European Journal of Cancer (2013) **49**, 769–781

Available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.ejcancer.info

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

Carine A. Bellera [a,b,*], Marina Pulido [a,b], Sophie Gourgou [c], Laurence Collette [d], Adélaïde Doussau [b,e,f], Andrew Kramar [g], Tienhan Sandrine Dabakuyo [h], Monia Ouali [d], Anne Auperin [i], Thomas Filleron [j], Catherine Fortpied [d], Christophe Le Tourneau [k], Xavier Paoletti [l], Murielle Mauer [d], Simone Mathoulin-Pélissier [a,b,f,o], Franck Bonnetain [m,n,o]

[a] *Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Centre, Bordeaux, France and Data Center for Cancer Clinical Trials, CTD-INCa, Bordeaux, France*
[b] *INSERM, Centre d'Investigation Clinique – Épidémiologie Clinique CIC-EC 7, F-33000 Bordeaux, France*
[c] *Biostatistical Unit, Centre Val d'Aurelle, Comprehensive Cancer Centre, Montpellier, France and Data Center for Cancer Clinical Trials, CTD-INCa, Montpellier, France*
[d] *Statistics Department, EORTC (European Organization for Research and Treatment of Cancer), Bruxelles, Belgium*
[e] *CHU de Bordeaux, Pôle de santé publique, Service d'information médicale, F-33000 Bordeaux, France*
[f] *Univ. Bordeaux, ISPED, Centre INSERM U897 – Epidemiologie-Biostatistique, F-33000 Bordeaux, France*
[g] *Unité de Méthodologie et Biostatistique, Centre Oscar Lambret, Comprehensive Cancer Centre, Lille, France*
[h] *Biostatistics and Quality of Life Unit, EA4184, Centre Georges François Leclerc, Comprehensive Cancer Centre, Dijon, France*
[i] *Unit of Biostatistics and Epidemiology, Institut Gustave Roussy, Villejuif, France*
[j] *Biostatistical Unit, Institut Claudius Régaud, Comprehensive Cancer Centre, Toulouse, France*
[k] *Department of Medical Oncology, Clinical Trial Unit, INSERM U900, Institut Curie, Paris, France*
[l] *Department of Biostatistics, INSERM U900, Institut Curie, Paris, France*
[m] *Data Center for Cancer Clinical Trials, CTD-INCa, GERCOR, France*
[n] *Methodological and Quality of Life Unit in Oncology (EA3181), CHU Besançon, France*

Available online 2 November 2012

KEYWORDS
Guidelines as topic
Clinical trial surrogate

Abstract ***Introduction:*** In randomised phase III cancer clinical trials, the most objectively defined and only validated time-to-event endpoint is overall survival (OS). The appearance of new types of treatments and the multiplication of lines of treatment have resulted in the

* *Corresponding author*: Address: Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Centre, 229 Cours de l'Argonne, 33076 Bordeaux Cedex, France.
° These authors equally contributed to this work.

http://dx.doi.org/10.1016/j.ejca.2012.09.035

endpoints
Endpoint definitions
Survival analysis
Clinical protocol

use of surrogate endpoints for overall survival such as progression-free survival (PFS), or time-to-treatment failure. Their development is strongly influenced by the necessity of reducing clinical trial duration, cost and number of patients. However, while these endpoints are frequently used, they are often poorly defined and definitions can differ between trials which may limit their use as primary endpoints. Moreover, this variability of definitions can impact on the trial's results by affecting estimation of treatments' effects. The aim of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project is to provide recommendations for standardised definitions of time-to-event endpoints in randomised cancer clinical trials.

Methods: We will use a formal consensus methodology based on experts' opinions which will be obtained in a systematic manner.

Results: Definitions will be independently developed for several cancer sites, including pancreatic, breast, head and neck and colon cancer, as well as sarcomas and gastrointestinal stromal tumours (GISTs).

Discussion: The DATECAN project should lead to the elaboration of recommendations that can then be used as guidelines by researchers participating in clinical trials. This process should lead to a standardisation of the definitions of commonly used time-to-event endpoints, enabling appropriate comparisons of future trials' results.

1. Introduction

In randomised phase III cancer clinical trials, the most objectively defined and only validated time-to-event endpoint is overall survival (OS).[1] The combined effects of new therapies and the development of molecularly targeted agents (sometimes cytostatic rather than cytotoxic), the current context of strategic trials and the multiplication of lines of treatment have led to the use of surrogate endpoints of OS to measure treatment efficacy. In essence, these criteria are composite endpoints combining different events such as local and distant progressions, local and distant recurrences and occurrence of a second cancer, death or severe toxicity (Tox). Depending on the disease setting, commonly used criteria include disease-free survival (DFS), recurrence-free survival, progression-free survival (PFS), time to progression or cancer-specific survival.[2–4] The development of these endpoints has largely been motivated by the necessity of reducing clinical trial duration, cost and number of patients, as well as the difficulty to observe an OS benefit when patients receive multiple lines of treatment at progression. Currently, these types of potential surrogate endpoints are increasingly being used as replacements for OS in clinical trials.[5]

As recommended by the International Conference on Harmonisation (ICH) guidelines[6] and the CONSORT statement[7], each time-to-event endpoint should be precisely defined. This implies specifying the date of origin (time zero), the list of events to be considered as failures and the censoring process. However, most of these time-to-event endpoints currently lack standardised definition enabling a cross comparison of results from different clinical trials.[4]

In addition, the variability of definitions for a particular time-to-event endpoint can strongly impact the trial's conclusions by affecting both statistical power and estimation. This issue was recently highlighted by Birgisson et al.[8] in the context of colorectal cancer. The authors demonstrated that the inclusion of a second primary other than colorectal cancer as an event in the definition of DFS significantly impacted the results. The estimated DFS rate for patients with stage I–III disease was 62% after 5 years if this event was not counted as an event, compared with 58% if it was. The difference was larger for stage II (68 versus 60%) than for stage III (49 versus 47%). Again, for colon cancer, results of the PETACC 03 randomised study[9] were either significant or not significant depending on whether second primary tumours were accounted for in the DFS definition or not. Similarly, Nout et al. highlighted the significant impact of including or not including non-breast cancer-related deaths and contralateral breast cancer on the estimated outcome probability in early breast cancer.[10] Finally, this heterogeneity in time-to-event endpoint definitions also complicates trial design since the survival rates expected in the control group are usually estimated based on results of previous trials, which may have used potentially different definitions.

The variety of time-to-event endpoints and the variability of their definitions are recognised by the international community. This has been demonstrated by different publications recommending the definition of specific criteria and/or the preferred use of certain criteria in specific cancer sites such as colorectal cancer in the adjuvant setting,[11] hepatocellular carcinoma (HCC)[12] and in breast cancer.[13] To the best of our knowledge, these recommendations, however, were developed based on experts' opinions only, without formal consensus.

The formal consensus is a method initially aimed at developing practice guidelines, and more generally recommendations.[14,15] Since the consensus process could

be done through questionnaires, experts from various institutions and countries can participate, favouring the acceptability and generalisability of the recommendations. Because of the consensus process involved in the formal consensus methodology and the solicitation of international experts, recommendations established through such processes are more likely to be accepted by the scientific community compared to recommendations based only on experts' opinions, and could thus lead to a greater acceptance and application of these definitions.

The objective of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project is to provide recommendations to standardise definitions of time-to-event endpoints used in randomised cancer clinical trials, and as such to ensure the reproducibility of the endpoints between studies and to allow accurate comparisons of results from different trials. This research project was initiated by a group of methodologists involved in clinical trials with a particular interest in the quality of the reporting and the analysis of clinical trials.[4] It is currently underway across several cancer sites and the general methodology of the project is presented herein, along with some illustrations based on preliminary results for the cancer sites for which data are already available.

2. Methods

2.1. The formal consensus

The formal consensus method is used to develop practice guidelines.[14,15] This method is both a practice guideline method and a consensus method. As a consensus method, its purpose is to formalise the degree of agreement among experts by identifying and selecting, through iterative ratings with feedbacks, the points on which there is disagreement or uncertainty. The guidelines are subsequently based on agreement scores. As a guideline method, its purpose is to draft a small number of concise unambiguous recommendations, which address the question asked. This rigorous and explicit method is based on the involvement of professionals in the field to which the guidelines relate.

The formal consensus method involves the following steps (Fig. 1): (I) assessment of the evidence with regard to the research question; (II) elaboration and pre-testing 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.

2.2. Experts' committees

Four experts' committees are involved in this process: the Coordinating committee (CC), involved in all six steps; the Steering committee (SC), involved in steps II, V and VI; the rating committee (RC), involved in steps III and V; the Peer-review committee (PRC) involved in step VI. We first describe how each of these committees is constituted and then provide step-by-step details of the DATECAN project.

The CC is involved in the development and design of the DATECAN project. It is made up of a group of statisticians and epidemiologists involved in the design and conduct of cancer clinical trials, with the majority being familiar with the formal consensus methodology. Moreover, this committee ensures that the project is homogeneously conducted across cancer sites. All members of the CC were involved in the design of the DATECAN project. Next, two to three members of the CC are subsequently 'allocated' to a cancer site depending on their specific research interests and involvement. As detailed further, the first task of the CC consists in identifying the cancer sites of interest. Within each cancer localisation subgroup, CC members are responsible for identifying a list of time-to-event endpoints based on a literature review, contacting academic research groups, supporting the SC when elaborating the questionnaire, sending the questionnaire for the scoring process, analysing the questionnaires, leading the in-person meeting, drafting and disseminating recommendations.

The SC involves two or three representatives of the coordination committee and 6–8 additional experts. The SC usually involves two to three statisticians and/or cancer clinical methodologists as well as five to six expert clinicians from different medical specialties (pathologist, surgeon, oncologist, etc.) and research cooperating groups. These experts are selected based on the following criteria: they should have at least 15 years experience in the speciality; they should have at least one publication for a given cancer localisation; they should be the principal investigator of at least three cancer clinical trials, or they should have participated in at least three research projects. The principal tasks of the SC consist in validating the list of time-to-event endpoints of interest for each cancer site, elaborating and pre-testing the questionnaire, validating the guidelines before their diffusion and disseminating the guidelines.

The RC is constituted by 20–30 experts who are neither part of the CC nor the SC. RC experts are selected with the following eligibility criteria: they should have at least 10 years experience in the speciality (not mandatory); they should be the principal investigator of at least one cancer clinical trial or have participated in at least one research project; they should have published at least one article related to the localisation. As representative of academic research groups, the principal tasks of the RC consist in scoring the questionnaires, attending and participating in the in-person meeting, validating the guidelines before their diffusion and disseminating them.

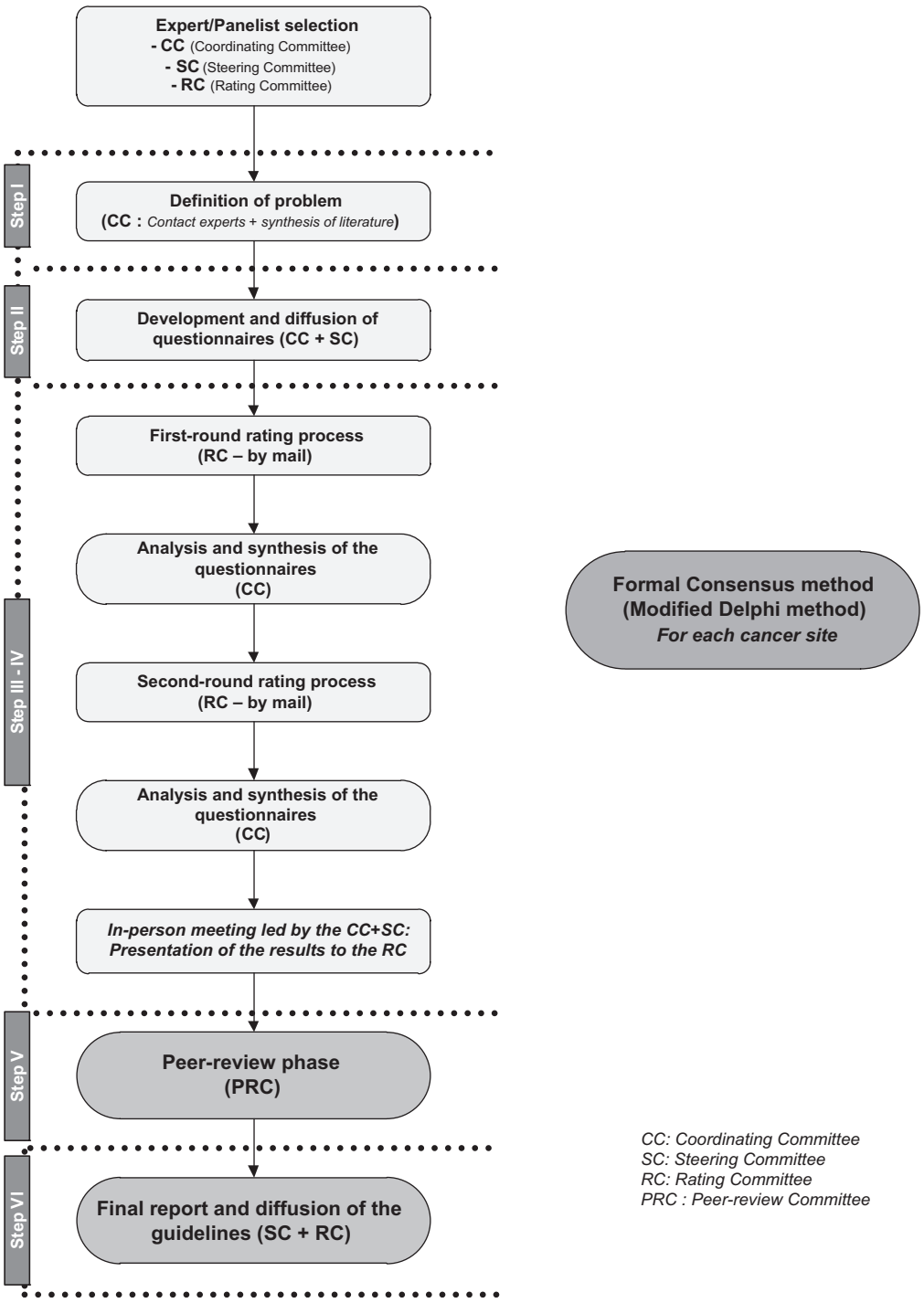


Fig. 1. The formal consensus of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project.

With regard to the constitution of the SC and the RC, experts of all specialties involved in clinical cancer trials and representatives from various collaborative groups (European, American, Australian, or other depending on the targeted cancer site) have been contacted and consulted to establish an initial list of scientific experts. For illustration, the expert group for pancreatic adenocarcinoma involved European experts only while this selection was extended to the International level for sarcomas and gastrointestinal stromal tumours (GISTs). Experts were contacted through academic and/or cooperative groups and on an individual basis (based on the SC recommendations). To avoid any imbalances, these experts were selected so as to be representative of multidisciplinarity and collaborative groups involved in cancer clinical trials: medical oncologists, surgeons, radiotherapists, pathologists, biostatisticians, epidemiologists, etc. Thereafter, an SC and an

RC were constituted for each cancer site as described above.

The PRC involves people involved in the topic, but the level of required expertise is not as high as for the SC and RC. This committee enables the group of researchers involved to be widened by contacting additional scientists. Its role is to provide a formal and advisory opinion on the content and form of the initial version of the guideline, in particular its readability, acceptability and applicability. With regard to DATECAN, PCR members can be researchers specialised in other cancers. In particular, for a given cancer site, the PRC can involve experts belonging to CC or SC groups from other cancer sites.

2.3. Step I: Comprehensive literature review to identify target cancer sites and time-to-event endpoints of interest

The first stage of the project, led by the CC, identifies cancer sites for which there are no recommendations for the definitions of time-to-event endpoints to be used in randomised trials that exist. The following cancer sites were initially identified: colorectal; hepatocellular carcinoma; pancreatic; breast; sarcoma/GIST; stomach/oesophagus; kidney, bladder, lymphoma and head and neck cancers. A comprehensive literature review was then performed independently for each cancer site using a common research algorithm adapted to each cancer site that was applied in PubMed to retrieve relevant articles dedicated to our topic: "*target cancer site*" [Mesh] AND (consensus OR recommendation OR guidelines OR standard* OR recommendations) AND (Search Endpoint OR evaluation OR outcome OR response criteria OR endpoints OR outcomes) AND (English[lang]). The search was not restricted to a specific time period. At the end of this literature search, we retained those cancer sites for which there was an absence of formal consensus-based recommendations. We excluded the case of lymphomas and prostate cancer since recommendations have been developed and are now widely used and accepted, although not based on a formal consensus methodology.[16,17]

For each of the cancer sites retained, the CC performed a second literature search aimed at identifying the time-to-event endpoints commonly reported in published randomised cancer trials. The PubMed research algorithm was the following: ("randomized controlled trial "[publication type] or "randomized controlled trials as topic"[mesh] or "meta-analysis "[publication type] or "meta-analysis as topic"[mesh]) and "*target cancer site*"[mesh] and published over the last five years and available in English. Depending on the rarity of the disease and the number of publications available, the period covered could be narrowed or extended to reach an appropriate number of publications. Based on this review, the CC established a list of time-to-event endpoints commonly reported in cancer trials for each of the seven sites. When the information was available in the publication, the CC also retrieved the various events that constituted these time-to-event endpoints.

2.4. Step II: Elaborating the initial questionnaire

Following this preliminary literature search, and independently for each cancer site, a teleconference is organised, which is led by representatives of the CC and all SC members. The objectives of this meeting are threefold. First, members of the CC validate the need for formal consensus guidelines for the cancer localisation based on the literature review. Second, they also validate the list of experts to be consulted during the rating process. Third, based on the literature review, the representatives of the CC present a list of the most widely used time-to-event endpoints that do not have standardised definitions and require recommendations. The SC can propose additional time-to-event endpoints at this point, or conversely remove some of them. A list of events that could potentially constitute these time-to-event endpoints is also submitted for validation. Finally, the SC decides whether time-to-event endpoints should be presented by disease setting (i.e. for example progression-free survival only for the advanced disease setting), and/or by treatment setting (adjuvant versus neoadjuvant setting). The SC can also be asked to choose between two layouts for the scoring questionnaires since two options were available: *single-time-to-event-endpoint-per-page* layout or *multiple-time-to-event-endpoints-per-page*. These two layouts are illustrated in Figs. 2, 3a and 3b which are extracts of the first questionnaires elaborated for pancreatic cancer and sarcomas, respectively.

As an illustration, the SC in charge of developing recommendations for pancreatic cancer retained 14 time-to-event endpoints, that is, the SC considered that it would be particularly useful to develop recommendations for the definitions of those endpoints. These time-to-event endpoints were the following: cancer-specific survival, disease-free survival, relapse-free survival (RFS), loco-regional relapse-free survival, time to local recurrence, distant metastases-free survival, time-to-treatment failure, failure-free survival, progression-free survival, time to progression, time to local progression (TLP), metastatic progression-free survival, time to performance status (PS) deterioration and time to quality of life (QoL) deterioration. The SC also suggested that some of these endpoints should be defined according to specific settings. For example, while failure-free survival would be discussed for all settings, time to local recurrence would be assessed only in the context of no detectable disease. Similarly, a list of events that could contribute to these endpoints was set up and included,

	Totally disagree								*Totally agree*
	1	2	3	4	5	6	7	8	9
Local progression									
Regional progression									
Progression of metastases/distant progression									
Appearance/occurrence of liver metastases									
Appearance/occurrence of non-liver metastases									
Second pancreatic cancer									
Second non-pancreatic cancer									
Death related to primary cancer/ to progression									
Death related to a second cancer									
Death related to protocol treatment									
Other cause of death									
Unknown cause of death									
End of treatment due to…									
Occurrence of grade 3-4 WHO PS									
Loss of follow up									
other : specify and score									

Legend: This is an illustration of a single-time-to-event-endpoint-per-page layout. On this extract, experts of the rating committee are asked to assess which clinical events (first column) should be included in the definition of progression-free survival

Fig. 2. Questionnaire (extract) for the first round of rating; Illustration with pancreatic cancer.

for example, appearance of metastases, death due to primary cancer and end of treatment due to toxicity.

2.5. Steps III and IV: The rating process (scoring and analysis of the questionnaires and preliminary report)

Rather than using the Delphi consensus method that involves an iterative consultation of experts until consensus is reached,[18,19] and can require several rounds without any guarantee of consensus, we rely on a modified Delphi consensus method and thus limit 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.[20] The questionnaires were scored following the RAND/UCLA scoring methodology that will be detailed shortly.[14]

3. First questionnaire

The questionnaire elaborated by the SC is sent to each expert of the RC. The questionnaire is sent electronically along with the project summary, the list of the participating experts (CC, SC and RC) and instructions for the rating process. For each time-to-event endpoint, the RC experts are asked to indicate on a scale ranging from one (totally disagree) to nine (totally agree) whether the clinical events should be regarded as events in the definition of the time-to-event outcomes. Once the questionnaire is completed, the RC experts return the forms by postal mail, fax or e-mail to the data centre of the DATECAN project. RC experts can also use the electronic database to complete the questionnaire online. Experts are reminded to complete the questionnaire on a regular basis (approximately every 3 weeks) and have on average up to three months to return it.

3.1. Analysis of the questionnaires of the first round

Once the first round of data from all experts is collected, the DATECAN data manager provides a statistical report. The CC representatives are then responsible for assessing whether consensus has been reached for each item. The DATECAN project relies on the RAND/UCLA scoring methodology.[14] The appropriateness or inappropriateness of including a given event in a time-to-event endpoint definition is the distribution of the scores provided by the experts who participated in the rating process (if an expert did not return the questionnaire s/he was excluded of the analyses of the first round, as well as the subsequent round and the final in-person meeting).

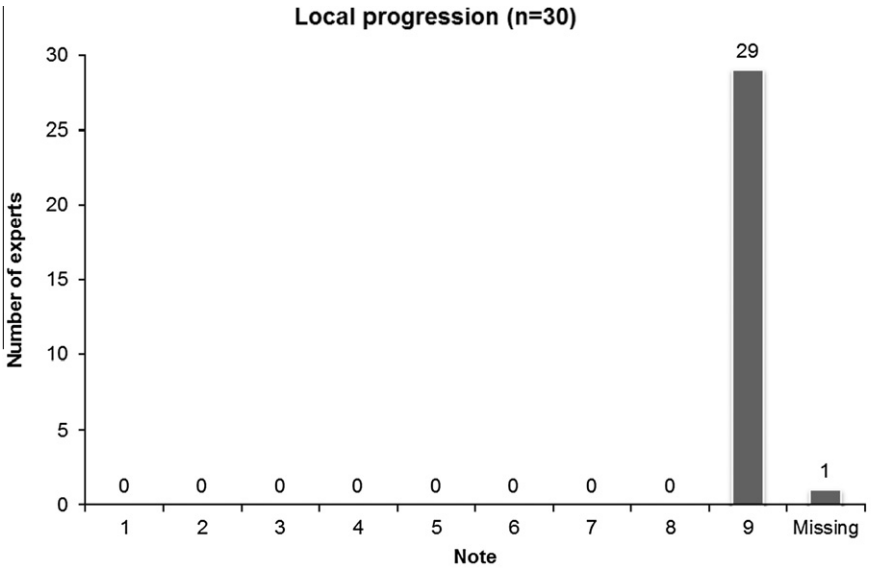
	Death related to primitive cancer / to progression	Death related to a second cancer	Death related to protocol treatment	Death related to other causes	Death related to unknown cause	End of treatment due to Tox. related to treatment	End of treatment due to Tox. unrelated to treatment	Lost to follow-up
Cancer-specific survival								
Disease-free survival								
Relapse-free survival								
Loco-regional relapse-free survival								
Distant metastases-free survival								
Failure-free survival								
Progression-free survival								
Local Progression-free survival								
Metastases Progression-free survival								
Time to progression								
Time to local progression								
Time to loco-regional progression								
Time to distant progression								
Time-to-treatment failure								

Legend: This is an illustration of a multiple-time-to-event-endpoints-per-page layout. On this extract, experts of the rating committee are asked to assess which clinical events (first column) should be included in the definitions of various time-to-event endpoints (next columns).

Fig. 3a. Questionnaire (extract – part 1) for the first round of rating; Illustration with sarcomas.

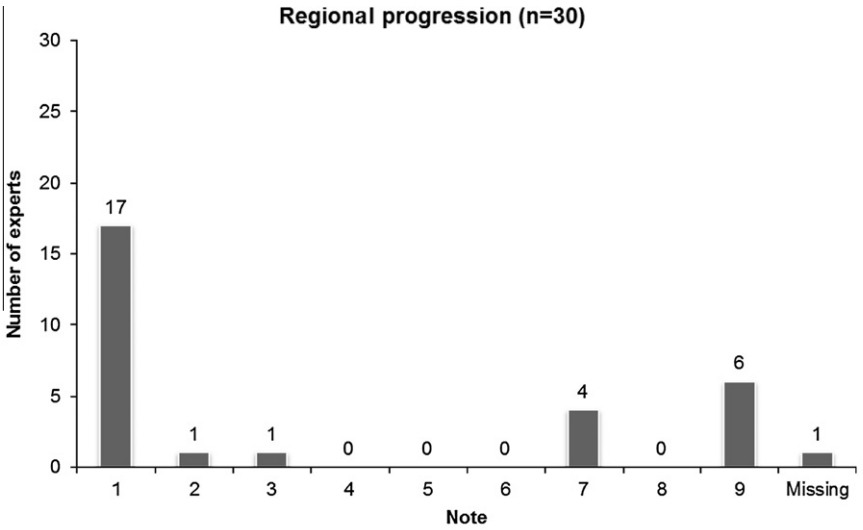
	Local relapse/ recurrence	Local progression	Regional Relapse/ recurrence	Regional progression	Appearance of metastases	Progression of metastases	Second sarcoma cancer	Second non sarcoma cancer	Other. Specify*: and score
Cancer-specific survival									
Disease-free survival									
Relapse-free survival									
Loco-regional relapse free survival									
Distant metastases-free survival									
Failure-free survival									
Progression-free survival									
Local Progression-free survival									
Metastases Progression-free survival									
Time to progression									
Time to local progression									
Time to loco-regional progression									
Time to distant progression									
Time-to-treatment failure									

Fig. 3b. Questionnaire (extract – part 2) for the first round of rating; Illustration with sarcomas.



Legend: Illustration with the 30 experts of the rating committee for sarcomas when rating whether local progression should be accounted for in the definition of time to local progression (TLP). Based on this distribution, there is strong consensus to include this event in the definition of TLP; there is no need for a second round of rating.

Fig. 4. Example of a consensual score distribution following the first round of rating; Illustration with sarcomas.



Legend: Illustration with the 30 experts of the rating committee for sarcomas when rating whether regional progression should be accounted for in the definition of time to local progression (TLP). Based on this distribution, consensus has not been reached and second round of rating is needed.

Fig. 5. Example of a non consensual score distribution following the first round of rating; Illustration with sarcomas.

The consensus for excluding or including an event is considered to be reached after the first round of rating if one of following conditions is satisfied:

– Inclusion of the event: the median of all scores is between 7 and 9, and all scores are greater or equal to 7 (i.e. range of the distribution between 7 and 9). In such cases, there is a strong consensus for including this event in the endpoint definition;
– Exclusion of the event: the median of all scores is between 1 and 3, and all scores are less or equal to 3 (i.e. range of the distribution between 1 and 3). In such cases, there is a strong consensus for excluding this event from the endpoint definition.

In all other cases (including missing data), the formal consensus method considers that there is no consensus and a second round of rating is required. Depending

	1st round			2nd Round								
	All experts		Your answer	Totally disagree								Totally agree
	Median	Min - Max		1	2	3	4	5	6	7	8	9
Death related to primitive cancer / to progression	9	1 - 9	9									
Death related to a second cancer	1	1 - 9	2									
Death related to protocol treatment	1	1 - 9	1									
Death related to other causes	1	1 - 9	1									
Death related to unknown cause	1	1 - 9	1									
End of treatment due to tox. related to treatment	1	1 - 6	5									
End of treatment due to tox. unrelated to treatment	1	1 - 4	2									
Lost to follow-up	1.5	1 - 9	2									
Local relapse / recurrence	9	1 - 9	8									
Local progression	9	7 - 9	9	Consensus reached: consider as an event - Scoring not needed								
Regional Relapse / recurrence	9	1 - 9	2									
Regional progression	9	7 - 9	9	Consensus reached: consider as an event - Scoring not needed								
Appearance of metastases	9	8 - 9	9	Consensus reached: consider as an event - Scoring not needed								
Progression of metastases	9	1 - 9	9									
Second sarcoma cancer	1	1 - 9	2									
Second non sarcoma cancer	1	1 - 9	1									

Legend: Experts are asked to rate whether various clinical events (first column) should be included in the definition of time-to-progression. Experts are presented with the results obtained at the first round (columns 2 and 3), as well as their initial score (column 4). Since consensus was reached for 3 items, those do not have to be scored.

Fig. 6. Questionnaire for the second round of rating; Illustration with sarcomas.

on the number of experts involved, up to one missing data may be accepted.

At the end of this first round, a report is produced that provides a list of events for which consensus has been reached (consensus to include or exclude), and a list of events for which a second round is required to reach consensus.

As an illustration, we present the distribution of the scores provided by the RC experts for the sarcoma group with regard to the events local progression (Fig. 4) and regional progression (Fig. 5) when studying time to local progression. Given the number of experts was important, one missing score was accepted. Since all scores were between 7 and 9 for local progression, we concluded that consensus for including this event in the definition of TLP was reached. Conversely, consensus was not reached for regional progression, and a second round of rating was required.

3.2. Second round of the rating process

When a second round of rating is required, the CC drafts a second questionnaire presenting all events for each time-to-event endpoint. Events for which consensus was reached appear only for information purposes and no scoring is requested. Experts are asked to score only those items for which consensus has still not been reached. They are provided with information about the distributions of scores obtained at the first round (the minimum, maximum and median scores are presented), as well their own initial score. Based on the initial answers provided by all experts and their own initial score, the experts are instructed to choose whether to maintain their initial score, or to modify it.

As an illustration, Fig. 6 is an extract of the second questionnaire that was sent to the experts participating in the DATECAN sarcoma project. Events for which a consensus was reached are highlighted and no scoring is requested.

3.3. Analysis of the questionnaires of the second round

Questionnaire responses are analysed according to the same RAND/UCLA methodology of the first round, however scoring rules can differ slightly.[14,15] The rule can be adapted according to the number of experts. When the rating group is large, (>20 experts), it is possible to accept up to two 'abnormal' scores defined as either missing data or an outlier. Hence, we consider that there is a strong consensus for including an event (i.e. median between 7 and 9) in three situations: (i) all scores are between 7 and 9 and 2 scores are missing; or (ii) all scores are between 7 and 9 and 2 scores are between 1 and 6, or (iii) all scores are

Table 1
Classifications after the second round of rating.

Opinions on end-points		Median	Distribution of responses after the second round (16–30 experts)
Appropriate	Strong consensus	$\geqslant 7$	All responses between 7 and 9, apart from up to 2 missing or outliers <7
	Relative consensus	$\geqslant 7$	All responses between 5 and 9, apart from up to 2, missing or <5 (2 missing or two responses <5 or one missing and one <5)
Inappropriate	Strong consensus	$\leqslant 3$	All responses between 1 and 3, apart from up to two missing or outliers >3
	Relative consensus	$\leqslant 3.5$	All responses between 1 and 5, apart from up to two missing or outliers >5
Uncertain	Indecision	between 4–6.5	Irrespective of responses.
	No consensus	$\geqslant 7$	At least three scores <5 or missing
		$\leqslant 3.5$	At least three scores >5 or missing

	1st Round		2nd round	2nd Round								
				Totally disagree								Totally agree
	Median	Min-Max	Median	1	2	3	4	5	6	7	8	9
Local relapse/recurrence	9	9		Consensus reached after first round: include event - Scoring not needed								
Regional Relapse/recurrence	9	1-9		Consensus reached after first round: include event - Scoring not needed								
Appearance/occurrence of distant metastases	9	1-9	9	0	0	0	0	0	0	1	3	26
Appearance/occurrence of liver metastases	9	1-9	9	0	0	0	0	0	0	1	3	26
Appearance/occurrence of non-liver metastases	9	1-9	9	0	0	0	0	0	1	1	2	26
Second pancreatic cancer	4	1-9	4	8	3	2	3	3	0	1	0	9
Second non pancreatic cancer	1	1-9	1	20	5	1	1	0	0	0	0	3
Death related to primary cancer	9	1-9	9	1	0	0	1	0	0	0	1	27
Death related to a second cancer	1	1-9	2	14	4	1	0	3	0	0	1	6
Death related to protocol treatment	2	1-9	2	12	5	1	0	2	0	0	1	8
Other cause of death	1	1-9	1	22	2	0	0	0	0	0	0	6
Unknown cause of death	1	1-9	1	16	2	1	0	4	1	1	0	6
End of treatment due toÖ	1	1-5	1	23	2	2	0	2	0	0	0	0
Occurrence of grade 3-4 WHO PS	1	1-9	1	24	3	1	0	0	0	0	0	1
Loss of follow up	1	1-9	1	25	2	1	0	2	0	0	0	0
Other : specify and score												

Legend: Illustration with relapse-free survival (RFS). Median, minimum and maximum scores are presented for the first round. At the end of the first round, consensus was reached for the first two events, and thus scoring was not needed anymore. For the second round, the median score, as well as the complete distribution of scores is presented. Events for which consensus was reached after the second round are presented in grey (8 events). Otherwise, events appear in white (5 events) and were subsequently discussed during the final in-person meeting to reach consensus.

Fig. 7. Results obtained following the second round of rating; Illustration with pancreatic cancer.

between 7 and 9, one score is missing and one score is between 1 and 6. Similarly, we consider that there is a strong consensus for excluding an event (i.e. median between 1 and 3) in three situations: (i) all scores are between 1 and 3 and 2 scores are missing; or (ii) all scores are between 1 and 3 and 2 scores are between 4 and 9, or (iii) all scores are between 1 and 3, one score is missing and one score is between 4 and 9. In addition to these two rules for the definition of strong consensus, the decision rules for assessing relative consensus, indecision or absence of consensus are summarised in Table 1.

For illustration, Fig. 7 presents the results obtained with regard to relapse-free survival for pancreatic

cancer after the second round. For this specific time-to-event endpoint, consensus was reached for two items after the first round, but was still lacking for 13. Of these 13 events, consensus was reached after the second round for eight events (in grey). The remaining five events (in white) were dicussed during the final in-person meeting.

3.4. In-person meeting

Items for which no strong consensus was reached at the end of the second round are discussed during an in-person meeting involving all experts. The aim of this meeting is to reach a consensus for these specific items. This meeting is led by a representative of the CC. All members of the CC and the SC are allowed to attend, however only RC experts are asked to provide their feedback. Preliminary to the meeting, an agenda is sent to the RC members, as well as all supporting documents, which include a brief reminder of the objectives of the DATECAN project, the decision rules to define the presence or absence of consensus, as well as a summary report of the analysis of the second round include highlighting time-to-event endpoints to be discussed.

3.5. Preliminary report

Based on the analysis of the two rounds of rating as well as the minutes of the in-person meeting, the CC elaborates a preliminary draft of the recommendations. Specifically, for each time-to-event endpoint, one of three possible recommendations is expressed: For the event "E", (i) it is recommended to include this event in the definition of the time-to-event endpoint "S", or (ii) it is recommended to exclude this event of the definition of "S", or (iii) the current state of knowledge does not allow one to provide specific recommendations regarding its inclusion or non-inclusion for the definition of "S".

Based on consensus proposals, this preliminary report is sent for review and content validation to experts of the CC and SC, as well as the RC experts who attended the in-person meeting.

3.6. Step V: Peer-review phase

Following the preliminary review by the CC, SC and RC committees, the first draft of the manuscript of recommendations is next sent (via email) to the PRC for Peer-review. Members provide a formal and advisory opinion on the content and form of the initial version of the guidelines, in particular its applicability, acceptability and readability. This review process is crucial as (i) it should lead to the final document that will be circulated to experts and as such should serve as a reference document and (ii) it is a first step towards dissemination and acquisition of the recommendations.

3.7. Step VI: disseminating the recommendations

Following remarks and comments of the Peer-review, a final document is proposed summarising the principle recommendations in the form of a manuscript for publication. This article will briefly describe the methodology of the DATECAN project, and, importantly, provide specific recommendations for definitions of time-to-event endpoints in cancer trials.

To ensure dissemination of these recommendations, and thus to reach our objective of standardisation of definitions, this manuscript in the form of guidelines will be sent for publication to an international cancer journal. Moreover, this document will be sent to cooperative groups involved in cancer trials.

4. Results and discussion

The DATECAN project is currently ongoing for several cancer sites as detailed in Table 2. The project is almost complete for pancreatic cancer (publication of the guidelines expected in 2012) and groups of experts are still being determined for a few other cancer sites.

The implementation of the DATECAN project is complex to set up for academic scientists, requiring both time and logistic support. Financial support was possible though research grants which were used to provide data management support and administrative assistance

Table 2
DATECAN project progress.

	Constitution of the expert groups (SC + RC)	Round 1	Round 2	In-person meeting	Publication of the guidelines
Pancreatic cancer	Done	Done	Done	Done	Expected 2012
Sarcoma/GIST	Done	Done	Done	Done	Expected 2013
Breast cancer	Done	Done	Done	Done	Expected 2013
Kidney	Done	Done	Ongoing		
Stomach/ Oesophagus	Done	Expected 2013			
Bladder	Ongoing				
Head and Neck	Ongoing				
Colo-rectal	Ongoing				

to the CC. Setting up meetings with the steering committees and contacting the experts (individually and through academic research groups) were, however, at the cost of the CC. Despite these difficulties, the project could be launched and has already been successful, since most of the experts contacted accepted to join the project, even though filling in the questionnaire can be time-consuming (up to one hour per round). Moreover it has been extended to the international level.

4.1. The need for recommendations

Results of the DATECAN project are awaited since it is now acknowledged that the variability of definitions for a time-to-event endpoint can strongly impact the trial's conclusions by affecting power and estimation, as can be seen with several examples including colorectal cancer[8,9] or breast cancer.[10] This heterogeneity limits the comparison of results across trials and can make the design of trials particularly complex since estimation of sample size are usually based on results from earlier trials. As such, standardising definitions will enhance trials' comparisons and design.

4.2. Advantages of the formal consensus process

Recommendations for the definitions of time-to-event endpoints in cancer trials have been elaborated for some cancer sites[16] however they are not yet widely used. Such recommendations were usually based on experts' opinions and thus do not appear to have sufficient legitimacy for the whole of the scientific community to be accepted and implemented. Recommendations developed in international collaborations and through a formal and validated consensus process, such as the DATECAN project, could increase chances of becoming widely adopted through a democratic process to reach consensus, and, as such, help in the standardisation process of these definitions. We preferred to rely on a formal consensus method to develop recommendations based on international collaboration through a rigorous and explicit approach. Because most of the consensus process is through questionnaires that can be emailed or faxed, experts from various institutions and countries can participate. Similarly, recommendations are developed in co-operation with many experts in the field of clinical trials from different scientific backgrounds (statisticians, oncologists, surgeons, etc.), again favouring the acceptability of the resulting recommendations. By providing feedback from previous rounds, the Delphi technique provides the advantage of a group process, building on the work and expertise of all panel members. Finally, this approach avoids issues that are commonly encountered in face-to-face group meetings, such as the dominance of key opinion leaders in the communication process. The identity and opinions of other panel members can be kept confidential from panel members, which allow them to express their personal views freely. These specificities, along with the involvement of a Peer-review group, tend to contribute to the generalisability and the acceptability of the resulting recommendations. Finally, efforts will be made to ensure, whenever relevant, consistency of definitions across cancer sites for common time-to-event endpoints, although experts' decisions made after consensus will never be overruled.

5. Conclusion

The purpose of the DATECAN project is to develop consensus-based recommendations to provide definitions of time-to-event endpoints commonly used in randomised cancer trials. The objective is to standardise the definitions of the most commonly used time-to-event endpoints, enabling appropriate comparisons of trial results. Recommendations will be published in an international cancer journal and disseminated to academic groups involved in cancer clinical research. Recommendations should be available in 2012 for pancreatic cancer and early 2013 for sarcomas/GISTs and breast cancer, and subsequently for the remaining cancer sites.

Conflict of interest statement

None declared.

Competing interests

None.

Authors' contributions

All authors developed the initial study protocol. All authors participated in the design and preparation of the study. C.B., M.P., S.M.P. and F.B. wrote the first draft of the manuscript. All other authors commented on this draft and contributed to the final manuscript.

Funding source

This research project is funded by a research grant from La Ligue Nationale Contre le Cancer and is also endorsed by the European Organization for Research and Treatment of Cancer (EORTC) and the French National Cancer Institute (INCa). The sponsor had no role in the design of the study, analysis of data or drafting of the manuscript.

Acknowledgement

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

References

1. Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist* 2008;**13**(Suppl. 2):19–21.
2. Chibaudel B, Bonnetain F, Shi Q, et al. Alternative end points to evaluate a therapeutic strategy in advanced colorectal cancer: evaluation of progression-free survival, duration of disease control, and time to failure of strategy – an Aide et Recherche en Cancerologie Digestive Group Study. *J Clin Oncol* 2011;**29**(31):4199–204.
3. Le Tourneau C, Michiels S, Gan HK, Siu LL. Reporting of time-to-event end points and tracking of failures in randomized trials of radiotherapy with or without any concomitant anticancer agent for locally advanced head and neck cancer. *J Clin Oncol* 2009;**27**(35):5965–71.
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**(22):3721–6.
5. Johnson JJ, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 2003;**21**(7):1404–11.
6. Food and Drug Administration H. International Conference on Harmonisation; Guidance on Statistical Principles for Clinical Trials; Availability. *Fed Regist* 1998;**63**(179):49583–98.
7. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;**152**(11):726–32.
8. 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.
9. Van Cutsem E, Labianca R, Bodoky G, et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol* 2009;**27**(19):3117–25.
10. Nout RA, Fiets WE, Struikmans H, Rosendaal FR, Putter H, Nortier JW. 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**(3):567–72.
11. 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**(13):998–1003.
12. Llovet JM, Di Bisceglie AM, Bruix J, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. *J Natl Cancer Inst* 2008;**100**(10):698–711.
13. 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**(15):2127–32.
14. Fitch K, Bernstein SJ, Aguilar MD, et al. *The RAND/UCLA appropriateness method user's manual*. Santa Monica: RAND; 2001.
15. Haute Autorité de santé. Bases méthodologiques pour l'élaboration de recommendations professionnelles par consensus formalisé; 2006.
16. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol* 2007;**25**(5):579–86.
17. Scher HI, Halabi S, Tannock I, et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. *J Clin Oncol* 2008;**26**(7):1148–59.
18. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;**311**(7001):376–80.
19. Linstone HA, Turoff M. *The Delphi method: techniques and applications*. Reading, Mass.: Addison-Wesley Pub. Co.; 1975.
20. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;**2**(3):i–88.