SPECIAL ARTICLE

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

By Bruce D. Cheson, John M. Bennett, Kenneth J. Kopecky, Thomas Büchner, Cheryl L. Willman, Elihu H. Estey, Charles A. Schiffer, Hartmut Doehner, Martin S. Tallman, T. Andrew Lister, Francesco Lo-Coco, Roel Willemze, Andrea Biondi, Wolfgang Hiddemann, Richard A. Larson, Bob Löwenberg, Miguel A. Sanz, David R. Head, Ryuzo Ohno, and Clara D. Bloomfield

Abstract: **An International Working Group met to revise the diagnostic and response criteria for acute myelogenous leukemia originally published in 1990, as well as to provide definitions of outcomes and reporting standards to improve interpretability of data and comparisons among trials. Since the original publication, there have been major advances in our understanding of the biology and molecular genetics of acute leukemia that are clinically relevant and warrant incorporation into response definitions. Differences from the 1990 recommendations included a category of leukemia-free state, new criteria for complete remission, including cytogenetic and molecular remissions and remission duration. Storage of viable blasts for correlative studies is important for future progress in the therapy of these disorders.**

J Clin Oncol 21:4642-4649.

IN 1988, a group of investigators interested in the design and conduct of clinical trials in acute myeloid leukemia (AML) met at the National Cancer Institute (United States) and developed a set of recommendations for response assessment.[1] The subsequent publication was widely adopted as a standardized means of designing and reporting trials,[1] although various study groups still used modifications of these definitions. In the ensuing decade, improvements in the diagnostic criteria and insights into the biology and genetics of AML made it apparent that revisions of these guidelines were needed. In addition, new therapeutic agents with different mechanisms of action and toxicities had become available. As a result, an international group of investigators met in Madrid, Spain, March 23–25, 2001, to develop a revised set of recommendations that incorporated new concepts of biology and therapy (Table 1).

The following guidelines were developed with the intent of being clinically relevant. Some of the recommendations are essential for the proper management of patients with AML, whereas others are, at present, relevant only for clinical research. There was unanimous agreement that phase I, II, and III studies addressing important clinical questions should have correlative laboratory studies if the outcome of AML is to be improved. Storage of viable leukemic cells should be an integral part of the overall strategy because of the increased interest in such technologies as genomics and proteomics.

DEFINITION AND DIAGNOSIS OF AML

AML describes a heterogeneous group of clonal hematopoietic progenitor cell disorders with a spectrum of morphologic, immunophenotypic, cytogenetic, and molecular characteristics. It may be possible to make the diagnosis on the basis of a peripheral blood examination; nevertheless, a bone marrow aspiration is strongly recommended. Bone marrow trephine biopsy is not routinely indicated, although it may be necessary if the aspirate is dilute, hypocellular, or inaspirable. Touch preparations and clot sections do not provide sufficient additional information to be recommended for general use but may be helpful if the aspiration and biopsy samples are inadequate.

The French-American-British (FAB) classification initially was based on morphology, cellularity, blast percentages, and cytochemistry.[2] Subsequently, the definition of AML, undiffer-

From the Department of Hematology/Oncology, Georgetown University, Washington, DC; University of Rochester Medical Center, Rochester, NY; Fred Hutchinson Cancer Research Center, Seattle, WA; University of Texas M.D. Anderson Cancer Center, Houston, TX; Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI; University of Münster, Münster; University of Grosshadern, Munich; University of Ulm, Ulm, Germany; University Tor Vergata of Rome; Clinica Pediatrica Università Milano-Bicocca, Ospedale S. Gerardo, Monza, Italy; Leiden University, Leiden; Erasmus Medical Center Rotterdam; Institute of Hematology, Rotterdam, Netherlands; University of Chicago; Northwestern University, Chicago, IL; Hospital Universitario La Fe, University of Valencia, Spain; University of New Mexico Cancer Treatment and Research Center, Albuquerque, NM; Aichi Cancer Center Hospital, Nagoya, Japan; St. Bartholomew's Hospital, London, UK; Vanderbilt University School of Medicine, Nashville, TN; and The Ohio State University Comprehensive Cancer Center, Columbus, OH.

Submitted April 3, 2003; accepted September 17, 2003.

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Address reprint requests to Bruce D. Cheson, MD, Head of Hematology, Georgetown University, Lombardi Cancer Center, 3800 Reservoir Rd, NW, Washington, DC 20007; e-mail: bdc4@georgetown.edu.

0732-183X/03/2124-4642/$20.00

DOI: 10.1200/JCO.2003.04.036

entiated type (M0), and acute megakaryocytic leukemia (M7), used immunologic markers.[3-5] The new WHO recommendations[6,7] updated and modified the FAB diagnostic criteria.

The terms de novo AML and secondary AML are widely used but are without standardized definitions. We recommend the following definitions on the basis of clinical history and availability of pathologic material, which are consistent with the intent of the WHO criteria and similar to those proposed by one of the coauthors[8]: (1) De novo AML should refer to AML in patients with no clinical history of prior myelodysplastic syndrome (MDS), myeloproliferative disorder, or exposure to potentially leukemogenic therapies or agents. (2) Secondary AML should refer to patients who have such clinical histories and should be further categorized as AML secondary to prior existing MDS, myeloproliferative disorder, or the development of AML secondary to proven leukemogenic exposure. A history of fatigue, bleeding, or recurrent infections that preceded the diagnosis of AML by 1 month or greater, although suggestive of a preleukemic state, should not by itself allow designation of a case in this category without confirmation of an existing peripheral blood film that demonstrates morphologic dysplasia.

Furthermore, there are two major categories that have been well described and are only mentioned briefly here for completeness: alkylating agent–related MDS/AML[9] and topoisomerase II–related AML.[10]

The use of both classes of agents, particularly in the bone marrow transplantation or dose-intensity setting, may confound the issues of this separation. It is also clear that the poor results of treatment with alkylating agent–related MDS/AML and primary AML with multilineage dysplasia imply that the biology of both types is very similar.

A careful history of exposures to potentially leukemogenic agents should be obtained in cases for which there is a question of a secondary AML. Such investigations are generally not revealing in individual cases and require careful epidemiologic evaluation of populations presumed to be at risk. The point at which a patient with AML is treated may also depend on the available therapeutic options.

A major departure by the WHO from the FAB criteria was to lower the threshold for the diagnosis of AML from 30% to 20% blasts in the peripheral blood and/or the bone marrow aspirate. Exceptions include AML with t(8;21), inv(16), or t(15;17), in which the diagnosis of AML is made regardless of the percentage of bone marrow blasts. Lowering the blast threshold to 20% eliminated the MDS category of refractory anemia with excess blasts in transformation.

The cytochemical and phenotypic criteria of the WHO were accepted by the present working group.[6,7,11] It is important to record the presence or absence of significant dysplasia.[12,13]

Routine cytochemical evaluation (peroxidase and esterase) should be carried out in conjunction with immunophenotyping by flow cytometry. Immunophenotyping is valuable to distinguish AML from acute lymphocytic leukemia, and for lineage determination. The use of multicolored flow cytometry is highly recommended as the preferred technique compared with immunohistologic methods but cannot be mandated where the diagnosis has clearly been established by routine morphologic and cytochemical criteria. The use of flow cytometry to evaluate minimal residual disease (MRD) is an area of active clinical investigation and is further discussed in subsequent sections of this report. However, the use of flow cytometry to replace standard differential counting is discouraged and should not substitute for an inadequate bone marrow aspirate.

Table 1. Revisions in Current AML Guidelines

1. Recommendations for storage of viable blasts
2. Definitions of de novo and secondary AML
3. Implications of dysplasia
4. Use of WHO definitions
5. Importance of flow cytometry
6. Prognostic relevance of bone marrow cytogenetics and molecular genetics (eg, FLT-3 mutations and PTD of *MLL* gene)
7. Molecular remission of APL
8. Indications for central pathology review
9. Leukemia-free state as a response criterion
10. CRc and CRm as response criteria
11. 1,000/μL neutrophils as threshold for CR
12. Elimination of 4-week requirement for CR
13. Relapse requiring 5%-20% blasts in the bone marrow
14. No requirement for cellularity in CR definition
15. Late MDS as criterion for recurrence
16. RFS and OS as primary end points

Abbreviations: AML, acute myelogenous leukemia; PTD, partial tandem duplication; APL, acute promyelocytic leukemia; CRc, cytogenic complete remission; CRm, molecular complete remission; CR, complete remission; MDS, myelodysplastic syndrome; RFS, relapse-free survival; OS, overall survival.

Cytogenetics confer the most important prognostic information in AML,[7,14,15] along with patient age, performance status, presenting WBC count, flt-3 mutation, *MLL* partial tandem duplication, and whether a patient has de novo or secondary disease. Therefore, detailed karyotypes should be performed on all patients with AML at diagnosis.[7,16] In addition, because specific, recurrent cytogenetic abnormalities confer varying prognoses, differing responses to chemotherapy, or transplantation regimens and will increasingly be used to direct patients to different targeted therapies, it is critical to report the specific chromosome abnormality, rather than reporting the broader prognostic categorizations frequently employed (eg, normal or abnormal, favorable or unfavorable).

In addition to the classification of AML patients by their cytogenetic abnormalities, molecular genetic studies in the last 10 years have also identified important clinical and biologic subsets of AML patients. Two prominent examples, both associated with a poorer prognosis in AML patients with normal cytogenetics, are (1) internal tandem duplications of FLT3 mutations,[17-19] and (2) partial tandem duplications of the *MLL* gene on 11q23.[20,21] Identification of patients with specific molecular genetic abnormalities may be important as therapies targeted to these molecular genetic lesions are developed.

Among the forms of AML with balanced reciprocal translocations, two groups deserve special mention. Acute promyelocytic leukemia (APL) has distinct biologic and clinical features, as well as unique treatment approaches resulting in 70% of patients being cured.[22] The clinical diagnosis of APL rests both

Table 2. Response Criteria in AML

Response Criterion	Time of Assessment	Neutrophils (μL)	Platelets (μL)	Bone Marrow Blasts (%)	Other
Early treatment assessment	7-10 days after therapy	NA	NA	< 5	
Morphologic leukemia-free state	Varies by protocol	NA	NA	< 5	Flow cytometry EMD
Morphologic CR	Varies by protocol	> 1,000	> 100,000	< 5	Transfusion EMD
Cytogenetic CR	Varies by protocol	> 1,000	> 100,000	< 5	Cytogenics—normal, EMD
Molecular CR	Varies by protocol	> 1,000	> 100,000	< 5	Molecular—negative, EMD
Partial remission	Varies by protocol	> 1,000	> 100,000	> 50 or decrease to 5-25	Blasts < 5% if Auer rod positive

Abbreviations: AML, acute myelogenous leukemia; EMD, extramedullary disease; CR, complete remission.

on morphology and detection of either the reciprocal translocation between chromosomes 15 and 17, or the PML-RARα fusion gene, which results from this translocation.

A second group of patients with AML of special note are those with functional inactivation of the core binding transcription factors (CBFs): AML1 and CBFβ. These cases include patients with AML and t(8;21)(q22;q22) or inv(16)(p13q22), two of the most frequent recurrent cytogenetic abnormalities in de novo AML in younger adults. Patients with t(16;16) should be regarded as clinically equivalent to inv(16). Those with 16q- deletions need to be molecularly determined if fusion transcripts are present.

DEFINITION OF RESPONSE IN AML

The definition of response in AML must be clinically relevant, practical, and reproducible to be easily used by investigators and clinicians from different institutions. Such criteria would permit comparison of results of studies from different institutions and study groups as well as provide important information to aid in decisions for the care of individual patients. Cooperative groups have used differing definitions in the past.

Problems that may be encountered in large cooperative group studies include the need for central pathology review of bone marrow.

Variability in the experience of pathologists at different institutions, compounded by variability in specimen quality, can also result in a wide range in the interpretation and significance of the bone marrow reports. Central review is generally not required except for studies in which the primary goal is to assess overall activity of a new agent, which would include clearance of bone marrow blasts and the establishment of a complete remission. In addition, central review should be required where molecular studies are being performed because of the need to know the number of blasts. Automated differential blood counts that describe a "few" blasts with no review of the peripheral blood smear by a pathologist are not acceptable.

The initial goal of therapy for AML is to achieve a complete remission (CR), given that a CR with currently available therapy is requisite, although not sufficient, for a cure. CR is the most important initial response reported in phase III trials because it is the sole outcome currently associated with improved survival. In phase I and II clinical trials, usually conducted in refractory patients or those who experience relapse, or in those who are in highly unfavorable risk groups, both complete and partial response rates should be recorded to avoid missing a therapeutic strategy with meaningful activity.

We recommend that each of the following be reported (Table 2): (1) Early treatment assessment: This evaluation is the first stage in some investigational studies to assess response. It is an evaluation made at approximately 7 to 10 days after completing the last dose of the initial course of treatment. Although such a sample is likely to be hypocellular, it provides an indication of antileukemic activity.[23,24] Early assessment is often required in clinical trials to guide subsequent treatment—for example, the need for or timing of a second induction course, or to meet a specific study objective.

(2) Morphologic leukemia-free state: This designation requires less than 5% blasts in an aspirate sample with marrow spicules and with a count of at least 200 nucleated cells. There should be no blasts with Auer rods or persistence of extramedullary disease. The presence of a unique phenotype (by flow cytometry) identical to what was found in the pretreatment specimen (eg, CD34, CD7 coexpression) should be viewed as persistence of leukemia. The timing of this determination varies from protocol to protocol and should be consistent with the objectives of the study. If there is a question of residual leukemia, a bone marrow aspirate should be repeated in a week. A biopsy allows more bone marrow tissue to be examined and should certainly be performed if spicules are absent from the aspirate sample. The biopsy also allows identification of clusters of blasts, which are rarely seen in normal hematopoiesis.

(3) Morphologic complete remission: A CR designation requires that the patient achieve the morphologic leukemia-free state and have an absolute neutrophil count of more than 1,000/μL and platelets of $\geq$ 100,000/μL. Hemoglobin concentration or hematocrit has no bearing on remission status, although the patient must be independent of transfusions. Recent series suggest that circulating blasts may be identified in small numbers and still be consistent with a similar disease-free survival (DFS) as patients without circulating blasts.[25] In general, however, persistent blasts in the peripheral blood correlate with persistent or recurrent AML and blast infiltration of the bone marrow. Occasionally, a rare peripheral blood blast may be identified during regeneration; however, if the patient is in CR, the bone marrow would have less than 5% blasts and no Auer rods.[26] Flow cytometry may also be useful to distinguish between leukemia and a regenerating bone marrow. There is no requirement for bone marrow cellularity. The implications of dysplasia after treatment are not clear. However, if the dysplasia was present at diagnosis, persistence suggests residual leukemia (except for mild megaloblastic change that could be secondary to

chemotherapy). Marrow cytogenetics and flow cytometry may be of help. There should be no residual evidence of extramedullary leukemia.

In the previous guidelines, a 4-week duration of complete response was required to qualify as a CR. However, some patients who fulfilled the other criteria for CR could not be considered CR because of the administration of postremission therapy before full recovery of blood counts within that time period, or because they had evidence of recurrent or persistent disease after more than 4 weeks but no documentation that all response criteria were satisfied for at least 4 weeks. Therefore, no duration of response is required in the current recommendations. The primary end points of phase III trials should be relapse-free survival (RFS) and overall survival (OS). Retrospective designation of a CR based on subsequent clinical course should be avoided. Other end points may be justified on the basis of a specific study design and these should be clearly defined in the protocol document and subsequent publication.

As newer and more sensitive technologies are developed to quantitate the level of leukemic burden beyond the sensitivity of the light microscope, definitions of CR will continue to evolve over the next several years. Three special categories of patients who fulfill the morphologic criteria for CR should thus be considered:

(3a) Cytogenetic complete remission (CRc). The majority of patients who achieve a CR are not cured. Therefore, the use of a morphologically determined cutoff of 5% blasts is arbitrary. On the basis of recent cytogenetic data, a separate category of CR is proposed to include reversion to a normal karyotype at CR because preliminary data suggest that patients with residual cytogenetic abnormalities have a much poorer prognosis than those in whom the chromosomal aberration is no longer detectable. However, because sufficient data are lacking from prospective trials, this category is recommended primarily for use in clinical research studies.[27,28] In such studies it will be important to describe explicitly how CRc is defined; for example, the minimum number of metaphases required to define a normal karyotype, or whether CRc is based on conventional banded studies or on more sensitive techniques such as fluorescence in situ hybridization.

(3b) Molecular complete remission (CRm). Clinical investigations of MRD using both molecular and multidimensional flow cytometric techniques have clearly demonstrated that the vast majority of AML patients in morphologic and cytogenetic CR have detectable residual disease. Automated quantitative reverse transcriptase polymerase chain reaction (RT-PCR) techniques are sensitive in detecting residual disease in AML cases with a specific genetic marker (eg, PML-RARα in t(15;17), AML1/ETO fusion in t(8;21), CBFβ-MYH11 fusion in inv(16), and others). Similarly, detection of aberrant phenotypes using multidimensional flow cytometric techniques is a sensitive approach to monitor MRD in a large percentage of AML patients.[29] Data from the Cancer and Leukemia Group B studies indicate a high frequency of immunophenotypic changes in AML patients between diagnosis and relapse, so that multiple antibody panels are needed for monitoring residual disease by multiparameter flow cytometry.[29] The prognostic significance of CRm achievement is clearly established for APL,[30,31] and CRm is recognized as a therapeutic objective in APL by most hematologists, whereas the significance of CRm in the other subsets of AML is still controversial (particularly for AML1/ETO patients). Other molecular targets, such as WT-1, can also be assessed using quantitative RT-PCR approaches and may be useful targets for MRD testing.[32] Using these quantitative MRD approaches in patients in clinical research studies, quantitative thresholds in the range of one leukemia cell in 1,000 to 10,000 cells can predict impending relapse compared with continuous CR. As these approaches are tested in the context of clinical trials, our definitions of useful response criteria will undoubtedly evolve and may vary with the specific genotypes of AML. Again, if CRm is used as an outcome measure in AML clinical trials, it must be defined precisely, including the marker or markers examined and the sensitivity of the quantitative PCR assay.

(3c) Morphologic complete remission with incomplete blood count recovery (CRi). After chemotherapy, some patients fulfill all of the criteria for CR except for residual neutropenia ($<$ 1,000/μL) or thrombocytopenia ($<$ 100,000/μL). This term is similar to the term CRp, used for studies with gemtuzumab ozogamicin.[33] The outcome for these patients does not seem to be comparable to that of patients with normalization of all counts, especially for those patients during initial therapy for their AML.[34] Although this category of response indicates activity, it should not be included with CR.

For the three CR categories, extramedullary leukemia, such as CNS or soft tissue involvement, must be absent. In the absence of symptoms, CNS surveillance is not recommended unless it is part of a clinical research study. Indeed, routine periodic bone marrow surveillance also probably is not essential, except for clinical research studies looking at MRD.[35]

(4) Partial remission (PR). PRs are relevant only for phase I and II trials evaluating the safety and activity of a new agent or approach. This designation requires all of the hematologic values for a CR but with a decrease of at least 50% in the percentage of blasts to 5% to 25% in the bone marrow aspirate. Thus, if the pretreatment bone marrow blast percentage was 50% to 100%, the percentage of blasts must decrease to a value between 5% and 25%; if the pretreatment blast percentage was 20% to less than 49%, they must decrease by at least half to a value of more than 5%. A repeat bone marrow aspiration after several weeks may be required to distinguish between a PR and increased blasts caused by bone marrow regeneration. A value of $\leq$ 5% blasts may also be considered a PR if Auer rods are present.

(5) Treatment failure (Table 3). Treatment failure includes those patients for whom treatment has failed to achieve a CR on a phase III trial or less than a PR on a phase I or II trial. Although not as important for phase III trials where RFS and OS are generally the primary end points, these definitions are important for evaluating new agents and are included in the assessment performed by the US Food and Drug Administration when considering an agent for approval. The following is recommended as a classification of treatment failures: (a) Treatment failure due to resistant disease includes appropriately treated patients who survive at least 7 days after completion of the final

Table 3. Treatment Failure in AML

Category	Definition
Resistant disease	Patient survives ≥ 7 days post-CT; persistent AML in blood or bone marrow
Aplasia	Patient survives ≥ 7 days post-CT; death while cytopenic, with aplastic bone marrow
Indeterminate cause	Patients who die < 7 days posttherapy Patients who die > 7 days posttherapy with no PB blasts, but no bone marrow examination Patients who do not complete the first course of therapy
Morphologic relapse	Reappearance of blasts post-CR in PB or bone marrow
Molecular or cytogenetic relapse	Reappearance of molecular or cytogenetic abnormality

Abbreviations: AML, acute myelogenous leukemia; CT, chemotherapy; PB, peripheral blood; CR, complete remission.

dose of the initial course of treatment but whose last posttreatment peripheral blood smear and/or bone marrow sample showed persistent AML. If the protocol includes a predetermined second induction attempt for patients with evidence of persistent disease after the first attempt, the two are considered the "initial course of treatment." (b) Treatment failure due to complications from aplasia includes patients who survive at least 7 days after the final dose of the initial course of treatment and die while cytopenic, but whose last posttreatment bone marrow was aplastic or hypoplastic, as determined by the institutional morphologist or pathologist, without evidence of leukemia, provided that marrow was obtained within 7 days of death. (c) Treatment failure of indeterminate cause includes three categories of patients: (1) those who die less than 7 days after conclusion of the initial course of treatment, and (2) patients who die 7 or more days after the conclusion of treatment whose most recent peripheral blood smear did not show persistent leukemia and who did not have a bone marrow examination subsequent to therapy. This category also includes (3) patients who die without completing the first course of therapy.

(6) Recurrence: Morphologic relapse. Relapse after CR is defined as a reappearance of leukemic blasts in the peripheral blood or ≥ 5% blasts in the bone marrow not attributable to any other cause (eg, bone marrow regeneration after consolidation therapy). The appearance of new dysplastic changes should also be considered relapse. In the setting of recent treatment, if there are no circulating blasts and the bone marrow contains 5% to 20% blasts, a repeat bone marrow performed at least a week later is necessary to distinguish relapse from bone marrow regeneration. In such instances the date of recurrence is defined as the first date that more than 5% blasts were observed in the marrow. The reappearance or development of cytologically proven extramedullary disease also indicates relapse. Molecular and/or genetic relapse is characterized by reappearance of a cytogenetic or molecular abnormality.

Unique to the treatment of APL is that there is often no obligatory period of bone marrow aplasia after chemotherapy and ATRA. The bone marrow aspirate performed 7 to 14 days after induction therapy usually reveals a hypercellular specimen with the misleading impression of resistant disease. This finding is not an indicator for additional induction therapy; however, the first posttreatment bone marrow aspirate or biopsy need not be performed until 10 to 14 days after completion of ATRA therapy.

The therapeutic end point in APL is achievement of a molecular remission as defined by the absence of the PML-RARα fusion transcript using RT-PCR methods with a sensitivity threshold of 10^{-3} or 10^{-4}. Absence of the fusion transcript after consolidation therapy is associated with a prolonged remission duration, whereas its reappearance after repeat negative PCR assays is associated with a high likelihood of disease recurrence.[30,31,36] The use of automated quantitative RT-PCR assays on the peripheral blood in this form of AML[37] will likely improve the ability to predict relapses and to determine the quality of a remission. The recommendation is that RT-PCR for the fusion transcript should be performed every 3 months for the first 2 years of CR, then every 3 to 6 months for the following 2 to 3 years.

DEFINITIONS OF TREATMENT OUTCOMES

A variety of treatment outcomes (Table 4) are used to measure the effectiveness of treatment regimens in clinical trials for AML. Two important categories of outcome are indicators of response to therapy (described previously) and measures of the duration of survival or remission described here. Additional measures such as quality of life, treatment cost or cost-benefit assessments, or presence of MRD, may be appropriate for a given trial but are not considered further here.

Duration of survival or of response is measured from a defined starting point (eg, the date of entry onto the study or the date of response) to the end point of interest (eg, death or AML relapse). Such time-to-event data are characterized by the possibility of censoring (ie, by the possibility that the end point of interest will not be observed for some patients because of the end of study follow-up) or because other intervening events (so-called competing events) preclude the end point's occurrence. In the

Table 4. Definitions of End Points for Clinical Trials in AML

Outcome	Response Category	Point of Measurement	Definition
Overall survival	All patients	Entry onto trial	Death from any cause
Relapse-free survival	CR	Leukemia-free state	Disease relapse or patient death from any cause
Event-free survival	All patients*	Entry onto trial	Treatment failure, disease relapse, or patient death from any cause
Remission duration	CR	Date of CR	Disease relapse

NOTE. Complete blood counts should be evaluated at least monthly, or more often if clinically indicated, to establish the durability of responses.
Abbreviations: AML, acute myelogenous leukemia; CR, complete remission.
*Under circumstances where presentation of event-free survival may be appropriate for responders only, this point should be clearly stated.

presence of censoring by competing events, the product-limit method of Kaplan and Meier[38] cannot be relied on to provide estimates of probabilities of the end point of interest, and alternative statistical summaries, such as cumulative incidence,[39-41] should be used. However, standard techniques for comparing groups of patients defined by treatment assignment or other characteristics, such as the log-rank test[42] or the proportional hazards regression model of Cox,[43] remain valid for the end point of interest even if the observation is censored by competing events.

Several measures of the duration of survival and/or response have been used as outcomes in the clinical oncology literature over the years, and there have been no standard definitions or nomenclature for these outcomes. The following are the recommended standard names and definitions.

OS is defined for all patients in a trial, and measured from the date of entry onto a study until death from any cause. For a patient who is not known to have died by the end of study follow-up, observation of OS is censored on the date he or she was last known to be alive. Defined in this way, OS is ordinarily not subject to competing risks in clinical trials, and the product-limit method of Kaplan and Meier[38] can be used to calculate estimates of survival probabilities.

RFS is defined only for patients who achieve CR, and is measured from the date of attaining the leukemia-free state (as discussed previously) until the date of AML relapse or death from any cause, whichever occurs first. For a patient who is not known to have relapsed or died by the end of study follow-up, observation of RFS is censored on the date of his or her last follow-up examination. RFS is ordinarily not subject to competing risks in clinical trials. The protocol and the final report should indicate whether outcomes other than morphologic CR (eg, CRc, CRm) are included in DFS.

Event-free survival (EFS) is defined for all patients and measured from the date of entry on study. It is measured until treatment failure, relapse from CR, or death from any cause, whichever occurs first. The time point at which the patient is resistant to therapy or survives induction without a CR should be noted. Treatment failure should be defined explicitly in the protocol and the final report. For a patient with none of these events before the end of study follow-up, observation of EFS is censored at the date of his or her last follow-up examination. Like OS, EFS is ordinarily not subject to competing risks in clinical trials. If the patient does not achieve a CR, EFS is defined as the point of progression or death, whichever comes first.

Remission duration, like RFS, is defined only for patients who achieve CR, and is measured from the date of CR by blood count recovery and bone marrow examination (rather than the date of the confirmatory bone marrow), until the date of relapse. However, unlike DFS, it is measured only until the date AML relapse is detected. For patients who die without report of relapse, remission duration is censored on the date of death, regardless of cause. For a patient with no report of relapse by the end of the follow-up data collection, observation is censored on the date of his or her last follow-up examination. Note that unlike OS, EFS, and RFS, remission duration is subject to the competing risk of death without relapse. Therefore, the Kaplan-Meier[38] method does not provide estimates of probabilities of remaining relapse free, and estimates of the cumulative incidence of relapse (CIR) should be used instead.

When required by a trial's objectives, outcomes similar but not identical to the four defined above may be appropriate. Such alternative outcomes may be defined for subsets of patients, and may differ in the specification of the date from which time is measured; however, the end point and censoring definitions are unchanged. For example, in a study of postremission therapy for which patients enroll while in CR, times until relapse or death from any cause should be measured from the date of enrollment, not the date CR was achieved. As another example, consider a phase III trial comparing two postremission regimens in patients who all receive a common induction regimen. If patients initially enter the trial for induction therapy, but only remitting patients are randomly assigned between postremission arms after achieving CR, it is sometimes informative to measure survival and remission duration of the patients who were randomly assigned from the date of randomization. To distinguish these outcomes with alternative starting points from the standard outcomes defined above, the alternative outcomes should be described with reference to the starting point; for example, RFS from study entry or survival from postremission randomization.

REPORTING STANDARDS FOR THERAPEUTIC TRIALS

Clinical trials should be conducted expeditiously and reported as rapidly as possible. However, to be able to interpret clinical data or the report of a clinical trial, minimal reporting criteria are needed. The objectives of the study should be explicitly stated, along with the purpose of the analysis (final report or interim analysis). The age of the patients and the distribution of patients with de novo AML, prior MDS, and therapy-related AML should be clearly stated. Eligibility criteria should be explicitly described, as well as any reasons for excluding patients from the reported analyses. In multicenter clinical trials, the local institution's diagnosis of AML or of the appropriate subtype of AML should ordinarily suffice for purposes of determining eligibility. The decision to have central review by a recognized hematomorphology expert should be determined by the nature of each study undertaken. Laboratory methods and criteria used for diagnosis, or for other studies, such as cytogenetics, RT-PCR, and immunophenotyping, whether used to define eligibility, for description of patients, or for response assessment, should be reported. Central analysis of immunophenotyping is recommended because of the wide range of different antibodies available and the need for uniform interpretation. Review of karyotypes is also of great importance because of interpretative differences.

In clinical studies, patients should be observed until their deaths, and reasons for censoring patients should be delineated. Results of interim analyses should not be reported publicly unless there is a compelling reason, and they provide statistically definitive results as determined by a data safety and monitoring committee, where appropriate. For phase III studies, a minimal median duration of follow-up of at least 3 years should be required before reporting data because this corresponds to the

time when the risk of relapse sharply declines,[44] unless the outcomes are so striking that earlier reporting is appropriate. If the report describes an analysis that was not scheduled as part of the study's protocol, then the justification for the unplanned analysis should be explained.

Treatment plans should be fully explained, especially if not previously reported, including doses, schedule, methods of delivery, toxicity-related modifications, and supportive care. The schedule of follow-up examinations should also be described. The numbers of patients who completed planned therapy should be reported, and the reasons for failure to complete planned treatment (eg, toxicity, refusal, death, or relapse while receiving treatment) should be summarized. Significant deviations from treatment plans or follow-up schedules that occurred during the trial should be described. The number of courses actually delivered should be provided. The primary analyses of randomized comparative (phase III) trials should be based on the so-called intent-to-treat principle; that is, should include all patients (or all eligible patients) according to their assigned treatment arms and without regard to completeness of therapy received. If additional analyses that attempt to relate outcomes to treatment actually received are reported, the speculative nature of any resulting conclusions, which arises from the inability to evaluate the extent to which treatment influenced outcome and the extent to which the opposite is true, must be acknowledged.

The distributions of the following minimal list of demographic and clinical characteristics should be summarized in all reports, using frequency distributions for categoric variables and appropriate summary statistics (ie, means, medians, and ranges) for quantitative (so-called continuous) variables: age, sex, race or ethnicity, performance status (eg, Eastern Cooperative Oncology Group, WHO, or Zubrod score), liver and kidney function, clinical onset of AML (de novo AML *v* prior MDS *v* prior leukemogenic therapy), WHO classification, peripheral blood and bone marrow blast percentage, leukocyte count, peripheral blast percentage, platelets, hemoglobin, immunophenotype, and chromosomal analysis. Any other variables used to determine specifics of treatment or to stratify patients for randomization or analysis (eg, serum lactate dehydrogenase) should also be included in these summaries. In multiarm trials these distributions should be presented for each arm. If patients are classified into prognostic categories on the basis of cytogenetics, the categories should be defined in the report or by reference, given that no standard prognostic categories have been provided.

Definitions of response and treatment outcomes should be included, especially if they vary from the standard definitions of OS, RFS, EFS, and remission duration given in this article. Frequency distributions of response should be reported, and if the responses of any patients cannot be determined, the reasons should be explained. For OS, RFS, and EFS, distributions and summary statistics such as medians or probabilities at given times after study entry should be estimated using the method of Kaplan and Meier.[38] CIR should be used to describe remission duration. In analyses of OS, the distributions of survival times and of lengths of time since last follow-up of the living patients should be summarized. When probabilities of OS, EFS, or DFS, or CIR are described in text or tables, the specific lengths of time to which they refer should be stated as, for example, OS at 5 years or CIR within 2 years. CIs or bands and/or SEs should be used routinely to quantify the precision with which response and other outcome measures are estimated.

It is clear that these recommendations will require modification as more is learned about the biology and genetics of these disorders, as treatment modalities evolve, and as the techniques for monitoring the disease become more sensitive and widely available. Until that time, we hope that the current guidelines will provide better communication among investigators conducting clinical research in the group of disorders that constitute AML.

ACKNOWLEDGMENT

Other participants included Alan Burnett and Jean-Luc Harrousseau. We thank Janet Brandom for her assistance in this workshop.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The authors indicated no potential conflicts of interest.

REFERENCES

1. Cheson BD, Cassileth PA, Head DR, et al: Report of the NCI-sponsored workshop on definitions of diagnosis and response in acute myeloid leukemia. J Clin Oncol 8:813-819, 1990

2. Bennett JM, Catovsky D, Daniel MT, et al: Proposals for the classification of the acute leukaemias: French-American-British (FAB) Co-Operative Group. Br J Haematol 33:451-458, 1976

3. Bennett JM, Catovsky D, Daniel MT, et al: Criteria for the diagnosis of acute leukemia of megakaryocyte lineage (M7). Ann Intern Med 103:460-462, 1985

4. Bennett JM, Catovsky D, Daniel MT, et al: Proposed revised criteria for the classification of acute myeloid leukemia. Ann Intern Med 103:626-629, 1985

5. Bennett JM, Catovsky D, Daniel MT, et al: Proposal for the recognition of minimally differentiated acute myeloid leukemia (AML-M0). Br J Haematol 78:325-329, 1991

6. Jaffe ES, Harris NL, Stein H, et al: Tumours of Haematopoietic and Lymphoid Tissues, World Health Organization Classification of Tumours: Pathology & Genetics. Lyon, France, IARC Press, 2001

7. Vardiman J, Harris NL, Brunning RD: The World Health Organization (WHO) classification of the myeloid neoplasms. Blood 100:2292-2302, 2002

8. Bennett JM: Secondary acute myeloid leukemia [editorial]. Leuk Res 19:231-232, 1995

9. Smith SM, Le Beau MM, Huo D, et al: Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: The University of Chicago experience. Blood 102:43-52, 2003

10. Quesnel B, Kantarjian H, Bjergaard JP, et al: Therapy-related acute myeloid leukemia with t(8;21), inv(16), and t(8;16): A report on 25 cases and review of the literature. J Clin Oncol 11:2370-2379, 1993

11. Harris NL, Jaffe ES, Diebold J, et al: World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: Report of the clinical advisory committee meeting—Airlie House, Virginia. J Clin Oncol 17:3835-3849, 1999

12. Goasguen JE, Matsuo T, Cox C, et al: Evaluation of the dysmyelopoiesis in 336 patients with de novo acute myeloid leukemia: Major importance of dysgranulopoiesis for remission and survival. Leukemia 6:520-525, 1992

13. Gahn B, Haase D, Unterhalt M, et al: De novo AML with dysplastic hematopoiesis: Cytogenetic and prognostic significance. Leukemia 10:946-951, 1996

14. Grimwade D, Walker H, Oliver F, et al: The importance of diagnostic cytogenetics on outcome in AML: Analysis of 1,612 patients entered into the MRC AML 10 trial. Blood 92:2322-2333, 1998

15. Byrd JC, Mrózek K, Dodge RK, et al: Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative index of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: Results from Cancer and Leukemia Group B (CALGB 8461). Blood 100:4325-4336, 2002

16. Estey E, Keating MJ, Pierce S, et al: Change in karyotype between diagnosis and first relapse in acute myelogenous leukemia. Leukemia 9:972-976, 1995

17. Whitman SP, Archer KJ, Feng L, et al: Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of FLT3: A cancer and leukemia group B study. Cancer Res 61:7233-7239, 2001

18. Thiede C, Steudel C, Mohr B, et al: Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: Association with FAB subtypes and identification of subgroups with poor prognosis. Blood 99:4326-4335, 2002

19. Fröhling S, Schlenk RF, Breitruck J, et al: Prognostic significance of activating *FLT3* mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: A study of the AML Study Group Ulm. Blood 100:4372-4380, 2002

20. Döhner K, Tobis K, Ulrich R, et al: Prognostic significance of partial tandem duplications of the *MLL* gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetic: A study of the Acute Myeloid Leukemia Study Group Ulm. J Clin Oncol 20:3254-3261, 2002

21. Caligiuri MA, Strout MP, Lawrence D, et al: Rearrangements of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics. Cancer Res 58:55-59, 1998

22. Tallman MS, Andersen JW, Schiffer CA, et al: All-*trans* retinoic in acute promyelocytic leukemia: Long-term outcome and prognostic factor analysis from the North American Intergroup protocol. Blood 100:4298-4302, 2002

23. Büchner T, Hiddemann W, Wörmann B, et al: Double induction strategy for acute myeloid leukemia: The effects of high-dose cytarabine with mitoxantrone instead of standard-dose cytarabine with daunorubicin and 6-thioguanins—A randomized trial by the German AML Cooperative Group. Blood 93:4116-4124, 1999

24. Kern W, Haferlach T, Schoch C, et al: Early blast clearance during aplasia following induction therapy is a major independent prognostic factor for both achievement of complete remission and long-term outcome in acute myeloid leukemia: Data from the German AML Cooperative Group trial 1992. Blood 98:462a, 2001 (abstr 1932)

25. Estey EH, Kantarjian HM, Cortes JE, et al: Is the absence of circulating blasts a valid criterion for CR in AML? Blood 100:88a, 2002 (abstr 324)

26. Estey EH, Thall PF, Wang X, et al: Effect of circulating blasts at time of complete remission on subsequent relapse-free survival time in newly diagnosed AML. Blood 102:3097-3099, 2003

27. Marcucci G, Archer KJ, Mrózek K, et al: Abnormal karyotype during complete remission (CR) predicts short relapse-free survival (RFS) in acute myeloid leukemia (AML): Results from CALGB 8461. Blood 98:577a, 2001 (abstr 2421)

28. Freireich EJ, Cork A, Stass SA, et al: Cytogenetics for detection of minimal residual disease in acute myeloblastic leukemia. Leukemia 6:500-506, 1992

29. Baer MR, Stewart CC, Dodge RK, et al: High frequency of immunophenotypic changes in acute myeloid leukemia at relapse: Implications for residual disease detection (Cancer and Leukemia Group B study 8361). Blood 97:3574-3580, 2001

30. Diverio D, Rossi V, Avvisati G, et al: Early detection of relapse by prospective reverse transcriptase-polymerase chain reaction analysis of the PML/RAR-alpha fusion gene in patients with acute promyelocytic leukemia enrolled in the GIMEMA-AIEOP multicenter "AIDA" trial. Blood 92:784-789, 1998

31. Jurcic JG, Nimer SD, Scheinberg DA, et al: Prognostic significance of minimal residual disease and PML/RAR-alpha isoform type: Long-term follow-up in acute promyelocytic leukemia. Blood 98:2651-2656, 2001

32. Cilloni D, Gottardi E, De Micheli D, et al: Quantitative assessment of WT1 expression by real time quantitative PCR may be a useful tool for monitoring minimal residual disease in acute leukemia patients. Leukemia 16:2115-2121, 2002

33. Sievers EL, Larson RA, Stadtmauer EA, et al: Efficacy and safety of gemtuzumab ozogamicin in patients with CD33-positive acute myeloid leukemia in first relapse. J Clin Oncol 19:3244-3254, 2001

34. Lopez G, Giles F, Cortes J, et al: Sub-types of resistant disease in patients with AML, RAEB-t, or RAEB who fail initial induction chemotherapy. Blood 98:329a, 2001 (abstr 1388)

35. Estey E, Pierce S: Routine bone marrow exam during first remission of acute myeloid leukemia. Blood 87:3899-3902, 1996

36. Burnett AK, Grimwade D, Solomon E, et al: Presenting white blood cell count and kinetics of molecular remission predict prognosis in acute promyelocytic leukemia treated with all-trans retinoic acid: Result of the Randomized MRC Trial. Blood 93:4131-4143, 1999

37. Gallagher RE, Yeap BY, Bi W, et al: Quantitative real-time RT-PCR analysis of PML-RAR-alpha mRNA in acute promyelocytic leukemia: Assessment of prognostic significance in adult patients from intergroup protocol 0129. Blood 101:2521-2528, 2003

38. Kaplan EL, Meier P: Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457-481, 1958

39. Gooley TA, Leisenring W, Crowley J, et al: Why Kaplan-Meier fails and cumulative index succeeds when estimating failure probabilities in the presence of competing risks, in Crowley J (ed): Handbook of Statistics in Clinical Oncology. New York, NY, Marcel Dekker, 2001, pp 513-524

40. Pepe MS, Mori M: Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data. Stat Med 12:737-751, 1993

41. Gray RJ: A class of k-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat 16:1141-1154, 1988

42. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719-748, 1959

43. Cox DR: Regression models and life-tables. J R Stat Soc B 34:187-220, 1972

44. 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 90:4719-4724, 1997