

The Clinical Viewpoint: Definitions, Limitations of RECIST, Practical Considerations of Measurement

Liza C. Villaruz and Mark A. Socinski

Abstract

In selecting an endpoint in clinical trial design, it is important to consider that the endpoint is both reliably measured and clinically meaningful. As such, overall survival (OS) has traditionally been considered the most clinically relevant and convincing endpoint in clinical trial design as long as it is accompanied by preservation in quality of life. However, progression-free survival (PFS) is increasingly more prominent in clinical trial design because of feasibility issues (smaller sample sizes and shorter follow-up). PFS has the advantage of taking into account not only responsive disease, but stable disease as well, an issue of particular importance in the relapsed and refractory setting in which therapies are often associated with a minimal to nil response but may still confer a survival advantage. Finally, PFS has a significant advantage in molecularly selected populations, in whom OS advantages are difficult to detect due to the effects of crossover. With an understanding of the limitations and biases that are introduced with PFS as a primary endpoint, we believe that PFS is not only a viable but also a necessary alternative to OS in assessing the efficacy of selected novel-targeted therapies in molecularly defined cancer populations. Ultimately, the selection of a clinical trial endpoint should not be based on a one-size-fits all approach; rather, it should be based on the specifics of the therapeutic strategy being tested and the population under study. *Clin Cancer Res*; 19(10); 2629–36. ©2013 AACR.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interests were disclosed.

CME Staff Planners Disclosures

The members of the planning committee have no real or apparent conflict of interest to disclose.

Learning Objectives

On completion of this activity, the participant should understand the limitations and advantages associated with the use of progression-free survival (PFS) as an endpoint in large phase II and III randomized controlled clinical trials.

Acknowledgment of Financial or Other Support

This activity does not receive commercial support.

Introduction

The measurement of efficacy in oncology clinical trial design is rooted in a landmark study reported in 1976 by Moertel and Hanley (1). Their study was conducted on the premise that "the culmination of most experimental therapeutic trials for solid tumors occurs when a [physician] places a ruler or caliper over a lump and attempts to estimate its size," and with this measurement

comes the inevitable component of human error. This historical perspective underscores an issue in clinical trial design that has persisted in the interim 36 years, and that is that clinical trials seek to measure that which can be measured reliably (and do so with variable success) and do not necessarily measure that which is clinically meaningful and applicable to everyday clinical practice. Herein lie the fundamental considerations in the selection of a meaningful endpoint in the design of a clinical trial; the endpoint must be reliably measured and clinically meaningful.

Overall survival (OS) has traditionally been considered the most clinically relevant and convincing endpoint in clinical trial design, as long as it is accompanied by preservation in the quality of life (QOL). However, progression-free survival (PFS) is increasingly more prominent in clinical trial design. In a review of clinical trials assessing the efficacy of monoclonal antibody therapies

Authors' Affiliation: University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania

Corresponding Author: L.C. Villaruz, University of Pittsburgh Cancer Institute, Division of Hematology/Oncology, UPMC Cancer Pavilion, 5150 Centre Avenue, Fifth Floor, Pittsburgh, PA 15232. Phone: 412-648-6578; Fax: 412-648-6579; E-mail: VillaruzL@upmc.edu

doi: 10.1158/1078-0432.CCR-12-2935

©2013 American Association for Cancer Research.

to the EGF receptor (EGFR) and VEGF pathways, these targeted agents in unselected populations were associated with marginal benefits (2). In this review, we argue that it is not only the nature of the targeted therapy being tested, but perhaps of more importance, also the disease setting in which it is being studied. The value of PFS is greatest in specific settings where the disease and the drug being tested would preclude detection of an OS benefit, as in highly effective targeted therapies in molecularly selected populations, where crossover to the new therapy is frequent after progression.

Wilkerson and Fojo conducted an analysis of randomized clinical trials in metastatic cancer (all types) and showed a numerical correlation between the magnitude of the difference in PFS with that of OS (3). Our intention in this review is not to assess the validity of PFS as a surrogate for OS but rather to assess its validity as a measure of efficacy in and of itself. The use of PFS applies to both the large randomized phase II clinical trials, in which it functions as a preliminary assessment of efficacy (the go or no-go decision) and in phase III clinical trials for validation of efficacy. Most importantly, the selection of PFS as a primary endpoint should be made only after careful consideration of the limitations in using PFS, the disease setting being studied, and the agent being tested.

Definitions

OS is defined as the time from subject randomization to the time of death from any cause and traditionally represents the most clinically relevant and convincing endpoint in clinical trial design, as long as it is accompanied by a reasonable preservation of QOL (4). The advantage of OS as a clinical trial endpoint is clear: It represents an unequivocal endpoint that can be assessed with 100% accuracy both for its occurrence and timing, provided the number of patients with incomplete follow-up times is not substantial (right censoring; ref. Table 1).

On the other hand, PFS is defined as the time from randomization until objective tumor progression or death (4). It is distinguishable from other endpoints that are commonly used in clinical trial design: Time to treatment failure (TTF) is a composite endpoint measuring time from randomization to discontinuation of treatment for any reason, including progressive disease, treatment toxicity and death; time to progression (TTP) is defined as the time from randomization until objective tumor progression and does not include death; and disease-free survival is typically used in the adjuvant setting where the disease is considered to be eradicated. Overall response rate (ORR) is defined as the proportion of patients who have a partial or complete response to therapy; it does not include stable disease and is a direct measure of drug tumoricidal activity. In contrast, disease control rate is a composite of ORR and stable disease and is useful to measure the efficacy of therapies that have tumoricidal effects rather than tumoricidal effects.

Limitations of RECIST

Response evaluation criteria in solid tumors (RECIST) forms the basis for PFS determination and defines progressive disease as at least a 20% increase in the sum of diameters of up to 5 target lesions (2 lesions/organ), taking as reference the smallest sum on study and an absolute lesion increase of at least 5 mm or the appearance of new lesions (5). A complete response is the disappearance of all target lesions, and a partial response (PR) is defined as at least a 30% decrease in the sum of the target lesions. Stable disease is defined as fitting the criteria neither for progressive disease nor a PR.

The advantage of RECIST is first and foremost that it represents a common language of efficacy for clinical researchers across disease sites and clinical trial settings. The major disadvantage of RECIST is the reliance on human measurement. In the original study by Moertel and Hanley,

Table 1. Advantages and disadvantages of OS and PFS as clinical trials endpoints

Endpoint	Advantages	Disadvantages
OS	<ul style="list-style-type: none"> • Precisely measured • Clinically meaningful when accompanied by a QOL preservation • Assessed on a continual basis 	<ul style="list-style-type: none"> • May be affected by crossover and subsequent therapies • May involve larger studies and longer follow-up • Includes noncancer deaths • Subject to right censoring
PFS	<ul style="list-style-type: none"> • Measurement of stable disease included • Smaller sample size • Shorter follow-up • Generally based on objective and quantitative assessments • Not affected by crossover or subsequent therapies 	<ul style="list-style-type: none"> • Definitions vary among studies • Frequent radiologic or other assessments • Involves balanced timing of assessments among treatment arms • Not precisely measured; subject to evaluation bias particularly in open-label studies • Subject to interval censoring • May be subject to informative censoring due to central review • Not statistically validated as a surrogate for survival in all settings

Downloaded from http://aascijournals.org/clinicalcancerres/article-pdf/19/10/2630/2011700/2629.pdf by guest on 22 May 2024

16 investigators were asked to assess the size of 12 solid spheres ranging from 1.8 to 14.5 cm covered by soft mattresses measuring 0.5 to 1.5 inches in thickness; 2 pairs of these "tumors" were identical in size (1). When the response criteria were set as a 25% reduction in the product of the perpendicular diameters of the tumor, 20% to 25% of the objective responses were erroneous, whereas a response defined as a 50% reduction in the product of the perpendicular diameters of the tumor was associated with a 5% to 10% erroneous objective response rate. The latter were the basis for the first formal response criteria in clinical trial design established by the World Health Organization in 1981 (6). Although the RECIST v1.1 guidelines published in 2009 represent an evolution of these radiographic criteria, they remain fundamentally rooted in a measurement that is prone to human error.

The use of RECIST is limited when tumor measurements on cross-sectional imaging are difficult or uninformative, as is the case in malignant pleural mesothelioma, which grows as a pleural rind rather than as spherical lesions with measurable diameters. The modified RECIST developed in 2004 were specifically designed to address the unique growth pattern of pleural malignant mesothelioma (7). Using these criteria, the tumor thickness is measured perpendicular to the chest wall or mediastinum and in clear relationship to anatomic landmarks to allow reproducible assessments at later timepoints. These criteria were applied to 2 clinical trials assessing the efficacy of cisplatin and gemcitabine in malignant pleural mesothelioma and predicted not only for survival but also for an increase in the forced vital capacity during treatment (7). This is an important consideration in mesothelioma, a disease in which QOL preservation is often defined by our therapeutic ability to control dyspnea.

Traditional RECIST are of limited use in clinical trials assessing immunotherapeutic agents. Anti-CTLA4 immunotherapy with ipilimumab is associated with an unprecedented OS advantage of 10 months in patients with advanced metastatic melanoma and is also associated with 4 distinct response patterns: shrinkage in baseline lesions without new lesions; durable stable disease; a response after an initial increase in the total tumor burden; and a response in the presence of new lesions (8, 9). Although these response patterns were all associated with favorable survival, the latter 2 response patterns represent progressive disease by traditional RECIST. As such, the immune-related response criteria were developed to more accurately reflect the true antitumor activity of immunotherapeutic agents. Using these criteria, new lesions do not necessarily define disease progression, and a determination of progressive disease necessitates a 25% increase in tumor burden by measurements taken 4 weeks apart (9).

Finally, there is the issue of nonmeasurable disease. This includes smaller lesions with diameters less than 10 mm and lymph nodes between 10 to 15 mm, and also truly immeasurable disease by parametric parameters, such as leptomeningeal disease, ascites, pleural or peri-

cardial effusion, inflammatory breast disease, and lymphangitic involvement of skin or lung (5). These potential sites of disease progression might be identified by physical examination, but are not measurable by reproducible imaging techniques.

Practical Considerations in the Use of PFS as a Primary Endpoint

Limitations

Understanding the limitations of PFS is a key to its proper use as a clinical trial endpoint. The BMS-099 clinical trial, which evaluated the addition of cetuximab to platinum-based chemotherapy in the treatment of chemotherapy-naïve advanced stage non-small cell lung cancer (NSCLC), cetuximab was associated with a PFS benefit when determined by the investigators but not when determined by an independent radiologic review committee (IRRC; ref. 10). Interestingly, the ORR by the IRRC was statistically different, but not different when assessed by the investigators.

This discrepancy between the ORR and PFS is due to how tumor measurements are taken (in the setting of the subjectivity of RECIST) and also when these measurements are made (Fig. 1). Assessments may be carried out "out of turn" or more frequently in one arm compared with the other, either from concerns of lack of efficacy or signs of clinical progression, so-called evaluation bias. Evaluation bias is especially prominent in trials such as BMS-099, in which the open-label trial design may have led to a lower investigator-driven threshold for radiographic imaging in the chemotherapy-alone arm. The converse of this is that patients receiving weekly infusions, as is the case with cetuximab, might be evaluated by a physician more frequently, which might trigger more frequent testing for progressive disease. In any clinical trial evaluating a novel therapy or combination of therapies, progressive disease may be reported later in the more toxic arm due to treatment delays or missed assessments from adverse events, a source of bias termed measurement bias.

Patients lost to follow-up or who drop out of a study and who typically do not have the same risk trajectory as those who experience progressive disease may introduce attrition bias. Informative censoring may also be introduced when a patient is censored because of clinical progression without meeting protocol defined progression, symptomatic deterioration, or toxicity, that is, the censoring time may be a reflection of a poor clinical outcome, thus biasing the resulting analysis (Fig. 2). Sridhara and colleagues in the *Focus* section of *CCR* point out that the judicious use of sensitivity analyses can provide some reassurance of the observed results and how sensitive the results are to missing data assumptions (11).

The use of sensitivity analyses is typified by the FLEX clinical trial, which showed that the addition of cetuximab to platinum doublet chemotherapy was associated with a 1-month OS benefit, but an identical PFS (TTP or death) to that of chemotherapy alone (12). A *post hoc* sensitivity

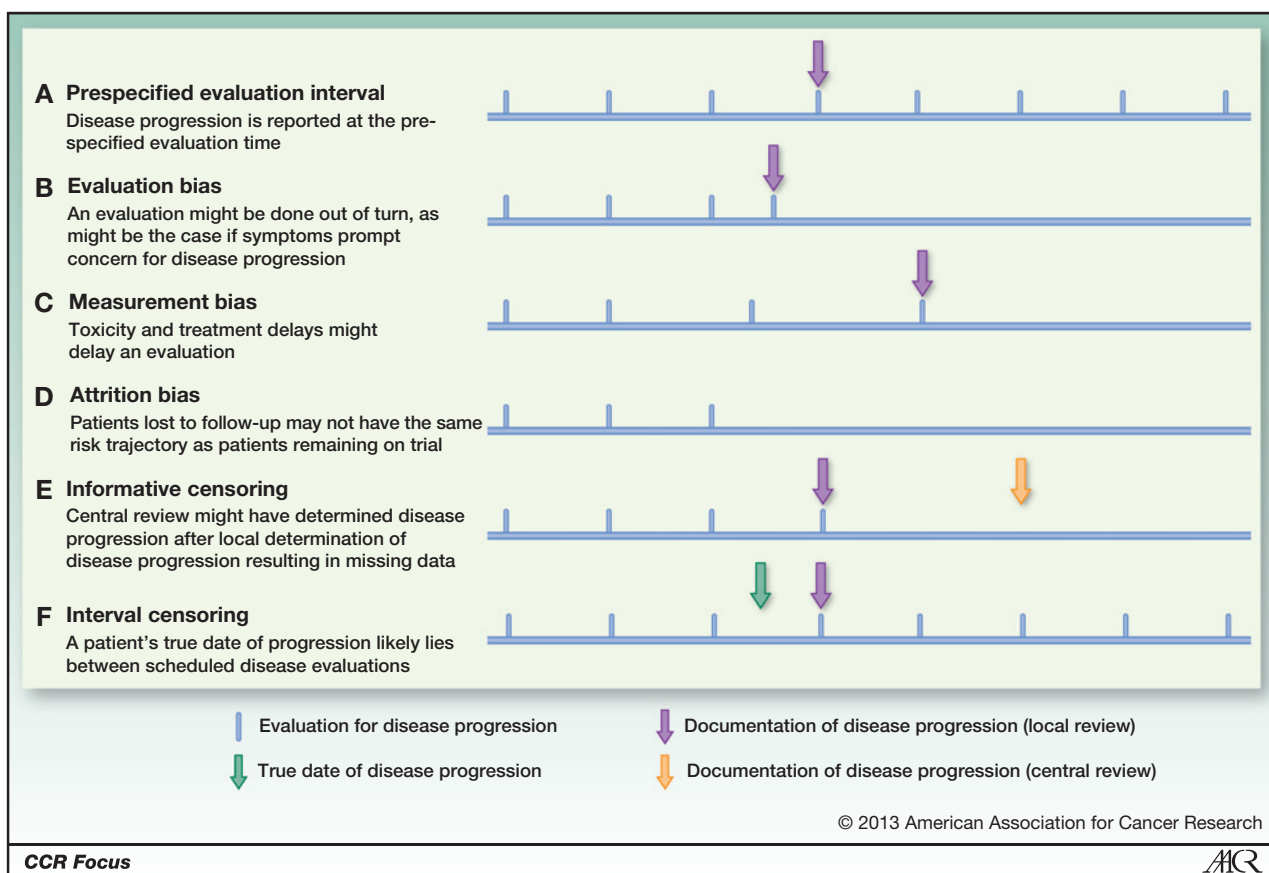


Figure 1. Bias and censoring in clinical trials. A, if disease evaluations are conducted at the correct time, the date of a patient's disease progression will coincide with the prespecified evaluation interval. However, this documented date of the disease progression is susceptible to (B) evaluation bias and (C) measurement bias. D, patients lost to follow-up may not have the same risk trajectory as those remaining on trial, resulting in attrition bias. Results can also be biased by (E) informative and (F) interval censoring.

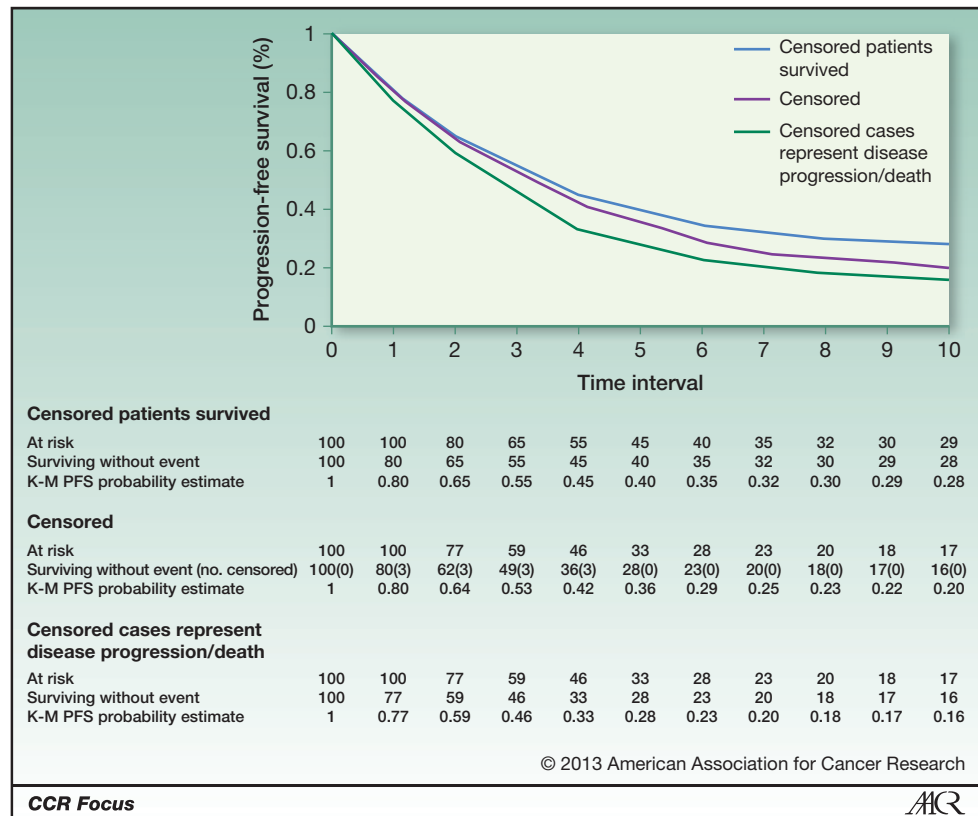
analysis showed that TTF (time to discontinuation of treatment from any cause) was statistically, albeit modestly prolonged (by about 2 weeks) with the addition of cetuximab to chemotherapy. This subtle distinction was the result of more patients in the chemotherapy-alone arm having started "salvage therapy" without documented disease progression or toxicity. This gives an investigator pause; slight changes in an endpoint definition can be the difference between a non-statistically significant and a statistically significant result.

Blinded independent central review (BICR) may abrogate the issues of bias revolving around PFS and has been advocated by regulatory agencies. However, BICR is most relevant if it can happen in real time, and in reality, it takes place in a largely retrospective manner. The review of all the scans for all patients participating in a clinical trial is not only expensive and time-consuming; it also does not eliminate the bias of informative censoring (Fig. 1). BICR may in fact exacerbate informative censoring, as the IRRC may overturn a local progression call and cause inappropriate censoring of a patient's data, thus biasing the resulting survival analysis, a scenario illustrated by Sridhara and colleagues (11). Auditing strategies

of selected cases rather than a complete case review, as described by Zhang and colleagues in this *CCR Focus* section, is an alternative to BICR and potentially feasible as a form of real-time review to abrogate the effects of informative censoring (13).

Another consideration is that an analysis that uses the detection date as the date of progression introduces interval censoring and invariably results in an overestimation of PFS (Fig. 1; ref. 14). In a study conducted by Qi and colleagues, individual patient data from 14 North Central Cancer Treatment Group (NCCTG) clinical trials and 7 Southwest Oncology Group (SWOG) clinical trials were used to determine the date of disease progression based on 4 common methods: the reported disease progression date (M1), 1 day after the last progression-free scan (M2), the midpoint between the last progression-free scan and the date of reported disease progression (M3), or interval censoring (M4; ref. 15). The magnitude of difference in the PFS estimates as determined by these 4 methods was large enough to alter trial conclusions in patients with advanced lung cancer; the midpoint method was the least influenced by the assessment schedule. Principally, the use of sensitivity analyses

Figure 2. Kaplan–Meier (K–M) curves for progression-free survival illustrating the bias associated with censoring. The middle curve represents the "real-world" scenario in which censored patients are considered to have the same risk trajectory as surviving patients; top, the curve that would have resulted if censored patients actually survived without event; bottom, the curve that would have resulted if censored patients actually experienced an event (progression or death) in the time interval in which they were censored.



is a key in evaluating the robustness of conclusions based on PFS.

Advantages

PFS, first of all, has the advantage of taking into account not only responsive disease, but stable disease as well (Table 1; ref. 16). This is particularly important in the relapsed and refractory setting, where therapies are often associated with a minimal to nil response but may still confer a survival advantage, as is illustrated in the relapsed metastatic NSCLC setting. In the randomized phase III clinical trial that contributed to the U.S. Food and Drug Administration (FDA) approval of docetaxel in previously treated NSCLC, docetaxel was associated with an ORR of 6% (all PRs) and a stable disease rate of about 40%. Despite the modest response rate, this translated to a TTP prolongation of almost 4 months and an OS prolongation of 2.4 months compared with best supportive care (17). An assessment of efficacy based on ORR alone would not have been an accurate reflection of the survival advantage associated with second-line docetaxel therapy. Second, clinical trials that use PFS require smaller sample sizes and shorter follow-up compared with studies that use OS. This has important implications with regard to the feasibility of conducting clinical trials both with accrual and the financial burden associated with a high sample size requirement.

Finally and perhaps the most important point, PFS, in contrast with OS, has the advantage of not being affected by

crossover or subsequent therapies (16, 18–20). It is critical to note that OS has the potential to be attenuated by any second-line therapy, but the degree to which it is attenuated is directly related to the population under study and the effectiveness of subsequent lines of therapy in modulating survival postprogression (SPP). This is illustrated in Fig. 3, which represents selected phase III clinical trials in lung cancer. E4599 was a phase III clinical trial of unselected nonsquamous NSCLC treatment-naïve patients in whom the addition of bevacizumab to platinum-based chemotherapy was associated with about a 2-month survival prolongation in both PFS and OS (21). Similarly, BR21 was a phase III clinical trial of unselected patients with previously treated NSCLC in whom erlotinib was associated with about a 2-month benefit in both PFS and OS (22). In both trials, patients were limited with regard to the number of postprogression therapies; therefore the degree to which OS is attenuated is small and an OS advantage was detected with the trial interventions.

On the other hand, in molecularly selected populations for whom targeted therapies are tested, our ability to detect an OS advantage is severely attenuated by the effects of crossover. In the EURTAC clinical trial, patients with activating mutations in *EGFR* were treated with targeted therapy with erlotinib (23). This was associated with a 4-month prolongation in PFS, but no benefit in OS (Fig. 3). Seventy-seven percent of patients on the chemotherapy arms of these trials went on to receive *EGFR*-targeted therapy at the time of progression. Four points are

Downloaded from http://aacrjournals.org/clinccancerres/article-pdf/19/10/2633/2011700/2633.pdf by guest on 22 May 2024

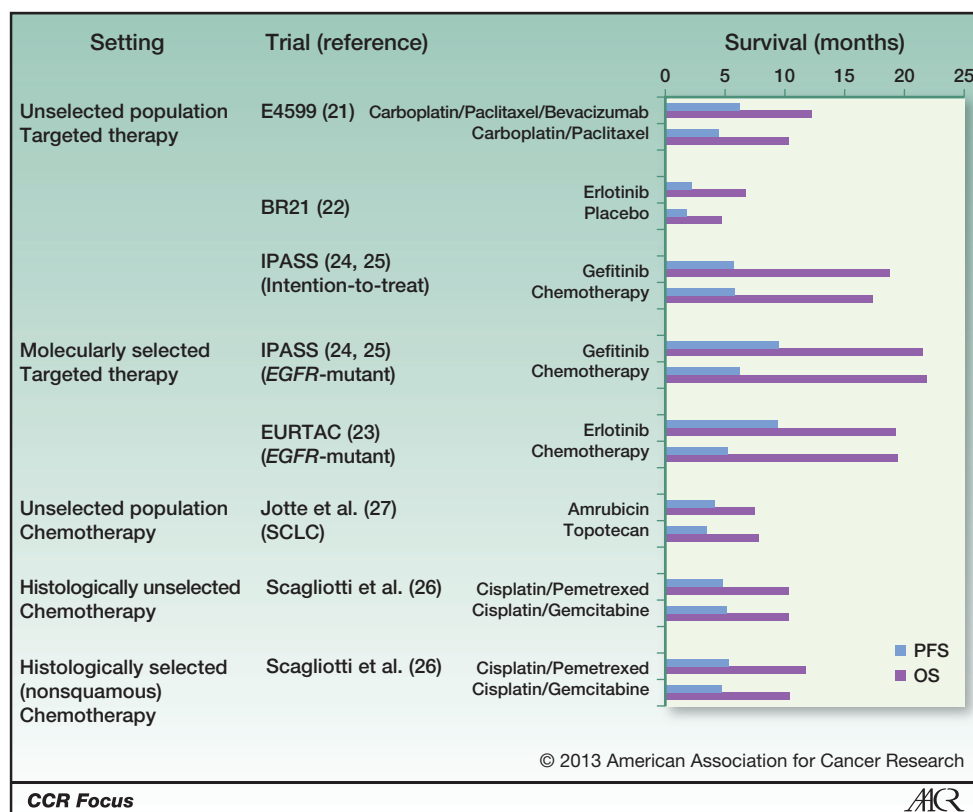


Figure 3. PFS and OS in selected phase III clinical trials in lung cancer. The ability to detect either a PFS or an OS advantage is dependent on the nature of the therapy being tested and the selection criteria for the population under study. SCLC, small-cell lung cancer.

underscored by this clinical trial: (i) crossover is a critical component of clinical trial design both as an ethical necessity and a tool for investigators to maximize accrual; (ii) not every patient in a clinical trial has the opportunity to cross over to the targeted agent (in this trial 23% patients did not cross over), as is the case with patients who have rapid progression of their disease whose performance status might preclude further therapy; and (iii) the assumption that an OS advantage will be negated in a crossover trial design requires that the SPP associated with a targeted therapy must be substantially large enough to blur the OS advantage (the effect of the targeted therapy must be similar regardless of whether it is given in the first or second line). In molecularly selected patients in whom targeted therapies are equally effective in the first or second line, the efficacy of the targeted therapy is such that an OS benefit is not only attenuated but rather obliterated by the effects of crossover.

The fourth lesson from the EURTAC clinical trial is also underscored by the IPASS clinical trial, and it is that the selection of an endpoint is dependent not only on the drug or the regimen being tested but also highly dependent on the population under study. The efficacy of gefitinib is lost in the unselected IPASS intention-to-treat population and is found again in the preplanned analysis of patients with activating *EGFR* mutations, in which a 3-month PFS benefit was seen with gefitinib over chemotherapy (Fig. 3; refs. 24, 25). As with the other trials of *EGFR* tyrosine kinase inhibitor therapy in the *EGFR*-mutant population, no

benefit was seen in OS; 64% of *EGFR*-mutant patients who received chemotherapy in the first-line setting went on to receive *EGFR* TKI therapy postprogression.

The selection of a population can also be based on histology, as with the Scagliotti and colleagues clinical trial, in which a pemetrexed-based platinum doublet chemotherapy regimen was associated with neither a PFS nor an OS benefit in an unselected population but was associated with a survival benefit in the subpopulation of NSCLC of nonsquamous histology (Fig. 3; ref. 26). About 50% of the patients on each treatment arm went on to receive postdiscontinuation therapy, with a higher proportion of patients in the non-pemetrexed arm going on to receive pemetrexed in the second line. In this instance, OS is not attenuated by postprogression therapies for 2 reasons: The proportion of patients who went on to receive second-line therapy was not as high as in molecularly driven clinical trials, and the effectiveness of second-line pemetrexed therapy was not great enough to completely attenuate the OS advantage associated with first-line pemetrexed therapy.

In contrast, in recurrent or relapsed small-cell lung cancer, amrubicin was associated with a paltry 2-week benefit in PFS, which was associated with no benefit in OS, reflective of minimal treatment effect and an underlying disease state in which all therapies to date are associated with marginal benefit (Fig. 3; ref. 27). In short, the validity of PFS as a primary endpoint depends on the drug being tested and the clinical and

molecular characteristics of the cancer population under study (28).

One final point is the necessity of incorporating QOL measures in the assessment of therapeutic efficacy. In the ZODIAC clinical trial that evaluated the addition of the multityrosine kinase inhibitor vandetanib to docetaxel as second-line therapy in advanced NSCLC, vandetanib was associated with a statistically significant prolongation in PFS by about a month, no improvement in OS, and a statistically significant increase in time to symptom deterioration (vandetanib, 3.5 months vs. placebo, 2.7 months; ref. 29). In treating patients with advanced cancer, it is important to consider that a benefit in OS is only clinically meaningful if it can be achieved with a reasonable preservation in QOL. Along the same lines, in patient populations in which a therapy is not associated with an OS advantage, a clinically meaningful endpoint might include a prolongation in PFS that is associated with preservation in QOL.

An example of this can be drawn from the metastatic pancreatic cancer literature. In a study of gemcitabine in the first-line treatment of advanced pancreatic cancer, the primary efficacy measure was clinical benefit response (a composite measurement of analgesic consumption and pain intensity, Karnofsky performance status, and weight; ref. 30). Gemcitabine was associated with a clinical benefit response rate of 24% and a 12-month survival rate of 18% compared with 5-fluorouracil, which was associated with a 5% clinical benefit response and a 2% 12-month survival rate. Gemcitabine confers a modest survival advantage, but more importantly, confers more effective alleviation of disease-related symptoms.

Discussion

Compared with randomized controlled trials between 1995 and 2004, clinical trials conducted between 2005 and 2009 were less likely to use OS as a primary endpoint (41% vs. 66%, $P < 0.001$) and were more likely to use other surrogate endpoints, such as disease-free survival (21% vs. 13%, $P = 0.092$; ref. 31). The proportion of randomized controlled trials evaluating targeted therapies between 2005 and 2009 was significantly greater (29%) compared with the earlier time period (4%, $P < 0.001$), and trials using OS as the primary endpoint were less likely to report a statistically significant result compared with trials using other time-to-event endpoints (32% vs. 70%, $P < 0.001$). Finally, in a review of oncology drugs approved by the FDA between 2005 and 2007, among 44 indications that received regular approval, a quarter were on the basis of PFS or TTP (32).

This past year, crizotinib was granted accelerated approval by the FDA for the treatment of *ALK*-positive NSCLC on the basis of 2 single-arm multicenter studies that enrolled 255 patients and showed an ORR ranging from 50% to 61% (33). In a retrospective analysis of

OS in patients with *ALK*-positive NSCLC, crizotinib therapy was associated with a dramatic improvement in OS compared with the survival of crizotinib-naïve *ALK*-positive historical control patients, suggesting that crizotinib alters the natural history of *ALK*-positive NSCLC (34). This is an OS benefit that will unlikely be confirmed in the ongoing randomized phase III clinical trial being conducted outside the United States, which has a crossover design and directly compares crizotinib with platinum-based chemotherapy in the first-line setting (PROFILE 1014). Indeed, in the randomized phase III crossover clinical trial comparing crizotinib with chemotherapy in the *ALK*-positive NSCLC second-line setting (PROFILE 1007), crizotinib was associated with a 4.7-month prolongation in PFS ($P < 0.0001$) compared with chemotherapy and no benefit in OS at the time of an interim analysis (35). Although ORR represents an extreme in the continuum of clinical trial endpoints that might be used for FDA approval, it illustrates that in the modern molecular era of cancer therapy, nontraditional endpoints can and should be used as indicators of effectiveness and regulatory endpoints.

In considering PFS as a clinical trial endpoint in large randomized phase II and phase III clinical trials, we must be mindful of the limitations with regard to the potential for bias and informative and interval censoring, the necessity of sensitivity analyses in evaluating the robustness of conclusions based on PFS, and the potential utility of BICR or selected case audits. OS historically is an endpoint most appropriate for the phase III clinical trial, and PFS was used as a screening endpoint best fitted to the phase II setting. We propose that the selection of the right drug for the right patient population makes the most sense; the selection of an endpoint should not be predicated on the phase of the trial but by the characteristics of the drug and population under study.

We are fortunate enough to be in an era of cancer therapeutics when highly effective targeted agents are successfully being developed in well-defined molecularly selected cancer populations. In these molecularly driven clinical trials, PFS is not only a valid but also a necessary measure of efficacy as it is the very effectiveness of targeted agents in these populations that prevents the detection of OS benefits.

Authors' Contributions

Conception and design: L.C. Villaruz, M.A. Socinski
Development of methodology: L.C. Villaruz, M.A. Socinski
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L.C. Villaruz, M.A. Socinski
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L.C. Villaruz, M.A. Socinski
Writing, review, and/or revision of the manuscript: L.C. Villaruz, M.A. Socinski
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L.C. Villaruz, M.A. Socinski
Study supervision: L.C. Villaruz, M.A. Socinski

Received November 26, 2012; revised March 15, 2013; accepted March 15, 2013; published online May 13, 2013.

References

- Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer* 1976;38:388–94.
- Fojo T, Parkinson DR. Biologically targeted cancer therapy and marginal benefits: are we making too much of too little or are we achieving too little by giving too much? *Clin Cancer Res* 2010;16:5972–80.
- Wilkerson J, Fojo T. Progression-free survival is simply a measure of a drug's effect while administered and is not a surrogate for overall survival. *Cancer J* 2009;15:379–85.
- U.S. Food and Drug Administration. Guidance for industry. Clinical trial endpoints for the approval of cancer drugs and biologics. Rockville, MD: U.S. Department of Health and Human Services; 2007.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
- Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;47:207–14.
- Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann Oncol* 2004;15:257–60.
- Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 2010;363:711–23.
- Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 2009;15:7412–20.
- Lynch TJ, Patel T, Dreisbach L, McCleod M, Heim WJ, Hermann RC, et al. Cetuximab and first-line taxane/carboplatin chemotherapy in advanced non-small-cell lung cancer: results of the randomized multicenter phase III trial BMS099. *J Clin Oncol* 2010;28:911–7.
- Sridhara R, Mandrekar SJ, Dodd LE. Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clin Cancer Res* 2013;19:2613–20.
- Pirker R, Pereira JR, Szczesna A, von Pawel J, Krzakowski M, Ramlau R, et al. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. *Lancet* 2009;373:1525–31.
- Zhang JJ, Zhang L, Chen H, Murgo A, Dodd LE, Pazdur R. Evaluation of audit methodologies for tumor progression in oncology clinical trials. *Clin Cancer Res* 2013;19:2637–45.
- 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.
- Qi Y, Allen Ziegler KL, Hillman SL, Redman MW, Schild SE, Gandara DR, et al. Impact of disease progression date determination on progression-free survival estimates in advanced lung cancer. *Cancer* 2012;118:5358–65.
- Mandrekar SJ, Qi Y, Hillman SL, Allen Ziegler KL, Reuter NF, Rowland KM Jr, et al. Endpoints in phase II trials for advanced non-small cell lung cancer. *J Thorac Oncol* 2010;5:3–9.
- Shepherd FA, Dancey J, Ramlau R, Mattson K, Gralla R, O'Rourke M, et al. Prospective randomized trial of docetaxel versus best supportive care in patients with non-small-cell lung cancer previously treated with platinum-based chemotherapy. *J Clin Oncol* 2000;18:2095–103.
- Buyse M, Squifflet P, Laporte S, Fossella F, Georgoulis V, Pujol J, et al. Prediction of survival benefits from progression-free survival in patients with advanced non small cell lung cancer: evidence from a pooled analysis of 2,838 patients randomized in 7 trials. *J Clin Oncol* 2008;26:15s, 2008 (suppl; abstr 8019).
- Foster NR, Qi Y, Shi Q, Krook JE, Kugler JW, Jett JR, et al. Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. *Cancer* 2011;117:1262–71.
- Lara PN Jr, Redman MW, Kelly K, Edelman MJ, Williamson SK, Crowley JJ, et al. Disease control rate at 8 weeks predicts clinical benefit in advanced non-small-cell lung cancer: results from Southwest Oncology Group randomized trials. *J Clin Oncol* 2008;26:463–7.
- Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med* 2006;355:2542–50.
- Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med* 2005;353:123–32.
- Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2012;13:239–46.
- Fukuoka M, Wu YL, Thongprasert S, Sunpawaravong P, Leong SS, Sriuranpong V, et al. Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *J Clin Oncol* 2011;29:2866–74.
- Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;361:947–57.
- Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* 2008;26:3543–51.
- Jotte R, Von Pawel J, Spigel DR, Socinski MA, O'Brien M, Paschold E, et al. Randomized phase III trial of amrubicin versus topotecan (Topo) as second-line treatment for small cell lung cancer (SCLC). *J Clin Oncol* 2011;29:15s, 2011 (suppl; abstr 7000).
- Amir E, Seruga B, Kwong R, Tannock IF, Ocana A. Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? *Eur J Cancer* 2012;48:385–8.
- Herbst RS, Sun Y, Eberhardt WE, Germonpre P, Saijo N, Zhou C, et al. Vandetanib plus docetaxel versus docetaxel as second-line treatment for patients with advanced non-small-cell lung cancer (ZODIAC): a double-blind, randomised, phase 3 trial. *Lancet Oncol* 2010;11:619–26.
- Burris HA 3rd, Moore MJ, Andersen J, Green MR, Rothenberg ML, Modiano MR, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *J Clin Oncol* 1997;15:2403–13.
- Kay A, Higgins J, Day AG, Meyer RM, Booth CM. Randomized controlled trials in the era of molecular oncology: methodology, biomarkers, and end points. *Ann Oncol* 2012;23:1646–51.
- Sridhara R, Johnson JR, Justice R, Keegan P, Chakravarty A, Pazdur R. Review of oncology and hematologic drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. *J Natl Cancer Inst* 2010;102:230–43.
- U.S. Food and Drug Administration. FDA approves Xalkori with companion diagnostic for a type of late-stage lung cancer. [cited 2011]. Available from: <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm269856.htm>.
- Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA, et al. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis. *Lancet Oncol* 2011;12:1004–12.
- Shaw AT. Phase III randomized study of crizotinib versus pemetrexed or docetaxel chemotherapy in advanced, ALK-positive non-small cell lung cancer (NSCLC) [abstract]. In: Proceedings of the ESMO 2012 Congress; 2012; Vienna, Austria. Abstract #LBA1.