EUROPEAN JOURNAL OF CANCER 45 (2009) 281–289

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Recommendations for the assessment of progression in randomised cancer treatment trials

J.E. Dancey[a,*], *L.E. Dodd*[b], *R. Ford*[c], *R. Kaplan*[d], *M. Mooney*[b], *L. Rubinstein*[b], *L.H. Schwartz*[e], *L. Shankar*[b], *P. Therasse*[f]

[a]*National Cancer Institute of Canada – Clinical Trials Group, 10 Stuart Street, Queen's University, Kingston, ON, Canada*
[b]*Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD, USA*
[c]*RadPharm, Princeton, NJ, USA*
[d]*National Cancer Research Network, Leeds, UK*
[e]*Memorial Sloan Kettering Cancer Center, New York, NY, USA*
[f]*GlaxoSmithKline Biologicals, Rixensart, Belgium*

ARTICLE INFO

Article history:
Received 17 October 2008
Accepted 29 October 2008

Keywords:
Progression
Solid tumours
Randomised clinical trials

ABSTRACT

Progression-free survival (PFS) is an increasingly important end-point in cancer drug development. However, several concerns exist regarding the use of PFS as a basis to compare treatments. Unlike survival, the exact time of progression is unknown, so progression times might be over-estimated (or under-estimated) and, consequently, bias may be introduced when comparing treatments. In addition, the assessment of progression is subject to measurement variability which may introduce error or bias. Ideally trials with PFS as the primary end-point should be randomised and, when feasible, double-blinded. All patients eligible for study should be evaluable for the primary end-point and thus, in general, have measurable disease at baseline. Appropriate definitions should be provided in the protocol and data collected on the case-report forms, if patients with only non-measurable disease are eligible and/or clinical, or symptomatic progression are to be considered progression events for analysis. Protocol defined assessments of disease burden should be obtained at intervals that are symmetrical between arms. Independent review of imaging may be of value in randomised phase II trials and phase III trials as an auditing tool to detect possible bias.

1. Introduction

Traditionally, objective response has been used as the primary end-point of phase II trials to screen for the activity of anticancer drugs and improvement in overall survival (OS) is the gold standard for definitively demonstrating clinical benefit of a new cancer therapy in a randomised phase III trial. Increasingly, however, progression-free survival (PFS) or time-to-tumour progression (TTP) has been used as the primary end-points for both types of studies.

Interest in these alternative end-points reflects both the types of agents being developed, the numbers of effective treatments currently available for cancer patients, and the hope that useful clinical outcome data can be generated more quickly. A number of agents such as those targeting tumour angiogenesis or tumour growth factor receptors mainly delay tumour growth (at least initially), thus, objective tumour response (OR) may be uncommon in phase II trials of these agents. For phase III trials, overall survival remains the most objective and quantitative assessment of patient benefit.

* *Corresponding author:* Tel.: +1 613 533 6430; fax: +1 613 533 2411.
E-mail address: jdancey@ctg.queensu.ca (J.E. Dancey).
0959-8049/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2008.10.042

However, the use of OS as the primary end-point of phase III trials is problematic. First, if survival is prolonged, the demonstration of a statistically significant difference in OS may require large numbers of patients and several years of accrual and follow-up, especially if targeted treatment effect is moderate in magnitude or when the natural disease history is lengthy. Second, survival analysis may be confounded because of attenuation of the differential effect by subsequent therapies administered to patients in both treatment arms after study therapy is discontinued.

Commonly used alternative primary end-points to OR rate in phase II and OS in phase III cancer trials are TTP or PFS. Delay in tumour growth may delay death and thus PFS and TTP may be reasonable surrogates for OS in certain settings.[1] Also, it is argued that delaying tumour progression independently confers clinical benefit since it delays physical and psychological tumour-associated morbidity and, thus, improves patient quality of life. Additional advantages of these end-points over OS have been proposed: (a) date of progression may occur well before death so that a smaller study population and shorter trial duration are required to detect an improved outcome; (b) date of progression is not confounded by the effects of subsequent therapies. However, TTP and PFS end-points also have several drawbacks: (1) the end-points are subject to measurement error and certain types of bias; (2) TTP and PFS are not reliable surrogates for overall survival in certain cancer settings, which raises questions about the relevance of showing a change in these measures in these settings[2]; (3) definition of progression, TTP and PFS are not consistently defined across studies, which may complicate cross study comparisons; (4) frequent radiological or other assessments and balanced timing of assessments among treatment arms are required. The selection of the primary end-point for a trial must balance the relative strengths and weaknesses of OS, PFS and TTP for the proposed disease setting and, for registration trials, the regulatory requirements.

Randomised cancer clinical trials may be conducted by a number of different organisations and for different purposes. The major groups conducting trials are academic cancer institutions, cancer cooperative groups and the pharmaceutical industry. These groups may perform phase II studies to identify promising agents and regimens for phase III trials, and phase III trials to define the clinical benefit of investigational therapies compared to standard treatments for the purpose of achieving regulatory approval for investigational agents. Academic investigators within cooperative groups and consortia may also conduct trials of standard, approved therapies to define the optimal treatment regimen in a particular disease setting. While such cooperative group and consortia sponsored trials may define the preferred standard amongst available therapies, they may not lead to a new indication for a particular agent. For these trials, some elements of the approach to data collection, review and analysis may differ from those for registration trials. The purpose of the trial may directly determine the design, data collection and analysis.

For randomised trials of investigational agents that may potentially lead to new indications, input and review of the proposed study are required from government regulatory agencies. The United States Food and Drug Administration (FDA) and European Medicines Agency (EMEA) have issued guidance documents for end-points in cancer clinical trials.[3,4] The role of PFS and TTP as primary end-points to support licensing approval varies in different cancers. Whether an improvement in PFS is considered a direct clinical benefit or a surrogate for clinical benefit depends on the magnitude of the effect and the risk-benefit of the new treatment compared to available therapies. Because the results of the trial may lead to the commercial availability of a new drug, there may be additional data collection, as well as analysis and review required by regulatory agencies. The latter may include independent review of imaging studies to determine or confirm dates of progression. While there have been public discussions and multiple trials have been submitted for regulatory review that have included charters describing the method and means for performing independent review of imaging, there is, as yet, no regulatory guidance that describes the preferred method(s) for conducting such a review, nor means of resolving discrepancies that inevitably arise when different readers (either among independent reviewers or between site and independent reviewers) evaluate the same images. Also, PFS and TTP may be determined in part by clinical progression events that are challenging to confirm independently (see discussion below). Although limited, currently available data suggest that, in phase III trials, the calculated progression rates are unlikely to be significantly changed when using site versus independent reviewer assessments (Ford et al., Lessons Learned Paper). This is important, since the resources required to assess TTP and PFS may be substantial, particularly in the registration trials of investigational agents.

The purpose of this article is to describe the issues of definition, measurement and analysis of progression end-points and to provide practical guidance for the design and conduct of phase II and phase III trials to reduce error and bias in its assessment of progression in multi-centre randomised trials.

2. General considerations for trials with progression as primary end-point

A clear definition of progression and a consistent and symmetrical assessment schedule for patients enrolled on all study arms will ensure that the determination of dates of progression is likely to be reasonably accurate and unbiased. However, multiple factors may arise in the course of a clinical trial that may introduce variability and bias into the assessment of progression (Table 1). Measurement variability in assessment may reduce the ability to detect a difference between treatment arms by increasing the 'noise' around the estimates of treatment effects. Treatment-related investigator bias in non-blinded or imperfectly blinded studies may skew results in favour of one treatment arm versus another. Efforts to improve accuracy of data and control potential biases are important in the conduct of both phase II and III trials.

3. Study design

Study design elements to be considered for a trial include randomisation and blinding. The rapidity of disease progression may be influenced by patient prognostic factors and tumour

Table 1 – Sources of measurement variability and bias.

Bias	Description	Consequence	Recommendation
Evaluation bias	Patient's response/progression status is influenced by knowledge of treatment arm	Biased evaluation of treatment effect	Auditing using central review will reassure the lack of such a bias
Evaluation-time bias	Systematic differences in evaluation times according to treatment arm	Apparent difference in PFS/TTP due to time of assessments	Consistent interval assessments across arms. Interval should capture the rate of progression with reasonable precision (i.e. fast growing tumours more frequently and slow growing tumours less frequently) within what is clinically practical
Attrition bias	Patient drop-out varies by treatment arm	Patient may be censored for progression although time-of-death may be ascertained; potential bias from informative censoring	If possible continue to follow patient for progression
Variability			
Sources of variability	Two readers may disagree on progression status, especially in borderline cases or complex cases or due to lesion selection or identification of new lesions	Discrepancies do not indicate systematic bias. Treatments may appear more similar than they really are, and therefore this leads to reduced power to detect true treatment effects	Consistent application of progression criteria, imaging modality and methodology, same lesions, and same reader for baseline and follow-up assessments. Since there is no treatment-related bias, a statistically significant difference can be relied upon to represent a treatment effect, likely under-estimated

Table 2 – Recommendations for randomised trials with progression-free survival as primary end-point.

Trial element	Recommendation
Design	Randomised and, when feasible, double-blinded
Patient eligibility	Ideally, all patients should have measurable disease at baseline
Definitions	• Definition of events (loco-regional, distant, second primary, death, etc.) • Definition of progression for non-measurable disease if relevant • Determination of the date of progression • Definition of what is an adequate assessment for purposes of evaluation at the scheduled time points (i.e. all baseline target lesions measured)
Assessments	• Disease burden assessments should be obtained at intervals that are symmetrical between arms • Ideally, patients should be followed until progression is objectively documented even if protocol therapy is discontinued • Intervals between assessments should be chosen as to avoid possible bias in interpretation of primary outcome
Data collection	• Case-report forms should collect ALL elements required to determine progression and date of progression accurately • Adapt CRFs specifically to collect elements that are critical to support the clinical/symptomatic progression if patients with only non-measurable disease are eligible and/or clinical, or symptomatic progression are to be considered progression events for analysis
Analysis	• Methods for handling missing data, censoring should be pre-specified • Independent review of imaging may be of value as an auditing tool to detect possible bias

biology. As a result, randomised controlled trials are needed for definitive assessment of either PFS or TTP. Because of the subjectivity that may be introduced in end-point assessment, blinding of trials is recommended. Blinding is particularly important when subjective investigator assessments are included as the components of the progression end-point or when the assessment schedule is subject to modification by the physician or patient due to the occurrence of toxicity, morbidity or inconvenience. Physician (or patient) bias that the experimental treatment is beneficial may lead to earlier identification of progression in the control arm than in the experimental arm, if only through an inadvertently more aggressive assessment schedule. Blinding the individual assessing progression to treatment assignment, while eliminating outright assessment bias, will not completely mitigate this form of bias at the site. There are circumstances that preclude blinding clinicians and patients to treatment allocation due to the differences in administration or toxicity. In these situations, additional measures or analyses may be undertaken to minimise or evaluate for bias as will be described in detail below.

4. Selection and definition of end-points

4.1. *Time to progression and progression-free survival*

Time-to-tumour progression (TTP) is defined as the time from a defined date, usually randomisation, to time of progressive

disease. Progression-free survival (PFS) duration is defined as the time from defined date, e.g. randomisation, to tumour progression or death. The time of progression is usually assigned to the date of the clinical visit or imaging assessment at which progression is identified. Factors that may affect the determination of the time to progression include (1) whether it is measured from diagnosis, randomisation or first dose; (2) whether progression is defined by a consistently applied algorithm; (3) the intervals between assessments and possibly (4) different qualities of response as progression is determined from the nadir tumour measurement. In the latter situation, it is not clear that progression from complete response (which must occur through the reappearance of 'detectable' disease), partial response or stable disease, or progression from very small lesions versus larger lesion size represents the same change in disease burden or carry the same clinical prognosis. Although each of the above may alter the calculated duration to progression, as long as the definitions and assessments are consistently applied across the treatments arms the study results should not be biased.

Both TTP and PFS are composite end-points. For both endpoints, progression includes radiologic progression, and, in some instances, non-radiologic criteria such as 'symptomatic progression' or 'clinical deterioration'. PFS also includes the occurrence of death from any cause. PFS is generally considered a better surrogate for clinical benefit than other endpoints such as time to disease progression alone in patients with incurable locally advanced or metastatic cancer, as the PFS end-point includes death from all causes, so that adverse effects of treatment or disease on survival are included in the end-point. For TTP, patients are censored at the time-of-death or at an earlier visit which gives rise to informative censoring (non-random pattern of loss from the study), unless patient deaths prior to disease progression are not correlated with progression. In situations where many deaths are unrelated to cancer or treatment, TTP can be an acceptable and preferred end-point.

Censoring of patients may introduce bias in the analysis. Date of progression may not be identified on all patients or date of clinical deterioration may precede the date of identifiable radiological progression. In both these instances, bias may be introduced in the progression analysis. Patients may discontinue protocol therapy or follow-up without progression due to toxicity, change to another anti-cancer treatment due to lack of apparent benefit of protocol therapy, or personal choice. Patients withdrawn from study may not be followed up to determine progression status and are censored at the time of event associated with withdrawal from protocol. If patient withdrawals from protocol are balanced across protocol treatment arms, bias may be avoided, but the power to detect a difference between treatments may be reduced due to a reduction in the number of progressions observed. However, if patients are preferentially removed from one protocol treatment arm, there may be bias in the progression analysis due to informative censoring.

If the definition of progression includes both radiological and clinical-symptomatic criteria, the situation may be problematic if only radiological progression is reviewed independently as may be required for registration trials. If patients are removed from study for symptomatic progression, radiological progression may or may not be identified after symptomatic progression. If identified by central review, then the date of radiological progression may be considered the date of progression for PFS analysis. If radiological progression is not identified, the patient may be censored. This type of informative censoring may result in bias in the PFS analysis. Thus, as will be discussed later in further detail, the definition of progression and its independent verification by central independent review require careful consideration and planning, particularly for registration trials.

PFS is a combined end-point that includes progression, which is subject to censorship through loss to follow-up, and death, which generally is not. If the time-of-death is determined at a considerable interval after the last radiologic or clinical assessment, the inclusion of late-deaths as PFS events, rather than censoring them at the time of last follow-up, may attenuate estimates of treatment effectiveness and reduce power, or even give rise to treatment-related bias, if such loss to follow-up is not balanced. The FDA has recommended that an appropriate maximum time interval between last scan and death (for including the death time as an event) should be specified to minimise bias from this type of missing data.[3]

5. Issues with defining progression

If TTP or PFS is the primary end-point of the trial, then all eligible patients should be assessable for progression. As noted above, there may be a regulatory preference that the determination of progression be based on imaging and that imaging results be verified by independent central review in registration trials of investigational agents within an indication. While there are a number of response criteria that define progression for specific indications such as RECIST for solid tumours, within an indication the regulators do not have a reported preference for the criteria.

In general, progression may be determined to have occurred based on (1) the appearance of one or more new lesions, (2) an increase in the size of target measurable lesions (greater than or equal to 20% of the sum of the longest diameters by RECIST), (3) a clear, unequivocal increase in non-target disease, which may be measurable or non-measurable, (4) worsening of the symptoms and signs of disease that are not evident on radiological assessments. Revised RECIST criteria clearly define the first two situations. A few general comments are warranted on the appearance of lesions and progression of target lesions. In both these situations and indeed all cases of progression, progression should be 'unequivocal', i.e. not attributable to differences in scanning technique within a modality (e.g. slice thickness change at computer tomography (CT) or magnetic resonance imaging (MRI)), change in imaging modality (e.g. CT at baseline and MRI at follow-up), within the range of measurement error, or findings thought to represent something other than tumour growth (e.g. inflammation or oedema). For example, an increase in size from a nadir measurement that is small should be greater than the technical variation in measurement to be deemed disease progression. As another example, tumour necrosis and cystic formation do not always repre-

sent an increase in tumour burden and thus should not be considered evidence of progression – indeed these may be harbingers of subsequent tumour regression. A new lesion should be unequivocally a metastatic deposit as it necessarily must be classified as progression even if the imaging test done 'on study' to identify the lesion was not performed at baseline. Therefore, baseline anatomic scanning to detect the presence of disease should include all areas likely to be the sites of metastases based on the known biology of the specific cancer. Consistent follow-up imaging assessments of the known disease sites in terms of modalities, methodology, reading/interpretation can improve the accuracy of the determination of tumour burden for study patients. It may be argued that follow-up imaging include imaging positive at baseline, and surveillance of the likely sites for the development of metastases to improve the accuracy of the determination of progression date by potentially identifying earlier sites of new disease (which might otherwise be identified only if lesions are symptomatic or other reasons); however, the additional scans must be balanced by concerns of cost containment and radiation dose. If assessment of progression is equivocal or uncertain, it is better, if medically appropriate, to continue the patient on study for additional follow-up until progression is unequivocal than to remove the patient from study early. The protocol analysis plan should specify whether the progression date may be subsequently back dated to the first equivocal finding or the date of the unequivocal finding.

Measurement variability may arise when lesions appear or are perceived to be different (due to the differences in technology, modality or technique), or measured differently. In addition, different radiologists may select different target lesions at baseline to follow, which may result in different dates of progression for individual patients. To minimise variability, the same imaging modality and method should be used at each patient assessment. Because progression of disease is judged from the nadir measurement, when lesions may be tiny, a minimum absolute change in size should be specified. Discrepancies due to measurement variability, lesion selection and perception of new lesions are expected when multiple radiologists review images. Therefore, the occurrence of discrepancies per se between site and independent readers does not impugn the quality of the data obtained from sites, and does not imply bias in site assessment of progression (particularly if the site radiologists are blinded to the patient's treatment). Measuring all lesions or using central review in real time as the determinant of progression is not necessary and may not be practical for many clinical trials. Analyses from the RECIST database indicate that these discrepancies would not substantially alter the estimated hazard ratio for the study arms.[5]

Defining criteria for progression in non-target measurable disease and non-measurable disease are more complex. Generally, when measurable disease is also present, progression of non-target lesions should rarely be the sole criterion for calling progression. As noted in the revised RECIST criteria[6] in this circumstance, to achieve 'unequivocal progression' there must be an overall level of substantial worsening in non-target disease (whether measurable or non-measurable) of a magnitude that, even in the presence of stable disease (SD) or partial response (PR) in target disease, the treating physician would feel it clinically important to change therapy. Both imaging and clinical status of the patient may influence this determination. A simple increase in one or more non-target lesions alone is not considered 'unequivocal progression'. For patients with increases in the sizes of non-target lesions that are measurable, an additional guideline could be considered: that the overall increase in non-target disease be such that, if the lesion(s) are included with the other target lesions, then the overall increase in measured disease would clearly meet or exceed the definition of PD. The practicalities of implementing this guideline within a clinical trial would need to be carefully considered as previously considered non-target (and thus unmeasured lesions) may need to be measured and 'added-in' to the sum of diameters from baseline and through follow-up assessments. The 'adding-in' would change both the numerator and denominator of the sum of diameters used to calculate change from baseline or nadir. The designation of the overall progression solely on the basis of change in non-target disease in the face of SD or PR of target disease should be rare.

For RECIST response categorisation, non-measurable (non-target) disease generally factors only into decisions about complete response or, on rare occasions, about clear progression in non-measurable disease such that the overall progression of disease has occurred based on the judgment of the investigator. Clearly no better determinant of treatment decisions for individual patients can be that of their treating physician. However, within a clinical trial, investigator judgments (1) may be influenced by bias if a particular regimen is favoured and the trial is not blinded, (2) may be inconsistent for patients with similar disease states and (3) may not be subject to independent review. Defining a worsening of disease burden from a non-measurable baseline is largely a subjective determination. It would be highly desirable to have an operational definition of progression of non-measurable disease as some patients may only have non-measurable disease at baseline and thus may be otherwise excluded from studies, and some patients may have progression of non-measurable disease that cannot be independently reviewed. One consideration would be to define progression of non-measurable disease solely as the appearance of new lesions, which is the most common means that progression is identified in patients. Defining additional criteria for worsening of non-measurable disease that can be consistently used within a clinical trial and independently verified require specifying the type and magnitude of change from nadir in the protocol. If it is to be independently verifiable, the supporting information must be collectable in case report forms. The additional data elements might include documentation of new or worsening adverse events solely attributable to cancer and not due to benign, incurrent illness, cancer treatment or changes in supportive care, tumour markers, pathology/cytology and imaging results. Such criteria would need to be relevant to the disease setting, the elements would need to be clearly defined, collected in case report forms and appropriately included in the analysis plan. The unattractive alternative is to consider limiting eligibility only to patients with measurable disease and/or accepting the imprecision of the determination of progression

in patients with non-measurable disease and that these patients may be censored if the PFS analysis is based on independent review.

In phase III trials that do not require measurable disease as an entry criterion, some patients may only have non-measurable disease. In this situation, the designation of progression should also require unequivocal progression, i.e. either new lesions or substantial worsening in non-target disease that is of a magnitude that the treating physician would feel it important to change therapy. Examples of progression of baseline non-measurable disease are worsening of lymphangitic disease in lung, the increase of malignant ascites/effusions; progressive pancytopaenia due malignant infiltration, significant worsening of pleural or peritoneal carcinomatosis and worsening bone metastases, not due to healing or flare phenomena. Definitive documentation of all manifestations of worsening disease from baseline may be needed to provide independent verification of progression and it might have more demonstrated value in certain indications where clinical PD is more common such as in breast and ovarian cancer. For patients that experience worsening adverse events that may be due either to toxicity of treatment, disease progression or change in supportive care (such as symptomatic deterioration with fatigue and decreased performance status or increased pain due to changes medications), follow-up assessment may determine the final decision, i.e. toxicity resolves, deterioration from disease may continue. In these instances, patients should continue to be followed on protocol until a final determination of disease progression status can be made. Retrospective analyses of phase III datasets have shown that it is relatively uncommon to have progression of only non-measurable disease among patients that have measurable disease (as such patients are more likely to experience progression of disease through the development of new lesions or through the progression of measured lesions.[5] Such occurrences, due to their infrequency, appear to rarely influence the overall trial outcome. However, the trials from which this conclusion is derived required that patients have measurable disease for entry. The frequency of occurrence of progression in non-measurable disease is likely to be higher if patients with only non-measurable disease at baseline are also eligible for specific clinical trials.

In summary, criteria for progression should include the definition of measurable disease and the increase from baseline that constitutes a clear increase in tumour burden per the current version of RECIST guidelines. The definition of progression of non-measurable disease or clinical deterioration should similarly specify the magnitude of increase from baseline or nadir that is considered progression if these events are to be used. For the latter, an operational definition relevant to the disease setting that can be consistently applied within and across studies is desirable. The magnitude of change should be unambiguous and clearly reflect an overall increase in tumour burden such that in the treating clinician's judgment, the patient should receive alternate treatment. Data collection and review to independently verify such cases, and whether review should be conducted retrospectively or prospectively are additional considerations.

6. Determination of assessment intervals

The exact date of progression for a patient is not known; rather it is imputed based on the types and timing of assessments. To determine date of progression, patients must be evaluated on a regular and balanced basis across treatment arms. When criteria for progression have been met, such as a tumour having grown by a certain percentage, the true progression time may be somewhere within the time interval between two assessments. In general, the date of progression is defined as the date at which progression was first evident, i.e. the 'upper time limit' of the assessment interval although alternatives such as using the date of the scan before the one at which progression was identified, i.e. 'lower limit' of the progression interval or using the midpoint of the interval between scan dates as the progression date has been proposed.[5] Thus, one problem with PFS is that any interval can be chosen for measuring tumour burden and the interval between assessments determines when progression is likely to be identified.[7] Additionally, the assessment intervals may influence the interpolation of median PFS from PFS Kaplan–Meier curves and can artificially suggest the differences between treatment arms within a trial or across trials.

Both the frequency and the symmetry of assessments across treatment arms are important considerations in trial design and conduct. Very broad intervals between assessments may lead to treatments appearing very similar; however, more frequent assessments impose increased burden on patients and clinical trial staff without necessarily improving accuracy of the determination of progression date.[7,8] If the treatments being assessed in a clinical trial are administered at different intervals, or if toxicity, missed assessments or unexpected and non-protocol specified assessments lead to asymmetrical treatment delays, and tumour assessments are defined by the numbers of cycles administered, tumour assessment intervals may differ between treatment arms. If there are unequal patient assessment intervals, the perceived benefit might be due to the difference in timing of measurements, as was suggested in the FDA's review of the bcl2 inhibitor oblimersen sodium.[9] This bias can magnify an apparent outcome difference if a systematic assessment difference of a week or two occurs in a study of large sample size and small effect size.

A final consideration in the selection of assessment intervals is their consistency with other studies within the same disease setting. Heterogeneity of tumour assessment intervals may complicate cross trial comparisons and when historic controls are used to inform the design of future studies within a disease setting.[7] Although consistency of intervals for cross study comparisons may not be a primary consideration in the selection of interval assessments, it may be particular value in settings where meta-analyses of trial results may be contemplated.

In summary, to minimise bias and error, timing of assessments across treatment arms should be symmetrical and the same evaluation tests should be used. The interval should be consistent with clinical practice for monitoring patients with the disease based on 'usual' progression expectations for the tumour type with a reasonable margin for scheduling around

these specified time points to accommodate the clinical realities. Consistent interval assessments within a trial are critical and across trials in a specific disease setting are recommended.

7. Data collection and analysis

The amount and type of data that are collected on cancer clinical trials depend on the organisation conducting the trial and the trial's specific purpose: i.e. whether the trial will support a regulatory submission, whether the drug is being tested for a similar indication to one previously conducted, the amount of data available from other sources on the safety of the agent or combination regimen. The data collected to determine the efficacy of the cancer therapy under investigation depend on the end-points and disease setting. The schedule for the collection of baseline and follow-up data for full evaluation of efficacy should be specified in the protocol. In addition to the investigator's categorisation of response, all data collected for evaluating efficacy should be recorded on the case report forms. For the measurement of PFS, documentation of tumour measurements by imaging at baseline and follow-up are needed. If clinical or symptomatic progression is considered as part of the definition of progression, additional documentation that may include recording baseline and degree of deterioration of performance status, development of moderate or severe tumour related adverse events requiring intervention and worsening of tumour markers may also be required.

PFS and TTP are generally calculated by Kaplan–Meier method and progression differences between treatment arms are determined by the application of an appropriate statistical method such as the log-rank test. However, an analysis of this end-point is complicated as patients may go off the study for various reasons, including treatment toxicity, worsening of disease symptoms, or change to alternative non-protocol therapy due to lack of benefit without documentation of progression. In these cases, further tumour measurements may not be performed or may not be reported. Whenever disease evaluations have been discontinued without progression being documented, there is the potential for censoring to be informative.[10] Censoring is informative if the progression risk for withdrawn patients is different from those remaining in the study. In fact, using central review as the primary determinant of progression may introduce bias as patients with symptomatic progression would be censored.[11] If those patients who are withdrawn due to worsening symptoms are at higher risk for progression, then the Kaplan–Meier estimation from PFS censored at the time of withdrawal would overestimate the PFS tail probabilities, and, similarly, if those withdrawn are at lower risk, the PFS tail probabilities would be under-estimated.[10] Informative censoring may lead to bias in the estimation of the PFS. As a consequence of such censoring, comparisons of PFS between treatment groups and conclusions may be misleading. Censoring at the time of study withdrawal due to toxicity or initiation of non-protocol therapy prior to reaching a confirmed progression event may actually lead to a result that favours the less efficacious, more toxic treatment. Sensitivity analysis methods have been proposed to assess the potential effect of informative censoring on outcome.[8,11] Adopting the intention to treat (ITT) approach for the analysis of survival, whereby all the patients are followed for a documented evidence of progression regardless of when and why they stop taking randomised treatment, would provide a better basis for comparing treatments. If desired, a supportive analysis could still be conducted censoring dropouts in the absence of documented progression.[8,11]

Additional considerations influencing the precision of the progression-free survival estimate include the proportion of the missing data (missing assessments) and the conventions that may be chosen to impute a missing value. The protocol should define an adequate assessment visit for each patient (i.e., a visit when all scheduled tumour assessments have been done and target lesions measured). The analysis plan should outline a comparison of the adequacy of follow-up in each treatment arm. Methodology for analysing incomplete and/or missed assessments and censoring should be specified in the protocol. The FDA has proposed that the impact of a missing assessment be evaluated by the use of a conservative convention, such as censoring at the time of the previous completed assessment.[3] The analysis plan should specify the primary analysis and one or more sensitivity analyses to evaluate the robustness of the results.

8. Phase II versus phase III trials

In general, the issues and recommendations for assessing progression apply to both randomised phase II and III trials where PFS or TTP is used as the primary end-point. Some may question whether the same rigour should be applied to both phase II and III trials. Traditional phase II trials of cancer therapies have been non-randomised trials that assessed objective response rate. While this approach is reasonable for agents expected to produce tumour regression, in the recent years many drugs under development are more likely to stabilise disease rather than to induce substantial tumour shrinkage. Also, for studies of combinations, PFS may be considered a more clinically relevant end-point. As a result, randomised phase II trials with PFS as a primary end-point are increasingly favoured.[12,13] Such trials are of greatest value in settings in which PFS is relatively short and the definitive phase III trial end-point is considered to be overall survival. Phase III trials are larger randomised controlled trials to definitively demonstrate patient benefit, establishing standard of care. Some may argue that greater emphasis and resources should be expended on conducting robust phase III trials since the implications of a positive result from a phase III trial are greater than for a phase II. However, the number of new agents and combinations in development is continuing to increase, producing a substantial need for methods providing greater confidence for prioritising the resources required for large phase III trials, which are also increasingly expensive to organise and conduct.

The value of phase II trials for cancer therapies is to screen for activity, which if substantial would lead to a phase III trial. At best, a small trial provides a limited estimate of antitumour activity of a treatment rather than assessing actual patient benefit. The phase II end-point should reasonably correlate

with patient benefit and the magnitude of the targeted treatment effect of interest should be sufficiently large as to limit the sample size of the study. If improvement in PFS is considered a generally accepted surrogate or is considered actual clinical benefit for patients in the specific disease setting then it may be a more appropriate end-point for a phase III trial. A phase II would have to target a larger difference, higher type 1 error (larger alpha), or type 2 error (larger beta), i.e. if the target difference is moderate, the level of certainty in the result is reduced to limit sample size. A 'positive' phase II result, unless unusually extreme, cannot definitively demonstrate benefit in PFS, nor should its presentation be allowed to compromise the ability to conduct a definitive trial. However, the perception among clinical investigators and patients that the phase II result indicates the superiority of treatment may compromise the ability to accrue to the phase III trial. Thus, it may be preferable that randomised phase II trials not use PFS as the primary end-point, if PFS is also considered the preferred end-point for a phase III trial.

Some might assume that phase II trials may not require the same rigour of design and execution as phase III trials. However, patient numbers typically enrolled on a phase II trial are limited and the 'go or no go' decision at the end of phase II is critical in drug development due to resources required to plan and launch a phase III trial. Variability in measurement and assessment and bias may have a greater potential to adversely affect the outcome of a phase II trial due to the limited numbers of enrolled patients. Blinding to treatment and independent confirmation of progression may be of greater value in phase II trials as these measures may improve the reliability of the result with a relative increase in resource that is modest compared to those required for a phase III trial. Thus, improving the quality of the information obtained from a phase II trial may be worth the additional investment.

9. Randomised phase III registration and non-registration trials

When PFS is a desired end-point, trials should be double-blind where possible to minimise bias in the estimate of the treatment effect. When double-blinded trials are not feasible, independent review of imaging to determine progression may be used as an auditing tool to detect bias in local evaluations to increase confidence in the trial's conclusions. Bias may have occurred if independent review identifies differential rates of unconfirmed progression according to treatment arm and/or higher rates of earlier progression in one treatment arm.[11] However, independent review as the primary determinant of progression may introduce informative censoring and bias. Phase III trials that use PFS should target treatment effects that are sufficiently large to be clinically important and robust. When feasible, all patients should be assessed until progression, regardless of intervening events. For registration trials, regulatory authorities may require additional data collection, analysis and review and these additional needs should be identified prior to proceeding with a phase III trial. Alternatively, non-registration trials that may be comparing available standard treatments may not require as much data collection. For example, toxicity of treatment is likely to have been defined in prior studies and need not be as intensively collected although all serious adverse events should be reported.

10. Conclusions and recommendations

Progression-free survival is an attractive end-point for phase II clinical trials to test agents that generally prolong stable disease rather than induce tumour response and for phase III clinical trials when an overall survival end-point may be confounded by additional treatments administered after progression or when it may occur long after the progression event. Our recommendations for assessing progression when PFS or TTP is the primary endpoint of the trial are listed in Table 2. Ideally trials with PFS as the primary end-point should be randomised and double blinded. All patients eligible for study should be evaluable for the primary end-point and thus, in general, have measurable disease at baseline. If patients with only non-measurable disease are eligible and/or clinical, or symptomatic progression are to be considered progression events for analysis, appropriate definitions should be provided in the protocol and data collected on the case-report forms. Protocol defined assessments of disease burden should be at consistently obtained at intervals that are symmetric between arms. Independent review of imaging may be of value in randomised phase II trials and phase III trials as an auditing tool to detect possible bias.

A final consideration of using PFS or TTP as a primary endpoint is the magnitude of difference between rates of progression between standard and experimental treatment that is considered important. The use of progression rates as primary endpoints may either limit sample size or allow the detection of smaller difference in progression rates than if trials were using overall survival as the primary endpoint. In designing the trial, it is important to consider that the magnitude of benefit should be both of clinical and statistical significance and sufficient to warrant the additional efforts and expenditure to determine the dates of progression in a reasonably accurate and unbiased fashion.

Conflict of interest statement

None declared.

Acknowledgements

The authors thank Carla Kurkjian, MD and Grant Williams, MD for careful review and comment on the manuscript.

REFERENCES

1. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007;**25**:5218–24.

2. Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol* 2008;**26**:1987–92.
3. US Department of Health and Human Services FaDA. *Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics*. <http://www.fda.gov/cder/guidance/7478fnl.pdf>; 2007 [accessed 19.08.08].
4. (CHMP) CfMPfHU. *Guideline on the evaluation of anticancer medicinal products in man*. European Medicines Agency 7 Westferry Circus, Canary Wharf, London E14 4HB, UK; 2005.
5. Bogaerts J, Ford R, Sargent D, et al. Individual patient data analysis to assess modifications to the RECIST criteria. *Eur J Cancer* 2009;**45**:248–60.
6. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumors: Revised RECIST guideline (version1.1). *Eur J Cancer*, 2009;**45**:228–47.
7. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *J Natl Cancer Inst* 2007;**99**:428–32.
8. Carroll KJ. Analysis of progression-free survival in oncology trials: some common statistical issues. *Pharm Stat* 2007;**6**:99–113.
9. Kane R, Yang P. *FDA presentation: genasense (Oblimersen) for metastatic melanoma* <http://www.fda.gov/ohrms/dockets/ac/04/slides/4037s1.htm>; 2004 [accessed 19.08.08].
10. Ruan PK, Gray RJ. Sensitivity analysis of progression-free survival with dependent withdrawal. *Stat Med* 2008;**27**:1180–98.
11. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol* 2008;**26**:3791–6.
12. Stone A, Wheeler C, Carroll K, Barge A. Optimizing randomized phase II trials assessing tumor progression. *Contemp Clin Trials* 2007;**28**:146–52.
13. Stone A, Wheeler C, Barge A. Improving the design of phase II trials of cytostatic anticancer agents. *Contemp Clin Trials* 2007;**28**:138–45.