available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Progression-free survival as an end-point in clinical trials of biotherapeutic agents

Lothar Bergmann[a,*], *Steven Hirschfeld*[b], *Charles Morris*[c], *Sergio Palmeri*[d], *Andrew Stone*[e], *Biotherapy Development Association (BDA)*

[a]*Medical Clinic II, Cancer Center Rhein-Main, J.W. Goethe University, Theodor-Stern-Kai 7, D-60590 Frankfurt, FRG, Germany*
[b]*Clinical Research, National Institute of Child Health and Human Development, 31 Center Drive, Room 2A03, MSC 2425, Bethesda, MD 20814, USA*
[c]*AstraZeneca International, Wilmington, Delaware, USA*
[d]*Chair of Medical Oncology, University of Palermo, Via del Vespro, 129, 90127 Palermo, Italy*
[e]*Oncology-Statistical Expert, AstraZeneca Ltd., Parklands SE2 A/2 Alderley Park, Macclesfield, Cheshire SK10 4TG, UK*

ARTICLE INFO

Keywords:
Anticancer agents
Biologics
Drug approval
Drug registration
End-point
Progression-free survival
Surrogate
Targeted agents

ABSTRACT

Progression-free survival (PFS), the time from registration or randomisation of a patient until objective disease progression or death, can be considered as an outcome for clinical research and also as a basis for regulatory approval. Current experience suggests that greater standardisation and consistency are needed for clinical trials utilising PFS end-points. To this end, the Biotherapy Development Association (BDA) convened a breakout session on the topic of PFS during its Third Alpine Meeting held 14–16 March 2007. Representatives of the pharmaceutical industry, regulatory agencies, academia, and patient advocacy groups identified challenges, developed recommendations, and worked to build consensus regarding the conduct of clinical trials utilising PFS as an end-point to help speed new targeted biologics to the patient bedside.

1. Introduction

Overall survival (OS) is recognised by both the European Medicines Agency (EMEA) and the US Food and Drug Administration (FDA) as the gold standard for clinical benefit in oncology clinical trials. However, trials relying upon OS endpoints can be long in duration and require large numbers of patients. Additionally, the formal demonstration of the underlying survival benefit of a particular therapy can be diluted or confounded by effective later line therapies. Clinical benefit may also manifest in ways other than an increase in survival – for example, pain relief or improved quality of life or delaying progression may be clinically beneficial even if overall survival is not impacted.

Endpoints based on disease progression include disease-free survival (DFS) in either the adjuvant setting or for complete responders, and time to progression (TTP), and progression-free survival (PFS), which can apply in many settings. Typically, disease progression is assessed via medical imaging (e.g. computed tomography, magnetic resonance imaging, positron emission tomography) or other modalities at scheduled intervals.

Several methodological approaches exist regarding the optimum assessment of progression endpoints and no consensus exists on how to minimise any bias in the assessment. For this reason, the Biotherapy Development Association (BDA) convened a breakout session during its Third Alpine Meeting held 14–16 March 2007 to specifically address some

* *Corresponding author*: Tel.: +49 69 6301 5121; fax: +49 69 6301 3876.
E-mail address: L.Bergmann@em.uni-frankfurt.de (L. Bergmann).
1359-6349/$ - see front matter
doi:10.1016/j.ejcsup.2007.09.002

of the issues around the use of PFS as an end-point in clinical trials of anticancer therapies.

Several different progression endpoints are reported, the most widely used regulatory end-point is progression-free survival (PFS), which is defined by the EMEA as 'the time from randomisation (or registration, in non-randomised trials) to objective tumour progression or death from any cause'.[1] The FDA defines the term similarly: 'the time interval between study enrolment or randomisation and the documentation of disease progression based on a prospective definition or death (although all causes of death are counted in the analysis, the interpretative assumption is that death is due to disease progression)'.[2] The US National Cancer Institute's definition of PFS incorporates the notion of probability by stating that PFS 'refers to the probability that a patient will remain alive, without the disease getting worse'.[3] The terms TTP and PFS are often used interchangeably but in general they differ in terms of the statistical handling of patients who die in the absence of progression, counted as events at the time of death for PFS or censored for TTP.

2. PFS: a regulatory view

Survival is considered the most reliable cancer end-point and, when studies can be conducted to adequately assess survival, it is the end-point generally preferred by regulators. As part of its Critical Path Initiative,[1,4] the FDA joined the US National Cancer Institute, the pharmaceutical industry and academia in efforts to identify surrogate markers and validate endpoints for clinical trials. To date, overall response rate (ORR) has been the most commonly used surrogate end-point in support of accelerated approval.[1,5] In May 2007, the FDA issued its final version of Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics,[1,6] which offers recommendations to sponsors on endpoints for cancer clinical trials submitted to the FDA. Specifically, the guidance document indicates that PFS may serve as a surrogate for accelerated approval or as a basis for regular approval and notes that randomised studies are essential, blinding is preferred, and blinded review is recommended.

The FDA identifies several advantages of using PFS as an end-point. PFS is a measure of activity that can be assessed prior to change in therapy in patient populations with either responding or stable tumours, thereby avoiding confounding by subsequent therapies—a common problem in trials based on survival endpoints. Usually trials with PFS endpoints based on imaging or other objective assessments encounter fewer problems with missing data than do trials based on patient reported outcomes such as symptom endpoints. Perhaps most important, PFS can be assessed earlier, and in smaller studies of shorter duration, compared with survival endpoints. This advantage can help ensure that patients have access to effective treatments earlier while minimising the number of patients exposed to potentially ineffective therapies.

The FDA guidance, however, cautions that PFS has disadvantages. First, progression is in the eye of the beholder; consider that although a variety of assessment tools and measurement criteria are available, regulators have not established standards for defining tumour progression. Instead, the agency relies on sponsors to define tumour progression in the protocol according to specified criteria. Regulators generally recommend a central, blinded process for verifying tumour end-point assessments who are blinded to study treatment. Second, PFS is not necessarily a direct measure of benefit nor a validated surrogate for survival. Third, PFS cannot be as precisely measured as survival endpoints and can be subject to assessment bias. Fourth, PFS trials usually entail frequent assessments such as radiologic or other imaging studies, and the resulting data are voluminous and complex compared to survival data. Fifth, the interpretation can be complicated by missing data. The FDA recommends assigning the progression date to the earliest time when any progression is observed without prior missing assessments and censoring at the date when the last assessment determined a lack of progression. Finally, PFS is an estimate of a continuing process based on intermittent sampling; thus to be interpretable, PFS assessments must have balanced timing and be symmetric among treatment arms.[7]

Table 1 provides four examples of drugs approved by the FDA based primarily on trials relying on PFS endpoints. PFS data have been reported in other applications, but these are four recently licensed agents for which approval was primarily based on studies using prospectively defined PFS endpoints.

The EMEA, in its draft guideline, deems it appropriate to use PFS as an end-point in phase III trials for registration if it measures clinical benefit, as might be the case when further lines of therapy modify OS or when PFS is much smaller than OS and no major differences in toxicity can be documented between the compared therapies. In all the cases, EMEA underscores the importance of establishing the definition of a progression event within the protocol through clear, prospectively defined, objective criteria.

Table 1 – Drugs approved by the FDA based on trials using PFS endpoints

Agent	Year	Indication	Study design	N	PFS for agent v. control
Sorafenib (Nexavar)	2005	2nd line renal cancer	Placebo design	769	167 versus 84 days
Gemcitabine (Gemzar)	2006	2nd line ovarian cancer	Add-on	356	8.6 versus 5.8 months (about 258 versus 174 days)
Panitumumab (Vectibix)	2006	2nd line colorectal cancer	Add-on	463	96 versus 60 days
Lapatanib (Tykerb)	2007	2nd line HER2 overexpressing breast cancer	Add-on	399	27.1 versus 18.6 weeks (about 190 versus 130 days)
Mean value				497	Mean PFS advantage = 66 days

In situations where PFS is used as the primary end-point in a confirmatory trial to support a marketing authorisation application, the EMEA advises trialists to seek scientific advice from the Committee for Medicinal Products for Human Use (CHMP) on methodologic issues related to the definition and assessment PFS or DFS, and handling of deviations.[2]

3. PFS: the clinician's view

Many new anticancer agents (e.g. tyrosine kinase inhibitors) are not cytotoxic; rather, they are cytostatic. Because the predominant goal of such therapies is not tumour shrinkage but inhibition of tumour growth, establishing their safety and efficacy must rely on a different type of end-point, such as PFS, to define patient's benefit.

PFS would be a suitable end-point in phase II trials if responses, principally complete responses, are not inducible or suitable as an end-point or if no surrogate markers for survival exist, as is the case for acute leukaemias. PFS endpoints are also appropriate if no curative treatment is available, or if standard or other therapy options exist. If, however, if the median OS is likely to be less than 12 months, PFS may not be an appropriate primary end-point.

Trials relying on PFS as an end-point have some important limitations, especially those that are not randomised. Because PFS might not be a surrogate for OS, OS should always be included as a secondary outcome until an unambiguous correlation with PFS is established. It is possible that rapid disease progression could occur after cessation of treatment (e.g. rebound after treatment with tyrosine kinase inhibitors), leading to no benefit or poorer survival.

It is important to bear in mind the clinical relevance of increasing PFS by a few days or weeks, acknowledging that patients have different outlooks when it comes to balancing the benefit of extending PFS against the risk of increased toxicities. The most relevant factor for some patients who are weighing therapy options is OS, but others seek to delay progression regardless of a therapy's effect, if any, on OS. Many progressions are silent; that is, symptoms are not exacerbated. In such cases, the progression event might be in-apparent and not clinically significant to the patient.

4. Measuring progression-free survival

Among the issues that must be addressed when designing studies using PFS as an end-point are: (1) the definition of clinical benefit, (2) the effect size that would be required to see a difference in clinical benefit and (3) methodologic considerations to avoid bias, such as the use of independent review. It is essential that PFS be assessed via randomised studies per the FDA's recently issued final guidance.[3] Blinded trials are preferred, and blinded review is recommended by some Health authorities, although the additional value from blinded review remains a topic of discussion.

The EMEA cautions clinical trialists about relying on endpoints that involve censoring data from subjects who withdraw before it has been determined that their disease progressed. The principles of intention-to-treat (ITT) should be followed as far as possible for the primary analysis of PFS/DFS. In particular, for all randomised patients, outcome data should be collected according to the intended schedule of assessment, and the date of progression or recurrence should be assigned based on the time of the first evidence of objective progression or recurrence regardless of violations, discontinuation of study drug or change of therapy. If, for a particular study, analysis in accordance with the principles of ITT is considered inappropriate, a justification is expected and European Union regulatory agreement is recommended at the planning stage.[7]

The decision to use PFS as a primary end-point should be made carefully when designing clinical trials, and investigators focused on a particular disease should develop consensus standards and strive for consistent surveillance intervals.[8]

5. PFS in phase II decision making

Historically, clinical development of cytotoxic agents involved a single-arm phase II evaluation of response rates. This approach, however, is generally inadequate for determining the efficacy of cytostatic agents. In addition to identifying active therapies, the objective of a phase II trial should be extended to include identifying those agents that are likely to be successful in later phase or confirmatory trials.[9,10] PFS can be considered as a basis for go/no-go decisions in development of new anticancer agents because it can provide preliminary evidence of efficacy in modestly sized randomised phase II studies, supply important information about dosing and provide an initial estimate of effect as the basis of statistical design for subsequent phase III studies.

For example, phase II time-to-progression data helped predict phase III survival outcomes in a trial of paclitaxel (Taxol) plus carboplatin with or without bevacizumab (Avastin) in non-small-cell lung cancer. The phase II study involved only 67 patients, but the results were strongly predictive of the phase III finding that the addition of bevacizumab translated into clinical benefit and increased survival for patients.[11]

A randomised discontinuation design based on a PFS endpoint was used to determine activity and provide supportive clinical efficacy information of sorafenib (Nexavar) for metastatic renal cell carcinoma.[13] Under this scheme, all patients initially receive the study drug. After a defined period, patients are evaluated, and those who respond to treatment continue on the treatment, those whose disease progresses are taken off the study and patients with stable disease are randomised to either continue drug or to observation/placebo for specified time period. The randomised discontinuation design is not as efficient as up-front randomisation if the treatment has a fixed effect on tumour growth rate or if clinical benefit is restricted to slower-growing tumours. On the other hand, the randomised discontinuation design can be advantageous if only a subset of patients expressing the molecular target is sensitive to the agent. In the case of sorafenib, a marked improvement in PFS was seen in the group receiving the sorafenib-containing regimen.[12,13] Consequently, patients who were receiving placebo and whose disease was progressing were allowed to crossover to the sorafenib group.

Randomised, blinded phase II studies are ideal ones, but an alternative approach is to proceed directly to a phase III study and embed the phase II component at the front end with a futility analysis. This approach, although sound, does entail some logistical and statistical challenges, and the sponsor would be ethically committed to completing the phase III study. Confidence intervals would be large until some estimates about sample size and power are available. Such a phase II/III trial would still need a go/no-go decision point, and the phase III portion would have to be powered sufficiently to achieve objectives.

PFS has an important and growing role in robust phase II decision making. Phase II extension of PFS from within randomised, controlled trials consistently seems predictive of phase III PFS, with no known published counter examples. Increasingly, pharmaceutical firms are committing to modestly sized phase II studies with PFS endpoints to increase confidence as their new agents enter phase III. For example, based on modelling a randomised trial of just 100 patients could lead to the termination of development of 90% of inactive agents, whereas at least 80% of agents with a meaningful and realistic increase in PFS would be tagged for an extension or separate confirmatory study. Randomised studies with PFS endpoints could be a powerful and economical means of establishing the clinical efficacy of new cytostatic agents.[13]

6. PFS design solutions

Methodologic solutions can enhance the development decision at the end of phase II in the context of progression endpoints used in randomised trials.[14]

Because PFS is assessed via radiologic studies at periodic intervals or after a fixed number of treatment cycles, the date of the radiologic evaluation at which progression is detected serves as a proxy for the true progression time, which actually lies somewhere within the time interval between two assessments. The result is interval-censored data. The following comments apply only to uncontrolled or poorly designed studies. Reliance upon PFS endpoints can introduce other sources of bias that must be controlled for or minimised. One example is time assessment bias, which can occur if patients in one group are assessed more frequently or earlier than the other. This can also be problematic if there are more unscheduled assessments in one group compared to the other or the protocolled assessment frequency differs between arms.[15]

Concerns over time assessment bias are greatest if trials are not properly blinded but alternative design and analysis strategies may be employed. One approach is comparison of PFS rates between study arms at a predefined time point. A fixed time point may allay concerns of time assessment bias, but it has some limitations as well. Ignoring events that occur after some arbitrary time point tends to reduce statistical power and concern may exist over the extrapolation of results taken from a single point in time. One possibility is to compare two landmarks, as described recently by Freidlin et al.[15] Whitehead et al.[16] have described an approach that extends this approach and groups events into time intervals between scheduled assessments.

Another strategy to minimise bias is event count, which simply involves counting events that have occurred by a given time point regardless of how long the patients have been on the trial.[17] The event count approach captures all observed events and could be more powerful than PFS assessed at a fixed time point and removes the potential for time assessment bias. In general, it is better to use progression endpoints that make use of all available progression data rather than early, fixed-time-point analyses.

Many investigators are uncertain about how often radiologic assessments should be made. In general there is little to be gained by assessing progression status any more frequently than would be required in routine clinical practice and specifically, if patients are assessed at a frequency that is 50% of the comparator median, there is a negligible increase in power in assessing patients more frequently.[18]

Censoring—a statistical manoeuvre for handling data from patients whose disease does not progress—can also introduce bias. It is critical to consider censoring and the assumptions behind it carefully because it reduces the power of analysis, which depends on the number of events. Every time one censors a patient, power is sacrificed. Trialists must prospectively define censoring rules. Bias can result if the rate of censoring differs between the arms and is related to patient prognosis. Trials with PFS endpoints are also jeopardised by missed patient visits, which makes it difficult to ascertain when disease progression occurred. Other questions related to censoring arise when patients stop therapy or start another therapy prior to progression. Rates of such occurrences are likely to depend on relative tolerability of the study drug. Censoring those patients' data could cause bias because those who switch treatments are likely to be the ones who are doing poorly on the study regimen. One simple solution is to follow all patients to progression regardless of intervening events.

Ascertainment bias is also a potential problem in clinical trials that rely upon radiologic assessment of disease progression particularly in inadequately blinded trials. Measurement of tumour dimensions and choice of target lesions is not an exact science, and the resulting data can pose analytical problems, as well. However, an open question is whether an independent, central review is more likely to get closer to the true relative benefits of therapy in a blinded trial. Particularly problematic is the handling of patients who the investigator believes has progressed but the independent reviewers do not. The analysis of the independent reviewers' data has no alternative to censor such patients and thus the manoeuvre that was intended to remove bias can introduce new sources of bias as described in the previous paragraph.

Nevertheless, if there is an effective therapy, the nature of the analysis would have a negligible impact on the study's conclusion; the success will not be disguised according to either investigator or independent assessment.

In summary, for clinical trials with PFS as an end-point, randomised trials should be the norm and blinded trials are strongly preferred. Alternative endpoints such as fixed time points or event counts can be used when blinding is a concern. Patients should not be assessed any more frequently than in routine clinical practise. Trials involving PFS endpoints must be statistically robust and adequately powered.

7. Conclusions

Several points of consensus can be made:

There is increased interest and utilisation in progression-free survival by clinical trialists, sponsors and regulatory authorities.

The BDA panellists strongly advocated randomisation and blinding (where feasible) in clinical trials, including phase II studies, that use PFS endpoints.

Recently published information indicates that it is not necessary to perform imaging studies at intervals more frequently than standard clinical practice to retain sufficient statistical power. Normal clinical monitoring yields sufficient statistical power under most scenarios.

Regulators, sponsors and patient advocates who participated in the BDA discussions questioned the cost-effectiveness of independent review for studies with effective blinding. Independent reviews have proved to be logistically challenging, costly and have not been demonstrated to alter the overall conclusions about whether an intervention is effective. The panel speculated that the considerable resources required for such reviews could be put to better use in the quest to develop and test new anticancer agents for safety and efficacy.

The BDA panellists advocated that PFS be considered more than just a radiologic outcome. PFS could be a composite of two or more parameters, including patient symptoms (e.g. neurologic cancers) or patient-reported outcomes.

The BDA panellists called for greater standardisation to ensure that trials based on PFS endpoints are done in meaningful and consistent ways that will lead to development and approval of new anticancer agents. Specifically, they identified these issues for further development:

1. Collaborative work is needed to consider areas where expansion of the definition of disease progression from tumour measurements to incorporate clinical status and symptom control would be beneficial. By integrating these two facets, we could better understand what PFS means in terms of patient benefit.
2. More standardisation is warranted regarding sensitivity analysis and data censoring to avoid bias. It is important to retain the principles of ITT to capture valuable data on study-withdrawal events.
3. The role of PFS as a robust phase II decision-making endpoint must be considered. Some studies have relied upon PFS endpoints for phase II studies as a basis for go/no-go decisions, but further standardisation and refinement could enhance its utility in this regard by allowing a preliminary estimate of effect size for later studies.
4. Explore use of alternative methodolgies such as fixed-time-points, multiple fixed time points and event counts for PFS, in clinical trials.
5. Due to second-line therapies becoming more effective, the BDA panellists recommended exploring the relationships of lines-of-treatment, indications and appropriate endpoints. For example, would a clinical trial of a compound as a first-line agent use a different end-point than a second-line study? In the United States, at least, the vast majority of submissions are for advanced cancers, and very few address untreated populations. No matter what the setting is, however, the unifying principle is that the patient must have clinical benefit.
6. Some therapies lead to marked effects on PFS and appear to be effective in achieving stable disease or slowing tumour growth, although they might not be associated with large numbers of conventional responses or prolongation of overall survival. The ultimate clinical question is this: is it always good to delay progression or, conversely, is it always bad to hasten progression?

Certainly, flexible approaches are needed, the BDA panellists unanimously agreed that continued collaboration of pharmaceutical companies, regulators, academia, clinicians and patient advocates is the best way forward to increase consistency, streamline clinical trials utilising PFS as an end-point and speed new anticancer biologics to the patient bedside.

Conflict of interest statement

The four authors of this paper can confirm that there is no conflict of interest involved in this paper, nor in their participation in this entire event.

REFERENCES

1. Committee for Medicinal Products for Human Use (CHMP). CPMP/EWP/205/95/Rev.3. Guideline on the Evaluation of Anticancer Medicinal Products in Man. 14 Dec 2005. London, England. <www.emea.europa.eu/pdfs/human/ewp/020595en.pdf> [accessed 25.03.2007].
2. Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. 2007 May. Rockville (MD): US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. <http://www.fda.gov/cder/guidance/7478fnl.pdf> [accessed 14.05.2007].
3. National Cancer Institute, U.S. Department of Health and Human Services. Dictionary of Cancer Terms. <www.cancer.gov/Templates/db_alpha.aspx?CdrID=44782> [accessed 24.03.2007].
4. <www.fda.gov/oc/initiatives/criticalpath/>.
5. Luri AJ, Djulbegovic B, Nebeker JR, Angelotta C, Gordon LI, Raisch DW, et al. Do post-approval phase III trials for accelerated approved cancer drugs violate equipoise? *J Clin Oncol, ASCO Proc* 2006;**24**:6026.
6. <http://www.fda.gov/cber/gdlns/clintrialend.pdf>.
7. Palmeri S, Bergh J, Bergmann L, Hansen SW, Moraleda JM, Marty M, et al. The European medicines Agency/EMEA) guidelines on oncology drug development. JCO 2007, 25 [abstr 6636].
8. 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.
9. Stone A, Wheeler C, Barge A. Improving the design of phase II trials of cytostatic anticancer agents. *Contemp Clin Trials* 2007;**28**:138–45.

10. Eisenhauer EA. Phase I and II trials of novel anti-cancer agents: Endpoints, efficacy and existentialism. *Ann Oncol* 1998;**9**:1047–52.
11. Sandler AB, Gray R, Perry MC, et al. Randomized phase II/III trial of paclitaxel (P) plus carboplatin (C) with or without bevacizumab (NSC #704865) in patients with advanced non-squamous non-small cell lung cancer (NSCLC): An Eastern Cooperative Oncology Group (ECOG) Trial–E4599. ASCO 2005 [abstract LBA4].
12. Escudier B, Eisen T, Stadler WM, et al. TARGET Study Group. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med* 2007;**356**:125–34.
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.
14. Stone A, Wheeler C, Carroll K, Barge A. Optimizing randomized phase II trials assessing tumor progression. *Contemp Clin Trials* 2007;**28**:146–52.
15. Freidlin B, Simon R. Evaluation of randomised discontinuation design. *J Clin Oncol* 2005;**23**: 5094–8.
16. Whitehead J. The analysis of relapse clinical trials, with application to a comparison of two ulcer treatments. *Stat Med* 1989;**8**:1439–54.
17. Carroll KJ. Analysis of progression-free survival in oncology trials: Some common statistical issues. *Pharmaceut Stat* 2007;**6**:99–111.
18. Stone A, Fernandes N, Morris C. Myths in the assessment of progression. How frequently should we assess patients in clinical trials? *Proc Am Soc Clin Oncol*:6521.