

Issues in Using Progression-Free Survival When Evaluating Oncology Products

Thomas R. Fleming, Mark D. Rothmann, and Hong Laura Lu

From the Department of Biostatistics, University of Washington, Seattle, WA; and the Division of Biometrics V, US Food and Drug Administration, Silver Spring, MD.

Submitted October 4, 2008; accepted January 5, 2009; published online ahead of print at www.jco.org on May 4, 2009.

Supported by National Institutes of Health/NIAID Grant No. R37 AI 29168 entitled the Statistical Issues in AIDS Research (T.R.F.); and US Food and Drug Administration internal Grants No. RSR 05-07 and RSR 06-11 (M.D.R.).

The opinions expressed in this article are those of the authors and not necessarily those of the US Food and Drug Administration.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Thomas R. Fleming, PhD, Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232; e-mail: tfleming@u.washington.edu.

© 2009 by American Society of Clinical Oncology

0732-183X/09/2717-2874/\$20.00

DOI: 10.1200/JCO.2008.20.4107

ABSTRACT

Several challenging and often controversial issues arise in oncology trials with the use of the end point progression-free survival (PFS), defined to be the time to detection of progressive disease or death. While this end point does not directly measure how a patient feels, functions, or survives, it does provide insights about whether an intervention affects the tumor burden process, the intended mechanism through which it is hoped that most anticancer agents will provide benefit. However, simply achieving statistically significant effects on PFS is insufficient to obtaining reliable evidence of important clinical benefit, and even is insufficient to justifying the conclusion that the experimental intervention is "reasonably likely to provide clinical benefit." The magnitude of the effect on PFS in addition to the statistical strength of evidence is of great importance in interpreting the reliability of the evidence regarding clinical efficacy. PFS has several important properties, including being a direct measure of the effect of treatment on the tumor burden process, being sensitive to cytostatic as well as cytotoxic mechanisms of interventions, and incorporating the clinically relevant event of death, increasing its sensitivity to influential harmful mechanisms and avoiding substantial bias that arises when deaths are censored. To obtain reliable evidence about the effect of an intervention on PFS and patient survival, randomized trials should be conducted where all patients are followed to progression and death, and where patients in a control arm do not cross-in at progression unless the experimental regimen has already been established to be effective rescue treatment.

J Clin Oncol 27:2874-2880. © 2009 by American Society of Clinical Oncology

INTRODUCTION

When evaluating interventions in oncology patients, a frequently used outcome measure is progression-free survival (PFS), defined to be the time to the detection of progressive disease or to the patient's death. Several challenging and often controversial issues arise in use of this end point. Regulatory authorities from the United States, Europe, and Canada have provided guidance regarding many of these issues.¹⁻⁴

In the section Outcome Measures in Oncology Trials in this article, the distinction between true clinical efficacy measures and measures of biologic activity is discussed as well as reasons why PFS belongs to the latter category. The challenges in validating a biomarker as a surrogate for clinical efficacy are recognized and an overview is provided of recent randomized clinical trials submitted for registration in metastatic colorectal cancer (CRC) that used an "add-on" design. This overview provides direct evidence to assess the role in this setting of PFS as an end point for screening trials and as a surrogate end point for

overall survival (OS). Specific challenges are also addressed when using PFS as an intermediate end point for accelerated approval in a trial that also is designed to definitively evaluate the effect of the intervention on the clinical end point, OS.

In the section Censoring, the importance of observing essentially all randomly assigned patients to the end point or to the end of study is considered, and reasons are provided for why patients should not be censored at the time other treatments are initiated when analyzing the PFS end point. Reasons also are provided regarding why PFS is preferred to an alternative time to disease progression (TTP) end point where deaths are censored, and why adjudication must be provided in a timely manner when independent review committees are in place to validate the timing of PFS events.

In the section Cross-In/Subsequent Therapies, scientific and ethical reasons are provided for why the protocol should not indicate that patients will cross-in to the experimental treatment at disease progression unless the experimental therapy has already been established to be effective rescue

treatment. However, such cross-in could be acceptable if it is performed on a randomized basis.

OUTCOME MEASURES IN ONCOLOGY TRIALS

Oncology patients are interested in achieving clinically meaningful beneficial effects on disease-related symptoms, on ability to carry out normal activities, and on OS. Direct measures about how a patient feels, functions, or survives are called clinical efficacy end points, and effects on such measures by interventions that have a favorable safety profile unequivocally reflect tangible benefit to patients.

Most interventions received by oncology patients intend to provide tangible benefit by favorably impacting the tumor burden process. In addition to PFS, some other commonly used biomarkers that reflect the biologic activity of anticancer interventions include the objective response rate (ORR) and the TTP. By the definition of TTP, patients who die without progression are censored in the analysis. Such censoring creates important complexities that will be discussed in the section Censoring.

Arguments that PFS, ORR, and TTP are clinical efficacy end points are not persuasive. While a 4- to 6-week delay in the occurrence of progressive disease could provide the patient a short-term tangible benefit of increased peace of mind, such benefit would be meaningfully offset by even moderate adverse effects of therapy that would diminish quality of life, and by the cost and inconvenience associated with the use of that therapy.

While single arm trials have some usefulness in evaluating effects of single agent regimens on ORR, randomized trials provide the preferred design to obtain a reliable determination of effects on PFS, TTP, OS, and on measures capturing how a patient feels or functions. Except for settings where huge treatment effects can be expected, intention-to-treat analyses of data from randomized trials are required to distinguish the effects of treatment from the influence of prognostic factors.⁵

Surrogates

While biomarkers such as ORR and PFS are not clinical efficacy end points, they could be used as replacement outcome measures (ie, surrogate end points) in registrational clinical trials if effects on these biomarkers reliably predict effects on clinical end points.^{6,7} However, even though tumor burden biomarkers will be correlated with OS and measures of how a patient functions or feels, (eg, responders tend to live longer than nonresponders), it does not follow that establishing effects on these biomarkers provides a reliable prediction of the effects of the intervention on clinical efficacy outcome measures. For example, it is undesirable to receive a treatment that lengthens time to progression but shortens time from progression to death through unintended mechanisms that induce either long-term harm on tumor burden or adverse effects on survival through mechanisms that are independent of effects on tumor burden.

The plausibility that a biomarker will be a reliable surrogate for benefits on clinical efficacy measures varies greatly across settings. While there is strong evidence that prevention of recurrent disease in the colon adjuvant setting is a reliable predictor of effects on OS when using fluorouracil-potentiated interventions,^{7,8} it is much less clear that a transient effect on tumor burden in advanced disease should predict effects on OS. Furthermore, the meaning of

progressive disease is different from patient to patient. Why should a 20% increase in the sum of the largest diameters of target lesions for a patient having a small tumor burden be equated with a 20% increase in the sum of the largest diameters of target lesions for a patient having a large tumor burden?

The reliability of any surrogate, including PFS, is dependent on the magnitude of effect on that measure. For example, a statistically significant 1 month improvement in PFS would provide much less persuasive evidence of a survival benefit than a statistically significant and clinically important 3- to 4-month improvement in PFS in a patient with 6- to 8-month median OS. Furthermore, a modest relative benefit in PFS in a setting where median PFS is much less than median OS (such as metastatic melanoma) probably is weak evidence of benefit on OS.

Documenting the Relationship Between Effects on PFS and Effects on OS

Table 1 provides an overview of randomized clinical trials in metastatic CRC that used an add-on design and that were submitted to the US Food and Drug Administration over the past decade.⁹⁻²² While the table does not directly address the relationship between estimated effects on PFS and true effects on OS, it does provide important insights into whether favorable estimated effects on PFS predict favorable estimated effects on OS.

It is apparent that there is correlation between estimated effects on PFS and estimated effects on OS in the 17 randomized comparisons presented in Table 1. However, even though we are considering a collection of large registrational trials, the estimated effects on PFS do not reliably predict whether the estimated effects on OS will be clinically meaningful. Although it is an oversimplification, for illustration suppose a 1.25-fold relative improvement in OS (ie, a treatment/control OS hazard ratio [HR] ≤ 0.80) is defined to be the threshold for a clinically important effect. If a 1.4-fold relative improvement in PFS (ie, a PFS HR ≤ 0.714) is selected to be the threshold for predicting clinical benefit, then clinically important estimated survival effects were seen in one (33%) of three trials failing to achieve the PFS threshold and in only seven (50%) of 14 trials successfully achieving that PFS threshold. Clinically important estimated survival effects are seen in only approximately 60% of trials successfully achieving the PFS threshold even if that threshold for predicting clinical benefit is set at PFS HR ≤ 0.60 (ie, in five of eight trials), or even at PFS HR ≤ 0.55 (ie, in four of seven trials). This level of reliability might be acceptable as a screen to justify the conduct of confirmatory trials, but not as substantial evidence of clinical benefit. Only by using a PFS HR ≤ 0.50 does it follow that clinically important estimated survival effects are obtained in all three trials meeting the PFS threshold. These results illustrate that biomarkers are more likely to be reliable when effects on these measures are very large.

Several issues should be considered in interpreting these results. Limiting to trials submitted to US Food and Drug Administration likely provides an over-representation of trials having favorable estimated effects on either PFS or OS. Further, the reliability of insights from these 17 randomized comparisons in CRC would be increased by a larger experience. Finally, to generalize these insights, one needs data from other clinical settings since evidence regarding validity of surrogate end points only applies to the clinical setting and class of agents represented by the trials in

Table 1. Overview of Recent Clinical Trials in Metastatic Colorectal Cancer That Used an “Add-On” Design

Study Arm	PFS			Overall Survival		
	HR	Months	95% CI	HR	Months	95% CI
EFC2962 ^{9,10}	0.67		0.54 to 0.83	0.83		0.65 to 1.06
Control median		5.9			14.7	
EFC2961, oxaliplatin ^{9,10}	0.74		0.54 to 1.00	1.10		0.78 to 1.55
Control median		4.2			19.2	
0038 ¹¹⁻¹³	0.75 TTP		0.60 to 0.92	0.80		0.65 to 0.99
Control median		4.3			12.6	
V303, irinotecan ¹¹⁻¹³	0.60		0.47 to 0.77	0.77		0.60 to 0.98
Control median		4.4			14.1	
AVF2107g ¹⁴	0.54		0.45 to 0.66	0.66		0.54 to 0.81
Control median		6.2			15.6	
AVF0780g ¹⁴	0.44		0.25 to 0.78	0.52		0.25 to 1.07
Control median		5.2			13.6	
AVF0780g ¹⁴	0.69		0.39 to 1.21	1.01		0.53 to 1.91
Control median		5.2			13.6	
PACCE ¹⁵	1.44		1.12 to 1.85	1.55		1.12 to 2.14
Control median		10.5			NP	
EMR-007 ¹⁶	0.54 TTP		0.42 to 0.71	0.91		0.68 to 1.21
Control median		1.5			6.9	
EFC4584 ^{17,18}	0.58		0.44 to 0.76	0.84		0.69 to 1.02
Control median		1.9			8.7	
E3200 ^{19,20}	0.61		0.51 to 0.73	0.75		0.63 to 0.89
Control median		4.7			10.8	
20020408 ²¹	0.54		0.44 to 0.66	0.94		0.77 to 1.27
Control median		1.7			6.0	
CA225025 ²²	0.68		0.57 to 0.80	0.77		0.64 to 0.92
Control median		1.8			4.6	
Study WW	0.70		0.57 to 0.87	0.90		0.72 to 1.11
Control median		4.2			12.0	
Study XX	0.70		0.58 to 0.83	0.89		0.72 to 1.09
Control median		8.5			18.9	
Study YY	0.40		0.31 to 0.52	0.59		0.39 to 0.90
Control median		1.4			8.1	
Study ZZ	0.41		0.33 to 0.52	0.79		0.66 to 0.95
Control median		2.7			10.2	

NOTE. The hazard ratios are experimental/control. The final four trials are listed as WW, XX, YY, and ZZ due to lack of references in the public domain. The AVF0780g trial appears twice, having two study arms compared with the control.

Abbreviations: PFS, progression-free survival; HR, hazard ratio; TTP, time to disease progression; PACCE, Panitumumab Advanced Colorectal Cancer Evaluation; NP, not provided.

that overview. Others have recently conducted related analyses in breast cancer²³ and CRC²⁴ settings.

PFS As an Intermediate Accelerated Approval End Point in Validation Trial Evaluating Effects on OS

Since 1995, the US Food and Drug Administration has approved some interventions in the oncology setting using the accelerated approval (AA) regulatory process. By this process, temporary marketing approval can be allowed if it has been clearly established that the effect of a product on a biomarker is reasonably likely to predict clinical benefit, and where such benefit would meaningfully exceed what can be obtained from established standard of care. While it is required that a validation trial evaluating effects on clinical efficacy measures must be completed in a timely manner, this often was not achieved in the early AA experiences in the oncology setting.⁷

In recent years, to achieve more timely completion of the validation trial, the design in Figure 1 frequently has been implemented. The same trial that provides data on the AA end point, such

as PFS, is then continued to provide validation through data subsequently collected on the clinical end point, such as OS. Such trials, by being fully powered to detect a survival effect, usually are overpowered for evaluating the effect on the biomarker. While having enhanced precision regarding treatment effect on the biomarker should be a favorable feature, it is problematic if one inappropriately considers a trial to be positive simply because the hypothesis of no effect on the biomarker end point can be ruled out at traditional levels of statistical significance.

For illustration, consider the setting in Figure 1 where it is assumed control patients have median PFS and median OS of 2 months and 10 months, respectively. A validation trial with $L = 647$ deaths will achieve statistical significance (at two-sided $P = .04$) if the estimated experimental versus control HR for OS is 0.851, approximately corresponding to an estimated 8-week improvement in survival. Such a trial is properly powered for OS given that this duration of survival improvement approximates the threshold for a clinically meaningful benefit. Substantially earlier in calendar time, the trial would achieve

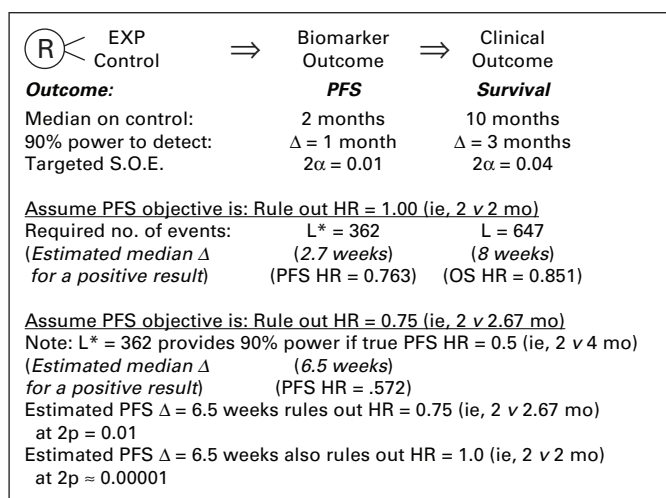


Fig 1. An approach is illustrated to achieve the timely completion of validation trials when pursuing accelerated approval in the oncology setting. EXP, the experimental regimen; L^* , the number of patients experiencing progression-free survival (PFS) events; L , the number of patients experiencing overall survival (OS) events; SOE, strength of evidence required to rule out a hypothesis.

its targeted $L^* = 362$ PFS events, where AA could be considered if results are reasonably likely to predict the 8-week improvement in OS. Such evidence would not be provided simply by ruling out that treatment has no effect on PFS, since a two-sided $P = .01$ would be obtained with an estimated HR of 0.763, corresponding to a very modest increase in PFS of approximately 2.7 weeks. It is difficult to justify that a 2.7-week improvement in PFS makes it reasonably likely that subsequent data will establish at least an 8-week improvement in survival. In contrast, a PFS increase of at least 6.5 weeks would allow one to rule out a modest one-third improvement in PFS (ie, 2 v 2.67 months) at a two-sided $P < .01$ and also to rule out no improvement at two-sided $P \approx .00001$. At least as importantly, this 6.5-week improvement in PFS would be an effect size that makes achievement of an 8-week improvement in OS much more likely. It should also be noted that the correlation between the results on PFS and OS depends on the number of events for each analysis. An OS analysis based on 500 deaths will be more correlated with the results of the PFS analysis after 400 events than after 100 events.

The effect on PFS or any other biomarker being used to pursue AA should be both robust statistically and also sufficiently impressive in magnitude to justify the argument that clinically meaningful effects on the clinical efficacy measure are likely to be achieved. More detailed criteria for use of a biomarker, such as PFS, for AA have been provided recently.⁷

CENSORING

For time-to-event end points, such as PFS and OS, the intent-to-treat principle requires that essentially all randomly assigned patients be observed to the end point or to the end of the study. When a patient's follow-up is censored x number of months after random assignment, then in the computation of Kaplan-Meier estimates or in log-rank, Cox regression or other standard methods of statistical analysis, that

patient's outcome after x number of months is imputed by using the longer term outcomes of other patients in their treatment group who also are free of the outcome at x number of months and remain under follow-up beyond x number of months. Thus, unless the reason for being censored is independent of that patient's prognosis, failure to observe that patient until occurrence of his study end points will lead to significant bias as well as increased variability in the evaluation of treatment effects.

In order to obtain a proper evaluation of PFS and OS end points, patients who receive other treatments should continue to be observed for progression and death events. An important illustration of the principles discussed above arises by the practice of censoring a patient at the time other treatments are initiated (denoted above as time x). This induces strongly dependent censoring because the true time to progression for that patient is replaced by the true time to progression of other patients who also were free of progression at month x , but did not need other treatments at that time. As pointed out by Carroll,²⁵ "...the regularly recommended maneuver to censor PFS time at drop-out due to toxicity or on the initiation of additional anticancer therapy is likely to favor the more toxic, less efficacious treatment and so should be avoided whenever possible." An alternative approach of calling the patient a failure at time x changes the end point to one involving delay in time to initiation of alternative treatments. Such an end point is at best a surrogate for true clinical benefit since the goal for cancer patients usually is not to avoid other therapies; rather their goal is to improve how they feel, function, or survive.

Suppose there are intermittent missing assessments of disease progression, such that a patient is alive without progression at x_1 , and dies much later at x_2 with evidence of disease progression but with no intervening evaluation of disease progression status. One certainly obtains a biased underestimate or overestimate of PFS by defining the event to be at x_1 or at x_2 , respectively. We do know the PFS event did occur in the interval (x_1, x_2) . A simple approach would be to define the event to be at $(x_1 + x_2)/2$. However, it is easy to configure situations where that estimate would be biased high, such as in melanoma where progression occurs much sooner than death. It could be biased low in a setting where death follows progression rather quickly. To be more sophisticated, one could use interval censoring methods.²⁶ The difficulties in addressing such irregularities in assessment of disease progression are further reason to favor OS as the primary end point, since OS does not provide the challenge of informative missingness in such situations.

PFS Versus TTP

PFS is a truly measurable and interpretable end point. In contrast, since patients who die without progression are censored in the analysis of the TTP end point, TTP is a "hypothetical" outcome (ie, TTP is estimating the time to progression in the hypothetical setting when patients are not at risk of death from any cause other than disease progression). Even if one could justify having genuine interest in such a hypothetical outcome, estimates are biased unless one can justify that deaths induce independent censoring. This is certainly implausible in that it requires justification that death at x number of months occurring to a patient who had not had progressive disease is not informative about the subsequent timing of that patient's progression in the hypothetical setting where he had not died at x number of months.

Statistical censoring of deaths, as done in the estimation of TTP, effectively results in attempts to estimate when, after the patient's death, progression would have occurred. Is that truly of interest? It also follows that median TTP could be multiple times longer than median OS. Would such a result be interpretable?

TTP is insensitive to effects on OS that are not mediated through effects on tumor burden. Hence, PFS is a more robust surrogate than TTP for treatment effect on OS. Suppose an agent's only impact would be to induce fatal toxicity in frail patients who likely would have had early progression had they received standard of care (SOC). In that setting, TTP would inappropriately indicate the agent is superior to SOC while PFS would correctly show the agent to be harmful. Finally, in settings where death follows relatively soon after disease progression, PFS is less sensitive than TTP to the bias arising when progression events are missed.

Issues When Using Independent Review Committees to Assess PFS

Due to the subjective nature of the assessment for when disease progression is sufficient to trigger the PFS end point, an investigator-determined evaluation can have substantial bias, especially in an open label trial.²⁷ A blinded reassessment of disease progression by an Independent Review Committee (IRC) has been a frequent approach to address this bias.

In many recent trials that have used an IRC assessment of PFS as a primary end point, failure to conduct timely IRC assessments has introduced significant bias. At the time of an investigator-determined diagnosis of progression, the investigator usually discontinues obtaining radiologic scans. The IRC has frequently disagreed with the investigator's judgment that progression has occurred. Because IRC evaluations of radiographic scans rarely have been done in real time, such discordance has not been identified in time to ensure that the investigator continues obtaining radiologic scans until an IRC validated progression. This leads to informative censoring that can significantly bias the evaluation of PFS.²⁸

Whenever any tumor progression measure is the primary end point, one of two approaches should be used. Either the trial should be blinded, with investigator assessments used to determine time of disease progression, or an expedited IRC reassessment of investigator-defined progression should occur, with investigators continuing scans until IRC confirmation of progression is received. The investigator would still be allowed to make treatment-related decisions with the patient while waiting for confirmation of progression.

In a blinded trial, it should be justified that the evaluator of progression status has not been unmasked by toxicities or laboratory measurements. In an open label trial, in addition to having an IRC, it is important to ensure the timing of assessments of progression status is the same between the intervention groups.

CROSS-IN/SUBSEQUENT THERAPIES

Often, an important component of the effect of an experimental agent is its impact on the need for supportive or subsequent care, the ability to receive it, and what effect it might provide. Therefore, a weakness rather than strength of PFS (relative to OS) is that it does not incorporate the real world long-term impact of the exper-

imental regimen between the times of disease progression and death. The trial should be designed to ensure these important longer effects can be properly estimated.

A clinical trial should compare the benefit-to-risk profiles of a regimen involving the experimental agent against a SOC regimen. Hence, a trial designed to cross control patients into the experimental therapy at disease progression is only justified when that therapy already has been proven to be an effective rescue treatment, and it only remains to determine the optimal timing of its delivery. When benefit as a rescue treatment has not been established in advance of the trial, the cross-in would compromise the ability of the trial to provide interpretable evidence about the longer term effects of the experimental agent.

A cross-in design also provides an enhanced risk that patients will be provided a misleading presentation during the informed consent process, through an indication that access to the experimental agent likely is in their best interest and participation in the trial will provide such access, either initially or at the time of disease progression. In trials not providing cross-in at time of disease progression, it is more likely patients would be properly informed regarding the equipoise about whether the patient should receive the experimental agent. Cross-in designs raise ethical and scientific concerns due to enhanced potential for coercion in the enrollment process and because they compromise the ability to reliably assess effects on longer-term outcome measures such as OS that meaningfully influence the benefit-to-risk profile of the experimental intervention.

Suppose cross-in of control patients to the experimental agent were routinely allowed at progression in a trial that established that agent substantially improved PFS, such as in Amgen's 20020408 trial evaluating panitumumab in third-line CRC patients.²⁹ In that trial, the primary analysis of OS had an HR of 1.00 after 380 deaths. When a statistically significant PFS advantage has been established and yet data fail to demonstrate a clinically meaningful difference in OS, a rationalization has been that the discordance is due to survival benefit provided to control patients when cross-in is allowed after disease progression. Such an argument presumes as truth (ie, that the experimental agent positively impacts OS) what has failed to be demonstrated to be true. If a regimen truly provides clinically meaningful prolongation of OS, it seems to be quite unlikely that substantially delayed access to the intervention by a fraction of the control patients (all of whom did not initially receive the anticancer therapy) would achieve the full survival benefit of immediate access by the entire cohort. Hence, if the immediate and delayed regimens have the same OS, there should be considerable skepticism that the intervention meaningfully improves OS. If such a therapy were to be used in clinical practice, delayed use should be considered since immediate access was not better than delayed access.

In settings where one wishes not only to provide a long-term comparison of the originally randomly assigned treatments, but also to assess the impact of a secondary intervention (including cross-in), patients could be randomly assigned to whether they receive this secondary intervention at the time it would be initiated. Statistical methods³⁰ can then be used to estimate the long-term relative efficacy of the originally randomized treatments for the setting when cross-in does not occur. Preferably, randomization to cross-in would be stratified by the time of cross-in and perhaps other factors.

CONCLUSION

Clinical efficacy outcomes that directly measure how a patient feels, functions, or survives, such as OS or symptom-based outcomes, provide direct evidence about the benefit-to-risk profile of an intervention. In contrast, biomarkers such as PFS, TTP, and ORR address proof-of-concept by providing important insights about whether an intervention affects the tumor burden process, the intended mechanism through which it is hoped that most anticancer agents will provide benefit.

It would be unethical to use an intervention that is inferior to an SOC regimen in its effects on major morbidity or mortality outcomes. Therefore, in oncology, regular approval of an intervention usually requires establishing the experimental regimen is superior or, on some occasions, noninferior to SOC with respect to such effects.³¹ If the evaluation of efficacy is based on a biomarker such as PFS, it must be known that the required level of efficacy on major morbidity or mortality outcomes is reliably established by the level of benefit established on the biomarker. However, in oncology, as presented in Table 1 in the setting of recent metastatic CRC randomized clinical trials used for registration, there is a frequent discordance between effects on PFS and effects on OS.

The goal of clinical research is not to obtain a statistically significant effect. Rather, “the primary goal should be to obtain a statistically reliable evaluation regarding whether the experimental intervention is safe and provides clinically meaningful benefit”.³² Important insight about this principle is provided in Table 1 where it is illustrated that it is not sufficient to establish a statistically significant effect on a biomarker such as PFS to have reliable evidence of clinically meaningful benefit. Figure 1 further illustrates that simply achieving statistical significance on the PFS measure does not even allow one to conclude that an intervention “is reasonably likely to provide clinical benefit”. The magnitude of the effect as well as the statistical reliability of the evidence is of great importance.

Among measures of biologic activity, PFS has many important characteristics. Unlike measures such as CA-125 or PSA, PFS is a direct measure of the effect of treatment on the tumor burden process.

Unlike ORR, it is sensitive to cytostatic as well as cytotoxic mechanisms of interventions. Unlike TTP, it incorporates the clinically relevant event of death, increasing its sensitivity to important harmful mechanisms and avoiding substantial bias that arises when deaths are censored by measures such as TTP.

To obtain reliable evidence about the effect of an intervention on PFS, randomized trials should be conducted. To maintain the integrity of random assignment, intention-to-treat analyses should be conducted where all patients are observed to progression and death, whether or not other treatments are initiated. When IRCs are used for validation of investigator-based assessments of the timing of PFS, procedures need to be in place to ensure nearly real-time validation when an investigator identifies a PFS event has occurred. Investigators can make treatment-related decisions with a patient but should not stop outcome evaluations before receiving IRC confirmation of disease progression. Finally, to enhance the ability of the trial to assess effects on longer-term clinical efficacy end points and to reduce the risk of coercion of patients during the recruitment process, patients in a control arm should not cross-in to the experimental treatment at progression unless the experimental therapy has already been established to be effective rescue treatment.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Thomas R. Fleming, Mark D. Rothmann, Hong Laura Lu

Collection and assembly of data: Thomas R. Fleming, Mark D. Rothmann, Hong Laura Lu

Data analysis and interpretation: Thomas R. Fleming, Mark D. Rothmann, Hong Laura Lu

Manuscript writing: Thomas R. Fleming, Mark D. Rothmann

Final approval of manuscript: Thomas R. Fleming, Mark D. Rothmann, Hong Laura Lu

REFERENCES

1. U.S. Food and Drug Administration: Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Silver Spring, MD, U.S. Food and Drug Administration, 2007. <http://www.fda.gov/cder/guidance/7478fnl.htm>
2. European Medicines Agency: Guideline on the Evaluation of Anticancer Medicinal Products in Man. London, United Kingdom, European Medicines Agency, 2005. www.emea.europa.eu/pdfs/human/ewp/020595en.pdf
3. European Medicines Agency: Methodological Considerations for Using Progression-Free Survival (PFS) As Primary Endpoint in Confirmatory Trials for Registration. London, United Kingdom, European Medicines Agency, 2008. <http://www.emea.europa.eu/pdfs/human/ewp/2799408en.pdf>
4. Health Canada: Issues Analysis Summary: The Use of Progression-Free Survival As the Efficacy Endpoint for Approval of Targeted and Chemotherapeutic Agents for Advanced Cancer. Ottawa, Canada, Health Canada, 2007. http://www.hc-sc.gc.ca/dhp-mps/prodpharma/activ/sci-com/onco/sacot_ias_ccs_req_2007-08-eng.php

http://www.hc-sc.gc.ca/dhp-mps/prodpharma/activ/sci-com/onco/sacot_ias_ccs_req_2007-08-eng.php

5. Ratain MJ, Humphrey RW, Gordon GB, et al: Recommended changes to oncology clinical trial design: Revolution or evolution? *European Journal of Oncology* 44:8-11, 2008
6. Fleming TR, DeMets DL: Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med* 125:605-613, 1996
7. Fleming TR: Surrogate endpoints and FDA's accelerated approval process. *Health Aff* 24:67-78, 2005
8. Sargent D, Wieand S, Haller DG, et al: Disease free survival versus overall survival as a primary endpoint for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 23:8664-8670, 2005
9. Briefing Document of the Effectiveness and Safety of Oxaliplatin in Combination with 5-FU-Based Chemotherapy in the Treatment of Advanced Colorectal Cancer, Sanofi-Synthelabo Research. Oncologic Drugs Advisory Committee Meeting. Bethesda, MD, March 16, 2000. <http://www.fda.gov/ohrms/dockets/ac/00/backgrd/3592b1cx.pdf>

10. FDA Statistical Review and Evaluation: Summary Statement. Oncologic Drugs Advisory Committee Meeting. Bethesda, MD, March 16, 2000. <http://www.fda.gov/ohrms/dockets/ac/00/backgrd/3592b1b.pdf>

11. Irinotecan Hydrochloride (CPT-11, Camptosar) Oncologic Drugs Advisory Committee Brochure, NDA #20-571, Supplement #9 (First-Line Therapy of Colorectal Cancer), Pharmacia & Upjohn Oncologic Drugs Advisory Committee Meeting. Bethesda, MD, March 16, 2000. <http://www.fda.gov/ohrms/dockets/ac/00/backgrd/3592b1f.pdf>

12. FDA Medical Officer Summary Statement for the Advisory Committee, Irinotecan Hydrochloride (Camptosar Injection) for Colorectal Cancer, (Application for Irinotecan as a Component of First-line Therapy of Patients with Metastatic Carcinoma of the Colon or Rectum). Oncologic Drugs Advisory Committee Meeting. Bethesda, MD, March 16, 2000. <http://www.fda.gov/ohrms/dockets/ac/00/backgrd/3592b1d.pdf>

13. FDA Summary Statement Statistical Review and Evaluation: Oncologic Drugs Advisory Committee Meeting. Bethesda, MD, March 16, 2000. <http://www.fda.gov/ohrms/dockets/ac/00/backgrd/3592b1d.pdf>

www.fda.gov/ohrms/dockets/ac/00/backgrd/3592b1e.pdf

14. FDA Statistical Review and Evaluation. http://www.fda.gov/cder/foi/nda/2004/STN-125085_Avastin_Statr.pdf

15. Vectibix product labeling. <http://www.fda.gov/cder/foi/label/2008/125147s026lbl.pdf>

16. FDA Statistical Review and Evaluation. http://www.fda.gov/cder/biologics/nda/125084_ERBITUX_STATR.PDF

17. FDA Medical Review. http://www.fda.gov/cder/foi/nda/2002/21-492_Eloxatin_medr_P2.pdf

18. Rothenberg ML, Oza AM, Burger B, et al: Final results of a phase III trial of 5-FU/leucovorin versus oxaliplatin versus the combination in patients with metastatic colorectal cancer following irinotecan, 5-FU, and leucovorin. *Proc Am Soc Clin Oncol* 22:252a, 2003 (abstr 1011)

19. Avastin product labeling. <http://www.fda.gov/cder/foi/label/2008/125085s145lbl.pdf>

20. Giantonio BJ, Catalano PJ, Meropol NJ, et al: Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: Results from

the Eastern Cooperative Oncology Group Study E3200. *J Clin Oncol* 25:1539-1544, 2007

21. FDA Statistical Review and Evaluation. http://www.fda.gov/cder/foi/nda/2006/125147s0000_StatR.pdf

22. Assessment Report for Erbitux. International Non-Proprietary Name/Common Name: Cetuximab. Procedure No. EMEA/H/C/000558/II/0020. <http://www.emea.europa.eu/humandocs/PDFs/EPAR/erbitux/H-558-II-20-AR.pdf>

23. 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* 26:1987-1992, 2008

24. Buyse M, Burzykowski, Carroll K, et al: Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 25:5128-5224, 2007

25. Carroll KJ: Analysis of progression-free survival in oncology trials: Some common statistical issues. *Pharmaceutical Statistics* 6:99-113, 2007

26. Sun J: *The Statistical Analysis of Interval-Censored Failure Time Data: Statistics for Biology and Health*. Springer, New York, NY, 2006, pp xvi, 302

27. Freidlin B, Korn EL, Hunsberger S, et al: Proposal for the use of progression-free survival in unblinded randomized trials. *J Clin Oncol* 25:2122-2126, 2007

28. 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* 26:3791-3796, 2008

29. Van Cutsem E, Peeters M, Siena S, et al: Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *J Clin Oncol* 25:1658-1664, 2007

30. Luncford JK, Davidian M, Tsiatis AA: Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics* 58:48-57, 2002

31. Fleming TR: Current issues in non-inferiority trials. *Stat Med* 27:317-332, 2008

32. Fleming TR: Standard vs. adaptive monitoring procedures: A commentary. *Stat Med* 25:3305-3312, 2006

