Ferrara fitness assessments if there was no history of pulmonary comorbidities and/or respiratory symptoms (see subset analysis in "Performance of Ferrara Unfitness Assessment in Distinct Patient Subsets"). With this approach, 655 (93%) and 642 (91%) patients could be classified based on Ferrara criteria and TRM score. One hundred ninety-seven of the 655 Ferrara criteria-classifiable patients (30%) were F-unfit, with 186 (28%) meeting one and 11 (2%) meeting more than one of the unfitness-defining criteria. Pulmonary function impairment was the most frequent unfitness criterion met (n = 79; Data Supplement), followed by age > 75 years (n = 42), active infections (n = 32), Eastern Cooperative Oncology Group performance status (PS) ≥ 3 not related to hematologic malignancy (n = 26), and cardiac comorbidities (n = 24). The TRM score ranged from 0.01-78.32 among 642 evaluable patients. We separated patients into TRM score low (n = 571; 89%) versus high (n = 71; 11%) using a cut-off of 13.1 per our local practice,[22,23] but also assessed the effect of the TRM score as a continuous variable.

Association Between Ferrara and TRM Score Fitness Classification and Outcomes

F-fit and F-unfit patients differed significantly with respect to age, PS, disease risk, laboratory findings at baseline, and

TABLE 1. Baseline Demographic and Clinical Characteristics of Study Cohort (N = 703)

Characteristic	All Patients	Missing Data (No.)
Age at diagnosis (range), years	60.5 (18.6-91.4)	0
Sex		0
Female	307 (44)	
Male	396 (56)	
Performance status		8
0-1	600 (86)	
2-3	95 (14)	
Disease		0
AML	567 (81)	
Other	136 (19)	
MDS-EB-2	122 (17)	
CMML-2	8 (1)	
MDS/MPN, unclassifiable	3 (0.5)	
Myelofibrosis	3 (0.5)	
Secondary disease		0
No	465 (66)	
Yes	238 (34)	
Therapy related	57 (8)	
Antecedent hematologic disorder	181 (26)	
Cytogenetic risk at initial disease diagnosis		42
Favorable	50 (8)	
Intermediate	407 (62)	
Adverse	204 (31)	
Disease status		0
Newly diagnosed	509 (72)	
Refractory	97 (14)	
Relapsed	97 (14)	
Prior HCT		0
No	652 (93)	
Yes	51 (7)	
Treatment		0
7 + 3	169 (24)	
CLAG-M	496 (71)	
Dose-reduced CLAG-M	38 (5)	
Laboratory finding at baseline		
WBC, $\times 10^9$/L (range)	5.5 (0.1-356.3)	58
Platelet count, $\times 10^9$/L (range)	51 (1-794)	58
Peripheral blood blasts, % (range)	9.5 (0-99)	59
Albumin, g/dL (range)	3.7 (1.8-5.1)	59
Creatinine, mg/dL (range)	0.87 (0.34-11.55)	59

(continued in next column)

TABLE 1. Baseline Demographic and Clinical Characteristics of Study Cohort (N = 703) (continued)

Characteristic	All Patients	Missing Data (No.)
Treatment year		0
2006-2014	248 (35)	
2015-2020	455 (65)	
TRM score (range)	3.22 (0.01-78.32)	61
TRM score category		61
Lower (≤ 13.1)	571 (89)	
Higher (> 13.1)	71 (11)	
Ferrara criteria category		48
F-fit	458 (70)	
F-unfit	197 (30)	
Ferrara criteria components		
Age, years		0
< 75	661 (94)	
≥ 75	42 (6)	
Performance status		7
< 3	670 (96)	
≥ 3	26 (4)	
Heart		59
Negative	620 (96)	
Positive	24 (4)	
Lungs		335
Negative	289 (79)	
Positive	79 (21)	
Kidneys		1
Negative	701 (100)	
Positive	1 (0)	
Liver		1
Negative	700 (100)	
Positive	2 (0)	
Infection		1
Negative	670 (95)	
Positive	32 (5)	
Mental illness		0
Negative	702 (100)	
Positive	1 (0)	
Other comorbidities		0
Negative	702 (100)	
Positive	1 (0)	

NOTE. Data are No. (%) unless otherwise indicated.

Abbreviations: AML, acute myeloid leukemia; CLAG-M, cladribine, cytarabine, granulocyte colony-stimulating factor, mitoxantrone; CMML-2, chronic myelomonocytic leukemia-2; HCT, hematopoietic cell transplantation; MDS-EB-2, myelodysplastic syndrome with excess blasts-2; MDS/MPN, myelodysplastic syndrome/myeloproliferative neoplasm; TRM, treatment-related mortality.

type of chemotherapy administered (Table 2). Likewise, patients with lower TRM score (ie, ≤ 13.1) differed significantly from those with higher TRM score (ie, > 13.1; Data Supplement). In many patients, Ferrara and TRM score classifications were concordant (Fig 1A): 398 of the 458 F-fit patients (95%) had lower TRM scores, and 52 of the 197 F-unfit patients (28%) had higher TRM scores. However, 19 of the F-fit patients (5%) had higher TRM scores, and 135 of the F-unfit patients (72%) had lower TRM scores (Fig 1A). There were 37 and 107 deaths within 28 and 100 days, respectively, of chemotherapy initiation in our cohort. Ninety-three of the 107 patients (87%) who died within 100 days received only 1 cycle of chemotherapy; 13 (12%) and one (1%) received 2 or 3 courses of chemotherapy. Primary causes of death by day 100 are summarized in the Data Supplement. Seven of 457 (2%) and 22 of 444 (5%) F-fit patients with follow-up data sufficient for endpoint evaluation died within 28 days or 100 days of initiation of chemotherapy, compared with 28 of 196 (14%) and 78 of 185 (42%) F-unfit patients. Both Ferrara unfitness criteria and the TRM score were associated with survival. As depicted in Figure 1B, F-unfit patients had statistically significantly shorter survival than F-fit patients ($P < .001$), with a median OS of 4.8 months (95% CI, 3.6 to 6.5 months) versus 36.8 months (95% CI, 27.4 to 73.0 months). Likewise, patients with a higher TRM score had significantly shorter survival than those with a lower TRM score ($P < .001$), with a median OS of 3.6 months (95% CI, 2.6 to 10.0 months) versus 18.4 months (95% CI, 14.3 to 23.2 months; Fig 1C). When outcome was stratified by Ferrara unfitness criteria and TRM score, F-unfit patients with higher TRM score had the shortest survival (median, 2.6 months; 95% CI, 2.1 to 3.9 months), whereas F-fit patients with a lower TRM score had the longest survival (median, 36.9 months; 95% CI, 27.4 to 75.8 months). The survival of patients with discordant fitness assessment results had outcomes between those with concordant results (F-fit/higher TRM score: median OS, 30.7 months; 95% CI, 13.5 months to infinity; F-unfit/lower TRM score: median OS, 6.2 months; 95% CI, 4.1-7.8 months; Fig 1D).

Prediction of Early Mortality and Survival With Intensive Induction or Reinduction Chemotherapy

In univariate analyses, several factors were associated with either day 28 and/or day 100 mortality, including age, PS, secondary disease, adverse cytogenetic risk, disease status, some laboratory parameters (platelet count, albumin), and type of chemotherapy, as were Ferrara fitness assessment and TRM score used either as continuous or binary variables (Table 3). Likewise, several factors (age, sex, PS, presence of secondary disease, cytogenetic disease risk, disease status, platelet count, albumin, type of treatment, TRM score, and Ferrara fitness assessment) were associated with OS. As summarized in Table 3, the ability of individual factors to predict day 28 or day 100 mortality was overall relatively limited with the exception of albumin and, to a lesser degree, PS and platelet count. The best univariate prediction ability for day 28 and day 100 mortality was obtained with the Ferrara fitness assessment, with AUCs of 0.76 and 0.79, respectively. In comparison, the TRM score's ability to predict day 28 and day 100 mortality was lower, with AUCs of 0.72 and 0.70, respectively, when using the score as a continuous variable and AUCs of 0.66 and 0.62 when using the score as a binary variable (Table 3). The ability of individual factors to predict OS was low, with c-statistic values not exceeding 0.59. The TRM score's ability to predict OS was only slightly better. The best predictive ability, although still limited, was seen with the Ferrara unfitness assessment, yielding a c-statistic of 0.67. Of note, we found no evidence that the relationship between Ferrara fitness assessment, TRM score, and outcome changed over time (period 2006-2014 *v* 2015-2020; eg, $P = .44$, 0.49, and 0.72, respectively, for interaction between Ferrara assessment, time, and day 28 mortality, day 100 mortality, or OS).

We built multivariable logistic and Cox regression models to determine to what degree the accuracy of shorter-term mortality and survival prediction can be improved by combining different factors. By including additional covariates such as PS and albumin, the predictive accuracy of the Ferrara unfitness assessment could be significantly improved, yielding AUCs as high as 0.84-0.85 for day 28 or day 100 mortality prediction. Similarly, inclusion of age, albumin, disease risk, disease stage, and TRM score improved the ability of the Ferrara unfitness assessment to predict OS, yielding a c-statistic as high as 0.75 (Table 4).

Performance of Ferrara Unfitness Assessment in Distinct Patient Subsets

As mentioned previously, we assumed patients who did not have baseline pulmonary function tests available to not have any pulmonary comorbidities for the purpose of Ferrara unfitness assessment in our overall analyses in the absence of documented history of pulmonary comorbidities and/or respiratory symptoms. To assess in what way this approach might influence our results, we performed a subset analysis of the 159 patients for whom results from baseline pulmonary function testing were available (Data Supplement). Consequently, this cohort was enriched in patients considered to be F-unfit, accounting for more than half of the patients in this subset (F-fit, n = 68 [44%]; F-unfit, n = 88 [56%]; Data Supplement). In this patient subset, the ability of the Ferrara unfitness criteria to predict day 28 and day 100 mortality was lower than when applied to the entire patient cohort when used as a single factor (AUCs, 0.67 and 0.69; Data Supplement). Although multicomponent prediction models could not be built for day 28 mortality because of the low number of deaths, multicomponent models for day 100 mortality were as accurate in this patient subset as in the entire patient cohort (AUCs

TABLE 2. Comparison of Baseline Demographic and Clinical Characteristics of Ferrara Score for Fit and Unfit Patients

Characteristic	F-Fit (n = 458)	F-Unfit (n = 197)	*P*	Missing Data (No.)
Age at diagnosis (range), years	59.1 (18.6-74.9)	64.0 (21.4-91.4)	< .001	0
Sex			.071	0
Female	213 (47)	76 (39)		
Male	245 (53)	121 (61)		
Performance status			< .001	3
0-1	434 (95)	128 (66)		
2-3	23 (5)	67 (34)		
Disease			.67	0
AML	364 (79)	160 (81)		
Other	94 (21)	37 (19)		
MDS-EB-2	86 (19)	32 (16)		
CMML-2	5 (1)	3 (2)		
MDS/MPN, uncclassifiable	0 (0.0)	2 (0.5))		
Myelofibrosis	3 (0.5)	0 (0.0)		
Secondary disease			.0031	0
No	318 (69)	113 (57)		
Yes	140 (31)	84 (43)		
Cytogenetic risk at initial disease diagnosis			< .001	34
Favorable	36 (8)	8 (4)		
Intermediate	280 (65)	104 (55)		
Adverse	117 (27)	76 (40)		
Disease status			.18	48
Newly diagnosed	346 (76)	136 (69)		
Refractory	59 (13)	29 (15)		
Relapsed	53 (12)	32 (16)		
Prior HCT			.13	48
No	431 (94)	179 (91)		
Yes	27 (6)	18 (9)		
Treatment			< .001	48
7 + 3	118 (26)	28 (14)		
CLAG-M	332 (72)	139 (71)		
Dose-reduced CLAG-M	8 (2)	30 (15)		
Laboratory finding at baseline				
WBC, $\times 10^9$/L (range)	5.3 (0.2-224.2)	6.8 (0.1-356.3)	.95	49
Platelet count, $\times 10^9$/L (range)	58.5 (2-794)	37.5 (1-650)	< .001	49
Peripheral blood blasts, % (range)	7 (0-99)	16 (0-97)	.0057	49
Albumin, g/dL (range)	3.8 (1.8-5.1)	3.5 (1.8-4.8)	< .001	49
Creatinine, mg/dL (range)	0.86 (0.39-7.94)	0.9 (0.34-11.55)	.22	49
Treatment year			.0068	0
2006-2014	169 (37)	51 (26)		
2015-2020	289 (63)	146 (74)		
TRM score	2.49 (0.01-37.7)	6.31 (0.09-78.3)	< .001	51

(continued on following page)

TABLE 2. Comparison of Baseline Demographic and Clinical Characteristics of Ferrara Score for Fit and Unfit Patients (continued)

Characteristic	F-Fit (n = 458)	F-Unfit (n = 197)	*P*	Missing Data (No.)
TRM score category			< .001	51
Lower (≤ 13.1)	398 (95)	135 (72)		
Higher (> 13.1)	19 (5)	52 (28)		
Best response to treatment			< .001	0
CR	327 (71)	70 (36)		
CRi	56 (12)	23 (12)		
MLFS	30 (7)	21 (11)		
Stable/progressive disease	38 (8)	52 (26)		
Death from indeterminate cause	7 (2)	31 (15)		

NOTE. Data are No. (%) unless otherwise indicated.

Abbreviations: AML, acute myeloid leukemia; CLAG-M, cladribine, cytarabine, granulocyte colony-stimulating factor, mitoxantrone; CMML-2, chronic myelomonocytic leukemia-2; CR, complete remission; CRi, CR with incomplete hematologic recovery; F-fit, met Ferrara criteria as fit for intensive chemotherapy; F-unfit, met Ferrara criteria as unfit for intensive chemotherapy; HCT, hematopoietic cell transplantation; MDS-EB-2, myelodysplastic syndrome with excess blasts-2; MDS/MPN, myelodysplastic syndrome/myeloproliferative neoplasm; MLFS, morphologic leukemia-free state; TRM, treatment-related mortality.

between 0.85 and 0.86), as was the prediction of OS (c-statistic, 0.78; Data Supplement).

We also performed an analysis of the 281 patients with previously untreated de novo AML, 266 of whom with data available for Ferrara fitness assessments (Data Supplement). In this more homogeneous patient subset, the Ferrara unfitness criteria predicted day 28 and day 100 mortality more accurately than in the entire patient cohort when used as a single factor (AUCs, 0.84 and 0.84) or in multicomponent models (AUCs, 0.93 and 0.91; Data Supplement). Finally, we analyzed the 338 F-fit and 110 F-unfit patients who received either 7 + 3 or CLAG-M (but not dose-reduced CLAG-M) for newly diagnosed disease separately (Data Supplement). In this patient subset, the ability of the Ferrara criteria to predict day 28 and day 100 mortality was similar to that in the entire patient cohort when used alone (AUCs, 0.74 and 0.80) or in multicomponent models (AUCs, 0.84 and 0.85; Data Supplement). As for the entire cohort, the accuracy to predict OS was lower than the accuracy to predict shorter-term mortality in these subset analyses.

DISCUSSION

Several quantitative scoring systems have been developed to predict early mortality after intensive AML chemotherapy and can guide treatment decision making.[10,11] Still, fitness evaluations based on conceptual criteria, many of which may not be quantitative, are common. The Ferrara criteria[19] are a prominent example for this approach.

In our cohort, we observed a day 28 and day 100 mortality of 2% and 5% for F-fit and 14% and 42% for F-unfit patients, respectively, as well as a median OS of > 3 years for F-fit versus < 6 months for F-unfit patients. Using AUC values, we found Ferrara criteria–based fitness assessments to have good to very good accuracy in predicting day 28 and day 100 mortality after intensive AML therapy. This prediction accuracy could be further increased by consideration of additional factors, in particular, albumin and additional PS information. Although our models offer no insight as to why these factors improve outcome predictions, it is interesting to speculate that they might capture patients affected by effects from less severe multiorgan dysfunction, which by themselves did not reach the level of severity required to meet Ferrara unfitness.

Not surprising, considering relapse risks and survival are substantially affected by genetic/molecular disease characteristics,[1,2,24,25] OS predictions were less accurate with the Ferrara criteria. Noteworthy, the TRM score did not separate survival outcomes as well as the Ferrara fitness assessment, possibly because many patients with lower TRM scores were F-unfit, whereas only a small proportion of F-fit patients had higher TRM scores. Likewise, the accuracy in predicting early mortality (and OS) was higher with the Ferrara criteria than the TRM score. This was true whether we used the TRM score as a continuous or dichotomized variable. Although any cut-off to dichotomize a continuous variable can be criticized, it is expected that any specific cut-off (here, ≤ 13.1 *v* > 13.1) would have an AUC/c-statistic that is no better (and likely worse) than the quantitative version of the variable.

Together, our findings indicate that the Ferrara criteria provide a useful tool for patient risk stratification with good to very good accuracy for the prediction of shorter-term mortality after intensive AML chemotherapy, which, in conjunction with molecular/genetic data, could serve as a basis for informed decision making, particularly in older patients and those with comorbidities. As a caveat, this approach will not completely eliminate bias in

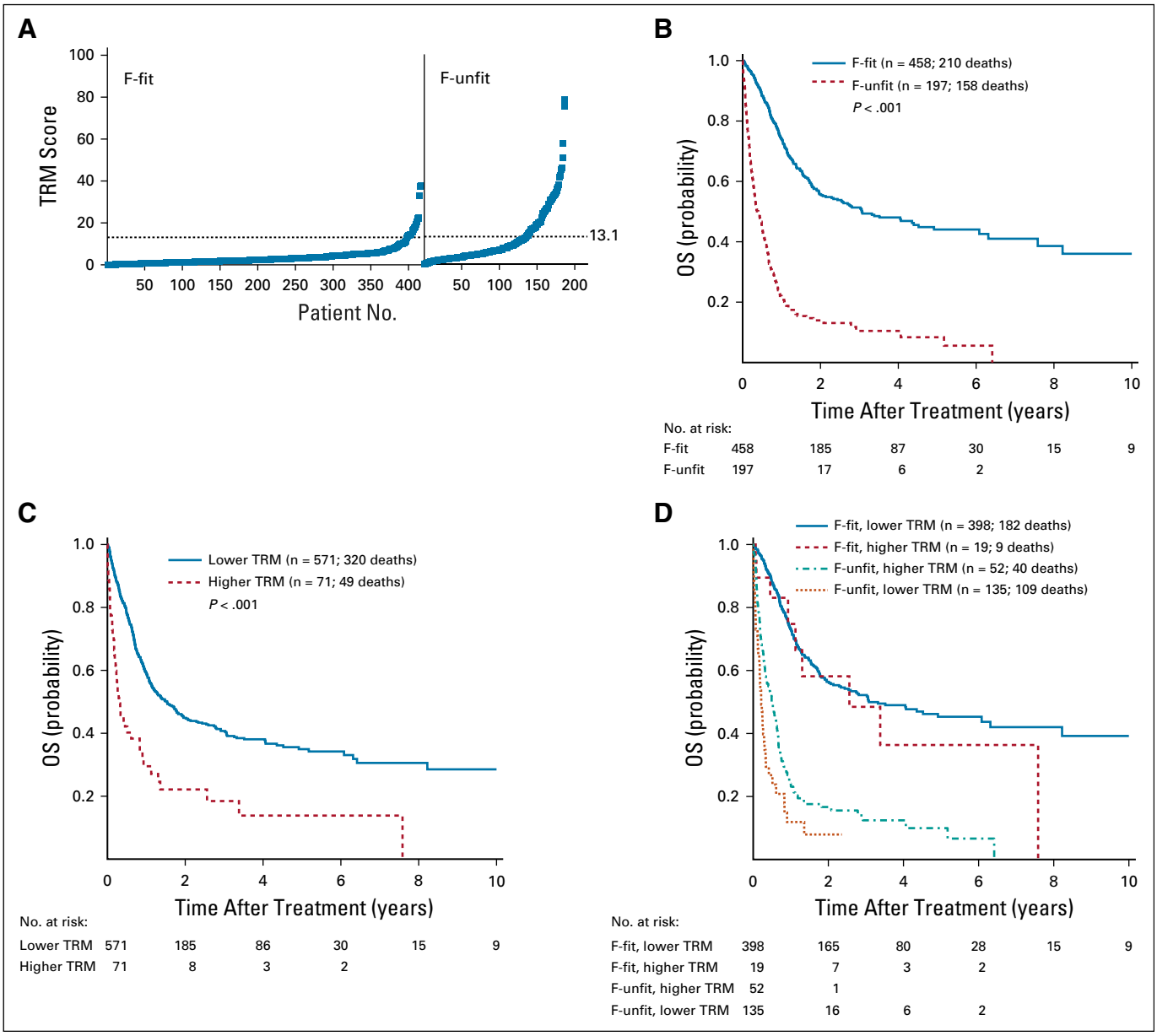

FIG 1. Relationship between Ferrara unfitness (F-unfit) assessment and treatment-related mortality (TRM) score, and overall survival (OS) after receipt of reinduction chemotherapy with 7 + 3, CLAG-M (cladribine, high-dose cytarabine, granulocyte colony-stimulating factor, and mitoxantrone), or dose-reduced CLAG-M. (A) Distribution of TRM scores among Ferrara-fit (F-fit) and F-unfit patients. (B-D) Kaplan-Meier estimates of OS of study population stratified by (B) F-unfit criteria (F-fit *v* F-unfit), (C) TRM score (lower *v* higher), and (D) both F-unfit criteria and TRM score.

decision making because the operational Ferrara criteria permit physician discretion to exclude patients from intensive therapy. Acknowledging this limitation, the Ferrara approach identifies a frail group of patients with poor survival expectations after intensive AML chemotherapy. Within this F-unfit subset, a majority of patients were considered fit based on a TRM score ≤ 13.1. Still, survival of these F-unfit/lower TRM score patients was only slightly longer than that of F-unfit/higher TRM score patients, even though the TRM score added some accuracy to outcome predictions made via Ferrara fitness assessments.

A strength of this study is that we had a large number of patients who received intensive AML chemotherapy with one of three regimens at our institution available for analysis. Because we offer intensive therapy to almost all adults with high-grade myeloid neoplasms, including those considered less fit, we could study a relatively large population of patients who received intensive chemotherapy despite being F-unfit. However, because many patients are referred to our institution for treatment, a bias in patient selection cannot be excluded in this retrospective analysis. As an important limitation of our retrospective study, not all patients had all information required for Ferrara and TRM

TABLE 3. Univariate Associations With Various Treatment Outcomes of Interest

Variable	28-Day Mortality OR (95% CI)	100-Day Mortality OR (95% CI)	Overall Survival HR (95% CI)
Age at diagnosis	1.02 (0.99 to 1.04), $P = .21$, AUC, 0.56	1.02 (1.00 to 1.04), $P = .02$, AUC, 0.58	1.02 (1.01 to 1.03), $P < .001$, c-stat, 0.57
Sex			
Female	1 (reference)	1 (reference)	1 (reference)
Male	0.91 (0.47 to 1.77), $P = .78$, AUC, 0.51	1.11 (0.73 to 1.68), $P = .64$, AUC, 0.51	1.24 (1.01 to 1.51), $P = .037$, c-stat, 0.52
Performance status			
0-1	1 (reference)	1 (reference)	1 (reference)
2-3	5.54 (2.77 to 11.05), $P < .001$, AUC, 0.66	6.45 (3.91 to 10.63), $P < .001$, AUC, 0.64	2.54 (1.95 to 3.32), $P < .001$, c-stat, 0.56
Disease			
AML	1 (reference)	1 (reference)	1 (reference)
Other	0.96 (0.41 to 2.24), $P = .93$, AUC, 0.50	1.16 (0.70 to 1.93), $P = .56$, AUC, 0.51	1.11 (0.87 to 1.41), $P = .41$, c-stat, 0.51
Secondary disease			
No	1 (reference)	1 (reference)	1 (reference)
Yes	1.05 (0.52 to 2.10), $P = .89$, AUC, 0.51	2.20 (1.45 to 3.35), $P < .001$, AUC, 0.59	1.42 (1.16 to 1.73), $P < .001$, c-stat, 0.55
Disease risk at diagnosis			
Favorable	0.48 (0.06 to 3.67), $P = .48$, AUC, 0.58	0.32 (0.08 to 1.38), $P = .13$, AUC, 0.61	0.43 (0.25 to 0.75), $P = .0031$, c-stat, 0.59
Intermediate	1 (reference)	1 (reference)	1 (reference)
Adverse	1.71 (0.84 to 3.47), $P = .14$	2.10 (1.35 to 3.26), $P = .001$	1.78 (1.44 to 2.19), $P < .001$
Disease status			
Newly diagnosed	1 (reference)	1 (reference)	1 (reference)
Refractory	1.40 (0.55 to 3.53), $P = .48$, AUC, 0.56	2.09 (1.22 to 3.57), $P = .0074$, AUC, 0.57	1.93 (1.47 to 2.52), $P < .001$, c-stat, 0.57
Relapsed	1.88 (0.82 to 4.34), $P = .14$	1.72 (0.98 to 3.00), $P = .057$	1.83 (1.41 to 2.38), $P < .001$
Prior HCT			
No	1 (reference)	1 (reference)	
Yes	2.09 (0.78 to 5.62), $P = .14$, AUC, 0.53	1.92 (0.99 to 3.74), $P = .055$, AUC, 0.53	
Treatment			
7 + 3	1 (reference)	1 (reference)	1 (reference)
CLAG-M	1.48 (0.60 to 3.66), $P = .39$, AUC, 0.57	1.96 (1.07 to 3.57), $P = .028$, AUC, 0.61	1.26 (0.99 to 1.59), $P = .059$, c-stat, 0.55
Dose-reduced CLAG-M	4.04 (1.16 to 14.03), $P = .028$	11.82 (4.88 to 28.63), $P < .001$	3.33 (2.14 to 5.17), $P < .001$

(continued on following page)

TABLE 3. Univariate Associations With Various Treatment Outcomes of Interest (continued)

Variable	28-Day Mortality OR (95% CI)	100-Day Mortality OR (95% CI)	Overall Survival HR (95% CI)
Laboratory finding at baseline			
WBC	1.01 (1.00 to 1.01), $P = .054$, AUC, 0.55	1.00 (1.00 to 1.01), $P = .066$, AUC, 0.52	1.00 (1.00 to 1.00), $P = .65$, c-stat, 0.51
Platelet count	0.99 (0.99 to 1.00), $P = .047$, AUC, 0.64	1.00 (0.99 to 1.00), $P = .013$, AUC, 0.64	1.00 (1.00 to 1.00), $P = .0046$, c-stat, 0.58
Peripheral blood blasts	1.00 (0.99 to 1.01), $P = .87$, AUC, 0.51	1.01 (1.00 to 1.01), $P = .13$, AUC, 0.53	1.00 (1.00 to 1.00), $P = .86$, c-stat, 0.52
Albumin	0.29 (0.16 to 0.52), $P < .001$, AUC, 0.73	0.30 (0.20 to 0.45), $P < .001$, AUC, 0.70	0.70 (0.58 to 0.83), $P < .001$, c-stat, 0.59
Creatinine	1.06 (0.70 to 1.60), $P = .78$, AUC, 0.50	1.14 (0.87 to 1.49), $P = .36$, AUC, 0.50	1.10 (0.98 to 1.24), $P = .11$, c-stat, 0.52
Treatment year			
2006-2014	1 (reference)	1 (reference)	1 (reference)
2015-2020	1.30 (0.63 to 2.67), $P = .48$, AUC, 0.53	1.42 (0.91 to 2.24), $P = .12$, AUC, 0.54	1.20 (0.97 to 1.47), $P = .09$, c-stat, 0.52
TRM score	1.06 (1.03 to 1.09), $P < .001$, AUC, 0.72	1.08 (1.06 to 1.11), $P < .001$, AUC, 0.70	1.03 (1.03 to 1.04), $P < .001$, c-stat, 0.65
TRM score category			
Lower	1 (reference)	1 (reference)	1 (reference)
Higher	7.09 (3.46 to 14.54), $P < .001$, AUC, 0.66	6.89 (3.94 to 12.06), $P < .001$, AUC, 0.62	2.58 (1.91 to 3.5), $P < .001$, c-stat, 0.55
Ferrara criteria category			
F-fit	1 (reference)	1 (reference)	1 (reference)
F-unfit	10.71 (4.59 to 24.99), $P < .001$, AUC, 0.76	13.98 (8.33 to 23.48), $P < .001$, AUC, 0.79	4.43 (3.58 to 5.49), $P < .001$, c-stat, 0.67

Abbreviations: AML, acute myeloid leukemia; AUC, area under the receiver operating characteristic curve; CLAG-M, cladribine, cytarabine, granulocyte colony-stimulating factor, mitoxantrone; c-stat, c-statistic; F-fit, met Ferrara criteria as fit for intensive chemotherapy; F-unfit, met Ferrara criteria as unfit for intensive chemotherapy; HCT, hematopoietic cell transplantation; HR, hazard ratio; OR, odds ratio; TRM, treatment-related mortality.

TABLE 4. Multivariable Logistic and Cox Regression Models for Day 28/100 Mortality and Overall Survival

Model	28-Day Mortality OR (95% CI)	100-Day Mortality OR (95% CI)	Overall Survival HR (95% CI)
Multivariable logistic regression models			
Model 1	AUC, 0.84	AUC, 0.85	—
PS 2-3 (ref: PS 0-1)	1.38 (0.50 to 3.80), $P = .53$	1.69 (0.78 to 3.67), $P = .19$	
Albumin	0.44 (0.22 to 0.87), $P = .019$	0.39 (0.24 to 0.64), $P < .001$	
F-unfit (ref: F-fit)	6.52 (2.61 to 16.27), $P < .001$	10.18 (5.79 to 17.91), $P < .001$	
TRM score	1.02 (0.98 to 1.05), $P = .37$	1.01 (0.98 to 1.04), $P = .69$	
Model 2	AUC, 0.84	AUC, 0.85	—
PS 2-3 (ref: PS 0-1)	0.94 (0.32 to 2.76), $P = .9$	1.33 (0.6 to 2.96), $P = .49$	
Albumin	0.43 (0.22 to 0.85), $P = .016$	0.4 (0.25 to 0.65), $P < .001$	
F-unfit (ref: F-fit)	6.45 (2.6 to 16.02), $P < .001$	10.12 (5.8 to 17.67), $P < .001$	
Higher TRM score (ref: lower TRM score)	2.81 (0.97 to 8.1), $P = .056$	1.78 (0.76 to 4.15), $P = .18$	
Model 3	AUC, 0.84	AUC, 0.85	—
PS 2-3 (ref: PS 0-1)	1.79 (0.79 to 4.04), $P = .16$	1.86 (1.00 to 3.45), $P = .051$	
Albumin	0.41 (0.21 to 0.80), $P = .0095$	0.38 (0.24 to 0.61), $P < .001$	
F-unfit (ref: F-fit)	6.87 (2.78 to 16.95), $P < .001$	1.4 (5.97 to 18.11), $P < .001$	
Multivariable Cox regression model			
Model 1	—	—	C-stat, .75
F-unfit (ref: F-fit)			3.80 (2.95 to 4.88), $P < .001$
TRM score			1.02 (1.01 to 1.03), $P = .0032$
Age			1.01 (1.00 to 1.02), $P = .018$
Adverse cytogenetic risk (ref: intermediate risk)			1.72 (1.37 to 2.16), $P < .001$
Favorable cytogenetic risk (ref: intermediate risk)			.54 (.30 to .96), $P = .038$
Refractory disease (ref: newly diagnosed disease)			2.09 (1.54 to 2.83), $P < .001$
Relapsed disease (ref: newly diagnosed disease)			2.04 (1.50 to 2.78), $P < .001$
Albumin			.75 (.61 to .91), $P = .004$

Abbreviations: AUC, area under the receiver operating characteristic curve; c-stat, c-statistic; F-fit, met Ferrara criteria as fit for intensive chemotherapy; F-unfit, met Ferrara criteria as unfit for intensive chemotherapy; HR, hazard ratio; OR, odds ratio; PS, performance status; ref, reference; TRM, treatment-related mortality.

fitness assessments. This limitation was particularly marked regarding results from pretreatment pulmonary function testing, which were missing in a significant number of patients. Although our data indicate the importance of pulmonary assessments to categorize patients as fit or unfit for intensive AML chemotherapy (because pulmonary abnormalities were the single most important criterion establishing F-unfitness), additional studies will be required to determine whether the absence of known pulmonary comorbidities and/or respiratory symptoms could be used as a surrogate for normal pulmonary function. If validated, the approach of mandating pulmonary function testing only for patients with known pulmonary compromise (perhaps including radiographic abnormalities) and/or symptoms—effectively modifying the Ferrara criteria—would simplify fitness assessments, particularly in institutions where lung function testing is not a standard of care for patients with AML.

Another limitation is that we were unable to determine the degree to which each of the criteria contributed to the predictive accuracy of the fitness assessment because some of them were only occasionally met. Substantially larger patient cohorts will be required to accomplish this. Although it is tempting to use our data as a justification to develop a simplified fitness assessment using a shorter criteria list, the frequency with which individual criteria are met will likely vary across patient populations. Removing those low-incidence groups to derive a new score would place the score potentially at high risk for not being able to be validated in a cohort with more of those patients.

Between 2017 and 2020, there have been nine new drugs approved for AML in the United States,[26] increasing lower-intensity treatment options substantially. Some of these agents (in particular, venetoclax) in combination with azanucleosides or low-dose cytarabine are emerging as new standards for unfit patients with AML, and there is now less separation between intensive and nonintensive AML therapies. Unfitness criteria similar to those proposed by Ferrara et al[20] are commonly used to select patients for lower-intensity treatments. However, how accurately they predict outcomes after these therapies is unknown. With only a small number of patients treated with lower-intensity AML treatments, including doublet therapies incorporating new drugs such as venetoclax, inhibitors of mutant IDH1/2, or glasdegib, at our institution, we were unable to determine the accuracy of Ferrara fitness assessments for early mortality and survival prediction for adults undergoing lower-intensity AML therapy. This will remain an important question to be addressed in future studies.

Finally, it is important to emphasize that our studies did not address the question of what the optimal treatment intensity is for patients classified as F-unfit. Although we found substantially worse outcomes for F-unfit versus F-fit patients with intensive chemotherapy, well-controlled, ideally randomized studies will be required to determine whether outcomes in F-unfit patients are better with alternative, less-intense therapies. Short of such studies, our data may serve as a historic benchmark for expected outcomes with intensive chemotherapy in such patients.

AFFILIATIONS

[1]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA
[2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA
[3]Department of Medicine, Division of Hematology, University of Washington, Seattle, WA
[4]Department of Laboratory Medicine & Pathology & Department of Epidemiology, University of Washington, Seattle, WA

CORRESPONDING AUTHOR

Roland B. Walter, MD, PhD, MS, Clinical Research Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, D2-190, Seattle, WA 98109-1024; e-mail: rwalter@fredhutch.org.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at DOI https://doi.org/10.1200/JCO.20.01392.

AUTHOR CONTRIBUTIONS

Conception and design: Raffaele Palmieri, Roland B. Walter
Administrative support: Roland B. Walter
Provision of study materials or patients: Anna B. Halpern, Roland B. Walter
Collection and assembly of data: Raffaele Palmieri, Anna B. Halpern, Roland B. Walter
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. Döhner H, Weisdorf DJ, Bloomfield CD: Acute myeloid leukemia. N Engl J Med 373:1136-1152, 2015
2. Döhner H, Estey E, Grimwade D, et al: Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 129:424-447, 2017
3. Othus M, Kantarjian H, Petersdorf S, et al: Declining rates of treatment-related mortality in patients with newly diagnosed AML given 'intense' induction regimens: A report from SWOG and MD Anderson. Leukemia 28:289-292, 2014
4. Percival ME, Tao L, Medeiros BC, et al: Improvements in the early death rate among 9380 patients with acute myeloid leukemia after initial therapy: A SEER database analysis. Cancer 121:2004-2012, 2015
5. Halpern AB, Culakova E, Walter RB, et al: Association of risk factors, mortality, and care costs of adults with acute myeloid leukemia with admission to the intensive care unit. JAMA Oncol 3:374-381, 2017
6. Walter RB, Estey EH: Management of older or unfit patients with acute myeloid leukemia. Leukemia 29:770-775, 2015
7. Loh KP, Klepin HD: Geriatric assessment in older patients with acute myeloid leukemia. Cancers (Basel) 10:225, 2018
8. Loh KP, Klepin HD: Geriatric assessment in acute myeloid leukemia: Current and future landscape. Blood Adv 2:2418, 2018
9. Klepin HD, Estey E, Kadia T: More versus less therapy for older adults with acute myeloid leukemia: New perspectives on an old debate. Am Soc Clin Oncol Educ Book 39:421-432, 2019
10. Palmieri R, Paterno G, De Bellis E, et al: Therapeutic choice in older patients with acute myeloid leukemia: A matter of fitness. Cancers (Basel) 12:120, 2020
11. Walter RB, Estey EH: Selection of initial therapy for newly-diagnosed adult acute myeloid leukemia: Limitations of predictive models. Blood Rev 10.1016/j.blre.2020.100679 [epub ahead of print on March 20, 2020]
12. Kim ES: Midostaurin: First global approval. Drugs 77:1251-1259, 2017
13. Kim ES: Enasidenib: First global approval. Drugs 77:1705-1711, 2017
14. Norsworthy KJ, Luo L, Hsu V, et al: FDA approval summary: Ivosidenib for relapsed or refractory acute myeloid leukemia with an isocitrate dehydrogenase-1 mutation. Clin Cancer Res 25:3205-3209, 2019
15. Dhillon S: Gilteritinib: First global approval. Drugs 79:331-339, 2019
16. Norsworthy KJ, By K, Subramaniam S, et al: FDA approval summary: Glasdegib for newly diagnosed acute myeloid leukemia. Clin Cancer Res 25:6021-6025, 2019
17. Guerra VA, DiNardo C, Konopleva M: Venetoclax-based therapies for acute myeloid leukemia. Best Pract Res Clin Haematol 32:145-153, 2019

18. Walter RB, Othus M, Borthakur G, 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. J Clin Oncol 29:4417-4423, 2011
19. Fred Hutchinson Cancer Research Center: Treatment related mortality (TRM) calculator. https://cstaging.fhcrc-research.org/TRM/
20. Ferrara F, Barosi G, Venditti A, 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 27:997-999, 2013
21. Arber DA, Orazi A, Hasserjian R, et al: The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127: 2391-2405, 2016
22. Walter RB, Othus M, Orlowski KF, et al: Unsatisfactory efficacy in randomized study of reduced-dose CPX-351 for medically less fit adults with newly diagnosed acute myeloid leukemia or other high-grade myeloid neoplasm. Haematologica 103:e106-e109, 2018
23. Halpern AB, Othus M, Gardner K, et al: Mini- vs. regular-dose CLAG-M (cladribine, cytarabine, G-CSF, and mitoxantrone) in medically less fit adults with newly-diagnosed acute myeloid leukemia (AML) and other high-grade myeloid neoplasms. Blood 134:1364, 2019 (suppl 1)
24. Walter RB, Othus M, Burnett AK, et al: Resistance prediction in AML: Analysis of 4601 patients from MRC/NCRI, HOVON/SAKK, SWOG and MD Anderson Cancer Center. Leukemia 29:312-320, 2015
25. Walter RB, Othus M, Paietta EM, et al: Effect of genetic profiling on prediction of therapeutic resistance and survival in adult acute myeloid leukemia. Leukemia 29:2104-2107, 2015
26. Patel SA, Gerber JM: A user's guide to novel therapies for acute myeloid leukemia. Clin Lymphoma Myeloma Leuk 20:277-288, 2020

ASCO® Meetings

ASCO offers premier scientific events for oncology professionals, patient advocates, industry representatives, and major media outlets worldwide.

View upcoming meetings and symposia at **meetings.asco.org**

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

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

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Megan Othus
Employment: Fred Hutchinson Cancer Research Center
Consulting or Advisory Role: Celgene, Glycomimetics, Cascadia Laboratories, Merck, Daiichi Sankyo
Other Relationship: Celgene, Glycomimetics

Anna B. Halpern
Research Funding: Pfizer (Inst), Nohla Therapeutics (Inst), Jazz Pharmaceuticals (Inst), Imago Pharma (Inst), Novartis (Inst), Bayer (Inst), Tolero Pharmaceuticals (Inst)

Mary-Elizabeth M. Percival
Consulting or Advisory Role: Genentech
Research Funding: Pfizer (Inst), Trillium Therapeutics (Inst), Nohla Therapeutics (Inst), Biosight (Inst), FLX Bio (Inst), Glycomimetics (Inst), Cardiff Biotechnology (Inst)

Colin D. Godwin
Research Funding: Pfizer, Immunogen

Pamela S. Becker
Consulting or Advisory Role: CVS Caremark, Pfizer (Inst)
Research Funding: Abbvie (Inst), Amgen (Inst), Bristol Myers Squibb (Inst), JW Pharmaceutical (Inst), Novartis (Inst), Pfizer (Inst), Glycomimetics (Inst), Trovagene (Inst), Invivoscribe (Inst), Aptose Biosciences (Inst), Trethera (Inst), Secura Bio (Inst), Cardiff Oncology (Inst)
Patents, Royalties, Other Intellectual Property: Provisional patent filed: Application 62/694,874 filed 7/6/2018 entitled: "High throughput drug screening of cancer stem cells"

Roland B. Walter
Stock and Other Ownership Interests: Amphivena Therapeutics
Consulting or Advisory Role: Covagen, Emergent Biosolutions, Pfizer, Agios, BiolineRx, Race Oncology, Jazz Pharmaceuticals, Argenx, BiVictriX, Boston Biomedical, Daiichi Sankyo, Kite Pharma, Astellas Pharma, Amgen, Newlink Genetics, Janssen, Macrogenics
Research Funding: Amgen (Inst), AbbVie (Inst), Stemline Therapeutics (Inst), Arog (Inst), ADC Therapeutics (Inst), Seattle Genetics (Inst), Pfizer, BiolineRx (Inst), Agios (Inst), Selvita (Inst), Jazz Pharmaceuticals (Inst), Aptevo Therapeutics, Celgene, Macrogenics (Inst), Immunogen (Inst)

No other potential conflicts of interest were reported.