

Definitions, methodological and statistical issues for phase 3 clinical trials in chronic myeloid leukemia: a proposal by the European LeukemiaNet

Joëlle Guilhot,¹ Michele Baccarani,² Richard E. Clark,³ Francisco Cervantes,⁴ François Guilhot,¹ Andreas Hochhaus,⁵ Sergei Kulikov,⁶ Jiri Mayer,⁷ Andreas L. Petzer,⁸ Gianantonio Rosti,² Philippe Rousselot,⁹ Giuseppe Saglio,¹⁰ Susanne Saussele,¹¹ Bengt Simonsson,^{12,13} Juan-Luis Steegmann,¹⁴ Andrey Zaritskey,¹⁵ and Rüdiger Hehlmann¹¹

¹Inserm CIC 0802, Centre Hospitalier Universitaire de Poitiers, Poitiers, France; ²Department of Hematology and Oncology "L. and A. Seràgnoli," S. Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy; ³Department of Hematology, Royal Liverpool University Hospital, University of Liverpool, Liverpool, United Kingdom; ⁴Hematology Department, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, Barcelona, Spain; ⁵Abteilung Hämatologie/Onkologie, Universitätsklinikum Jena, Jena, Germany; ⁶National Research Center for Hematology, Moscow, Russia; ⁷Department of Internal Medicine, Hemato and Oncology, University Hospital Brno, Central European Institute of Technology, Masaryk University, Brno, Czech Republic; ⁸Internal Medicine I, Hematology and Medical Oncology, Barmherzige Schwestern Hospital Linz, Linz, Austria; ⁹Department of Hematology and Oncology, University of Versailles Saint-Quentin-en-Yvelines, Le Chesnay, France; ¹⁰Department of Cellular Biotechnologies and Hematology, University at Orbassano, Turin, Italy; ¹¹III. Medizinische Klinik, Medizinische Fakultät Mannheim, Universität Heidelberg, Mannheim, Germany; ¹²Department of Medical Sciences, University of Uppsala and ¹³Department of Hematology, University Hospital, Uppsala, Sweden; ¹⁴Hematología, Hospital Universitario de la Princesa, Madrid, Spain; and ¹⁵Institute of Hematology, Almazov Federal Centre of Heart, Blood and Endocrinology, St Petersburg, Russia

The treatment policy of chronic myeloid leukemia (CML), particularly with tyrosine kinase inhibitors, has been influenced by several recent studies that were well designed and rapidly performed, but their interpretation is of some concern because different end points and methodologies were used. To understand and compare the results of the previous and future

studies and to translate their conclusion into clinical practice, there is a need for common definitions and methods for analyses of CML studies. A panel of experts was appointed by the European LeukemiaNet with the aim of developing a set of definitions and recommendations to be used in design, analyses, and reporting of phase 3 clinical trials in this disease.

This paper summarizes the consensus of the panel on events and major end points of interest in CML. It also focuses on specific issues concerning the intention-to-treat principle and longitudinal data analyses in the context of long-term follow-up. The panel proposes that future clinical trials follow these recommendations. (*Blood*. 2012;119(25):5963-5971)

Introduction

Applied statistics are important tools in medical evaluations. The relevance of statistical designs and statistical results in trials is based on concise definitions regarding diagnosis, management, and treatment strategies. The choice of adequate statistical tests depends on the parameters to be analyzed and on specific end points.

After the initial descriptions of chronic myeloid leukemia (CML) more than 160 years ago, little progress was made in its treatment for more than a century.¹ Survival prolongation was first achieved with drugs, such as hydroxyurea.¹ Then, major improvements were obtained with allogeneic hematopoietic stem cell transplantation. First priority at that time was to analyze survival.

The understanding of the pathogenesis of CML began with the discovery of the Philadelphia (Ph) chromosome followed later on by the recognition of its molecular counterpart, the *BCR-ABL* fusion gene. In CML, clinical trials performed during the past 2 to 3 decades have profoundly improved the outcome of patients with CML.² In Europe, the coordinators of national CML Study Groups (European Investigators on CML) joined forces to perform European trials and long-term observations,³ to conduct common meta-analyses,^{4,5} and to elaborate new prognostic scores.^{6,7}

With the advents of IFN- α and of tyrosine kinase inhibitors (TKIs), analyses of treatment response moved more into focus.

Thus, the relationship between responses and survival became of particular interest. The cytogenetic outcome is usually considered as a reliable surrogate marker for survival in CML,^{7,8} whereas the relationship between molecular response and survival is still under investigation.⁹ Molecular response is probably more relevant to the issue of treatment discontinuation without relapse and cure.¹⁰ On the other hand, because of the recent improvement of survival offered by the use of TKIs, other diseases and deaths from causes other than CML progression have become more frequent in CML patients. Analysis of survival may include these new parameters. Quality of life and compliance have also to be considered in the context of this chronic disease with an expected long duration of treatment.

Consequently, statistical methods in addition to Kaplan-Meier (KM)¹¹ analyses and initial Cox regression models¹² should be applied with respect to the intention-to-treat (ITT) principles. Responses to treatment are time-dependent variables needing specific analyses.^{13,14} Surrogate markers should be carefully selected.¹⁵ Adequate censoring and accounting for possible competing risks are critical to obtain unbiased estimates of time-to-event end points. Therefore, we propose a set of common definitions and methods to promote adequate studies and to allow comprehensive reviews and meta-analyses in CML.

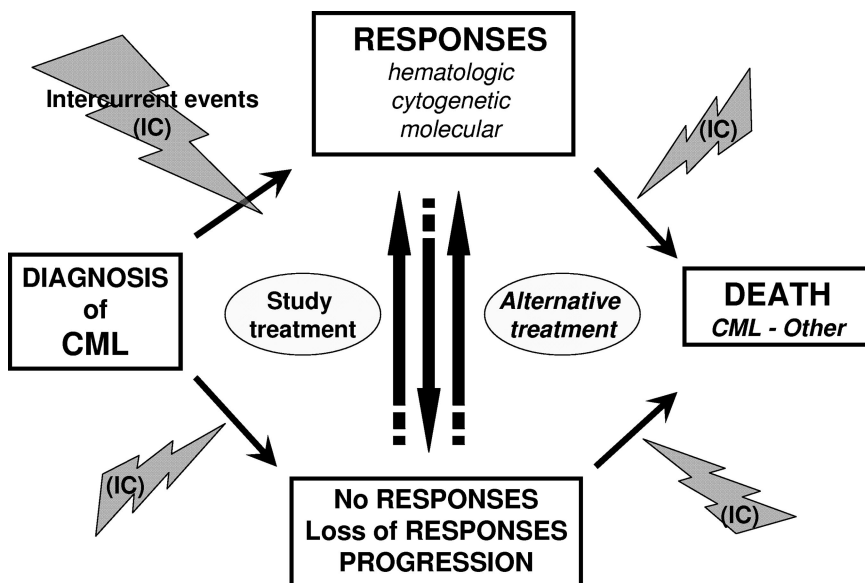


Figure 1. Outcomes and events in CML. Outcomes and events that may potentially occur during the course of CML disease are presented from diagnosis to death. Intercurrent events (IC) pertain to adverse events (ie, toxicities resulting from treatment) or events unrelated with the disease or its treatment.

Methods

The panel of experts that constitute the authors of this article was appointed by the European LeukemiaNet (ELN) and is composed of 17 researchers with well-recognized methodologic and clinical experience in CML. The scope of this publication is to propose a consensus concerning methods of analyses and reporting in phase 3 trials. Because of their peculiarities, other trials are not considered here. The statements of the panel are presented in this report.

Definitions of events in CML

“Events” in statistics relate to favorable or unfavorable outcomes, which both occur in CML patients. These events may be related to the disease, its treatment, or both. In some circumstances, intercurrent events unrelated to CML may also occur. Events are described in “Favorable events,” “Unfavorable events,” “Intercurrent events,” and summarized in Figure 1 and Table 1. To avoid bias in analyses and comparisons between treatment groups, the date of occurrence of events of interest has also to be clearly defined. Considerations about date of event are provided and summarized in Table 2.

Favorable events

As reports of spontaneous remissions are exceptionally rare in CML, favorable events refer to responses to treatment (Table 1). The assessment of the response is based on hematologic, cytogenetic, and molecular considerations, mainly complete hematologic response (CHR), partial (PCgR) and complete cytogenetic response (CCgR), and various levels of molecular response, such as major molecular response (MMR) or undetectable BCR-ABL1 transcript.¹⁶⁻¹⁹ The definitions that were proposed and published by the ELN^{1,20,21} have already been implemented into clinical trials and serve as an example of shared definition of events, although they are clearly subjected to periodic update.

The date of response (hematologic, cytogenetic, or molecular) is the date of the examination documenting the response. Usually, the assessment of response is unique. However, in some protocols, criteria, such as hematologic, cytogenetic, or molecular responses, have to be confirmed by a second analysis. For example, in the

Dasision study, a confirmed CCgR was defined as a CCgR documented at 12 months on 2 consecutive assessments at least 28 days apart.²² In these particular cases, and if the response is confirmed, the date of response to be considered for statistical analyses is the date of the first occurrence of the response, not the date of the confirmation.

Unfavorable events

Unfavorable events relate to lack of efficacy or loss of efficacy; it pertains to no response to treatment, loss of responses (hematologic, cytogenetic, and molecular), and progression to accelerated phase (AP)¹ or blast crisis (BC).¹ Adverse events and death from any causes should also be considered (Table 1).

1. No response and insufficient response. “No response” and “insufficient response” are not events per se. Because it takes time to obtain a response, the time at which “no response” or “insufficient” level of response is considered as a failure must be predefined. It depends on the treatment that is tested and on the

Table 1. Events

Type of events, category	Time point of assessment
Favorable events	
CHR	Any time
CgR	Any time
PCgR	Any time
CCgR	Any time
Molecular response	Any time
Unfavorable events	
No response	
Less than CHR	At predefined time
No CgR (Ph ⁺ > 95%)	At predefined time
Less than PCgR (Ph ⁺ > 35%)	At predefined time
Less than CCgR (Ph ⁺ > 0%)	At predefined time
Loss of CHR	Any time
Loss of CCgR	Any time
Accelerated phase	Any time
Blast crisis	Any time
Adverse events	Any time
Death	Any time (any cause)

CHR indicates complete hematological response; CgR, cytogenetic response; PCgR, partial cytogenetic response; and CCgR, complete cytogenetic response.

Table 2. No response: recommended time points of assessment

No response	Imatinib-treated patients, mo*
Less than CHR	3
No CgR (Ph ⁺ > 95%)	6
Less than PCgR (Ph ⁺ > 35%)	12
Less than CCgR (Ph ⁺ > 0%)	18

The time lines refer to the allocation of study treatment and were published for patients treated by imatinib frontline. For reasons of comparability in a given trial and between trials, these recommendations should also be applied to patients given a treatment other than imatinib.

*Adapted from Baccarani et al.²⁰

study design. For patients treated by imatinib at standard dose or by imatinib-based regimens, definitions were published by the ELN^{1,20,21} in which timelines at 3, 6, 12, and 18 months are recommended; no CHR at 3 months, no cytogenetic response (CgR) at 6 months, less than PCgR at 12 months, and less than CCgR at 18 months are unfavorable events.

For reasons of comparability, these recommended time points can be applied to patients treated by any TKI, not only imatinib, but also second-generation compounds, although it is recognized that the response to these agents is more rapid.⁹

For future protocols, it may be justified to also regard as failure “no MMR” at 12 or 18 months.

2. Loss of response. Loss of response (hematologic, cytogenetic, or molecular) may be recorded at any time during the treatment. In accordance with the failure definitions in Tables 1 and 2, sufficient time must have elapsed to be able to establish a certain response level; in addition, the date of loss of response is the date of the examination documenting the first occurrence of the loss of response.

3. Progression. Progression in this paper relates to AP or BC only, as previously defined by the ELN.¹ This definition was frequently modified.²³ The National Comprehensive Cancer Network has included loss of hematologic or cytogenetic response in the definition. In some studies, such as the IRIS study,²⁴ the definition of progression was broadened, including other criteria, such as an increase in the percentage of Ph⁺ metaphases, if the patients were still in CML chronic phase. Clonal evolution is also an unfavorable event indicating a lack of efficacy. However, the panel still recommends reserving the term “progression” specifically to the transformation of the disease (ie, AP and BC). Progression is the most important event before death because it still predicts (or heralds) death in most patients, with a median survival from progression to death of less than 1 year.

Despite the importance and despite the shared definitions of AP and BC,^{1,20} it is unfortunate that these definitions were frequently modified, without any reported and convincing evidence in support of the modification. Therefore, the panel suggests using another denomination when further events of interest are considered.

4. Adverse events. Adverse events have to be carefully checked in the context of the trial, according to their severity and based on defined rules. Our recommendation is to report these events according to the common and widespread definitions of the US National Cancer Institute²⁵ to allow adequate comparisons between trials.

Relevant additional information can be provided by tools, such as the MD Anderson Symptom Inventory, which was developed to assess patient-reported symptom severity and interference in patients with cancer.²⁶ A CML-specific variant of the MD Anderson Symptom Inventory is in development.²⁷ However, this tool has not yet been validated in all languages. The panel recommends to carefully check and ask for validation of scales before using

them in clinical trials. Serious and severe adverse events are of major interest, and the date of their first occurrence should be their date of event.

Dose reduction and permanent discontinuation of the treatment are usually consequences of adverse events and are thus seen as subsequent clinical measures and not as events by themselves. However, in case of chronic mild toxicities, the exact date of their occurrence is not easily assessable. In such cases, the date of their leading to major modification of the treatment, such as stopping the drug for another TKI, is proposed as the reference date of the event.

5. Death. Death related to CML disease, its treatment, or both, or death from other causes, are all unfavorable events. Death is the most important event and can be measured better than any other event. Only for particular calculations, deaths can be divided according to the cause, either the leukemia, or its treatment, or other, provided that the death reports are well detailed.

Intercurrent events

Other diseases or injuries may occur more frequently because the survival in CML was highly improved with the use of TKIs. However, it is usually not possible to show the independence between comorbidities and CML. Hence, it is recommended to consider such events as related to CML, unless robust evidence such as death resulting from natural disasters is demonstrated.

Definitions of end points

End point definitions are a key issue. In many CML trials, such as the French SPIRIT¹⁸ trial, estimated rates of responses that are provided at specific time points are usually lower than estimated rates provided by cumulative incidence analysis of responses up to this time point. In addition, different long-term definitions may result in perceived differences in outcomes, as recently demonstrated by Kantarjian et al.²⁸

The panel aims to provide relevant definitions in the following statements: to assess outcomes in CML, the study population must be analyzed at specific time points (cross-sectional analysis) and over defined periods of time (longitudinal analysis). Both methods of analyses are necessary and add information to each other. Major end points are described in the next sections and summarized in Table 3.

Response at specific time points

Time points of interest are dependent on the treatment that is tested and on the study design. For patients treated by imatinib, the following key issues in CML are recommended.

The achievement of CHR should be assessed after the first 3 months of treatment.

Key comparisons regarding cytogenetic responses are as follows: at 6 months, no CgR versus any CgR; at 12 months, less than PCgR versus PCgR versus CCgR and CCgR versus others; and at 18 months, less than PCgR versus PCgR versus CCgR, PCgR versus CCgR, and CCgR versus others.

For patients who have been treated with allogeneic hematopoietic stem cell transplantation or IFN, the proportion of patients who are in CCgR at 12 and 18 months, respectively, is the most solid and confirmed early surrogate marker of any measure of survival.³ For 18 months, this result has recently been confirmed for imatinib-treated patients, too.⁷

Table 3. End points in longitudinal analyses

Events	PFS	FFS	EFS	OS	ATFS
No responses, or insufficient response to study therapy, after a predefined period of time*		Yes	Yes		
Loss of responses to study treatment		Yes	Yes		
Progression (AP/BC)	Yes	Yes	Yes		
Adverse events leading to permanent discontinuation of study therapy (any grade)			Yes		
Treatment change†					Yes
Deaths	Yes	Yes	Yes	Yes	Yes

*Time point of assessment are defined in Table 1.

†Discontinuation of treatment for patients in molecular response is not considered as a "treatment change" in this analysis.

For patients treated by second-generation TKIs, early end points are currently discussed, as stated in "No response and insufficient response." Of note, concerning hematologic and cytogenetic response end points, it is not the definition of the response that is under exploration. It is the relevant date of assessment and consequently the time to response.

Again, the panel recommends keeping definitions of end points that were previously defined for imatinib, in addition to the new ones when comparisons between trials are expected.

The prognostic relevance of molecular response is also documented. However, both the optimum timing and the cut-off of the response are still a matter of investigation.

For patients treated with imatinib, the proportion of patients who are in MMR at 12 and 18 months is an important candidate for an early marker of late outcome. Strong evidence for MMR as a marker for survival comes from the German CML study IV.¹⁶

However, a recent update of the IRIS trial has shown no significant impact of MMR on overall survival (OS), but a slight statistically significant benefit in an event-free survival (EFS) analysis when MMR was assessed by 18 months.²⁹ In addition, several other studies have failed to show any advantage in response duration and long-term outcome in favor of patients who achieved an MMR at 12 or 18 months compared with those who achieved only a CCgR at different time points. It has been reported for imatinib-treated patients^{30,31} as well as for treatment with second-generation TKIs.⁹

Concerning the cut-off level, a ratio of BCR-ABL/ABL1 transcript less than or equal to 0.01% is currently under investigation. Reaching an undetectable transcript level is also a key end point; however, the definition of complete molecular response is still controversial.

Of note, the assessment of the ratio BCR-ABL at 3 months could predict outcomes of patients in first-line imatinib therapy³² or in second-line therapy with other TKIs.³³ Similarly, a recent subanalysis of the United Kingdom SPIRIT 2 study documented the predictive value of early molecular response on subsequent cytogenetic and molecular responses in patients treated with dasatinib front line.³⁴ However, the question whether it translates into better late outcomes and survival is not yet addressed. Consequently, at present, the panel can only recommend to carefully perform molecular monitoring at regular intervals to allow further investigations.

Analyses of responses over a period of time

These analyses are usually presented as "cumulative incidence of responses." The cumulative incidence of responses over a period of time, such as CCgR and MMR, are of interest to assess the efficacy of a drug. It describes the probabilities of a patient's first achievement of a certain level of response over time. These analyses are useful to estimate the proportion of responding

patients over time and to show the velocity of response to treatment. However, the cumulative incidence of response is not sufficient to judge the efficacy of the study treatment because it does not measure the response rate at a specific time point. The main reason is that relapses are not taken into account. For example, the cumulative incidence rate of MMR at 18 months does not indicate how many patients who reached an MMR before 18 months were still in MMR at 18 months. Consequently, the rate of responding patients, which is estimated by the cumulative incidence method at 18 months, may be more than the rate for patients still in response at 18 months. The panel members recommend reporting the duration of responses in addition to such analyses and to clearly state that it does not represent the response rate at any time point.

PFS and TTP. Progression to AP/BC signals the metamorphosis of the disease; it is usually not reversible unless a major treatment change is undertaken. Thus, it stands for a definitive failure of the current treatment. Soon after progression has started, it will be noticed by patients and physicians, and the date of its recording will be close to the date of its actual beginning. Consequently, this parameter is less sensitive to bias than remission parameters.

As indicated by the term "survival," progression-free survival (PFS) composes not only the events "AP" and "BC" but also "death." Up to now, it was recommended to also consider unrelated death as an event and consequently to use the PFS method, which includes deaths whatever their causes, according to guidelines of regulatory authorities. Within the era of TKIs, it turns out that the probability of BC is decreasing and the life expectancy of patients is considerably improving as a result of the increasing proportion of patients achieving major and complete molecular response. Thus, the frequency of unrelated CML deaths is increasing, too. As a consequence, deaths that are indisputably unrelated to CML or to its therapy might be regarded as competing events in the future but at present, the panel still recommends using the PFS definition, including death for any reason.

In the Food and Drug Administration Guidance for Industry,³⁵ time to progression (TTP) is defined as the time from randomization to objective tumor progression. TTP does not include deaths. As stated by Fleming et al, TTP is an estimate in the hypothetical setting in which patients are not at risk of death from any cause other than disease progression because patients who died without progression are censored.³⁶ Such analysis of TTP is sensitive to bias, and its results are difficult to interpret.

Indeed, the TTP is an important issue. However, the method of censoring that is used in this document is not recommended in CML by the panel. Instead, methods taking into account competing risks, as described in "How to handle analyses" should be considered.

FFS. Considering that "failure" focuses on responses to study treatment as defined by the ELN criteria, the events to be

considered here are as follows: no response according to the ELN definition and recommended timelines at 3, 6, 12, and 18 months (Tables 1 and 2); loss of responses; AP or BC at any time; and death at any time. If a patient experiences several successive events, the date of failure is the date of the first of these events.

Within this definition, patients who are considered as failure in such analysis are (1) patients experiencing primary and secondary resistance to study treatment, including progression; and (2) patients who did not achieve response, lost response, or who progressed to AP/BC because of dose reduction or transient discontinuation of the study treatment related to toxicity or poor adherence.

Some other patients may have to switch to alternative therapy because of toxicities whatever the severity, even though they were responding to study treatment. In this context, the adverse event should be considered as a competing risk as described in “How to handle analyses.”

The occurrences of adverse events are obviously not in favor of the study treatment. However, they are not considered as events of interest per se in this definition of “failure-free survival” (FFS). The reasons are that (1) adverse event does not necessarily mean “no response”; and (2) the severity of the adverse event does not necessarily correlate with dose intensity that the patient received before the assessment of the response. For example, even mild chronic toxicities may lead to poor compliance. Consequently, it is important to perform longitudinal analyses, including responses and duration of responses only, independently of the dose of treatment.

EFS. In an EFS analysis, the same events as for FFS analysis, plus drug discontinuation because of adverse events are considered, whichever comes first. The aim of this end point, which includes efficacy and toxicity considerations, is to assess the whole usefulness of the study treatment. However, the more events are considered in such composite end points, the more potential bias may arise when statistical analyses are performed. Therefore, adequate monitoring is mandatory, as for other end points, and could become a critical issue based on the number of parameters that have to be collected.

Survival. OS. With regard to OS, events are defined by death from any cause. As far as survival is concerned, OS remains the most important endpoint. As previously stated, the life expectancy of the CML patients is considerably improving and the frequency of potentially unrelated CML deaths is increasing. Consequently, other approaches are also discussed.

Disease-specific survival and CML mortality rate analysis. As it is usually not possible to fairly distinguish between death resulting from comorbidities and death resulting from CML or its treatment, such analyses are not recommended. (Suicide would be an example.)

Relative survival models. Relative survival attempts to separate mortality from the disease of interest from mortality resulting from all other causes. To do this, the ratio of the observed (all-cause) survival in the cohort of interest and the expected survival in a similar group in the general population is calculated as follows³⁷: Relative survival = [observed (all-cause) survival in cohort studied]/[expected survival based on rates in a comparator population].

The cohort of interest is the sample of persons with CML. The comparator group is obtained from routine data, matched to the cohort of interest by age, sex, and other potentially important covariates. Among them, age distribution is an important factor for any calculation, including survival, because, in many clinical trials, age is lower than in the general population. In a relative survival model, the observed mortality rate within the cohort of interest is

made up of the background mortality rate in the general comparator population (ie, deaths from all causes) plus the excess mortality rate associated with the condition of interest (ie, more additional deaths resulting from CML). When the comparator population is available, this method offers a useful tool for additional analyses.

Alternative treatment-free survival (ATFS). Discontinuation of the study treatment by patients still in chronic phase, in accelerated or in blastic phase, is a key issue. In clinical trials, the reasons for discontinuation mainly relate to objective failure or side effects. However, it may also relate to subjective inclinations, such as lack of compliance or unsatisfactory effects combining mild toxicities and suboptimal responses. In such situations, the panel strongly recommends not including subjective events in a failure or in an EFS analysis. Definition, assessment, and date of such events are not reliable; the date of discontinuation of the study treatment may also be questionable because of transient discontinuation.

In such a situation, the panel considers that the ATFS method, as documented by Zackova et al,³⁸ is a reasonable approach to assess the usefulness of a study treatment. Some reasons for switching to an alternative treatment may be subjective in some patients, but the switch per se is a robust marker; the study treatment is replaced, and the date of switch is easily assessable. As indicated by the term “survival,” alternative treatment-free survival includes “death” also.

Of note, discontinuation of treatment for patients with sustained undetectable BCR-ABL transcripts has been recently proposed.¹⁰ It could be considered a goal in future protocols. However, the discontinuation of therapy should not be considered as an event in all previous analyses. The favorable event is the achievement of the molecularly undetectable disease and then, if present, the unfavorable event would be the molecular relapse.

Quality of life

Health-related quality of life has become an issue of major interest in CML. It was recently analyzed and published by Efficace et al.³⁹ As stated by the authors, the limitation of their study was the lack of internationally validated measures for CML patients. A cross-cultural initial development of an EORTC health-related quality of life questionnaire for patients with CML is ongoing.⁴⁰

How to handle analyses

The importance of the ITT analysis

ITT and per-protocol principles include the following statements:

The ITT analysis of a clinical trial is defined as the analysis of all enrolled patients (ie, the ITT population), in accordance with the treatment group to which they were prospectively assigned. The panel agreed that ITT analysis should be performed in all randomized clinical trials but also in nonrandomized prospective CML studies.

ITT analysis in CML has the following consequences: (1) with regard to time-to-event end points, the time at risk of a patient should not be censored because he/she did not receive or stopped receiving the study treatment assigned at enrollment; and (2) furthermore, the time at risk of a patient should not be censored because of a minor event before a time to event end point, which was defined by a more serious event.

Keeping all prospectively enrolled patients in the trial, ITT analysis is the only safe way to deal with protocol violations, such as noncompliance to study treatment intake or the reception of treatment(s) other than the assigned one.⁴¹ Any other handling of

protocol violation will depend on subjective decision and is thus likely to introduce bias.

Per-protocol analyses include the following: (1) Analyses on the ITT population taking into account the treatment the patients received. Cross-sectional and longitudinal analyses of patients on study treatment add information to the analyses based on the ITT principles. However, the panel stressed that these analyses are valid only if all enrolled patients are considered in the estimations. For example, the denominator of any calculation, such as the proportion of patient in CCgR at 18 months and still under study treatment, should be the total number of patients who were assigned to receive the study treatment. (2) Analyses on subgroup of interest. The design of additional analyses may vary according to studies; however, the panel reminds that subgroup analyses are highly sensitive to bias. Therefore, results from subgroup analyses should not be accepted unless they were planned in advance.

The panel stressed that key end points should be analyzed and reported on the ITT principle first, and then per-protocol, on the ITT population. Relevant ITT and per-protocol end points in CML are described in Table 3.

If for a particular study, analyses in accordance with the ITT principle are considered as insufficient or inappropriate, a justification is expected at the planning stage and not a posteriori.

Prognostic score

At diagnosis, it is important to collect all data relevant to calculate the CML risk scores (Sokal,^{42,43} Euro,⁶ and EUTOS scores⁷) as they have an important impact on the outcome of CML patients. Analyses stratified and/or adjusted for risk score support the understanding and interpretation of trial results. Accordingly, these analyses are recommended. The EUTOS score⁷ was the only one developed in a sample of patients treated with TKIs, and it is suggested to put it into focus when CCgR at 18 months and PFS are analyzed.

Competing risks

Patients may experience an event that either precludes the occurrence of the event under investigation or fundamentally alters the probability of occurrence of the event of interest. Events influencing the probabilities of observing the event of interest are known as competing risk events.

As far as the analysis of (multiple) influence variables is concerned, there are 2 analytic main options available for addressing competing risks: the cause-specific hazard and the subdistribution hazard.⁴⁴⁻⁴⁶ The cumulative incidence function (CIF) graphically displays probabilities influenced by competing risks. Detailed descriptions and discussions of these methods are beyond the scope of this report. However, it has to be mentioned that both approaches should be applied to the data.

Missing values and unbalanced assessment within groups

Missing values in clinical trials are a critical issue. Some missing values are intermittent, and others relate to dropout from the study. Consequently, minimizing the subject attrition is a challenge when collecting longitudinal data in prospective studies. If the nonrandom attrition of study subjects is closely tied to the patterns or outcomes that are the object of the investigation, bias of unknown sign and size may be introduced.⁴⁷⁻⁴⁹

Each protocol should clearly define in the statistical plan how missing values will be handled in analyses, and the documentation of the reasons of missing values is essential. As previously stated, adequate reference dates are crucial for relevant analyses. Except for death, the reference date for the continuation of an event-free

status is not automatically given by the date of the last contact unless a relevant examination of the patient under investigation was performed. For example, let us assume that a patient achieves a CCgR at month 12 and that this result was confirmed by a new cytogenetic test at month 18. At month 24, no cytogenetic analysis was performed and the molecular test that was performed was not evaluable. The relevant assessment is that the patient was alive at month 24, but it cannot be stated that he was still in CCgR at month 24.

In addition, evaluations linked to the end points of the trial need to be closely monitored. It is important to keep in mind that the precision regarding the estimation of the time to events (favorable or unfavorable) and the duration of responses are dependent on the number of assessments that are performed.

Multiple testing

In case of multiple reassessments or comparisons, the adjustment of the *P* value derived from statistical tests must be considered. This relates in particular to confirmatory hypotheses testing. The reason is that there is an increased finding of false significance because of chance when multiple outcomes measures are used and when multiple simultaneous hypotheses are tested. To accommodate for this, methods are proposed to adjust the *P* value of individual tests upward to ensure that the overall risk of finding significant differences remains equal to .05. In CML, this applies, for example, to trials with more than 2 treatment arms and when all possible paired arm comparisons are considered; it also applies when differences in outcome, such as PFS or OS, are reanalyzed at different periods of time.

To accommodate multiple hypotheses testing, a hierarchical ordering of the hypotheses to be tested could be prospectively determined in the study protocol. Further methods are the Bonferroni correction and/or the closed test procedure.^{50,51} For repeated analyses, an option to keep the significance level is a group sequential design.⁵²

Analyses at specific time points

The statistical analysis should be performed when all subjects enrolled in the study have reached the required follow-up required by the study design. No patients should be censored because of insufficient follow-up. Nonassessable cases, whatever the reason, must be presented and adequately taken into account in the statistical analyses. Estimates should be presented with confidence intervals and, when relevant, with adjustment on validated prognostic variables. Analyses at specific time points should be performed with respect to the ITT principle first, and then per-protocol, on the ITT population (Table 4).

In accordance with the ITT analysis principle, a patient who is assigned to receive imatinib in a trial, had unsatisfactory response and was subsequently switched to a second-line therapy leading to achieving the desired response level, has to be considered as a responding patient on the whole.

As previously stated, this is the method of choice to deal with some potential protocol violations when treatment arms are compared. In addition, some patients responding to a second-line therapy may have benefitted from the previous treatment that was under investigation. However, if a patient who had failed the treatment under investigation is switched to another treatment and is responding, the response is up to the second-line treatment in most of the cases. A typical example was illustrated by the IRIS trial.²⁴ Of the 553 patients originally assigned to IFN- α plus

Table 4. ITT and per-protocol analyses

	PFS	FFS	EFS	OS	ATFS
ITT principle (ITT population, whatever the treatment)	+++	–	–	+++	–
Per-protocol analyses					
Within the ITT population, and according to study therapy	Optional	+	+	Optional	+
Within subgroups			Optional		

Optional indicates that these analyses are not recommended by ELN (if these analyses are required in some studies, reasons and definitions have to be explicitly provided) + + +, highly recommended; +, to be considered; and –, not relevant, based on ELN definitions.

cytarabine, 65% crossed over to imatinib and improvement of the rate of hematologic and cytogenetic responses was observed thereafter.⁵³

Consequently, adequate per-protocol analyses are also strongly recommended to provide estimates of responses achieved by the treatment that was initially allocated. These analyses should be performed on the ITT population.

Additional analyses of subgroups are optional and should be presented with caution, as previously stated.

Longitudinal analyses: PFS, FFS, EFS, ATFS, and OS

The probabilities of these time-to-event end points are usually described using the KM method¹¹ and compared between treatment groups through the log-rank test. It should be stressed that censoring is only permissible when it is strictly independent of the event that is considered. To illustrate this issue, imagine a patient whose PFS is under investigation. Events defined for this endpoint are AP, BC, and death. If the patient loses response under the study treatment and is switched to an alternative therapy before one of the events linked to PFS occurs, his PFS time may not be censored at the time of the loss of response. From a clinical viewpoint, if a patient fails the initial treatment, the test treatment, then is switched to another treatment, progresses, and dies, these events are a consequence of the initial treatment and must be counted in a PFS analysis. In addition, the ITT principle avoids the potential bias link to subjective decision for switching to another treatment when different arms are compared.

Both end points, OS and PFS, allow analyses on the ITT principle and per-protocol. The panel recommends analyzing these end points on the ITT principle first. If investigators are interested in the influence of multiple variables on PFS or OS, the application of Cox regression models is suitable in many cases.¹²

Other analyses are optional (Table 3). The FFS and EFS end points, as proposed by ELN, take into account the study treatment in the definition of some events. Consequently, these are per-protocol analyses only. The panel recommends performing these per-protocol analyses on the ITT population. Unfortunately, composite end points harbor methodologic problems. The more events are considered, the more potential bias may arise. It has to be made sure that the evaluations are performed at regular and similar intervals in all groups of treatments that are compared. Differences in documentation are particularly problematic when different studies are compared. In any case, endpoint definitions are a key issue. As recently demonstrated by Kantarjian et al, different long-term definitions may result in perceived differences in outcomes.²⁸ If, for a particular study, definitions proposed by ELN have to be modified, the panel strongly recommends to state it and to document these definitions to avoid inadequate meta-analyses of trials.

Longitudinal analyses: analyses of response over time

Instead of the KM method, the application of the CIF needs to be applied.^{44-46,54} In CML, time-to-response analyses are a typical situation in which competing risks may be present. For instance, let us consider time from randomization to a first CCgR. A competing risk to this event of interest is death before a (possible) observation of a first CCgR. It was repeatedly observed that the analyses of the probabilities of the achievement of a first CCgR over time were performed by calculating reversed KM curves. “Death before a possible observation of a first CCgR” was censored at the time it occurred. However, it is a prerequisite of the application of the KM method that censoring is “noninformative.” This means that censoring carries no information on the subsequent probabilities of the event of interest. It is obvious that, in the situation of death, the prerequisite of noninformative censoring is not met. Whereas noninformative censoring implies that a first CCgR may be observed in the future (after a longer follow-up), death reduces the probability of a later CCgR to zero and is thus informative. For example, a patient in PCgR who died from a road accident after 6 months of treatment has definitely no chance to reach a CCgR, whereas a patient in PCgR with only 6 months of follow-up and alive still has this opportunity. In the presence of competing risk, the application of the KM method is incorrect.

The probabilities of the achievement of a first CCgR should be calculated by a CIF instead.⁴⁴⁻⁴⁶ With relation to the achievement of a first CCgR, Pfirrmann et al presented an example that demonstrates the overestimation bias if the probabilities of response achievement over time are calculated through the inappropriate KM approach.⁴⁶ Their paper also provides the opportunity to follow the CIF calculation of the correct probabilities. For the comparison of CIF estimates between 2 or more patient groups, the Gray test⁵⁵ may be applied. The use of the log-rank test would be inadequate. Further questions and problems connected with CIF estimations⁴⁴⁻⁴⁶ as well as a software recommendation^{56,57} were previously addressed.

In any case, the panel stressed that competing risks, if present, have to be adequately defined with regard to the respective ITT and per-protocol principles. In addition, the panel strongly recommends not limiting analyses to cumulative incidence of response. The analyses of response at specific time points and the analyses of responses over time should always be both presented, as they add information to each other.

Conclusion

The present expert recommendations summarize and aim to clarify current definitions and end points in CML. End points and analyses are discussed in the context of unfavorable and favorable events statistical analyses. The panel strongly recommends analyses at specific time points in addition to longitudinal analyses. Both

approaches add information to each other. Key analyses should be carried out according to the ITT principle and on the ITT population when per-protocol analyses are needed. The pragmatic interpretation of the OS and PFS probabilities is mainly proposed in the context of the allocated first-line therapy and whatever subsequent therapies were used.⁵⁸ The analysis of cumulative response should rather be understood as a supportive tool to illustrate the velocity of a first achievement of a certain level of response over time. As the life expectancy of the CML patients is considerably improving, modeling in the presence of competing risk events is considered, such as relative survival. The panel recommends following the recommendations proposed in this manuscript for analyzing and reporting results in future clinical trials in CML. Although the panel considers these recommendations as essential for analyzing current trials, it is acknowledged that they could be revised in the future. The time point at which the end points of interest are evaluated is one aspect that may change in the future. However, the methods described in this paper for the analyses of the end points are still applicable for the analysis of such subsequent changes.

References

- Baccarani M, Saglio G, Goldman J, et al. Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood*. 2006;108(6):1809-1820.
- Hehlmann R, Hochhaus A, Baccarani M, et al. Chronic myeloid leukaemia. *Lancet*. 2007;370(9584):342-350.
- Bonifazi F, de Vivo A, Rosti G, et al. Chronic myeloid leukemia and interferon-alpha: a study of complete cytogenetic responders. *Blood*. 2001;98(10):3074-3081.
- Chronic Myeloid Leukemia Trialists' Collaborative Group. Interferon alfa versus chemotherapy for chronic myeloid leukemia: a meta-analysis of seven randomized trials. *J Natl Cancer Inst*. 1997;89(21):1616-1620.
- Chronic Myeloid Leukemia Trialists' Collaborative Group. Hydroxyurea versus busulphan for chronic myeloid leukaemia: an individual patient data meta-analysis of three randomized trials. *Br J Haematol*. 2000;110(3):573-576.
- Hasford J, Pfirrmann M, Hehlmann R, et al. A new prognostic score for survival of patients with chronic myeloid leukemia treated with interferon alfa: Writing Committee for the Collaborative CML Prognostic Factors Project Group. *J Natl Cancer Inst*. 1998;90(11):850-858.
- Hasford J, Baccarani B, Hoffmann V, et al. Predicting complete cytogenetic response and subsequent progression-free survival in 2060 patients with CML on imatinib treatment: the EUTOS score. *Blood*. 2011;118(3):686-692.
- Hasford J, Pfirrmann M, Shepherd P, et al. The impact of the combination of baseline risk group and cytogenetic response on the survival of patients with chronic myeloid leukemia treated with interferon alpha. *Haematologica*. 2005;90(3):335-340.
- Jabbour E, Kantarjian HM, O'Brien S, et al. Front-line therapy with second-generation tyrosine kinase inhibitors in patients with early chronic phase chronic myeloid leukemia: what is the optimal response? *J Clin Oncol*. 2011;29(32):4260-4265.
- Mahon FX, Réa D, Guilhot J, et al. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. *Lancet Oncol*. 2010;11(11):1029-1035.
- Kaplan EL, Meier P. Non parametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457-481.
- Cox DR, Oakes D. *Analysis of Survival Data*. London, United Kingdom: Chapman and Hall; 1984.
- Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Process*. New York, NY: Springer; 1993.
- Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983;1(11):710-719.
- Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605-613.
- Hehlmann R, Lausker M, Jung-Munkwitz S, et al. Tolerability-adapted imatinib 800 mg/d versus 400 mg/d versus 400 mg/d plus interferon- α in newly diagnosed chronic myeloid leukemia. *J Clin Oncol*. 2011;29(12):1634-1642.
- Simonsson B, Gedde-Dahl T, Markev m B, et al. Combination of pegylated interferon- α 2b with imatinib increases molecular response rates in patients with low or intermediate risk chronic myeloid leukemia. *Blood*. 2011;118(12):3228-3235.
- Preudhomme C, Guilhot J, Nicolini FE, et al. Imatinib plus pegylated interferon-alpha2a in chronic myeloid leukemia. *N Engl J Med*. 2010;363(26):2511-2521.
- Jabbour E, Saglio G, Hughes TP, Kantarjian H. Suboptimal responses in chronic myeloid leukemia: implications and management strategies. *Cancer*. 2012;118(5):1181-1191.
- Baccarani M, Cortes J, Pane F, et al. Chronic myeloid leukemia: an update of concepts and management recommendations of European LeukemiaNet. *J Clin Oncol*. 2009;27(35):6041-6051.
- Baccarani M, Castagnetti F, Gugliotta G, et al. Response definitions and European LeukemiaNet Management recommendations. *Best Pract Res Clin Haematol*. 2009;22(3):331-341.
- Kantarjian H, Shah NP, Hochhaus A, et al. Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med*. 2010;362(24):2260-2270.
- Ansstas G, Vij R. Evolution of definitions of response, progression-free survival, and event-free survival in front-line studies of chronic myeloid leukemia [published online ahead of print, February 13, 2012]. *Leuk Lymphoma*. doi: 10.3109/10428194.2011.653786.
- O'Brien SG, Guilhot F, Larson RA, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med*. 2003;348(11):994-1004.
- U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE), Version v4.03. http://eversusnci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14/QuickReference_5x7.pdf. Accessed June 14, 2010.
- Cleeland CS, Mendoza TR, Wang XS, et al. Assessing symptom distress in cancer patients: the M. D. Anderson Symptom Inventory. *Cancer*. 2000;89(7):1634-1646.
- Pinilla-Ibarz J, Cortes J, Mauro MJ. Intolerance to tyrosine kinase inhibitors in chronic myeloid leukemia: definitions and clinical implications. *Cancer*. 2011;117(4):688-697.
- Kantarjian H, O'Brien S, Jabbour E, et al. Impact of treatment end point definitions on perceived differences in long-term outcome with tyrosine kinase inhibitor therapy in chronic myeloid leukemia. *J Clin Oncol*. 2011;29(23):3173-3178.
- Hughes TP, Hochhaus A, Branford S, et al. Long-term prognostic significance of early molecular response to imatinib in newly diagnosed chronic myeloid leukemia: an analysis from the international randomized study of interferon versus ST1571 (IRIS). *Blood*. 2010;116(19):3758-3765.
- Kantarjian HM, Talpaz M, O'Brien S, et al. Survival benefit with imatinib mesylate versus interferon-alpha-based regimens in newly diagnosed chronic-phase chronic myelogenous leukemia. *Blood*. 2006;108(6):1835-1840.
- Marin D, Milojkovic D, Olavarria E, et al. European LeukemiaNet criteria for failure or sub-optimal response reliably identify patients with CML in early chronic phase treated with imatinib whose eventual outcome is poor. *Blood*. 2008;112(12):4437-4444.
- Marin D, Ibrahim AR, Lucas C, et al. Assessment of BCR-ABL1 transcript levels at 3 months is the only requirement for predicting outcome for patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors. *J Clin Oncol*. 2012;30(3):232-238.
- Milojkovic D, Apperley JF, Gerrard G, et al. Responses to second-line tyrosine kinase inhibitors are durable: an intention-to-treat analysis in chronic myeloid leukemia patients. *Blood*. 2012;119(8):1838-1843.
- Marin D, Hedgley C, Clak RE, et al. The predictive value of early molecular response in chronic

Acknowledgments

The European LeukemiaNet is supported by the European Union, Sixth Framework Program (contract LSHC-CT-2004-503216).

Authorship

Contribution: J.G. coordinated the work and wrote the manuscript; and M.B., R.E.C., F.C., F.G., A.H., S.K., J.M., A.L.P., G.R., P.R., G.S., S.S., B.S., J.-L.S., A.Z., and R.H. fully participated in this consensus paper and revised the manuscript.

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

Correspondence: Jo lle Guilhot, Inserm CIC 0802, University Hospital of Poitiers, 2 rue de la Mil trie, 86021 Poitiers, France; e-mail: j.guilhot@chu-poitiers.fr.

- phase CML patients treated with dasatinib first-line therapy. *Blood*. 2011;118(21):785.
35. FDA: Clinical trial endpoints for the approval of cancer drugs and biologics: U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Clinical/Medical. <http://www.fda.gov/downloads/drugs/GuidanceComplianceRegulatoryInformation/Guidance/UCM071590.pdf>. Accessed May 2007.
 36. Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol*. 2009;27(17):2874-2888.
 37. Lambert PC, Dickman PW, Nelson CP, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Stat Med*. 2010;29(7):885-895.
 38. Zackova D, Klamova H, Dusek L, et al. Imatinib as the first-line treatment of patients with chronic myeloid leukemia diagnosed in the chronic phase: can we compare real life data to the results from clinical trials? *Am J Hematol*. 2011;86(3):318-321.
 39. Efficace F, Baccarani M, Breccia M, et al. Health-related quality of life in chronic myeloid leukemia patients receiving long-term therapy with imatinib compared with the general population. *Blood*. 2011;118(17):4554-4560.
 40. Efficace F, Breccia M, Saussele S, et al International development of an EORTC measure to assess patient-reported quality of life (QoL) and symptoms in chronic myeloid leukemia (CML). *Blood*. 2011;118(21):3132.
 41. Altman DG. *Practical Statistics for Medical Research*. London, United Kingdom: Chapman & Hall; 1991.
 42. Sokal JE, Cox EB, Baccarani M, et al. Prognostic discrimination in "good-risk" chronic granulocytic leukemia. *Blood*. 1984;63(4):789-799.
 43. Sokal JE, Baccarani M, Tura S, et al. Prognostic discrimination among younger patients with chronic granulocytic leukemia: relevance to bone marrow transplantation. *Blood*. 1985;66(6):1352-1357.
 44. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*. 1999;18(6):695-706.
 45. Varadhan R, Weiss CO, Segal JB, Wu AW, Scharfstein D, Boyd C. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Med Care*. 2010;48(6 suppl):S96-S105.
 46. Pffirmann M, Hochhaus A, Lauseker M, Sauflele S, Hehlmann R, Hasford J. Recommendations to meet statistical challenges arising from endpoints beyond overall survival in clinical trials on chronic myeloid leukemia. *Leukemia*. 2011;25(9):1433-1438.
 47. Brown CH. Protecting against non-randomly missing data in longitudinal studies. *Biometrics*. 1990;46(1):143-157.
 48. Diggle BP, Kenward MG. Informative drop-out in longitudinal data analysis. *Appl Stat*. 1994;43:49-72.
 49. Ziogas DC, Zintaras E. Analysis of the quality of reporting of randomized controlled trials in acute and chronic myeloid leukemia, and myelodysplastic syndromes as governed by the CONSORT statement. *Ann Epidemiol*. 2009;19(7):494-500.
 50. Bauer P. Multiple testing in clinical trials. *Stat Med*. 1991;10(6):871-889; discussion 889-890.
 51. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655-660.
 52. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549-556.
 53. Guilhot F, Druker B, Larson RA, et al. High rates of durable response are achieved with imatinib after treatment with interferon alpha plus cytarabine: results from the International Randomized Study of Interferon and ST1571 (IRIS) trial. *Haematologica*. 2009;94(12):1669-1675.
 54. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, NY: Wiley; 1980:167-171.
 55. Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat*. 1988;16(3):1141-1154.
 56. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26(11):2389-2430.
 57. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplant*. 2007;40(4):381-387.
 58. Korn EL, Freidlin B, Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol*. 2011;29(17):2439-2442.