

Guidelines for time-to-event end point definitions in sarcomas and gastrointestinal stromal tumors (GIST) trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials)[†]

C. A. Bellera^{1,2*}, N. Penel³, M. Ouali^{4,5}, S. Bonvalot⁶, P. G. Casali⁷, O. S. Nielsen⁸, M. Delannes⁹, S. Litière⁴, F. Bonnetain¹⁰, T. S. Dabakuyo¹¹, R. S. Benjamin¹², J.-Y. Blay^{13,14,15}, B. N. Bui¹⁶, F. Collin¹⁷, T. F. Delaney¹⁸, F. Duffaud¹⁹, T. Filleron⁵, M. Fiore²⁰, H. Gelderblom²¹, S. George²², R. Grimer²³, P. Grosclaude²⁴, A. Gronchi²⁰, R. Haas²⁵, P. Hohenberger²⁶, R. Issels^{27,28}, A. Italiano¹⁶, V. Jooste²⁹, A. Krarup-Hansen³⁰, C. Le Péchoux³¹, C. Mussi³², O. Oberlin³³, S. Patel¹², S. Piperno-Neumann³⁴, C. Raut³⁵, I. Ray-Coquard¹³, P. Rutkowski³⁶, S. Schuetze³⁷, S. Sleijfer³⁸, E. Stoeckle³⁹, M. Van Glabbeke⁴, P. Woll⁴⁰, S. Gourgou-Bourgade⁴¹ & S. Mathoulin-Pélissier^{1,2}

¹Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Centre, Bordeaux; ²Clinical Epidemiology Unit, INSERM CIC 14.01 (Clinical Epidemiology), Bordeaux; ³Department of Medical Oncology, Centre Oscar Lambret, Comprehensive Cancer Centre, Lille, France; ⁴Department of Biostatistics, European Organisation for Research and Treatment of Cancer, Brussels, Belgium; ⁵Biostatistics Unit, Institut Claudius Regaud, Comprehensive Cancer Centre, Toulouse; ⁶Department of Surgery, Institut Gustave Roussy, Comprehensive Cancer Centre, Villejuif, France; ⁷Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; ⁸Faculty of Health Sciences, Aarhus University, Aarhus, Denmark; ⁹Department of Radiotherapy, Institut Claudius Regaud, Comprehensive Cancer Center, Toulouse; ¹⁰Methodological and Quality of Life Unit in Oncology (EA3181), CHU Besançon, Besançon; ¹¹Biostatistics and Quality of Life Unit (EA4184), Centre Georges-François Leclerc, Comprehensive Cancer Centre, Dijon, France; ¹²Division of Cancer Medicine and Sarcoma Center, The University of Texas M.D. Anderson Cancer Center, Houston, USA; ¹³Department of Medical Oncology, Centre Léon Bérard, Comprehensive Cancer Centre, Lyon; ¹⁴Claude Bernard Lyon I University, Lyon; ¹⁵Medical Oncology Unit, Edouard Herriot Hospital, Lyon; ¹⁶Department of Medical Oncology, Institut Bergonié, Comprehensive Cancer Centre, Bordeaux; ¹⁷Department of Biology and Pathology, Centre Georges-François Leclerc, Comprehensive Cancer Centre, Dijon, France; ¹⁸Department of Radiation Oncology and Center for Sarcoma and Connective Tissue Oncology, Massachusetts General Hospital, Boston, USA; ¹⁹Department of Medical Oncology, La Timone Hospital University, Marseille, France; ²⁰Department of Surgery and Sarcoma Unit, Sarcoma Service, Department of Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; ²¹Department of Clinical Oncology, Leiden University Medical Center, Leiden, The Netherlands; ²²Department of Medical Oncology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA; ²³Royal Orthopaedic Hospital NHS Trust, Birmingham, UK; ²⁴Cancer Registry of Tam, Institut Claudius Regaud, Comprehensive Cancer Centre, Toulouse, France; ²⁵Department of Radiation Oncology, The Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands; ²⁶Division of Surgical Oncology and Thoracic Surgery, Mannheim University Medical Center, Mannheim; ²⁷Sarcoma Center, Ludwig-Maximilian University Munich, Munich; ²⁸Department of Internal Medicine, Klinikum Grosshadern Medical Center, University of Munich, Munich, Germany; ²⁹Burgundy Digestive Cancer Registry, INSERM U866, University of Burgundy, Dijon, France; ³⁰Department of Oncology, Herlev Hospital–University Copenhagen, Herlev, Denmark; ³¹Department of Radiotherapy, Institut Gustave Roussy, Comprehensive Cancer Centre, Villejuif, France; ³²Department of Surgery, Humanitas Clinical and Research Center, Rozzano, Milan, Italy; ³³Department of Surgery and Department of Pediatric and Adolescent Oncology, Institut Gustave Roussy, Comprehensive Cancer Centre, Villejuif; ³⁴Department of Medical Oncology, Institut Curie, Comprehensive Cancer Centre, Paris, France; ³⁵Department of Surgery, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA; ³⁶Department of Soft Tissue/Bone Sarcoma and Melanoma, Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland; ³⁷Department of Medical Oncology, University of Michigan, Ann Arbor, USA; ³⁸Department of Medical Oncology, Erasmus University Medical Center, Daniel den Hoed Cancer Center, Rotterdam, The Netherlands; ³⁹Department of Surgical Oncology, Institut Bergonié, Comprehensive Cancer Centre, Bordeaux, France; ⁴⁰Department of Oncology, Sheffield Cancer Research Centre, Weston Park Hospital, Sheffield, UK; ⁴¹Montpellier Cancer Institute, Comprehensive Cancer Centre, Montpellier, France

Received 30 January 2014; revised 27 June 2014; accepted 23 July 2014

Background: The use of potential surrogate end points for overall survival, such as disease-free survival (DFS) or time-to-treatment failure (TTF) is increasingly common in randomized controlled trials (RCTs) in cancer. However, the definition

*Correspondence to: Dr Carine Bellera, Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Centre, 229 cours de l'Argonne, INSERM CIC 14.01 (Clinical Epidemiology), F-33000 Bordeaux, France. Tel: +33-5-56-33-04-95; Fax: +33-5-56-33-04-85; E-mail: c.bellera@bordeaux.unicancer.fr

[†]This work was presented as poster communications at the following meetings: ASCO (Chicago, June 2013), ESMO/ESMO (Amsterdam, September 2013) and CTOS (Connective Tissue Oncology Society, New York, November 2013).

of time-to-event (TTE) end points is rarely precise and lacks uniformity across trials. End point definition can impact trial results by affecting estimation of treatment effect and statistical power. The DATECAN initiative (Definition for the Assessment of Time-to-event End points in CANcer trials) aims to provide recommendations for definitions of TTE end points. We report guidelines for RCT in sarcomas and gastrointestinal stromal tumors (GIST).

Methods: We first carried out a literature review to identify TTE end points (primary or secondary) reported in publications of RCT. An international multidisciplinary panel of experts proposed recommendations for the definitions of these end points. Recommendations were developed through a validated consensus method formalizing the degree of agreement among experts.

Results: Recommended guidelines for the definition of TTE end points commonly used in RCT for sarcomas and GIST are provided for adjuvant and metastatic settings, including DFS, TTF, time to progression and others.

Conclusion: Use of standardized definitions should facilitate comparison of trials' results, and improve the quality of trial design and reporting. These guidelines could be of particular interest to research scientists involved in the design, conduct, reporting or assessment of RCT such as investigators, statisticians, reviewers, editors or regulatory authorities.

Key words: guidelines, randomized controlled trial, time-to-event end point, efficacy measure, sarcoma, gastrointestinal stromal tumors

introduction

In randomized clinical trial (RCT) in cancer, the validated and most objectively defined evaluation criterion is overall survival (OS), defined as the time interval between randomization and death (all causes). Therapeutic progress, which in certain contexts has significantly reduced mortality, the development of new types of cytostatic drugs as well as the multiplication of lines of treatment have resulted in the need for end points that measure efficacy with precision and that are available earlier than OS. Such alternative composite time-to-event (TTE) end points include for example progression-free survival (PFS) in second-line treatment. These potential surrogate end points are increasingly being used to replace OS [1] and their development is strongly influenced by the goal to reduce the number of patients in clinical trials, the duration and ultimately the cost of RCT.

As recommended by the International Conference on Harmonisation (ICH) guidelines [2] and the CONSORT statement [3], each TTE end point should be precisely defined. This implies specifying the date of origin, the list of events to be considered as failures and the censoring process. In addition, we expect that, for a given disease and setting, each TTE end point should encompass the same set of events across trials. However, most TTE end points are often poorly defined, and when a definition is provided, it can vary between publications, as underlined by recent works [4, 5] and by the Food and Drug Administration [6]. For illustration, Mathoulin-Pélissier et al. showed that a precise definition of TTE end points was reported for only 52% of cancer RCT published in major journals [4]. Toulmonde et al. showed that clear definitions of primary and secondary outcome measures were reported in only 33% of published RCT reporting on the treatment of sarcomas [5]. Similar results have been highlighted for phase II sarcoma trials [7].

The heterogeneity of TTE end point definitions has been highlighted by the international community, as demonstrated by recent publications recommending the definition of specific criteria and/or the preferred use of certain criteria in specific localizations such as for colorectal cancer [8], hepatocellular carcinoma [9], breast cancer [10] or lymphoma [11]. Most of these recommendations were based on experts' opinions, without a formal international consensus process, which could explain

why some have not yet been widely adopted in practice. To our knowledge, no definition of TTE end points has yet been proposed for sarcomas and gastrointestinal stromal tumors (GIST). This lack of standardized definitions clearly constitutes a limitation to the use of alternative TTE end points as efficacy end points.

It is important to distinguish between the process of selecting the relevant end point and the process of defining this end point. The selection of TTE end points to assess a therapeutic strategy will depend on the characteristics of the trial including settings (adjuvant versus metastatic) and treatments (systemic, local or any combination thereof). As such, the choice of the primary end point will be trial-specific. On the other hand, once the end point is identified, it has to be appropriately defined, ideally using a standardized definition to ensure future comparisons.

Based on a formal consensus process, we developed the international DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials) [12], which aims to obtain standardized consensus definitions for multiple cancer sites: sarcomas/GIST, pancreas, breast, stomach/esophagus, head and neck, colon/rectum, kidney/bladder and lung cancers. Here, we report guidelines for the definitions of TTE end points used in cancer RCT in sarcoma and GIST treatment.

methods

The DATECAN project was initially launched in 2010 for three cancer sites: sarcoma/GIST, pancreatic and breast cancers. The coordinating committee (CC) for sarcoma and GIST included three experts (CB, SMP, MO). In addition to the CC members, the steering committee included sarcoma/GIST experts in medical oncology (NP, PC), radiation oncology (OSN, MD) and surgery (SB). An independent group of experts, the rating committee (RC), was in charge of scoring the questionnaires.

consensus process

A formal consensus method was used to develop these guidelines [13, 14]. We relied on a modified Delphi consensus method and limited ourselves to two rounds of questionnaires with a final in-person meeting to discuss items for which

consensus has not been reached after two rounds of rating. The purpose of the formal consensus method is to formalize the degree of agreement among experts, using iterative ratings with feedback to identify and select points on which there is disagreement or uncertainty. The guidelines are subsequently based on agreement scores. The formal consensus method involves the following steps (Figure 1): (i) assessment of the evidence with regard to the research question; (ii) elaboration and pretesting of the questionnaire to collect experts' opinions; (iii) scoring of the questionnaires; (iv) analysis of the experts' opinions and drafting of the final report; (V) peer review phase; (vi) diffusion of the recommendations. An overview of these steps is provided in supplementary Data S1, available at *Annals of Oncology* online. We refer the reader to an earlier publication for a full description of the methodology of the consensus process [12].

literature reviews

We conducted literature reviews (i) to assess whether guidelines had been previously developed for TTE end points used in trials

for the treatment of sarcomas and GIST and (ii) to list TTE end points commonly reported, either as primary or secondary end points. The used research algorithms are available in supplementary Data S2, available at *Annals of Oncology* online.

questionnaires

In the first round, all experts of the RC received the same questionnaire (one for sarcoma, one for GIST) to score each TTE end point on a scale of 1 (totally disagree) to 9 (totally agree), according to whether various clinical events should be regarded as events in the definition of TTE outcomes. In the second round, each expert of the RC received a personalized questionnaire (supplementary Data S3, available at *Annals of Oncology* online). Items for which strong consensus had been reached after the first round were highlighted. For items without consensus, the distributions of scores obtained at the first round were summarized (minimum, maximum and median scores), and the initial score of the RC expert was indicated. Experts were asked to rescore items for which consensus had not been reached. By construction,

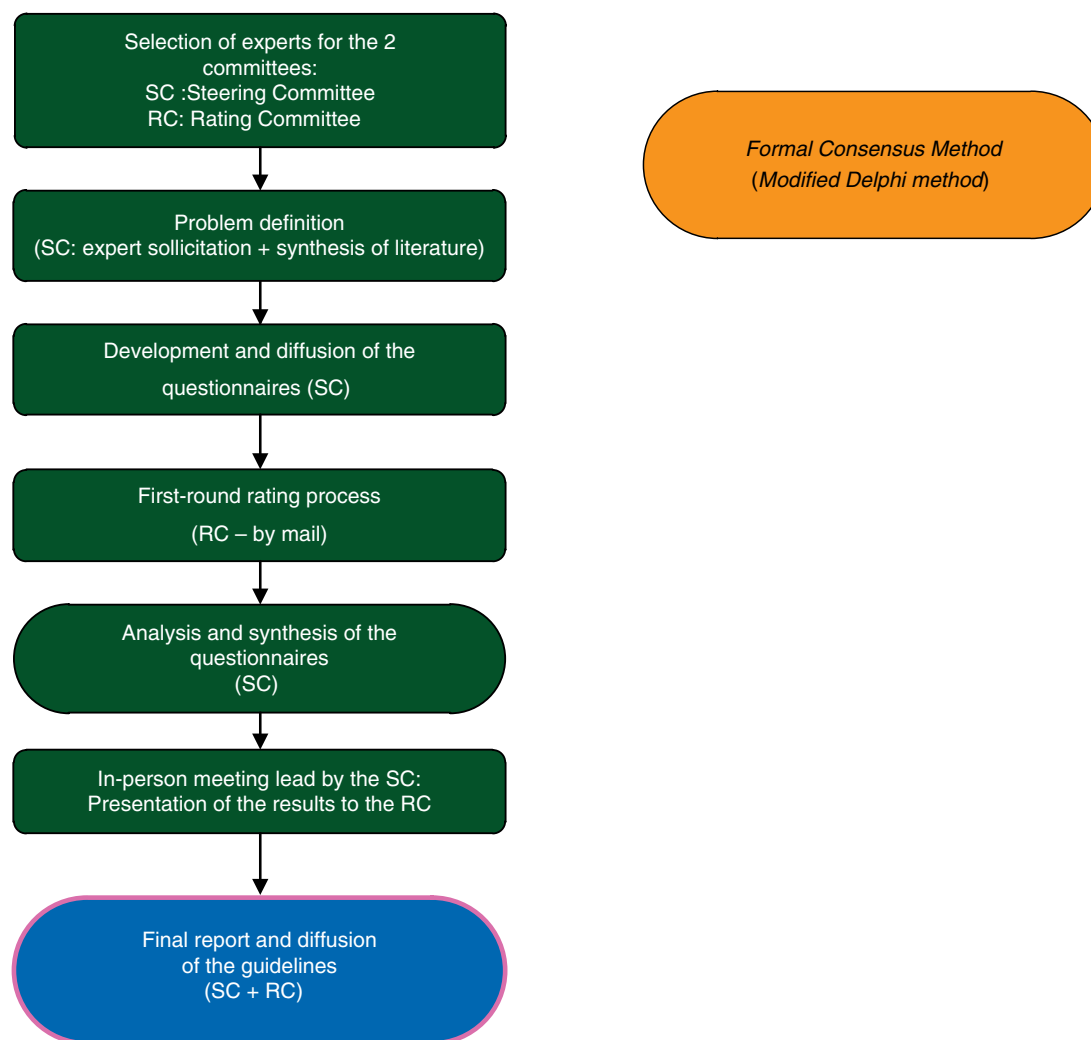


Figure 1. Modified Delphi method used in the DATECAN initiative to reach consensus for time-to-event end points in randomized controlled trials for sarcomas/GIST.

the formal consensus process aims to guide experts to position themselves, while allowing them to maintain their opinion.

results

selection of TTE end points to be defined and clinical events of interest

We identified 14 TTE end points from the literature review and established a list of clinical events that could be included in the definition of these end points (Table 1).

experts for the scoring process

The CC drafted a list of 57 experts for the RC. Of the 32 experts who filled in the first questionnaire, 28 (87%) also participated in the second round of rating. The 28 experts (supplementary Data S4–S6, available at *Annals of Oncology* online) were specialists in medical oncology ($n = 14$; 50%), radiation oncology ($n = 3$; 11%), surgery ($n = 7$; 25%), epidemiology/biostatistics ($n = 3$, 11%) and pathology ($n = 1$, 3%). They were from institutions in various

countries: France, Germany, Italy, Netherlands, Poland, United States and UK. Each expert was a member of one or more co-operative groups including ASCO (American Society of Clinical Oncology); ASTRO (American Society for Radiation Oncology); CTOS (Connective Tissue Oncology Society); EORTC (European Organization for Research and Treatment of Cancer); GSF (French Sarcoma Group); ISG (Italian Sarcoma Group); RTOG (Radiation Therapy Oncology Group); SARC (Sarcoma Alliance for Research through Collaboration); SFRO (French Society for Radiation Oncology); SWOG (South-West Oncology Group) and UK NCRI (UK National Cancer Research Institute).

consensus rates after the two rounds of rating

The consensus process involved two rounds of rating (first round: November 2010 to November 2011; second round March 2012 to May 2012) and one face-to-face meeting (3 June 2012, Chicago, IL). In the sarcoma questionnaire, overall, experts rated a total of 224 events pertaining to the 14 TTE end points. The same number of events had to be rated for GIST, leading to a total of 448 events. The scoring process is summarized in Supplementary Data S1 (available at *Annals of Oncology* online). After the first round, strong consensus was reached for 37 (8%) of the events. After the second round, strong consensus was reached for 306 events (68%), and strong or relative consensus was reached for 327 events (73%).

face-to-face meeting

After the rating process, strong consensus was not reached for 142 events covering all 14 end points for sarcomas and GIST. Discussion was therefore needed for these events and took place during the face-to-face meeting. Before discussing each event on an individual basis, and with the aim of harmonization, experts present at the meeting first made a number of decisions. The aim of this preliminary discussion was fourfold: to select the items that would be discussed, to maintain the consensus process even though not all experts attended the meeting and were therefore unavailable for discussion, to decide on the procedure to adopt in case of absence of consensus and to attempt to adopt homogeneous rules for sarcomas and for GIST. A summary of these preliminary decisions is available online (supplementary Data S7, available at *Annals of Oncology* online).

standardized definitions of TTE end points

Table 2 summarizes the events to be included in the definition of each TTE end point following the consensus process with a longer version available online (supplementary Data S8, available at *Annals of Oncology* online). Of the 14 end points initially proposed, 2 were considered ambiguous and/or redundant with other end points (failure-free survival (FFS), and relapse-free survival, RFS). As such, experts discouraged their use. Common definitions were adopted for sarcomas and GIST. The reference date is usually the date of randomization, but it can also be the date of diagnosis or treatment initiation depending on the study. End points were defined according to the setting. Two end points were specifically defined for the adjuvant setting (or disease no longer detectable): disease-free survival (DFS) and (distant) metastasis-free survival. For the metastatic setting, the following end points were defined: PFS, local progression-free

Table 1. Time-to-event end points considered for the elaboration of guidelines for their definitions, and clinical events to be considered for inclusion in definitions

Time-to-event end points
Disease-specific survival (DSS)
Disease-free survival (DFS)
Relapse-free survival (RFS)
Locoregional relapse-free survival (LRRFS)
(Distant) metastasis-free survival (DMFS)
Failure-free survival (FFS)
Progression-free survival (PFS)
Local progression-free survival (LPFS)
Metastatic progression-free survival (MPFS)
Time-to-treatment failure (TTF)
Time to progression (TTP)
Time-to-local progression (TTLP)
Time-to-locoregional progression (TTLRP)
Time-to-distant progression (TTDP)
Clinical events
Death
Death related to primary cancer/to progression
Death related to a second cancer
Death related to protocol treatment
Death related to other causes
Unknown cause of death
End of treatment
Due to toxicity related to treatment
Due to toxicity unrelated to treatment
Loss of follow-up
Relapse/recurrence/progression
Local
Regional
Metastatic
Second sarcoma cancer (or second GIST)
Second nonsarcoma cancer (or second non-GIST)

Table 2. DATECAN guidelines for standardized definitions of time-to-event end points in randomized controlled trials assessing treatment of sarcomas and GIST

Time-to-event end points	Clinical events to be included in definition of the time-to-event end points								
	Death due to primary cancer (primary site)	Death due to primary cancer (meta. disease)	Death due to second cancer	Death due to protocol treatment	Death due to other causes	Death due to unknown cause	Local events	Regional events	Metastatic events
All settings									
Disease-specific survival	X	X		X					
Locoregional relapse-free survival	X	X	X	X	X	X	X	X	
Time to progression	X	X					X	X	X
Time-to-local progression	X						X		
Time-to-locoregional progression	X						X	X	
Time-to-distant progression		X							X
Time-to-treatment failure	X	X		X			X	X	X
Adjuvant setting									
Disease-free survival	X	X	X	X	X	X	X	X	X
(Distant) metastasis-free survival	X	X	X	X	X	X			X
Metastatic setting									
Progression-free survival	X	X	X	X	X	X	X	X	X
Local progression-free survival	X	X	X	X	X	X	X		
Metastatic progression-free survival	X	X	X	X	X	X			X

NOTE 1: It was recommended NOT to include the following events in any of the time-to-event end points: end of treatment due to toxicity related to treatment; end of treatment due to toxicity unrelated to treatment; loss to follow-up; second cancer.

NOTE 2: Relapse-free survival and failure-free survival were considered irrelevant/ambiguous end points and the use of other TTE end points is recommended.

survival and metastatic progression-free survival. Finally, definitions were provided for end points that can be used in any setting: disease-specific survival, locoregional relapse-free survival, time to progression (TTP), time-to-local progression, time-to-locoregional progression, time-to-distant progression, time-to-treatment failure.

validation of the guidelines and peer review

Minutes of the in-person meeting which included the final guidelines were validated by email by all participating experts of

the RC and next submitted to a peer review group for external review who provided a formal and advisory opinion on the content and form of the initial version of the guidelines (supplementary Data S4, available at *Annals of Oncology* online).

discussion

We propose guidelines for the definitions of TTE end points to be used in RCT evaluating treatments on sarcomas and GIST. Following the first round of rating which included 28

international experts from various medical specialties, strong consensus was reached for only 8% of the items and strong or relative consensus was reached for 12%. This lack of initial consensus highlighted the disparity of experts' opinions and the need for harmonization. These consensus rates improved to 68% (strong consensus) and 73% (strong or relative consensus) at the second round. The primary reason for observing an improvement in the consensus rates is related to the design of the consensus process. By construction, the formal consensus process aims to guide experts to take position, while allowing them to maintain their opinion at each scoring round. Another reason for explaining such improvement is the different rules used to define consensus after each round. Rules for the second round are slightly relaxed to ensure that the systematic exclusion of a proposal by a rater blocks the consensus process [12, 13].

Questionnaires did not make the distinction between adjuvant and metastatic settings. Yet, this absence of distinction between settings cannot explain entirely the divergence of opinions between experts. Indeed, although the strong consensus rate was extremely low at the first round (8%), it went up to 68% at the second round without adding any specification of the setting, suggesting that absence of specification of the setting was not responsible for the low consensus rate observed initially. Discussion of the setting was introduced at the meeting which allowed to distinguish those items to be used in one specific setting or in both, and thus to provide more concise definitions. For example PFS should be preferred for metastatic settings, while DFS should be kept for adjuvant situations only.

The consensus process underlined that two end points were not particularly relevant, either because of redundancy or ambiguity. Results of the second rating round suggested that the same events should be included for RFS (defined for all settings), DFS (defined for the adjuvant setting) and PFS (defined for the metastatic setting). To reduce ambiguity and to ensure precision, experts proposed not to use the TTE end point RFS, but instead rely on DFS and PFS, respectively, for the adjuvant and metastatic settings.

FFS was also subject to debate. Specifically, following the second round of rating, there was no consensus for the inclusion of second cancers, and experts did not reach a consensus at the meeting either. Other events to be included in FFS were similar to those included in the definition of DFS and PFS. As a result, it was judged that this TTE end point did not add any relevant information and should not be used as an end point in RCT.

Finally, disease-specific survival (DSS) was initially named cancer-specific survival. Since no consensus was reached after the two scoring rounds, experts at the meeting considered that this term was ambiguous and propose to change it to DSS.

Our initial list of TTE end points was established following a literature review of recent randomized trials. The review highlighted that a majority of RCT for sarcomas and GIST assessed two or more TTE end points, usually OS, and either PFS, DFS or TTP, as also discussed in recent reviews [15, 16]. Although not all of the 12 end points defined in these guidelines are used on a frequent basis, they can be relevant (as primary or secondary end points) in specific trials depending on the setting and treatment investigated. Our objective was to propose a large panel of definitions, so that researchers can find a standardized

definition for the outcome that best suits the objective of their study.

The initial literature review also showed that the generic term, event-free survival (EFS), was commonly used but could have multiple definitions [17–23]. For illustration, in a trial of children with osteosarcoma without clinically detectable disease, EFS was defined as the time from study entry until adverse event or last patient contact, whichever came first [18]. Adverse events included disease progression, diagnosis of second malignant neoplasm, or death before disease progression. On the other hand, EFS included relapse after complete response, death from any cause, and nonresponse (taken as date of change to 'off-protocol' therapy) in a trial of nonmetastatic rhabdomyosarcoma in children [17]. Again, this reflects the need for concise and precise definitions of TTE end points. We can also wonder if the use of alternative definitions for a same outcome could have affected the conclusions of these trials, since it has been shown that different definitions can lead to different trial conclusion, as highlighted for colorectal and breast cancer [24, 25].

We have provided some general guidance for defining multiple TTE end points; nevertheless, some of them have to be refined according to the clinical setting. For example in the case of retroperitoneal sarcoma, when we evaluate the role of perioperative radiotherapy based on locoregional relapse-free survival, it must be precisely defined if peritoneal lymph nodes or liver metastases are part of locoregional relapse. This example illustrates the importance of the dialog between clinicians and statisticians when designing clinical trials.

Similarly, several tools are available to measure progression and/or relapse (e.g. RECIST [26] or Choi criteria [27]). It is however important to distinguish between the 'concept' and the 'measures', that is, between, 'which' outcomes to measures and 'how' to measure them. Both are crucial when designing trials, but complex enough so that they could not be addressed at once in this work. Measurement tools (but also the calendar of surveillance, imaging techniques etc.) must be defined study by study while accounting for the evolution of the conceptual elements to be included in the definitions of the end points. Just as the development of standardized definitions, these issues could be addressed with consensus processes using independently different trials' scenarios (disease, setting, treatment).

We deliberately did not include recommendations about the censoring process. When a clinical event is not included in a definition, it can be censored, ignored or accounted for (using competing-risk analysis) in the statistical analysis and the selected method will be study-specific depending on objectives (Additional discussion in supplementary Data S9, available at *Annals of Oncology* online).

Using composite TTE end points requires that the specific events be collected during patients' follow-up. When designing the trial, this implies clearly defining the TTE end points used as primary and secondary end points, the events that constitute these end points, the schedule and duration of patient follow-up as well as the tools to measure these events (e.g. RECIST [26] or Choi criteria [27]). Since an end point is constituted by several events, it is important that events beyond the first one observed be collected. For example assuming PFS is the primary end point, local progression can preclude distant progression, but

only the date of the local progression will be used for the estimation of PFS. Patient follow-up and data collection beyond local progression can provide complementary information with regards to treatment effect on distant progression. Hudis et al. provide an extensive discussion on this issue of follow-up and data collection [10] and emphasize that patient follow-up and data collection should go beyond the occurrence of the first event. Similarly, the cause of death and the site of first metastatic progression should be collected, including their timing. When reporting study results, a breakdown of the type of events that constitute each TTE end point in addition to the usual estimated median survival time should be reported whenever possible.

Finally, formal consensus techniques are recognized methods for developing core outcome sets to measure in clinical trials [28]. The formal consensus methodology and the participation of international experts should increase the validity and the utility of these guidelines, and as such should contribute to the generalizability and the acceptability of the resulting recommendations and their wide-scale implementation in future research.

conclusion

The formal consensus process was used to elaborate standardized definitions of TTE end points in RCT for sarcomas and GIST. The availability of these guidelines should improve international comparisons of trials' results. These recommendations could be disseminated for acquisition and endorsement by researchers and academic groups participating in clinical research. In addition, such guidelines could be of interest to other potential users including reviewers and editors of scientific journals which have recently shown increased interest in the quality of reporting clinical trials [4, 5, 29], regulatory authorities [6] or any research scientist or initiative interested in improving outcome measurements and reporting of clinical trials [30]. We therefore emphasize that it is particularly important to collect and publish detailed data on those distinct clinical events that contribute to these TTE end points, including events that follow occurrence of a first event.

Future perspectives of the DATECAN project and ongoing work include extensions to additional cancer sites (breast cancer, pancreatic cancer, stomach/esophagus, kidney, bladder, head and neck, colon, lung cancers), assessment of the impact of these definitions on academic cancer RCT as well as evaluation of the statistical properties of the newly defined intermediate end points (in particular PFS and DFS) as surrogates for OS. This ongoing work should inform us on the performance of these end points to adequately capture treatment effect depending on the disease, setting (adjuvant, metastatic) and treatment (local, cytotoxic, cytostatic etc.).

acknowledgement

The authors thank Pippa McKelvie-Sebileau for medical editorial assistance in English.

funding

This work was supported by grants from the French National League Against Cancer (Ligue Nationale Contre le Cancer; 2009

national grant for Clinical Research) and the French National Cancer Institute (Institut National du Cancer; financial support for organizing the in-person meeting held in Chicago).

disclosure

The authors have declared no conflicts of interest.

references

1. Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 2003; 21: 1404–1411.
2. International Conference on Harmonisation; guidance on statistical principles for clinical trials; availability—FDA. Notice. *Fed Regist* 1998; 63: 49583–49598.
3. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010; 152: 726–732.
4. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol* 2008; 26: 3721–3726.
5. Toulmonde M, Bellera C, Mathoulin-Pelissier S et al. Quality of randomized controlled trials reporting in the treatment of sarcomas. *J Clin Oncol* 2011; 29: 1204–1209.
6. Food and Drugs Administration DoHaHS. Guidance for industry clinical trial endpoints for the approval of cancer drugs and biologics 2007.
7. Penel N, Cousin S, Duhamel A, Kramar A. Activity endpoints reported in soft tissue sarcoma phase II trials: quality of reported endpoints and correlation with overall survival. *Crit Rev Oncol Hematol* 2013; 88: 309–317.
8. Punt CJ, Buyse M, Kohne CH et al. Endpoints in adjuvant treatment trials: a systematic review of the literature in colon cancer and proposed definitions for future trials. *J Natl Cancer Inst* 2007; 99: 998–1003.
9. Llovet JM, Di Bisceglie AM, Bruix J et al. Design and endpoints of clinical trials in hepatocellular carcinoma. *J Natl Cancer Inst* 2008; 100: 698–711.
10. Hudis CA, Barlow WE, Costantino JP et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol* 2007; 25: 2127–2132.
11. Cheson BD, Pfistner B, Juweid ME et al. Revised response criteria for malignant lymphoma. *J Clin Oncol* 2007; 25: 579–586.
12. Bellera CA, Pulido M, Gourgou S et al. Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project: Formal consensus method for the development of guidelines for standardised time-to-event endpoints' definitions in cancer clinical trials. *Eur J Cancer* 2013; 49: 769–781.
13. Fitch K. The Rand/UCLA appropriateness method user's manual 2001.
14. Haute Autorité de la Santé. Élaboration de recommandations de bonne pratique Méthode (Recommandations pour la pratique clinique) 2010.
15. Schuetz SM, Baker LH, Benjamin RS, Canetta R. Selection of response criteria for clinical trials of sarcoma treatment. *Oncologist* 2008; 13(Suppl 2): 32–40.
16. Chmielowski B, Federman N, Tap WD. Clinical trial end points for assessing efficacy of novel therapies for soft-tissue sarcomas. *Expert Rev Anticancer Ther* 2012; 12: 1217–1228.
17. Stevens MC, Rey A, Bouvet N et al. Treatment of nonmetastatic rhabdomyosarcoma in childhood and adolescence: third study of the International Society of Paediatric Oncology—SIOP malignant mesenchymal tumor 89. *J Clin Oncol* 2005; 23: 2618–2628.
18. Meyers PA, Schwartz CL, Krailo MD et al. Osteosarcoma: the addition of muramyl tripeptide to chemotherapy improves overall survival—a report from the Children's Oncology Group. *J Clin Oncol* 2008; 26: 633–638.
19. Paulussen M, Craft AW, Lewis I et al. Results of the EICESS-92 Study: two randomized trials of Ewing's sarcoma treatment—cyclophosphamide compared with ifosfamide in standard-risk patients and assessment of benefit of etoposide added to standard treatment in high-risk patients. *J Clin Oncol* 2008; 26: 4385–4393.
20. Le Deley MC, Guinebreteiere JM, Gentet JC et al. SFOP OS94: a randomised trial comparing preoperative high-dose methotrexate plus doxorubicin to high-dose

- methotrexate plus etoposide and ifosfamide in osteosarcoma patients. *Eur J Cancer* 2007; 43: 752–761.
21. Meyers PA, Schwartz CL, Krailo M et al. Osteosarcoma: a randomized, prospective trial of the addition of ifosfamide and/or muramyl tripeptide to cisplatin, doxorubicin, and high-dose methotrexate. *J Clin Oncol* 2005; 23: 2004–2011.
22. Worden FP, Taylor JM, Biermann JS et al. Randomized phase II evaluation of 6 g/m² of ifosfamide plus doxorubicin and granulocyte colony-stimulating factor (G-CSF) compared with 12 g/m² of ifosfamide plus doxorubicin and G-CSF in the treatment of poor-prognosis soft tissue sarcoma. *J Clin Oncol* 2005; 23: 105–112.
23. Miser JS, Krailo MD, Tarbell NJ et al. Treatment of metastatic Ewing's sarcoma or primitive neuroectodermal tumor of bone: evaluation of combination ifosfamide and etoposide—a Children's Cancer Group and Pediatric Oncology Group study. *J Clin Oncol* 2004; 22: 2873–2876.
24. Birgisson H, Wallin U, Holmberg L, Glimelius B. Survival endpoints in colorectal cancer and the effect of second primary other cancer on disease free survival. *BMC Cancer* 2011; 11: 438.
25. Nout RA, Fiets WE, Struikmans H et al. The in- or exclusion of non-breast cancer related death and contralateral breast cancer significantly affects estimated outcome probability in early breast cancer. *Breast Cancer Res Treat* 2008; 109: 567–572.
26. Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228–247.
27. Choi H, Charnsangavej C, Faria SC et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007; 25: 1753–1759.
28. Sinha I, Smyth R, Williamson P. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Med* 2011; 8: e1000393.
29. You B, Gan HK, Pond G, Chen EX. Consistency in the analysis and reporting of primary end points in oncology randomized controlled trials from registration to publication: a systematic review. *J Clin Oncol* 2012; 30: 210–216.
30. COMET Initiative. <http://www.comet-initiative.org> (1 June 2014, date last accessed).