Endpoints in cancer clinical trials

F. Fiteni, V. Westeel, X. Pivot, C. Borg, D. Vernerey, F. Bonnetain

Abstract
Endpoints are measurable clinical and biological findings that are used for the development and assessment of treatment options. In the treatment of cancer, endpoints can be classified into two categories: "patient-centered clinical endpoints" including overall survival (OS) and health-related quality of life (QoL), and "tumor-centered clinical endpoints" such as progression-free survival. Surrogate endpoints are tumor-centered clinical endpoints that can be used as substitutes for patient-centered clinical endpoints, particularly OS. The choice of endpoints in oncology trials is a major problem. The published Consolidated Standards of Reporting Trials (CONSORT) best-practice guidelines encourage the reporting of clearly defined primary and secondary outcome measures. OS is the gold standard of endpoints but as increasing numbers of effective salvage treatments become available for many types of cancer, much larger numbers of patients are included; this requires a longer follow-up period and increases the cost of clinical trials. Thus, tumor-centered clinical endpoints that can be assessed earlier and used as surrogates for overall survival are increasingly studied, but most of them currently lack standardized definitions to enable cross comparison of results among different clinical trials and they have not been validated as surrogate endpoints. In addition, the variability of their definition can strongly impact the trial's conclusions by affecting both statistical power and estimation. In this context, QoL constitutes an available and useful surrogate endpoint for trials to ensure treatment benefit from both the patient and public health points of view. Methodological research should be pursued to develop standard outcome definitions for use in cancer clinical trials and to define a standardized longitudinal analysis of QoL data.

Introduction
Primary or secondary endpoints in oncologic clinical trials can be classified into two main categories: "patient-centered clinical endpoints" including overall survival (OS) and health-related quality of life (QoL), and "tumor-centered clinical endpoints" such as progression-free survival [1–3]. The use of OS as a primary endpoint is the reference...
standard to demonstrate patient clinical benefit, but the increasing number of effective salvage treatments for many types of cancer (i.e., second-line salvage treatments) has resulted in the need for a larger number of patients to be included and/or the need of a more prolonged observation period to attain sufficient data that can achieve planned statistical power; this increases the cost of clinical trials and requires a longer duration to obtain results. Consequently “tumor-centered clinical endpoints” such as progression-free survival are often used because they can be assessed earlier (i.e., intermediate endpoints). However, they do not reflect a direct clinical benefit for the patient, there is a lack of consistency in their definitions, and they have not been systematically validated as surrogate endpoints for OS. In this context, health-related quality of life (QoL) constitutes an alternative endpoint that allows earlier assessment of direct clinical benefit for the patient.

Here, we focus on the different problems raised by the choice of primary endpoint in cancer clinical trials.

**Patient-centered clinical endpoints**

Patient-centered clinical endpoints are variables that reflect a patient’s feeling of well-being or his survival [1]. They assess the effects of a therapeutic intervention for the patients in clinical trials. In oncologic clinical trials, such endpoints include OS and QoL. They assess a direct patient clinical benefit [2,3].

**Overall survival**

OS is considered as the gold standard and the most clinically relevant primary endpoint in oncology clinical trials. The US Food and Drug Administration (FDA) has considered OS as a universally accepted measure of direct treatment benefit and the preferred endpoint of Phase III clinical trials.

OS is defined as the time from randomization or initiation of treatment to patient death, irrespective of cause. Patients who are alive or lost to follow-up at the cut-off date are excluded.

Another endpoint is cancer-specific survival which is defined as the time from randomization or treatment initiation to patient death caused by the index cancer, whether due to the original tumor or to a second primary of the same cancer type. Deaths from other cancers, non-cancer-related deaths, treatment-related deaths, patients lost to follow-up and those alive at the cut-off date are excluded. The difficulty lies in identifying the specific cause of death. Moreover, this is a composite endpoint and therefore somewhat difficult to define.

However, with the increasing number of effective salvage treatments available for many types of cancer, much longer follow-up is necessary to observe the number of deaths required to achieve significant statistical power and then to establish treatments that may improve OS. In this context, we could be faced with the risk of a larger number of patients lost to follow-up (information bias). Moreover, for good-prognosis cancers, the expected difference between the two arms of treatment is often small thereby increasing the number of patient deaths needed in order to define a statistically significant difference in OS. The prolonged follow-up and the higher number of patients included increase the cost of clinical trials. Moreover, given the availability of other effective therapies administered after participation in the trial, interpretation of results can be confounded by the effects of post-trial therapy. Current management of metastatic breast cancer illustrates this problem; it becomes increasingly difficult to demonstrate an OS benefit for first-line therapy, unless post-trial administration of other effective therapies is controlled for [33]. However, taking into account randomization, we could also speculate that administration of standard-of-care subsequent second- and third-line therapies the will result in equal distribution between treatment arms, thereby avoiding confounding impact on OS.

All these reasons support the need for surrogate clinical endpoints that could be assessed earlier in the clinical course.

**Health-related Quality of life (QoL)**

In cancer research, the number of studies that incorporate QoL assessment has been rising over the last decade. Two events have combined to promote QoL as an endpoint: the creation of the World Health Organization (WHO) and the establishment of a multidimensional definition of health that encompasses physical, emotional, and social well-being [4]. QoL could be considered as a patient-defined measure using the WHO health definition of QoL. There is general agreement concerning the multidimensional concept of QoL taking into account levels of physical, mental, social, and patient satisfaction with treatment as well as treatment outcome [5,6]. QoL reflects direct clinical benefit of treatments for the patient. Many validated self-completion QoL questionnaires for cancer patients are available: e.g. the European Organization for Research and Treatment of Cancer QoL Questionnaire (EORTC QLQ-C30), the Functional Assessment of Cancer Therapy (FACT), the Rotterdam Symptom Checklist (RSLC), and the Functional Living Index-Cancer (FLIC). The QLQ-C30 is the most widely used QoL measurement tool in Phase III cancer clinical trials. It is composed of 30 items that generate 15 scores: five functional scales, three symptom scales, a global health status/QoL scale, and six single items. It can be supplemented by different modules for specific cancer types. EORTC recommends a minimum of three measurements during clinical trials: before, during, and after treatment [7]. A more intensive longitudinal assessment of QoL is generally encouraged to ensure the capture of meaningful clinically-important differences.

The FDA considers QoL as an endpoint in assessing clinical benefit. Consequently QoL has emerged as a primary endpoint candidate in oncology clinical trials, offering two advantages: a shorter study duration and an ability to assess clinical patient benefit. Opponents of QoL as a primary endpoint note its limitations due to its subjective nature, incomplete data, and the lack of standardization for analyzing and reporting QoL results. Moreover, it is difficult to provide analysis of clinically interpretable QoL results.

The evidence that QoL analysis presents methodological and statistical difficulties due to the type of data generated and the multidimensional nature of the measurement tools should be addressed to ensure comparability of trial results. One of the major concerns is missing data. Indeed, in longitudinal studies, patient assessments can be missed at specified time points because the patients miss visits or fail to complete prescribed questionnaires [8–10]. Consequently, potential biases limit confidence in QoL results, even though many statistical solutions have been robustly tested and proposed, such as multiple imputation and logistical prevention of missing QoL questionnaires. There is also
a need to define a standardized analysis of longitudinal QoL data and to develop ways of presenting results that are clinically meaningful in order to assist clinical decision-making.

**Tumor-centered clinical endpoints**

Tumor-centered clinical endpoints include biological markers. In most oncologic clinical trials, these endpoints include laboratory or histology markers to define response to a therapeutic intervention [1]. For instance, tumor response (RECIST), circulating tumor cells, disease-free survival (DFS) and progression-free survival (PFS) are tumor-centered clinical endpoints. Nevertheless, these criteria might not directly reflect clinical benefit to the patient. Conclusions of trials using such endpoints remain questionable and complementary studies to assess patients’ QoL are often requested by regulatory agencies. Their use as surrogate endpoints for patient-centered clinical endpoints must be validated.

**Surrogate endpoints**

A surrogate endpoint is a tumor-centered clinical endpoint used as a substitute for a patient-centered clinical endpoint. In oncology, endpoints that could be considered for use as surrogates for OS but assessable earlier in the patient’s course have been increasingly studied. Statistical validation of a surrogate endpoint is a difficult but necessary process. For example, Shindoh et al. [11] demonstrated that optimal morphologic response (tumor shrinkage) in response to pre-operative chemotherapy before undertaking resection of hepatic colorectal metastases is correlated with improved OS; they concluded that optimal morphologic response can be considered as a surrogate endpoint of OS. Nevertheless, a simple correlation between the proposed surrogate and OS is not sufficient since we need to demonstrate that the treatment effect on the candidate surrogate endpoint is correlated to the treatment effect on OS and also that the surrogate endpoint correlates with OS.

To achieve these requirements, two main methods of validation have been developed: one allowing validation through a single clinical trial and the other requiring a meta-analytic approach.

Single trial validation includes two approaches: the use of Prentice criteria and the proportion of treatment effect attributable to a substitution criteria [12].

In 1989, Prentice explained that a surrogate endpoint should "yield unambiguous information about differential treatment effects on the true endpoint" and he defined a valid surrogate endpoint as a ‘response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint’ [13]. Prentice proposed four operational criteria:

- treatment must have a significant effect upon the surrogate endpoint;
- treatment must have a significant effect upon the true endpoint;
- the surrogate endpoint must have a significant effect upon the true endpoint;
- the full effect of treatment upon the true endpoint must be mediated by the surrogate.

However, the demonstration of the last criterion is difficult because failure to reject the null hypothesis may be due to lack of power.

Freedman et al. [14] proposed the quantification of the proportion of the treatment effect explained by the surrogate. The quantification of proportion is expressed as: $PE = 1 - (\hat{B}/\hat{B})$ where $\hat{B}$ and $\hat{B}$ are the estimates of the effect of the treatment on the true endpoint without and with adjustment for the surrogate endpoint. The most serious difficulty is the possibility for $PE$ to be >1 or to be negative, which can hardly be justified for a proportion [14,15].

The meta-analytic method is now the gold standard for validation of a surrogate endpoint. Based on the results of previously completed trials [8,13] and using individual data, this approach jointly estimates:

- the correlation between the candidate surrogate and the final endpoints;
- the correlation between the treatment effect on the candidate surrogate and its effect on OS.

Thus, the surrogate endpoint must first be a prognostic marker, without reference to any specific intervention. This aspect of surrogacy is called ‘individual-level’ surrogacy, which means that for an individual patient, the surrogate endpoint must correlate with the final endpoint of interest, such as OS. Secondly, the effect of treatment on the candidate surrogate endpoint must be correlated with the effect of treatment on the true clinical endpoint. This aspect of surrogacy is called ‘trial-level’ surrogacy since it must be demonstrated for a group of patients in a clinical trial [16].

Nevertheless, a major difficulty for the validation of surrogate endpoints arises from the fact that they must be validated with respect to a specific therapeutic class, for a specific disease, at a specific stage (i.e., 5-FU chemotherapy for first line treatment in metastatic colorectal cancer). It is uncertain if the same surrogacy relationship will be applicable for a different treatment or at another stage of the disease [16]. Then, since trials must be performed in order to validate a surrogate for a new therapeutic class, clinical benefit in trials using such endpoints remains uncertain.

Currently, in cancer clinical trials of 5FU-based chemotherapy for advanced colorectal cancer, only disease-free survival has been validated as a surrogate for OS [17]. In T3 and T4 rectal cancers, while pathological complete response to neo-adjuvant therapy and R0 resection are important endpoints for local control, they have not been validated as surrogate endpoints of OS [18]. The ARCAD Clinical Trials Program is a worldwide collaboration of clinicians, statisticians and scientists specializing in gastrointestinal malignancies who have collaboratively constructed the ARCAD Advanced Colorectal Cancer Database, integrating patient-level data from most recent major trials in advanced colorectal cancer into a single resource [19–21]. New results regarding surrogate endpoints from the ARCAD database including biotherapies will be presented in 2013.

**Tumor-centered endpoints: the example of disease-free survival (DFS) and progression-free survival (PFS)**

While overall survival is the most clinically relevant endpoint in cancer clinical trials, there is a need for surrogate endpoints that could be assessed earlier. Tumor-centered endpoints are composite endpoints, based on tumor assessment, and they combine different events such as local and
distant progression, local and distant recurrence, development of metachronous cancer, death, or severe toxicity. PFS is generally the time between random treatment assignment and tumor progression or death resulting from any cause in the metastatic stage. DFS is generally defined as the time from randomization until tumor recurrence or any-cause death after treatments given with curative intent. These endpoints are frequently used by oncologists as primary outcomes in oncologic clinical trials because they are available earlier than OS, less influenced than OS by competing causes of death, and not influenced by treatments administered after progression. Nevertheless, they suffer from important limitations:

- there are a variety of composite endpoints used as primary outcomes in oncologic clinical trials. Time to tumor progression (TTP) differs from PFS in that the event of interest is only disease progression, while patients who die of other causes are not included. PFS and TTP have often been used interchangeably in the recent literature. Consequently cross comparison of results from different clinical trials is often difficult;
- there is no international consensus standard for the definition of PFS and DFS; consequently the definition of the same ‘endpoint’ is variable in different studies of the same disease. For example, the four randomized trials of adjuvant trastuzumab for HER-2 positive breast cancer illustrate the confusion that may arise from the use of heterogeneous definitions of DFS (Table 1) [22–25]. Moreover, Birgisson et al. [26] demonstrated that the inclusion of a secondary primary cancer other than the incident colorectal cancer as an event in the definition of disease-free survival (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%). Another example is the PETACC 03 randomized study [27] where results were either significant or non-significant depending on whether or not second primary tumors were accounted for in the DFS definition [28];
- baseline tumor size measurement can be determined before treatment or after optimal tumor response. According to RECIST 1.1 criteria, tumor progression is defined as an increase of at least 20% in the sum of diameters of target lesions compared with the smallest sum on record (this includes the baseline sum if that is the smallest on record);
- tumor progression is variably defined in different trials (clinical or radiological progression (RECIST criteria) or laboratory measures of progression). This complicates comparisons of results between different trials. Moreover, symptomatic or non-radiologic progression is often subjective, typically lacks a standard definition and is a potential source of bias [29];
- disease progression is assessed on the basis of radiologic testing at scheduled times but the true moment of progression actually lies somewhere within the time interval between two radiologic studies; in addition, surveillance intervals vary widely between and even within trials [30];
- ascertaining radiological disease progression is subject to measurement error and bias;
- PFS has been validated as a surrogate endpoint only for 5-FU-based chemotherapy in advanced colorectal cancer [18]. Nevertheless, PFS is frequently used as the primary outcome in other types of cancer and many trials have demonstrated improved PFS with no improvement of OS.

For instance, the EZ100 trial [34] studied bevacizumab plus weekly paclitaxel in metastatic breast cancer and showed a 6-month improvement in median PFS compared with paclitaxel alone. As a consequence, the FDA accelerated approval for its use. Subsequent double-blind studies with other chemotherapies confirmed more modest but still statistically significant improvements in DFS. None of these studies showed improvement in overall survival. In view of the absence of survival benefit, the modest PFS benefits noted in later studies together with toxic effects from all disease indications, the FDA took the unusual step of withdrawing approval for bevacizumab in the treatment of metastatic breast cancer. Conversely, the European Medicines Agency (EMA) approved the use of bevacizumab with paclitaxel and capecitabine [31]. This shows that the demonstration of clinical benefit in oncology trials using these endpoints has been increasingly questioned and will remain the subject of debate [32].

The variety of composite time-to-event tumor-centered endpoints and the variability of their definitions, particularly PFS and DFS, are recognized as a major methodological problem and efforts must be pursued to improve their reliability and reproducibility.

**Future perspectives**

Given all the problems raised by the choice of endpoints in oncologic clinical trials, the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project has been developed to provide recommendations and to standardize definitions of time-to-event endpoints for each specific diseases and at each specific stage by use of a formal consensus methodology [29]. This is the preliminary step before assessing their surrogate capabilities (DATECAN 2 project). Nevertheless, when evaluating a new class of therapy, trials must be performed to validate a surrogate endpoint; therefore the clinical benefit of these trials using such endpoints remains uncertain.

Consequently, health-related quality of life (QoL) constitutes an alternative endpoint that allows earlier assessment of direct clinical benefit for the patient. However, there is a need to define a standardized longitudinal analysis of QoL data and a method to present results in a clinically meaningful way in order to assist clinicians in their decision-making. One pathway under study is the development of an endpoint called ‘Time to deterioration of QoL,’ defined as the time from inclusion in the study to the time of deterioration of the QoL score [9,10]. Several definitions of ‘Time to QoL deterioration’ can be proposed depending on disease stage (adjuvant vs. metastatic chemotherapy); for this reason, practice guidelines similar to response evaluation criteria in solid tumors (RECIST) criteria for radiologic tumor evaluation should be defined, and standardized recommendations should be developed for application to QoL.

Finally, tumor-centered endpoints, for example, pathological complete response and local control in rectal cancers could be combined with QoL as co-primary endpoints to ensure a clinical benefit for patients [18]. However, methodological research must be pursued to clearly define decision-making rules for the design of such trials using two co-primary endpoints: one patient-centered and one tumor-centered.
Table 1  Example of inconsistent definitions of disease-free survival (DFS) in the four randomized trials of trastuzumab for HER-2 positive breast cancer in the adjuvant setting.

<table>
<thead>
<tr>
<th>Trials</th>
<th>Events used to define DFS as primary time-to-event endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Local/regional recurrence</td>
</tr>
<tr>
<td>HERA [24]</td>
<td>X</td>
</tr>
<tr>
<td>NSABP B-31/ NCCCTG N9831 [22]</td>
<td>X</td>
</tr>
<tr>
<td>BCIRG 006 [23]</td>
<td>X</td>
</tr>
<tr>
<td>FinHer [25]</td>
<td>X</td>
</tr>
</tbody>
</table>

DCIS: ductal carcinoma in situ.
Conclusion

Overall survival remains the gold standard endpoint in oncologic clinical trials even though OS is difficult to measure. For that reason, surrogate endpoints of OS have been developed but their statistical validation is a difficult process. PFS and DFS are frequently used as primary endpoints but suffer from important limitations, particularly heterogeneity and a lack of validation as surrogates of OS. In this context, QoL could serve as an endpoint to judge treatment efficacy, particularly in advanced cancer, but this too needs methodological improvements and standardization. Tumor-centered endpoints could be combined with QoL as co-primary endpoints to ensure a clinical benefit for the patients [18].

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

References