Guidance for Industry
Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2007
Clinical/Medical

Guidance for Industry
Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2007
Clinical/Medical

TABLE OF CONTENTS

Guidance for Industry[1]
Clinical Trial Endpoints for the Approval
of Cancer Drugs and Biologics

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I.   INTRODUCTION

This guidance is the first in a planned series of cancer endpoint guidances. It provides recommendations to applicants on endpoints for cancer clinical trials submitted to the Food and Drug Administration (FDA) to support effectiveness claims in new drug applications (NDAs), biologics license applications (BLAs), or supplemental applications.[2] It also provides background information and discusses general regulatory principles. The endpoints discussed in this guidance are for drugs to treat patients with an existing cancer. This guidance does not address endpoints for drugs to prevent or decrease the incidence of cancer.

The FDA is developing guidance on oncology endpoints through a process that includes public workshops and discussions before the FDA's Oncologic Drugs Advisory Committee (ODAC).[3] Each subsequent guidance will focus on endpoints for specific cancer types (e.g., lung cancer, colon cancer) to support drug approval or labeling claims.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

---

[1] This guidance has been prepared by the Division of Drug Oncology Products and the Division of Biologic Oncology Products in the Center for Drug Evaluation and Research (CDER) in cooperation with the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration.

[2] For the purposes of this guidance, all references to *drugs* include both human drugs and biological products unless otherwise specified.

[3] Transcripts are available at http://www.fda.gov/cder/drug/cancer_endpoints/default.htm.

II.   BACKGROUND

Clinical trial endpoints serve different purposes. In conventional oncology drug development, early phase clinical trials evaluate safety and identify evidence of biological drug activity, such as tumor shrinkage. Endpoints for later phase efficacy studies commonly evaluate whether a drug provides a clinical benefit such as prolongation of survival or an improvement in symptoms. The following sections discuss the general regulatory requirements for efficacy and how they have influenced endpoint selection for the approval of cancer drugs. Later sections describe these endpoints in more detail and discuss whether they might serve as measures of disease activity or clinical benefit in various clinical settings.

A.   Regulatory Requirements for Effectiveness

The requirement that new drugs show effectiveness is based on a 1962 amendment to the Federal Food, Drug, and Cosmetic Act. This law requires substantial evidence of effectiveness and specifies that this evidence must be derived from adequate and well-controlled clinical investigations. Similarly, the Public Health Service Act requires biological products to be safe, pure, and potent. Clinical benefits that have supported drug approval have included important clinical outcomes (e.g., increased survival, symptomatic improvement) but have also included effects on established surrogate endpoints (e.g., blood pressure, serum cholesterol).

The accelerated approval regulations (21 CFR part 314, subpart H and 21 CFR part 601, subpart E), promulgated in 1992, allow use of additional endpoints for approval of drugs or biological products that are intended to treat serious or life-threatening diseases and that either demonstrate an improvement over available therapy or provide therapy where none exists. In this setting, the FDA may grant approval based on an effect on a surrogate endpoint that is *reasonably likely* to predict clinical benefit ("based on epidemiologic, therapeutic, pathophysiologic, or other evidence"). Such surrogates are less well-established than surrogates in regular use, such as blood pressure or cholesterol for cardiovascular disease. A drug is approved under the accelerated approval regulations on condition that the manufacturer conducts clinical studies to verify and describe the actual clinical benefit. If the postmarketing studies fail to demonstrate clinical benefit or if the applicant does not demonstrate due diligence in conducting the required studies, the drug may be removed from the market under an expedited process. In the following discussion, the term *regular approval* denotes the longstanding route of drug approval based on the demonstration of clinical benefit. That term is distinguished from *accelerated approval*, which is associated with use of a surrogate endpoint that is reasonably likely to predict benefit.

The evidence critical for supporting drug approval, including the preferred number of clinical trials, is discussed in the guidance for industry *FDA Approval of New Cancer Treatment Uses for Marketed Drug and Biological Products*[4] and in the FDA Modernization Act of 1997.[5] In most

---

[4] We update guidances periodically. To make sure you have the most recent version of a guidance, check the CDER guidance Web page at http://www.fda.gov/cder/guidance/index.htm.

[5] http://www.fda.gov/cder/fdama/default.htm

cases, the FDA recommends at least two adequate and well-controlled clinical trials. In certain cases, evidence from a single trial can be sufficient (e.g., in cases in which a single multicenter study provides highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and in which confirmation of the result in a second trial would be practically or ethically impossible). For drugs approved for treatment of patients with a specific stage of a particular malignancy, evidence from one trial may be sufficient to support an efficacy supplement for treatment of a different stage of the same cancer.

B.   Endpoints Supporting Past Approvals in Oncology

For regular approval, it is critical that the applicant show direct evidence of clinical benefit or improvement in an established surrogate for clinical benefit. In oncology, survival improvement is considered an appropriate measure of clinical benefit. In addition, sponsors have used other endpoints for cancer drug approval. In the 1970s, the FDA usually approved cancer drugs based on objective response rate (ORR), determined by tumor assessments from radiological tests or physical examinations. In the early 1980s, after discussion with the ODAC, the FDA determined that cancer drug approval should be based on more direct evidence of clinical benefit, such as improvement in survival, improvement in a patient's quality of life (QOL), improved physical functioning, or improved tumor-related symptoms. These benefits may not always be predicted by, or correlate with, ORR.

Over the next decade, several endpoints were used as established surrogates for clinical benefit. Improvement in disease-free survival (DFS) supported drug approval in selected adjuvant settings, in which a large proportion of patients were expected to have cancer symptoms at the time of recurrence. Durable complete response was considered an established endpoint of clinical benefit in leukemia, where complete response is associated with less infection, bleeding, and blood product support. A high, substantiated ORR can support regular approval in select solid tumors, but that response duration, relief of tumor-related symptoms, and drug toxicity also should be considered when making the approval decision (O'Shaughnessy and Wittes et al., 1991, Commentary Concerning Demonstration of Safety and Efficacy of Investigational Anticancer Agents in Clinical Trials, J Clin Oncol, 9:2225-2232). For example, randomized trials for hormonal drugs for breast cancer have used ORR as an endpoint supporting regular approval. Improvement in tumor-related symptoms in conjunction with an improved ORR and adequate response duration has supported regular approval in several clinical settings.

Surrogate endpoints for accelerated approval must be reasonably likely to predict clinical benefit (21 CFR part 314, subpart H and 21 CFR part 601, subpart E). Such drugs also must provide a benefit over available therapy (21 CFR part 314, subpart H and 21 CFR part 601, subpart E).[6] ORR has been the most commonly used surrogate endpoint in support of accelerated approval. Tumor response is widely accepted by oncologists in guiding cancer treatments. Because ORR is directly attributable to drug effect, single-arm trials conducted in patients with refractory tumors where no available therapy exists provide an accurate assessment of ORR.

---

[6] See the guidance for industry *Available Therapy* (http://www.fda.gov/cder/guidance/index.htm).

III.   GENERAL ENDPOINT CONSIDERATIONS

This section provides an overview of general issues in cancer drug development. A discussion of commonly used cancer endpoints is followed by a discussion of pertinent issues in cancer clinical trial design using these endpoints (future guidances will discuss specific treatment indication endpoints). The endpoints that are discussed in this section include overall survival, endpoints based on tumor assessments (e.g., DFS, ORR, complete response, time to progression (TTP), progression-free survival (PFS)), and endpoints based on symptom assessment. Table 1 provides a comparison of endpoints in cancer drug approval. Many issues relating to the proper analysis of efficacy endpoints are addressed in the ICH guidance for industry *E9 Statistical Principles for Clinical Trials.*

Table 1. A Comparison of Important Cancer Approval Endpoints

Endpoint	Regulatory Evidence	Study Design	Advantages	Disadvantages
Overall Survival	Clinical benefit for regular approval	• Randomized studies essential • Blinding not essential	• Universally accepted direct measure of benefit • Easily measured • Precisely measured	• May involve larger studies • May be affected by crossover therapy and sequential therapy • Includes noncancer deaths
Symptom Endpoints (patient-reported outcomes)	Clinical benefit for regular approval	• Randomized blinded studies	• Patient perspective of direct clinical benefit	• Blinding is often difficult • Data are frequently missing or incomplete • Clinical significance of small changes is unknown • Multiple analyses • Lack of validated instruments
Disease-Free Survival	Surrogate for accelerated approval or regular approval*	• Randomized studies essential • Blinding preferred • Blinded review recommended	• Smaller sample size and shorter follow-up necessary compared with survival studies	• Not statistically validated as surrogate for survival in all settings • Not precisely measured; subject to assessment bias, particularly in open-label studies • Definitions vary among studies

continued

Table 1, continued

Endpoint	Regulatory Evidence	Study Design	Advantages	Disadvantages
Objective Response Rate	Surrogate for accelerated approval or regular approval*	• Single-arm or randomized studies can be used • Blinding preferred in comparative studies • Blinded review recommended	• Can be assessed in single-arm studies • Assessed earlier and in smaller studies compared with survival studies • Effect attributable to drug, not natural history	• Not a direct measure of benefit • Not a comprehensive measure of drug activity • Only a subset of patients who benefit
Complete Response	Surrogate for accelerated approval or regular approval*	• Single-arm or randomized studies can be used • Blinding preferred in comparative studies • Blinded review recommended	• Can be assessed in single-arm studies • Durable complete responses can represent clinical benefit • Assessed earlier and in smaller studies compared with survival studies	• Not a direct measure of benefit in all cases • Not a comprehensive measure of drug activity • Small subset of patients with benefit
Progression-Free Survival (includes all deaths) or Time to Progression (deaths before progression censored)	Surrogate for accelerated approval or regular approval*	• Randomized studies essential • Blinding preferred • Blinded review recommended	• Smaller sample size and shorter follow-up necessary compared with survival studies • Measurement of stable disease included • Not affected by crossover or subsequent therapies • Generally based on objective and quantitative assessment	• Not statistically validated as surrogate for survival in all settings • Not precisely measured; subject to assessment bias particularly in open-label studies • Definitions vary among studies • Frequent radiological or other assessments • Involves balanced timing of assessments among treatment arms

*Adequacy as a surrogate endpoint for accelerated approval or regular approval is highly dependent upon other factors such as effect size, effect duration, and benefits of other available therapy. See text for details.

A.   Overall Survival

Overall survival is defined as the time from randomization until death from any cause, and is measured in the intent-to-treat population. Survival is considered the most reliable cancer endpoint, and when studies can be conducted to adequately assess survival, it is usually the preferred endpoint. This endpoint is precise and easy to measure, documented by the date of

death. Bias is not a factor in endpoint measurement. Survival improvement should be analyzed as a risk-benefit analysis to assess clinical benefit.

Overall survival should be evaluated in randomized controlled studies. Data derived from historical trials are seldom reliable for time-dependent endpoints (e.g., overall survival, PFS). Apparent differences in outcome between historical controls and current treatment groups can arise from differences other than drug treatment, including patient selection, improved imaging techniques, or improved supportive care. Randomized studies minimize the effect of these differences by providing a direct outcome comparison. Demonstration of a statistically significant improvement in overall survival can be considered to be clinically significant if the toxicity profile is acceptable, and has often supported new drug approval.

Difficulties in performing and analyzing survival studies include long follow-up periods in large trials and subsequent cancer therapy potentially confounding survival analysis.

B.   Endpoints Based on Tumor Assessments

This section discusses several endpoints that are based on tumor assessments. These endpoints include DFS, ORR, TTP, PFS, and time-to-treatment failure (TTF). The collection and analysis of data on these time-dependent endpoints are based on indirect assessments, calculations, and estimates (e.g., tumor measurements). PFS data collection and analysis is supplemented by the tables shown in Appendix 3.

Tumor-assessment endpoints selection should include two judgments. First, a determination of whether the endpoint will support either accelerated approval or regular approval should be ascertained. Second, the endpoint should be evaluated for the potential of bias or uncertainty in tumor endpoint assessments. Drug applications using studies that rely on tumor measurement-based endpoints as sole evidence of efficacy may need confirmatory evidence from a second trial. Accuracy in measuring tumors can differ among tumor settings. Tumor measurements used in response rate determinations can be imprecise in locations where there is a lack of demarcated margins (e.g., malignant mesothelioma, pancreatic cancer, brain tumors).

When the primary study endpoint is based on tumor measurements (e.g., PFS or ORR), tumor endpoint assessments generally should be verified by central reviewers blinded to study treatments (see Appendix 4). This measure is especially important when the study itself is not blinded. It may be appropriate for the FDA to audit a sample of the scans to verify the central review process. Additional details regarding data collection are listed in Appendix 1. Centralized independent verification of tumor endpoint assessments (especially for PFS or DFS) may not be necessary when randomized trials are blinded (unless the adverse event profile would substantially *unblind* the trial in practice) or effect sizes are robust in large randomized trials where sensitivity analysis supports lack of observer bias (especially for DFS).

1.   *Disease-Free Survival*

Generally, DFS is defined as the time from randomization until recurrence of tumor or death from any cause. The most frequent use of this endpoint is in the adjuvant setting after definitive

surgery or radiotherapy. DFS also can be an important endpoint when a large percentage of patients achieve complete responses with chemotherapy. Although overall survival is a conventional endpoint for most adjuvant settings, DFS can be an important endpoint in situations where survival may be prolonged, making a survival endpoint impractical. DFS has been the primary basis of approval for adjuvant breast cancer hormonal therapy, adjuvant colon cancer, and adjuvant cytotoxic breast cancer therapy. Compared with standard cytotoxic therapies, hormonal therapies carry minimum side effects and thus a favorable risk-benefit relationship. DFS can be a surrogate for clinical benefit or it can provide direct evidence of clinical benefit. This determination is based on the magnitude of the effect, its risk-benefit relationship, and the disease setting. However, in disease settings where survival benefit has been already established, it is unlikely that DFS can be considered a clinical benefit. In December 2003, the ODAC consensus was DFS prolongation represented clinical benefit if the magnitude of this benefit outweighed the toxicity of the adjuvant treatment. In May 2004, the ODAC recommended that DFS be considered an acceptable endpoint for colon cancer drugs in the surgical adjuvant setting.[7]

Important considerations in evaluating DFS as a potential endpoint include the estimated size of the treatment effect and proven benefits of standard therapies. The protocol should carefully delineate both the definition of DFS and the schedule for follow-up studies and visits. Unscheduled assessments can occur for many reasons and differences between study arms in the frequency, timing, or reason for unscheduled assessments can introduce bias. Bias can be minimized by blinding patients and investigators to the treatment assignments. The potential effects of bias due to unscheduled assessments can be evaluated by a comparative analysis of the total number of events over the follow-up period regardless of when the events occurred.

The definition of DFS can be complicated, particularly when deaths are noted without prior tumor progression documentation. These events can be scored either as disease recurrences or as censored events. Although all methods for statistical analysis of deaths have some limitations, considering all deaths (deaths from all causes) as recurrences can minimize bias. DFS can be overestimated using this definition, especially in patients who die after a long period without observation. Bias can be introduced if the frequency of long-term follow-up visits is dissimilar between the study arms or if dropouts are not random because of toxicity. Some analyses count cancer-related deaths as DFS events and censor noncancer deaths. This method can introduce bias in the attribution of the cause of death. Furthermore, any method that censors patients, whether at death or at the last visit, assumes that the censored patients have the same risk of recurrence as noncensored patients.

2. *Objective Response Rate*

ORR is defined as the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period. Response duration usually is measured from the time of initial response until documented tumor progression. Generally, the FDA has defined ORR as the sum of partial responses plus complete responses. When defined in this manner, ORR is a direct measure of drug antitumor activity, which can be evaluated in a single-arm study. Stable disease should not be a component of ORR. Stable disease can reflect the natural history of disease,

---

[7] Transcripts are available at http://www.fda.gov/cder/drug/cancer_endpoints/default.htm.

whereas tumor reduction is a direct therapeutic effect. Also, stable disease can be more accurately assessed by TTP or PFS analysis (see below). If available, standardized criteria should be used to ascertain response. A variety of response criteria have been considered appropriate (e.g., RECIST criteria) (Therasse and Arbuck et al., 2000, New Guidelines to Evaluate Response to Treatment in Solid Tumors, J Natl Cancer Inst, 92:205-16). The response criteria should be predefined in the protocol before the start of the study. The significance of ORR is assessed by its magnitude and duration, and the percentage of complete responses (no detectable evidence of tumor).

3. *Time to Progression and Progression-Free Survival*

TTP and PFS have served as primary endpoints for drug approval. TTP is defined as the time from randomization until objective tumor progression; TTP does not include deaths. PFS is defined as the time from randomization until objective tumor progression or death. The precise definition of tumor progression is important and should be carefully detailed in the protocol.

a. TTP vs. PFS

Compared with TTP, PFS is the preferred regulatory endpoint. PFS includes deaths and thus can be a better correlate to overall survival. In TTP analysis, deaths are censored, either at the time of death or at an earlier visit representing informative censoring (nonrandom pattern of loss from the study). PFS assumes patient deaths are randomly related to tumor progression. However, in situations where the majority of deaths are unrelated to cancer, TTP can be an acceptable endpoint.

b. PFS as an endpoint to support drug approval

Table 1 provides advantages and disadvantages of using PFS as an endpoint. PFS can reflect tumor growth and be assessed before the determination of a survival benefit. Its determination is not confounded by subsequent therapy. For a given sample size, the magnitude of effect on PFS can be larger than the effect on overall survival. However, the formal validation of PFS as a surrogate for survival for the many different malignancies that exist can be difficult. Data are usually insufficient to allow a robust evaluation of the correlation between effects on survival and PFS. Cancer trials are often small, and proven survival benefits of existing drugs are generally modest. The role of PFS as an endpoint to support licensing approval varies in different cancer settings. Whether an improvement in PFS represents a direct clinical benefit or a surrogate for clinical benefit depends on the magnitude of the effect and the risk-benefit of the new treatment compared to available therapies.

c. PFS trial design issues

The methodology for assessing, measuring, and analyzing PFS should be detailed in the protocol and statistical analysis plan (SAP). It is also important to carefully define tumor progression criteria in the protocol. There are no standard regulatory criteria for defining progression. Applicants have used a variety of different criteria, including the RECIST criteria. The broad outline presented in most published PFS criteria should be supplemented with additional details in the protocol and SAP. Visits and radiological assessments should be symmetric between the

two study arms to prevent systematic bias. When possible, studies should be blinded. Blinding is particularly important when patient or investigator assessments are included as components of the progression endpoint. At a minimum, the assessments should be subjected to a blinded independent adjudication team, generally consisting of radiologists and clinicians. The FDA and the applicant should agree prospectively on the following items:

- The study design
- The definition of progression
- The data to be recorded on the case report form (CRF)
- The SAP
- The methodology for handling missing data and censoring methods
- The operating procedures of an independent endpoint review committee (IRC), if applicable (see Appendix 4)

> d.   Analysis of PFS

Missing data can complicate analysis of PFS. The protocol should define an adequate assessment visit for each patient (i.e., a visit when all scheduled tumor assessments have been done). The analysis plan should outline a comparison of the adequacy of follow-up in each treatment arm. Methodology for analyzing incomplete and/or missing follow-up visits and censoring methods should be specified in the protocol. The analysis plan should specify the primary analysis and one or more sensitivity analyses to evaluate the robustness of the results. Although any analyses with missing data can be problematic, the results can be strengthened by similar results in both the primary and the sensitivity analyses. The evaluation should include the number of deaths in patients who have been lost to follow-up for a prolonged time period. An imbalance in such deaths could bias the PFS measurement by overestimating PFS in the treatment arm with less follow-up.

Because progression data can be collected from multiple sources (including physical exams at unscheduled visits and radiological scans of various types) and at different times, data collection for each assessment visit should be limited to a specified short time interval around the scheduled visit. Difficulties can arise in determining the event date and censoring date when data are collected over a prolonged time period. We recommend 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 radiological assessment determined a lack of progression. Appendix 3 provides a set of tables for potential analyses of PFS that can be used for primary or sensitivity analyses. Plans for PFS data collection and analysis should be discussed with the FDA at end-of-phase 2 meetings and verified in special protocol assessments.

> 4.   *Time-to-Treatment Failure*

TTF is defined as a composite endpoint measuring time from randomization to discontinuation of treatment for any reason, including disease progression, treatment toxicity, and death. TTF is not recommended as a regulatory endpoint for drug approval. TTF does not adequately distinguish efficacy from these additional variables. A regulatory endpoint should clearly

distinguish the efficacy of the drug from toxicity, patient or physician withdrawal, or patient intolerance.

C.   Endpoints Involving Symptom Assessment

Symptomatic improvement is considered a clinical benefit. FDA drug approvals have used patient symptom assessments and/or physical signs representing symptomatic improvement (e.g., weight gain, decreased effusion) as the primary efficacy endpoint. However, measures of global health-related quality of life (HRQL) have not served as primary efficacy endpoints in oncology drug approvals. HRQL instruments and their validation are discussed in the draft guidance for industry *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*.[8] For the improvement of signs and symptoms or QOL assessments to be used as primary endpoints to support cancer drug approval, the FDA should be able to distinguish between improvement in tumor symptoms and lack of drug toxicity. An apparent effectiveness advantage based on a global HRQL instrument can simply indicate less toxicity rather than effectiveness.

1.   Specific Symptom Endpoints

Time to progression of cancer symptoms, an endpoint similar to TTP, is a direct measure of clinical benefit rather than a potential surrogate. As discussed earlier, problems in measuring progression (e.g., missing assessments) also exist in evaluating time to symptomatic progression. Because few cancer trials are blinded, assessments can be biased. A delay between tumor progression and the onset of cancer symptoms can occur. Often alternative treatments are initiated before achieving the symptom endpoint, confounding this analysis. Many cancer trials are performed in patients who may have minimal cancer symptoms. In addition, tumor symptoms can be difficult to differentiate from drug toxicity.

A *composite symptom endpoint* should have components of similar clinical importance and the results should not be exclusively attributed to one component. For example, drugs have been approved for treatment of patients with cancer metastases to the skeleton based on a composite benefit endpoint. Skeletal-related events are defined as pathological fractures, radiation therapy to bone, surgery to bone, and spinal cord compression.

Selection of the appropriate population can be critical for documenting symptom benefit. Patients symptomatic at study baseline can be evaluated with a categorical symptom response analysis. In asymptomatic patients at baseline, a time-to-first-symptom analysis can be used. If patients discontinue the study drug or begin a new drug, symptomatic progression can still be assessed if follow-up is continued until documentation of the first symptom.

2.   Problems Encountered with Symptom Data

Missing data and infrequent assessments can complicate the evaluation of symptom data especially in open-label studies. Withdrawing treatment because of drug toxicity or tumor

---

[8] When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a guidance, check the CDER guidance Web page at http://www.fda.gov/cder/guidance/index.htm.

progression is one cause of missing symptom data. Ideally, when patients stop treatment, data collection forms should continue to gather information to inform the analysis. Data collection on multiple symptoms should be addressed prospectively regarding multiplicity and the necessary statistical adjustments should be specified in the SAP.

D. Biomarkers

Generally, biomarkers assayed from blood or body fluids have not served as primary endpoints for cancer drug approval, although paraprotein levels measured in blood and urine have been used as part of myeloma response criteria. Further research is needed to establish the validity of available tests and determine whether improvements in biomarkers predict clinical benefit.

The FDA has accepted tumor markers as elements of a composite endpoint. The occurrence of certain clinical events (a significant decrease in performance status, or bowel obstruction) in conjunction with marked increases in CA-125 was considered progression in ovarian cancer patients. Alternatively, biomarkers can be useful in identifying prognostic factors and in selection of patients and stratification factors to be considered in study designs.

IV. CLINICAL TRIAL DESIGN CONSIDERATIONS

Per 21 CFR 314.126, the FDA approves drugs based on substantial evidence of efficacy from "adequate and well-controlled investigations," as described in the regulations. Studies must allow a valid comparison to a control and must provide a quantitative assessment of the drug's effect. The most reliable method for demonstrating efficacy is to show a statistically significant improvement in a clinically meaningful endpoint in blinded randomized controlled trials. The following sections discuss several issues related to the design of cancer trials.

A. Single-Arm Studies

In settings where there is no available therapy and where major tumor regressions can be presumed to be attributed to the tested drug, the FDA has sometimes supported ORR and response duration observed in single-arm studies as substantial evidence supporting accelerated approval. Response rates have been used in settings such as acute leukemia for regular approval where complete responses have been associated with decreased transfusion requirements, decrease in infections, and increased survival. Single-arm trials do not adequately characterize time-to-event endpoints such as survival, TTP, or PFS. Because of variability in the natural history of many forms of cancer, a randomized study is necessary to evaluate time-to-event endpoints.

B. Studies Designed to Demonstrate Noninferiority

A noninferiority (NI) trial should demonstrate the new drug's effectiveness by showing that the new drug is not less effective than a standard regimen (the active control) by a prespecified amount (*noninferiority margin*) (Temple and Ellenberg, 2000, Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments, Part 1: Ethical and Scientific Issues,

Ann Intern Med, 133(6):455-63).[9] This noninferiority margin should be a clinically acceptable loss that is not larger than the effect of the active control drug. The standard regimen should have a well-characterized clinical benefit (survival benefit). If the new drug is inferior to the active control by more than the noninferiority margin, it will be presumed to be ineffective.

NI trials rely on external (historical) data to establish the active control's treatment effect size. In cancer trials, this effect frequently has not been adequately characterized. NI trials also rely on constancy assumption. This assumption includes that the active-control effect has remained constant between the historical study and the current study. This assumes constancy of patient population characteristics, supportive care measures, and evaluation techniques between the current trial and the historical data from which the active-control effect was derived. The estimated size of the active-control's treatment effect should be based on a comprehensive meta-analysis of historical studies. These studies should reliably reproduce the active-control effect compared with placebo arm. Difficulties in conducting NI trials include the estimation of active-control effect and the determination of amount of effect (NI margin) to be retained. NI trials usually involve large sample sizes compared with superiority trials and involve replication of clinical trial results. Furthermore, subsequent therapies and crossover to the active-control arm can confound any NI analysis. NI trials with endpoints other than survival are problematic.

C.   Trial Designs for Radiotherapy Protectants and Chemotherapy Protectants

Radiotherapy protectants and chemotherapy protectants are drugs designed to ameliorate the toxicities of therapies. These trials usually have two objectives. The first is to assess the amelioration of cancer treatment toxicity. The second objective is to determine whether anticancer activity is compromised by the protectant. The second objective usually examines surrogate endpoints; for example, ORR or TTP, rather than overall survival.

V.   CONCLUSION

Although general principles outlined in this guidance should help applicants select endpoints for marketing applications, we recommend that applicants meet with the FDA before submitting protocols intended to support NDA or BLA marketing applications. The FDA will ensure that these meetings include a multidisciplinary FDA team of oncologists, statisticians, clinical pharmacologists, and often external expert consultants. Applicants can submit protocols after these meetings and request a special protocol assessment that provides confirmation of the acceptability of endpoints and protocol design to support drug marketing applications.[10]

---

[9] See also the ICH guidance for industry *E10 Choice of Control Group and Related Issues in Clinical Trials* (http://www.fda.gov/cder/guidance/index.htm).

[10] See the guidance for industry *Special Protocol Assessment* (http://www.fda.gov/cder/guidance/index.htm).

Ultimately, of course, marketing approval depends not only on the design of clinical trials, but on FDA review of the results and data from all studies in the drug marketing application.

APPENDIX 1:
TUMOR MEASUREMENT DATA COLLECTION[11]

The following are important considerations for tumor measurement data. We recommend that:

- The CRF and electronic data document the target lesions identified during the baseline visit before treatment. Retrospective identification of such lesions would not be considered reliable.

- Tumor lesions be assigned a unique identifying letter or number. This assignment provides differentiation among multiple tumors occurring at one anatomic site and the matching of tumors measured at baseline and tumors measured during follow-up.

- A mechanism be in place that ensures complete data collection at critical times during follow-up. The CRF should ensure that all target lesions are assessed at baseline and that the same imaging or measuring method is used for all tests required at baseline and follow-up.

- The CRF contains data fields that indicate whether scans were performed at each visit.

- A zero be recorded when a lesion has completely resolved. Otherwise, disappearance of a lesion cannot be differentiated from a missing value.

- Follow-up tests provide for timely detection of new lesions both at initial and new sites of disease. The occurrence and location of new lesions should be recorded in the CRF and in the submitted electronic data.

---

[11] For the purposes of this appendix, *tumor data* refers to data in SAS transport files, not images. Generally, images are not submitted to the NDA or BLA, but can be audited by the FDA during the review process.

APPENDIX 2:
ISSUES TO CONSIDER IN PFS ANALYSIS

The protocol and SAP should detail the primary analysis of PFS. This analysis should include a detailed description of the endpoint, appropriate modalities for evaluating tumors, and procedures for minimizing bias, such as procedures for an IRC. One or two secondary analyses should be specified to evaluate anticipated problems in trial conduct and to assess whether results are robust. The following important factors should be considered.

- **Definition of progression date.** In survival analyses, the exact death date is known. In PFS analyses, the exact progression date is unknown. The following two methods can be used for defining the *recorded progression date (PDate)* used for PFS analysis.

  1. PDate assigned to the first time at which progression can be declared.

     − For progression based on a new lesion, the PDate is the date of the first observation that the new lesion was detected.

     − If multiple assessments based on the sum of target lesion measurements are done at different times, the PDate is the date of the last observation or radiological assessment of target lesions that shows a predefined increase in the sum of the target lesion measurements.

  2. PDate as the date of the protocol-scheduled clinic visit immediately after all radiological assessments (which collectively document progression) have been done.

- **Definition of censoring date.** Censoring dates are defined in patients with no documented progression before data cutoff or dropout. In these patients, the censoring date is often defined as the last date on which progression status was adequately assessed. One acceptable approach uses the date of the last assessment performed. However, multiple radiological tests can be evaluated in the determination of progression. A second acceptable approach uses the date of the clinic visit corresponding to these radiological assessments.

- **Definition of an adequate PFS evaluation.** In patients with no evidence of progression, censoring for PFS often relies on the date of the last *adequate tumor assessment.* A careful definition of what constitutes an adequate tumor assessment includes adequacy of target lesion assessments and adequacy of radiological tests both to evaluate nontarget lesions and to search for new lesions.

- **Analysis of partially missing tumor data.** Analysis plans should describe the method for calculating progression status when data are partially missing from *adequate tumor assessment* visits.

- **Completely missing tumor data.** Assessment visits where no data are collected are sometimes followed by death or by assessment visits showing progression. In other cases, the subsequent assessment shows no progression. In the latter case, it may seem appropriate

to continue the treatment and continue monitoring for progression. However, this approach treats missing data differently depending upon subsequent events and can represent informative censoring. Another possible approach is to include data from subsequent PFS assessments. This can be appropriate when evaluations are frequent and when only a single follow-up visit is missed. Censoring at the last adequate tumor assessment can be more appropriate when there are two or more missed visits. The SAP should detail primary and secondary PFS analyses to evaluate the potential effect of missing data. Reasons for dropouts should be incorporated into procedures for determining censoring and progression status. For instance, for the primary analysis, patients going off-study for undocumented clinical progression, change of cancer treatment, or decreasing performance status can be censored at the last adequate tumor assessment. The secondary sensitivity analysis would include these dropouts as progression events. Although missed visits for progression can be problematic, all efforts should be made to keep following patients for disease progression irrespective of the number of visits missed.

- **Progression of nonmeasurable disease.** When appropriate, progression criteria should be described for each assessment modality (e.g., CT scan, bone scan). Scans documenting progression based on nonmeasurable disease should be verified by a blinded review committee and be available for verification by the FDA.

- **Suspicious lesions.** An algorithm should be provided for evaluating and following indeterminate lesions for assignment of progression status at the time of analysis.

APPENDIX 3:
EXAMPLE TABLES FOR PFS ANALYSIS

As discussed in section III.B., sensitivity analyses can be helpful in determining whether the PFS analysis is robust. However, these sensitivity analyses are exploratory and supportive of the results of the primary analysis, and efficacy may not be claimed based on sensitivity analysis alone. Different sensitivity analyses can be described in tables that specify how dates of progression events and dates for censoring of progression data can be assigned. The following three tables describe examples of three different sensitivity analyses.

a. Table A represents a sensitivity analysis that only includes well-documented and verifiable progression events. Other data are censored. In Table A, the progression dates are:

- Based only on radiological assessments verified by an IRC. *Clinical progression* is not considered a progression endpoint.

- Assigned to the first time when tumor progression was noted.

- The date of death when the patient is closely followed. However, deaths occurring after two or more missed visits are censored at the last visit.

Table A. PFS 1 (includes documented progression only)

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Earliest of: • Date of radiological assessment showing new lesion (if progression is based on new lesion); or • Date of last radiological assessment of measured lesions (if progression is based on increase in sum of measured lesions)	Progressed
No progression	Date of last radiological assessment of measured lesions	Censored
Treatment discontinuation for undocumented progression	Date of last radiological assessment of measured lesions	Censored
Treatment discontinuation for toxicity or other reason	Date of last radiological assessment of measured lesions	Censored
New anticancer treatment started	Date of last radiological assessment of measured lesions	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last radiological assessment of measured lesions	Censored

The sensitivity analysis in Table B corrects for potential bias in follow-up schedules for tumor assessment by assigning the dates for censoring and events only at scheduled visit dates. However, this approach can introduce bias if the progression occurred closer to the

last visit, particularly in an open-label study. This approach can be suitable in blinded, randomized studies.

Table B. PFS 2 (uniform progression and assessment dates)

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Date of next scheduled visit	Progressed
No progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for undocumented progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started	Date of last visit with adequate assessment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last visit with adequate assessment	Censored

b. The sensitivity analysis in Table C evaluates PFS according to the investigator's assessment. However, this approach can introduce bias if the progression occurred closer to the last visit, particularly in an open-label study. This approach can be suitable in blinded, randomized studies.

Table C. PFS 3 (includes investigator claims)

Situation	Date of Progression or Censoring	Outcome
No baseline assessment	Randomization	Censored
Progression documented between scheduled visits	Next scheduled visit	Progressed
No progression	Date of last visit with adequate assessment	Censored
Investigator claim of clinical progression	Scheduled visit (or next scheduled visit if between visits)	Progressed
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started with no claim of progression	Date of last visit with adequate assessment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits or after patient misses one assessment visit	Date of death	Progressed
Death after an extended lost-to-follow-up time (two or more missed assessments)	Last visit with adequate assessment	Censored

APPENDIX 4:
INDEPENDENT REVIEW OF TUMOR ENDPOINTS

Clinical trial results that support drug approval should be verifiable by applicants and the FDA. ORR determined in single-arm studies can be evaluated by reviewing a limited number of images. When drug approval is based on measurement of PFS, careful planning can minimize bias and enable the applicant and the FDA to verify results. An IRC can minimize bias in radiographic interpretation of the radiological findings and independent adjudication of assessments. A clearly written plan of the charter outlining the IRC function and process (independent review charter) should be agreed upon with the FDA before initiation of the study. The plan should describe the assurance of the committee's independence and procedure for collection, storage, and transportation of the results. The charter also should include the resolution of differences in interpretation and incorporation of clinical data in the final interpretation of data and audit procedures. The use of an IRC is discussed further in the guidance for industry *Developing Medical Imaging Drug and Biological Products, Part 3: Design, Analysis, and Interpretation of Clinical Studies*.