

# Evaluation of event-free survival as a robust end point in untreated acute myeloid leukemia (Alliance A151614)

Jun Yin,<sup>1</sup> Betsy LaPlant,<sup>1</sup> Geoffrey L. Uy,<sup>2</sup> Guido Marcucci,<sup>3</sup> William Blum,<sup>4</sup> Richard A. Larson,<sup>5</sup> Richard M. Stone,<sup>6</sup> and Sumithra J. Mandrekar<sup>1</sup>

<sup>1</sup>Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN; <sup>2</sup>Division of Oncology, Washington University School of Medicine, St. Louis, MO; <sup>3</sup>City of Hope, Duarte, CA; <sup>4</sup>Department of Hematology and Medical Oncology, Emory University, Atlanta, GA; <sup>5</sup>Comprehensive Cancer Center, University of Chicago, Chicago IL; and <sup>6</sup>Dana-Farber/Partners CancerCare, Boston, MA

## Key Points

- In randomized trials, treatment effect estimates based on the hazard ratios were unaffected by various EFS definitions.
- In single-arm trials, incorrect conclusions about efficacy could be made if the EFS definition is not consistent with the historical control.

Event-free survival (EFS) is controversial as an end point for speeding approvals in newly diagnosed acute myeloid leukemia (AML). We aimed to examine the robustness of EFS, specifically timing of complete remission (CR) in defining induction failure and impact of hematopoietic cell transplantation (HCT). The study included 1884 untreated AML patients enrolled across 5 trials conducted through Alliance for Clinical Trials in Oncology using anthracycline and cytarabine induction chemotherapy. EFS was defined as time from randomization/registration to induction failure, relapse, or death. Three definitions of induction failure were evaluated: failure to achieve CR by 60 days after randomization/registration, failure to achieve CR by the end of all protocol-defined induction courses, and failure to achieve CR by the end of all protocol-defined treatment. We considered either censoring or no censoring at time of non-protocol-mandated HCT. Although relapse and death are firm end points, the determination of induction failure was not consistent across studies. There was minimal impact of censoring at HCT on EFS estimates; however, median EFS estimates differed considerably based on the timing of CR in defining induction failure, with the magnitude of difference being large enough in most cases to lead to incorrect conclusions about efficacy in a single-arm trial, if the trial definition was not consistent with the definition used for the historical control. Timing of CR should be carefully examined in the historical control data used to guide the design of single-arm trials using EFS as the primary end point. Trials were registered at [www.clinicaltrials.gov](http://www.clinicaltrials.gov) as #NCT00085124, #NCT00416598, #NCT00651261, #NCT01238211, and #NCT01253070.

## Introduction

Acute myeloid leukemia (AML) is the most common acute leukemia in adults and is among the most lethal. In the United States, the annual incidence of AML is 19 000 cases, and the annual incidence of AML-associated deaths is 10 000.<sup>1</sup> Although there has been significant research effort aimed at improving outcomes in AML, standard therapy for most subtypes of newly diagnosed AML remains suboptimal.<sup>1,2</sup> Especially among patients age >60 years, outcomes are poor, with a 5-year overall survival (OS) of 10% to 20%; outcomes are even worse among older patients who are unfit for intensive chemotherapy, with a median OS of only 5 to 10 months.<sup>1,3,4</sup>

In parallel with research on new therapies, emphasis has been placed on new end points other than OS that may facilitate drug development and shorten the time to approval for use in AML.<sup>2,5</sup> OS in comparative oncology clinical trials remains the gold-standard end point to assess efficacy of drugs for

approval by the US Food and Drug Administration. However, use of OS as an end point requires following up participants until a sufficient number of deaths occur.<sup>2,6-8</sup> For example, midostaurin was recently approved for patients with newly diagnosed *FLT3*-mutated AML based on a clinically significant improvement in OS.<sup>9</sup> However, the protocol was amended to perform the primary analysis with OS as the end point without waiting for the 509 originally planned OS events to occur, because the original planned OS events had not yet been reached, likely because of the higher-than-expected rate of hematopoietic cell transplantation (HCT). This protocol amendment also promoted consideration of event-free survival (EFS) as the most clinically meaningful secondary end point for assessment of the effect of midostaurin on the natural course of *FLT3*-mutated AML. The molecular, immunophenotypic, and biologic heterogeneity of AML has allowed the use of targeted agents tailored to individual patients, but these patient subsets are small, and trial recruitment is slow as a result of patient selection.<sup>10,11</sup> This adds additional time to the clinical trial process. Therefore, there is an increased interest in surrogate end points that might provide faster assessments of novel agents to treat AML. EFS is a common alternative end point in AML. It is similar to progression-free survival, which has been validated as a surrogate end point in pivotal randomized trials for solid tumors.<sup>12-14</sup> In contrast to OS, where death is the only event of interest, EFS is a composite end point that includes death resulting from any cause, failure to achieve complete remission (CR), and relapse from CR as events. Most failures are from lack-of-remission events or from relapse from CR and occur within 1 year after treatment initiation.<sup>15,16</sup> Therefore, EFS can be assessed much faster than OS. In addition, EFS is less affected by HCT or salvage therapy after relapse and takes into account the entire study population.<sup>17-19</sup> These postfailure interventions are often unspecified in protocols and potentially bias or dilute the frontline treatment effect, because they can rescue some patients and prolong the survival of others. Therefore, EFS may be a more reliable end point than OS for assessment of the primary treatment effect.

Despite these attractive features, the EFS-based end point is controversial as a means of accelerating AML approvals. This results from the potential confounding factor of treatment failure components of EFS, such as events of induction failure and transplantation, as well as inconsistent definitions of these components. Therefore, in many protocols, alternative definitions, revisions of the primary definition, are used for supportive/sensitivity analysis. The objective of this study was to better understand the different components of EFS and examine the robustness of efficacy results using EFS as the primary endpoint, with varying induction failure definitions and censoring status at the occurrence of HCT.

## Materials and methods

### Trial identification

Since 2003, the Cancer and Leukemia Group B (CALGB; now part of the Alliance for Clinical Trials in Oncology) has completed five prospective trials using anthracycline and cytarabine chemotherapy for newly diagnosed patients with AML and reported their primary end point data. In the current study (A151604), we evaluated 1884 adults with newly diagnosed AML in these five Alliance (CALGB) trials (Table 1), including 2 randomized phase 3 trials (CALGB 10201, N = 506; CALGB 10603, N = 717) and 3 single-arm

phase 2 trials (CALGB 10503, N = 546; CALGB 10801, N = 61; CALGB 11001, N = 54).<sup>9,20-23</sup> CALGB 10201 enrolled AML patients age  $\geq 60$  years, and CALGB 11001 enrolled AML patients age  $\geq 60$  years with *FLT3*-mutated AML. CALGB 10503 and CALGB 10603 enrolled AML patients age  $< 60$  years, and CALGB 10603 enrolled patients age  $< 60$  years with *FLT3*-mutated AML. CALGB 10801 enrolled patients age  $\geq 18$  years with the favorable cytogenetic risk, core binding factor–positive AML. Cytarabine and daunorubicin were used for remission induction, together with other agents in some studies, and all 5 trials specified a second course of induction therapy for initial nonresponders.

### Statistical analysis

We evaluated 3 definitions of induction failure according to when CR was assessed: definition 1 (D1), failure to achieve CR by 60 days after registration/randomization; D2, failure to achieve CR by the end of all protocol-defined induction courses; and D3, failure to achieve CR by the end of all protocol-defined treatment. CR was defined, using the established response criteria for AML therapy,<sup>24</sup> as  $< 5\%$  blasts in cellular marrow with recovery of  $> 1000$  neutrophils per  $\mu\text{L}$  (but  $> 1500$  neutrophils per  $\mu\text{L}$  in CALGB 10201<sup>25</sup>),  $> 100\,000$  platelets per  $\mu\text{L}$ , and no red blood cell transfusion requirement. EFS was defined as the time from randomization/registration to induction failure using 1 of the 3 definitions, relapse, or death resulting from any cause. Patients who were last known to be alive without an EFS event were censored at the date of last contact. In addition to the 3 definitions of induction failure, we also considered either censoring or not censoring patients who underwent HCT at the time of HCT. EFS was separately estimated in each of the five Alliance (CALGB) trials using the Kaplan-Meier method,<sup>26</sup> with data pooled from across the arms for the randomized trials. Medians and 95% confidence intervals (CIs) as well as EFS estimates at predetermined time points were reported. Treatment effects in the randomized trials were assessed by using the different EFS definitions to further understand how these definitions affected trial outcomes.

Each participant signed an institutional review board–approved, protocol-specific informed consent document in accordance with federal and institutional guidelines. Data collection and statistical analyses were conducted by the Alliance Statistics and Data Center. Results analyzed were available in our database as of 13 February 2018.

## Results

In total, 1884 adult patients with newly diagnosed AML were included in this study from 3 single-arm and 2 randomized trials (Table 1), with a median follow-up of 57 months. Overall, 63% of patients died and 78% had an event (ie, induction failure within 60 days of treatment initiation, relapse, or death). Among the 5 studies, CALGB 10201 and CALGB 11001 enrolled older patients (median age, 67 and 69 years, respectively), whereas CALGB 10503 and CALGB 10603 enrolled younger patients (median age, 47 and 48 years, respectively; Table 2). Approximately half of the patients were men in all trials except for CALGB 10201 (61% male). Younger AML patients had better performance status. The percentage of deaths was higher in the AML trials with older patients (range, 70%-92%) compared with those with younger patients (range, 50%-56%), with varying length of follow-up on patients still alive. In addition, the percentage of EFS events

**Table 1. Trial characteristics**

Trial	N	Phase	Period of accrual	Treatment induction*	Population
CALGB 10201	506	3	2003-2006	7 + 3 ± oblimersen	Older AML
CALGB 10503	546	2	2007-2011	7 + 3 + 3	Younger AML
CALGB 10603	717	3	2008-2015	7 + 3 ± midostaurin	Younger <i>FLT3</i> -mutated AML
CALGB 10801	61	2	2011-2014	7 + 3 + dasatinib	CBF AML
CALGB 11001	54	2	2011-2014	7 + 3 + sorafenib	Older <i>FLT3</i> -mutated AML

CBF, core binding factor.

\*7 + 3, anthracycline or cytarabine on days 1-7 and daunorubicin on days 1-3; 7 + 3 + 3, anthracycline or cytarabine on days 1-7, daunorubicin on days 1-3, and etoposide on days 1-3.

was higher in the AML trials with older patients (range, 80%-94%) compared with younger AML patients (range, 72%-75%). The patients with favorable cytogenetic risk (core binding factor–positive AML) enrolled in CALGB 10801 had the lowest death rate (21%) and EFS event rate (36%).

Depending on when CR status was assessed, the induction failure rate differed across D1 to D3. Overall patterns were consistent among the 5 trials (Table 3), where D1 (CR by 60 days from randomization/registration) yielded the highest induction failure rate, D3 (CR by end of all protocol-defined treatment) yielded the lowest induction failure rate, and D2 (CR by end of protocol-defined induction therapy) yielded an intermediate induction failure rate.

Differences among definitions within individual trials ranged between 3% and 8% across the 5 trials. These differences were converted to EFS estimates.

Median EFS estimates differed considerably according to the timing for CR used to define induction failure (Table 3). Consistently, EFS estimates determined with D3 (end-of-treatment definition) were the highest, EFS estimates determined with D2 (end-of-induction definition) were intermediate, and EFS estimates determined with D1 (60-day definition) were the most conservative. This was expected, because by definition, a longer event-free time would yield better outcomes. Furthermore, when examining the effect over time, we observed that the differences in EFS estimates

**Table 2. Summary of studies**

	CALGB 10503 (N = 546)	CALGB 10603 (N = 717)	CALGB 10801 (N = 61)	CALGB 11001 (N = 54)	CALGB 10201 (N = 506)
<b>Age, y</b>					
Median	48	47	50	67	69
Range	17-60	18-60	19-85	60-82	60-88*
<b>Sex</b>					
Male	298 (55)	318 (44)	31 (51)	30 (56)	307 (61)
Female	248 (45)	399 (56)	30 (49)	24 (44)	199 (39)
<b>Performance score</b>					
0	187 (34)	306 (43)	28 (46)	23 (43)	151 (30)
1	310 (57)	327 (46)	25 (41)	27 (50)	280 (55)
≥2	48 (9)	84 (12)	7 (11)	4 (7)	75 (15)
Deaths	307 (56)	357 (50)	13 (21)	38 (70)	464 (92)
<b>Follow-up, mo†</b>					
Median	60	58	34	28	100
IQR	51-65	47-67	26-41	18-32	87-105
EFS events	395 (72)	541 (75)	22 (36)	43 (80)	476 (94)
No CR1‡	175 (32)	317 (44)	8 (13)	17 (31)	273 (54)
Early death	28 (5)	27 (4)	3 (5)	4 (7)	77 (15)
Early progression	3 (1)	1 (0)	0 (0)	1 (2)	16 (3)
Induction failure	144 (26)	289 (40)	5 (8)	12 (22)	180 (36)
Relapse after CR1§	181 (33)	182 (25)	13 (21)	20 (37)	164 (32)
Death without relapse in CR1§	39 (7)	42 (6)	1 (2)	6 (11)	39 (8)

All data are n (%) unless otherwise noted.

IQR, interquartile range.

\*One ineligible patient was age 48 y and excluded from the analysis; all other eligible patients were age ≥60 y in CALGB 10201.

†Defined as lower to upper quantile.

‡No CR1 indicates induction failure by 60 d after registration/randomization.

§CR1 indicates CR during induction (by 60 d after registration/randomization).

**Table 3. EFS estimates determined by using different induction failure definitions with and without censoring at HCT**

Induction failure definition	Induction failure, n (%)	Median (95% CI) EFS, mo	Patients undergoing HCT, %	Patients censored at time of HCT, %	Median (95% CI) EFS censoring at time of HCT, mo
<b>CALGB 10201 (N = 506)</b>			9		
D1	273 (54)	2.0 (NA)		5	2.0 (NA)
D2	253 (50)	3.0 (2.3-3.8)		4	3.0 (2.3-3.8)
D3	241 (48)	4.3 (3.4-5.4)		5	4.3 (3.3-5.3)
<b>CALGB 10503 (N = 546)</b>			35		
D1	175 (32)	9.8 (8.4-11.7)		16	10.6 (8.9-13.0)
D2	159 (29)	10.6 (9.3-12.4)		16	11.0 (9.5-13.5)
D3	155 (28)	11.2 (9.7-13.8)		17	11.4 (9.8-14.5)
<b>CALGB 10603 (N = 717)</b>			57		
D1	317 (44)	5.5 (2.6-6.7)		21	5.7 (2.6-7.1)
D2	280 (39)	6.9 (5.7-8.3)		21	7.1 (5.7-8.3)
D3	261 (36)	9.7 (8.3-11.7)		27	9.5 (8.2-10.8)
<b>CALGB 10801 (N = 61)</b>			8		
D1	8 (13)	NA (20.9-NA)		3	NA (20.9-NA)
D2	6 (10)	NA (NA)		3	NA (NA)
D3	5 (8)	NA (NA)		3	NA (NA)
<b>CALGB 11001 (N = 54)</b>			41		
D1	17 (31)	6.9 (3.4-11.5)		28	6.0 (3.4-10.2)
D2	15 (28)	8.3 (4.6-11.5)		28	7.8 (4.4-10.2)
D3	15 (28)	8.3 (4.6-11.5)		28	7.8 (4.4-10.2)

NA, not reached.

using the 3 definitions diminished over time (Figure 1). At 1 year after treatment initiation, the estimated EFS rates ranged from 21% to 24% in CALGB 10201, 45% to 48% in CALGB 10503, 37% to 46% in CALGB 10603, and 78% to 83% in CALGB 10801; the estimated EFS rates were 35% for all definitions in CALGB 11001 (Figure 1). For D3 compared with D1, the difference in 1-year EFS reached as much as 9%.

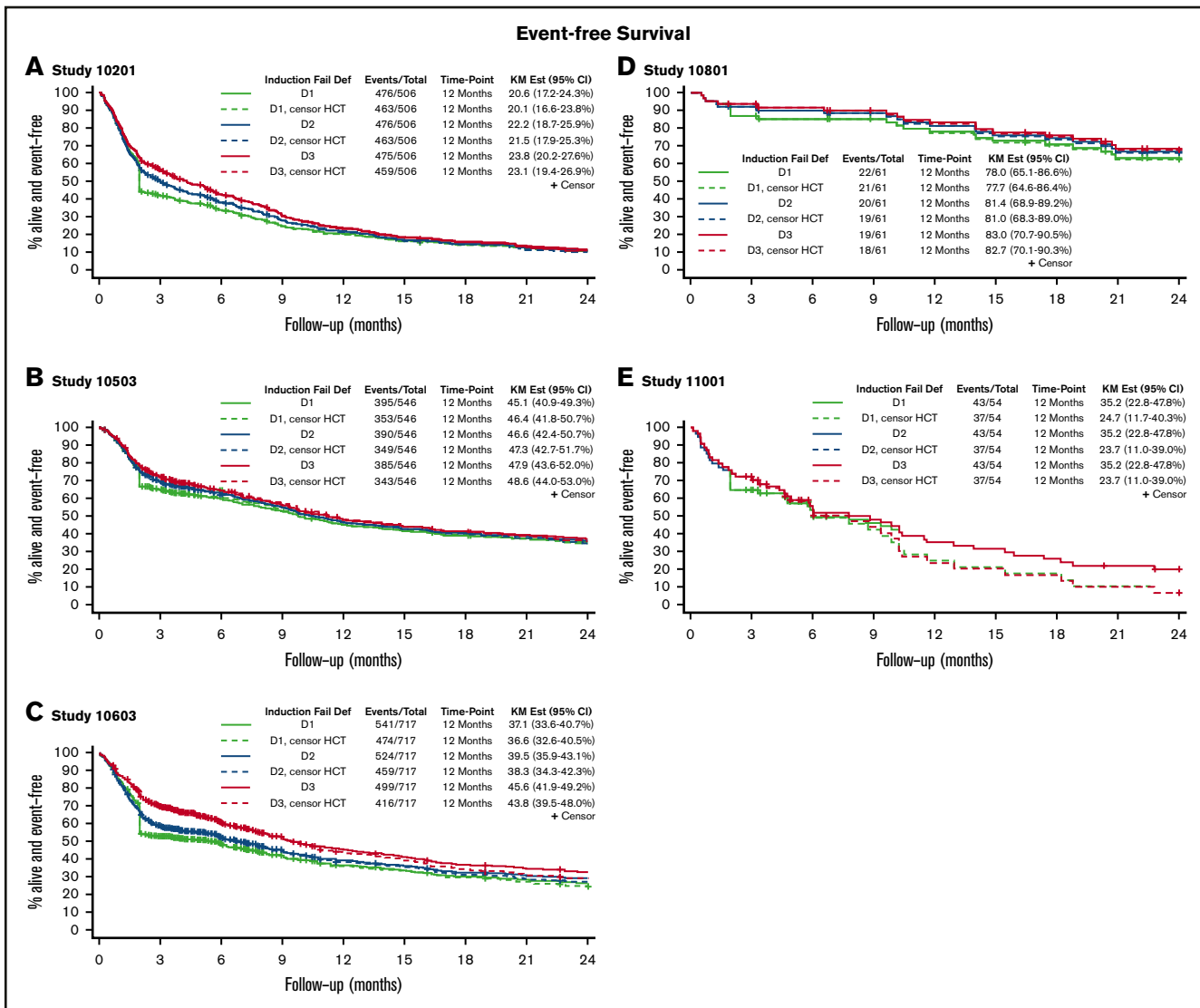
The effect of HCT on EFS estimation was also investigated. The percentages of patients who underwent HCT in first CR or after relapse were 9% in CALGB 10201, 35% in CALGB 10503, 57% in CALGB 10603, 8% in CALGB 10801, and 41% in CALGB 11001 (Table 3). However, the percentage of patients who were actually censored at the time of HCT was reduced by approximately half (range, 3%-28%), thereby indicating that an event of interest (relapse or death) occurred before transplantation in approximately half of the patients who underwent HCT. Given that a majority of patients were not affected by censoring of HCT (ie, their events happened before HCT), results on EFS were similar regardless of their HCT status (Table 3; Figure 1).

Additional sensitivity analyses were conducted by using D2 to calculate EFS, wherein the reported induction end date was replaced with the date of last clinical assessment during induction. Similarly, the last follow-up date was replaced with the clinical assessment date, which was defined as the last contact date and used to calculate EFS. Furthermore, to evaluate the effect of transplantation, only allogeneic HCT was considered instead of all types of HCT. All of these sensitivity analyses yielded similar results and consistent conclusions.

For the 2 randomized phase 3 trials (CALGB 10201, CALGB 10603), analysis with the 3 induction failure definitions resulted in the same overall conclusions about treatment effect. Furthermore, comparisons between the arms in each trial were consistent across the 3 definitions, and the conclusions about treatment effect were not affected by the 3 definitions (Table 4). For example, in CALGB 10603, it was concluded that the addition of multi-targeted kinase inhibitor midostaurin to standard chemotherapy significantly prolonged EFS among patients with AML and an *FLT3* mutation. In our analysis, EFS estimates were significantly longer in the midostaurin arm than in the placebo arm, with an HR ranging from 0.71 to 0.79 depending on the induction failure definition used. In contrast, the addition of oblimersen to standard chemotherapy failed to improve the outcomes of older AML patients in CALGB 10201, regardless of induction failure definition.

## Discussion

Appropriate sensitivity analyses for the primary efficacy end point and the key secondary efficacy end points are often required by regulatory agencies to evaluate the robustness of efficacy results.<sup>27</sup> For example, the potential bias caused by timing and scheduling of disease progression assessments has received much attention and is well documented.<sup>28</sup> However, specific to AML, no studies so far have systematically considered the potential confounding events; for example, non-protocol-mandated HCT and induction failure leading to changes in treatment. In this analysis, we examined the robustness of EFS in measuring clinical benefit in untreated AML using individual patient data across studies, and



**Figure 1.** Kaplan-Meier (KM) curves of EFS by induction failure definition (fail def) D1 to D3, with and without censoring at HCT. KM curves of EFS in CALGB 10201 (A), CALGB 10503 (B), CALGB 10603 (C), CALGB 10801 (D), and CALGB 11001 (E). Green curves represent EFS estimates (est) by D1 (no CR by 60 days after registration/randomization), blue curves represent EFS estimates by D2 (no CR by the end of all protocol-defined induction courses), and red curves represent EFS estimates by D3 (no CR by the end of all protocol-defined treatment). Solid curves represent EFS analysis with censoring at non-protocol-specified HCT, and dashed curves represent EFS analysis without censoring at HCT.

we provide recommendations on trial design using EFS as an end point.

Although relapse and death are firm end points, the determination of induction failure is not consistent across studies. Median EFS estimates differed considerably depending on the timing used to define induction failure, and the magnitude of the difference ranged from 14% to 115%. In all 5 studies of untreated AML patients who received standard intensive induction chemotherapies, EFS estimates determined by D3 (failure to achieve CR during the entire protocol treatment) were consistently the highest because of the length of EFS. EFS estimates determined by D1 (failure to achieve CR by 60 days) were the most conservative, and EFS estimates determined by D2 (failure to achieve CR after all induction therapies) were intermediate. This suggests that incorrect conclusions about

efficacy could be made in a single-arm trial if the definition of EFS used in the trial were inconsistent with the definition used to determine the historical control estimate. However, in randomized trials, HRs have been shown to be unsusceptible to such bias. Therefore, more emphasis should be placed on using HRs to measure clinical benefit instead of the Kaplan-Meier estimates of EFS medians in trials with a concurrent control.

To further explore the reason behind the discrepancy between EFS using D1 vs D2, we specifically studied patients who were considered to have induction failure by D1 but not by D2. For example, in CALGB 10603, 62 patients achieved a CR after 60 days postrandomization, but before their reported end date of all induction courses. Of these 62 patients, only 1 achieved a CR during the first induction; the other 61 patients achieved a CR

**Table 4. EFS estimates determined by using different induction failure definitions for randomized trials CALGB 10201 and CALGB 10603**

Induction failure definition	Median (95% CI) EFS by arm, mo		Comparison of EFS between arms	
	Arm 1	Arm 2	HR (95% CI)	P
<b>CALGB 10201</b>				
D1	2 (2.0-3.4)	2.0 (NA-NA)	1.03 (0.86-1.23)	.770
D2	3.3 (2.4-5.3)	2.7 (1.9-4.3)	1.02 (0.85-1.22)	.850
D3	4.5 (3.4-5.7)	3.8 (2.3-5.7)	1.04 (0.87-1.25)	.670
<b>CALGB 10603</b>				
D1	7.8 (4.7-10.6)	2.8 (2.0-5.9)	0.79 (0.67-0.94)	.005
D2	9.5 (7.3-13.1)	5.5 (3.0-6.7)	0.76 (0.64-0.90)	.002
D3	14.5 (10.6-17.3)	7.2 (6.0-8.9)	0.71 (0.60-0.85)	<.001

HR, hazard ratio; NA, not reached.

during their second induction therapy. Because of the timing of their CRs, using D1, these patients were considered to have induction failures with their EFS time equal to 60 days (1.97 months), whereas using D2, they were considered to be in induction remission until relapse or death, with a median EFS of 23.8 months (95% CI, 11.8 to not reached). These data are included in supplemental Table 3. Similarly, when comparing D2 vs D3, we identified 25 patients from CALGB 10603 who achieved a CR postinduction; half of the patients in this group achieved a CR 42 days postinduction (25% to 75% quantile, 23-85 days), suggesting that they were receiving protocol consolidation therapy when they achieved a CR.

Although we evaluated the effect of the timing of remission assessment, a limitation of the study is that we did not evaluate the effect of what constitutes a remission in the definition of EFS. For example, less stringent response criteria include CR without platelet recovery and CR with incomplete hematologic recovery. Another limitation of our study is that we did not take into account adherence to the protocol induction therapy. For example, in another AML trial (H049),<sup>6,29</sup> >90% of patients who did not have a CR after the first course of induction received a second course as required by the protocol, whereas in trial S0106,<sup>30</sup> the adherence rate was only ~55%. This may relate to differences in clinical practice; in the United States, physicians may remove a patient from protocol therapy on day 14 if bone marrow biopsy results indicate a high burden of disease, but in Europe, 2 cycles of induction are customarily administered regardless of poor early response. However, because induction failure is delayed when a second course of induction is received by initial nonresponders, the EFS becomes longer just by improving protocol adherence. In addition, subsequent treatment information, including HCT, was not systematically recorded in the clinical trials considered in this analysis, and this is another limitation. Statistical approaches, such as competing risk analysis and multistate models, are more appropriate for investigation of the effect of off-protocol transplantation on these outcomes.

Defining the clinical benefit of therapy in AML is complicated by disease- and treatment-specific considerations. Although CR has been uniformly defined in AML and is a more tractable end point, several recent studies have indicated a dissociation between CR and OS (ie, improvement in CR but no improvement in OS compared

with controls) under some conditions.<sup>31,32</sup> Minimal (or measureable) residual disease has been used as an alternative end point to support the regulatory approval of agents for both chronic myeloid leukemia and acute lymphoblastic leukemia. The use of minimal residual disease in AML has been hampered by the lack of standardized assessments. However, EFS has the advantage of reaching an end point sooner than OS, but unlike OS, EFS is not influenced by subsequent treatments administered after failure to achieve or maintain a remission and thus provides a more precise assessment of the efficacy of a particular drug.

The analysis presented here is different from a surrogacy analysis of EFS to OS, where 2 levels of correlations (ie, individual patient-level correlation and trial-level correlation) need to be formally established using a meta-analytical approach.<sup>33</sup> Ongoing work is focused on collecting data from trials conducted across the National Cancer Institute National Clinical Trials Network to perform a formal surrogacy analysis of EFS to OS.

To conclude, this is by far the largest study (N = 1884) to evaluate the robustness of EFS and consider potentially confounding events using individual patient data collected from 5 CALGB trials. Although analysis of the randomized trials revealed that the trial conclusions remained unaffected, median EFS estimates differed considerably based on the timing of CR used in defining induction failure. The magnitude of difference in these estimates could be large enough in most cases to lead to incorrect conclusions about efficacy in a single-arm trial, if the trial definition were not consistent with the definition used for the historical control. Therefore, the timing of CR should be carefully examined in the historical control data used to guide the design of single-arm trials using EFS as the primary end point.

## Acknowledgments

Supported by the National Cancer Institute, National Institutes of Health, under awards U10CA180821 and U10CA180882 (to the Alliance for Clinical Trials in Oncology), U10CA180833, U10CA180836, U10CA180850, and U10CA180867 and in part by funds from Bayer AG (CALGB 11001), Bristol-Myers Squibb (CALGB 10803), and Novartis (CALGB 10603).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Authorship

Contribution: J.Y. designed the study, performed the statistical analysis, and wrote the first draft of the manuscript; B.L. performed the statistical analysis and reviewed and revised the manuscript; G.L.U., G.M., W.B., R.A.L., and R.M.S. provided the data and reviewed and revised the manuscript; S.J.M. conceptualized and designed the study, coordinated data collection, and reviewed and revised the manuscript; and all authors participated in the editing of the manuscript and approved the final version.

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

ORCID profiles: J.Y., 0000-0003-3242-9664; R.A.L., 0000-0001-9168-3203.

Correspondence: Jun Yin, Mayo Clinic, 200 First St SW, Rochester, MN 55905; e-mail: yin.jun@mayo.edu.

## References

1. Kadia TM, Ravandi F, Cortes J, Kantarjian H. New drugs in acute myeloid leukemia. *Ann Oncol*. 2016;27(5):770-778.
2. Estey E, Othus M, Lee SJ, Appelbaum FR, Gale RP. New drug approvals in acute myeloid leukemia: what's the best end point? *Leukemia*. 2016;30(3):521-525.
3. Döhner H, Estey EH, Amadori S, et al; European LeukemiaNet. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010;115(3):453-474.
4. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. *N Engl J Med*. 2015;373(12):1136-1152.
5. Przepiorka D, Deisseroth A, Farrell AT. Acute myeloid leukemia response measures other than complete remission. *J Clin Oncol*. 2015;33(31):3675-3676.
6. Löwenberg B, Ossenkoppele GJ, van Putten W, et al; Swiss Group for Clinical Cancer Research (SAKK) Collaborative Group. High-dose daunorubicin in older patients with acute myeloid leukemia [published correction appears in *N Engl J Med*. 2010;362(12):1155]. *N Engl J Med*. 2009;361(13):1235-1248.
7. Nand S, Othus M, Godwin JE, et al. A phase 2 trial of azacitidine and gemtuzumab ozogamicin therapy in older patients with acute myeloid leukemia. *Blood*. 2013;122(20):3432-3439.
8. Burnett AK, Russell NH, Hills RK, et al. Optimization of chemotherapy for younger patients with acute myeloid leukemia: results of the medical research council AML15 trial. *J Clin Oncol*. 2013;31(27):3360-3368.
9. Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N Engl J Med*. 2017;377(5):454-464.
10. Gao W, Estey E. Moving toward targeted therapies in acute myeloid leukemia. *Clin Adv Hematol Oncol*. 2015;13(11):748-754.
11. Burnett AK, Knapper S. Targeting treatment in AML. *Hematology Am Soc Hematol Educ Program*. 2007;2007:429-434.
12. Herndon TM, Demko SG, Jiang X, et al. U.S. Food and Drug Administration approval: peginterferon-alfa-2b for the adjuvant treatment of patients with melanoma. *Oncologist*. 2012;17(10):1323-1328.
13. André T, Boni C, Mounedji-Boudiaf L, et al; Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) Investigators. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med*. 2004;350(23):2343-2351.
14. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *J Clin Oncol*. 2012;30(10):1030-1033.
15. de Lima M, Strom SS, Keating M, et al. Implications of potential cure in acute myelogenous leukemia: development of subsequent cancer and return to work. *Blood*. 1997;90(12):4719-4724.
16. Yanada M, Garcia-Manero G, Borthakur G, Ravandi F, Kantarjian H, Estey E. Potential cure of acute myeloid leukemia: analysis of 1069 consecutive patients in first complete remission. *Cancer*. 2007;110(12):2756-2760.
17. Burnett AK, Goldstone A, Hills RK, et al. Curability of patients with acute myeloid leukemia who did not undergo transplantation in first remission. *J Clin Oncol*. 2013;31(10):1293-1301.
18. Cornely OA, Maertens J, Winston DJ, et al. Posaconazole vs. fluconazole or itraconazole prophylaxis in patients with neutropenia. *N Engl J Med*. 2007;356(4):348-359.
19. Pagano L, Caira M, Candoni A, et al; SEIFEM Group. Evaluation of the practice of antifungal prophylaxis use in patients with newly diagnosed acute myeloid leukemia: results from the SEIFEM 2010-B registry. *Clin Infect Dis*. 2012;55(11):1515-1521.
20. Vredenburgh JJ, Madan B, Coniglio D, et al. A randomized phase III comparative trial of immediate consolidation with high-dose chemotherapy and autologous peripheral blood progenitor cell support compared to observation with delayed consolidation in women with metastatic breast cancer and only bone metastases following intensive induction chemotherapy. *Bone Marrow Transplant*. 2006;37(11):1009-1015.
21. Blum W, Sanford BL, Klisovic R, et al; Alliance for Clinical Trials in Oncology. Maintenance therapy with decitabine in younger adults with acute myeloid leukemia in first remission: a phase 2 Cancer and Leukemia Group B Study (CALGB 10503). *Leukemia*. 2017;31(1):34-39.
22. Frei E, Visco C, Xu-Monette ZY, et al. Addition of rituximab to chemotherapy overcomes the negative prognostic impact of cyclin E expression in diffuse large B-cell lymphoma. *J Clin Pathol*. 2013;66(11):956-961.
23. Uy GL, Mandrekar SJ, Laumann K, et al. A phase 2 study incorporating sorafenib into the chemotherapy for older adults with FLT3-mutated acute myeloid leukemia: CALGB 11001. *Blood Adv*. 2017;1(5):331-340.
24. Cheson BD, Bennett JM, Kopecky KJ, et al; International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia [published correction appears in *J Clin Oncol*. 2004;22(3):576]. *J Clin Oncol*. 2003;21(24):4642-4649.
25. Cheson BD, Cassileth PA, Head DR, et al. Report of the National Cancer Institute-sponsored workshop on definitions of diagnosis and response in acute myeloid leukemia. *J Clin Oncol*. 1990;8(5):813-819.
26. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.
27. US Food and Drug Administration. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics>. Accessed 28 February 2019.

28. Qi Y, Allen Ziegler KL, Hillman SL, et al. Impact of disease progression date determination on progression-free survival estimates in advanced lung cancer. *Cancer*. 2012;118(21):5358-5365.
29. Othus M, van Putten W, Lowenberg B, et al. Relationship between event-free survival and overall survival in acute myeloid leukemia: a report from SWOG, HOVON/SAKK, and MRC/NCRI. *Haematologica*. 2016;101(7):e284-e286.
30. Othus M, Appelbaum FR, Petersdorf S, Erba HP, Estey EH. Evaluation of which patients get a second course of 3+7 on cooperative group trials for newly diagnosed acute myeloid leukemia: a report from SWOG [abstract]. *Blood*. 2013;122(21). Abstract 3925.
31. Burnett AK, Russell NH, Hunter AE, et al; UK National Cancer Research Institute AML Working Group. Clofarabine doubles the response rate in older patients with acute myeloid leukemia but does not improve survival. *Blood*. 2013;122(8):1384-1394.
32. Burnett AK, Hills RK, Hunter AE, et al; UK National Cancer Research Institute AML Working Group. The addition of gemtuzumab ozogamicin to low-dose Ara-C improves remission rate but does not significantly prolong survival in older patients with acute myeloid leukaemia: results from the LRF AML14 and NCRI AML16 pick-a-winner comparison. *Leukemia*. 2013;27(1):75-81.
33. Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2004;167(1):103-124.