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

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

**December 2018
Clinical/Medical**

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

Additional copies are available from:

*Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov*

<https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>

and/or

*Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov*

<https://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>

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

**December 2018
Clinical/Medical**

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	BACKGROUND	2
A.	Statutory and Regulatory Requirements for Effectiveness	2
B.	Endpoints Supporting Past Approvals in Oncology	3
III.	GENERAL ENDPOINT CONSIDERATIONS	4
A.	Overall Survival	7
B.	Endpoints Based on Tumor Assessments	7
1.	<i>Disease-Free Survival (and Event-Free Survival)</i>	<i>8</i>
2.	<i>Objective Response Rate</i>	<i>9</i>
3.	<i>Complete Response</i>	<i>9</i>
4.	<i>Time to Progression and Progression-Free Survival</i>	<i>10</i>
5.	<i>Time to Treatment Failure</i>	<i>11</i>
C.	Endpoints Involving Symptom Assessment	11
1.	<i>Specific Symptom Endpoints</i>	<i>11</i>
2.	<i>Problems Encountered with Symptom Data</i>	<i>12</i>
D.	Blood or Body Fluid-Based Biomarkers	12
E.	Emerging Endpoints	12
IV.	CLINICAL TRIAL DESIGN AND ANALYSIS CONSIDERATIONS	13
A.	Single-Arm Studies	13
B.	Randomized Studies Designed to Demonstrate Noninferiority	13
C.	Trial Designs for Radiotherapy Protectants and Chemotherapy Protectants	14
D.	Clinical Trial Design Considerations	14
E.	Clinical Trial Analysis Issues	14
V.	CONCLUSION	15
	APPENDIX: INDEPENDENT REVIEW OF TUMOR ENDPOINTS	16

Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry¹

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA office responsible for this guidance as listed on the title page.

I. INTRODUCTION

This guidance 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.³ This guidance does not address endpoints for drugs to prevent or decrease the incidence of cancer.

This guidance is a revision of the final guidance of the same title that published in May 2007. This guidance replaces the May 2007 guidance of the same title.

In general, FDA's guidance documents 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 Oncology Center of Excellence, Center for Drug Evaluation and Research (CDER), and 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 See the guidance for industry *Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics* for recommendations specific to non-small cell lung cancer clinical trials. See the guidance for industry *Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval* for recommendations specific to high-risk early-stage breast cancer. We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA web page. The guidances mentioned in this document are available on the Biologics guidance web page at <http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>, and/or the Drugs guidance web page at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/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. Statutory and 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 (FD&C 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.⁴ Clinical benefits that have supported drug approval have included important clinical outcomes (e.g., increased survival, symptomatic improvement) but have also included effects on surrogate endpoints known to predict clinical benefit (e.g., blood pressure).

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 generally 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.⁵ Such surrogate endpoints are less well-established than those surrogate endpoints for traditional approvals, such as blood pressure for cardiovascular disease. FDA may also grant accelerated approval based on an effect on a clinical endpoint that can be measured earlier than irreversible morbidity or mortality that is reasonably likely to predict an effect on irreversible morbidity or mortality or other clinical benefit (i.e., an intermediate clinical endpoint).⁶ 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.⁷ If the postmarketing studies fail to demonstrate clinical benefit or if the applicant does not demonstrate due diligence in conducting the required studies, among other reasons, FDA may withdraw approval of the drug or indication.⁸ In the following discussion, the term *traditional approval* denotes the longstanding route of drug approval based on the demonstration of clinical benefit or an effect on a surrogate endpoint

⁴ Potency has long been interpreted to include effectiveness (21 CFR 600.3(s)).

⁵ Section 506(c)(1)(B) of the FD&C Act. 21 CFR 314.510 and 601.41 provide that the Agency may consider “. . . epidemiologic, therapeutic, pathophysiologic, or other evidence . . .” in determining whether an endpoint is reasonably likely to predict clinical benefit. The Food and Drug Administration Safety and Innovation Act of 2012 provides that FDA may consider “. . . epidemiological, pathophysiological, therapeutic, pharmacologic, or other evidence developed using biomarkers, for example, or other scientific methods or tools.” See guidance for industry *Expedited Programs for Serious Conditions – Drugs and Biologics*.

⁶ See guidance for industry *Expedited Programs for Serious Conditions – Drugs and Biologics*.

⁷ For a more comprehensive discussion of this condition of accelerated approval and a discussion of other conditions of approval, see the guidance for industry *Expedited Programs for Serious Conditions – Drugs and Biologics*.

⁸ See section 506(c)(3) of the FD&C Act and §§ 314.530(a) and 601.43(a).

Contains Nonbinding Recommendations

known to predict clinical benefit. That term is distinguished from *accelerated approval*, which is associated with use of a surrogate endpoint or intermediate clinical endpoint that is reasonably likely to predict clinical benefit to support drug approval.

The evidence needed to establish effectiveness is discussed in the guidance for industry *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* and the guidance for industry *FDA Approval of New Cancer Treatment Uses for Marketed Drug and Biological Products*. In many cases, at least two adequate and well-controlled clinical investigations are needed. In certain cases, evidence from a single adequate and well-controlled clinical investigation, with confirmatory evidence, 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). FDA may consider data from one adequate and well-controlled clinical investigation and confirmatory evidence as substantial evidence, if FDA determines that such data and evidence are sufficient to establish effectiveness.⁹ For example, for drugs approved for treatment of patients with a specific stage of a particular malignancy, evidence from one adequate and well-controlled clinical investigation with confirmatory evidence 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 traditional approval, applicants show direct evidence of clinical benefit or improvement in a surrogate endpoint known to predict 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 Oncologic Drugs Advisory Committee (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, improved physical functioning, or improved tumor-related symptoms. These benefits may not always be predicted by, or correlate with, ORR.

Over time, larger improvements in tumor reduction and delay in tumor growth have been seen, and tumor measurement endpoints have been used to support both traditional and accelerated approval. Improvement in disease-free survival (DFS) has 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 has supported traditional approval in leukemia, where complete response (CR) is associated with less infection, bleeding, and blood product support. A large improvement in progression-free survival (PFS) or high, substantiated durable ORR has been used to support traditional approval in select malignancies, but magnitude of effect, relief of tumor-related symptoms, and drug toxicity should also be considered when making the approval decision.^{10,11,12} For example, randomized trials for

⁹ See section 505(d) as amended by the Food and Drug Administration Modernization Act of 1997.

¹⁰ Blumenthal GM, PG Kluetz, J Schneider, KB Goldberg, AE McKee, R Pazdur, 2017, Oncology Drug Approvals: Evaluating Endpoints and Evidence in an Era of Breakthrough Therapies, *The Oncologist*, 22(7): 762-767.

¹¹ Blumenthal GM and R Pazdur, 2016, Response Rate as an Approval Endpoint in Oncology: Back to the Future, *JAMA Oncol*, 2(6): 720-721.

¹² 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.

Contains Nonbinding Recommendations

hormonal drugs for breast cancer and a single arm trial of a drug for ROS1-positive metastatic non-small cell lung cancer have used ORR as an endpoint supporting traditional approval. Improvement in tumor-related symptoms in conjunction with an improved ORR and adequate response duration has supported traditional approval in several clinical settings.

Surrogate endpoints for accelerated approval must be reasonably likely to predict clinical benefit (FD&C Act § 506(c)(1)(A); 21 CFR part 314, subpart H; and 21 CFR part 601, subpart E). While durable ORR has been used as a traditional approval endpoint in some circumstances, ORR has also 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. Whether tumor measures such as ORR or PFS are used as an accelerated approval or traditional approval endpoint will depend on the disease context and the magnitude of the effect, among other factors (see sections III.B.2 and 4).

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 (see the guidances referenced in footnote 3 regarding non-small cell lung cancer and high-risk early-stage breast cancer for more detailed discussion for these diseases). The endpoints that are discussed in this section include overall survival, endpoints based on tumor assessments (e.g., DFS, event-free survival (EFS), ORR, CR, time to progression (TTP), and PFS), endpoints involving symptom assessment, blood or body fluid-based biomarkers, and emerging endpoints. Tables 1 and 2 provide 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*. Recommendations regarding the use of placebos and blinding in randomized controlled clinical trials are described in the draft guidance for industry *Hematologic Malignancy and Oncologic Disease: Considerations for Use of Placebos and Blinding in Randomized Clinical Trials for Drug Product Development*.¹³ Cancer patient and caregiver experiences provide unique insights that contribute to important patient preference information for identifying relevant clinical trial endpoints to ultimately inform medical product development programs that best meet patient needs. Recommendations on how stakeholders can collect and submit patient experience data¹⁴ and other relevant information from patients and caregivers for medical product development and regulatory decision making are described in the draft guidance for industry, FDA staff, and other stakeholders *Patient-Focused Drug Development: Collecting Comprehensive and Representative Input*.¹⁵

¹³ When finalized this guidance will represent FDA current thinking on the topics it addresses.

¹⁴ 21st Century Cures Act: <https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf>.

¹⁵ When finalized this guidance will represent FDA current thinking on the topics it addresses.

Contains Nonbinding Recommendations

Table 1. A Comparison of Important Cancer Approval Endpoints

As noted in the table, several oncology endpoints can serve different purposes (i.e., clinical endpoint that represents clinical benefit for traditional approval, surrogate endpoint to support traditional approval, surrogate endpoint to support accelerated approval) based on the specific context of use. The determination is based on the specific diseases and is highly dependent upon factors such as effect size, effect duration, depth of response (e.g., number of CRs), available therapy, disease setting, location of disease, the clinical consequences of delaying or preventing disease progression or delaying administration of more toxic therapies, and the risk-benefit relationship. See text for details. See section V regarding recommendations for obtaining FDA feedback on endpoints and protocol design.

Endpoint	Type of Endpoint			Study Design Recommendations		
	Clinical Endpoint	Surrogate Endpoint for TA*	Surrogate Endpoint for AA**	Randomized	Single-Arm	Independent Blinded Review
Overall Survival	X			X		
Symptom Endpoints (patient-reported outcomes)	X			X		
Disease-Free Survival or Event-Free Survival	X	X	X	X		X***
Objective Response Rate	X	X	X	X	X	X
Complete Response	X	X	X	X	X	X
Progression-Free Survival or Time to Progression	X	X	X	X		X***

* TA - Traditional approval, ** AA - Accelerated approval, *** Not always recommended

Contains Nonbinding Recommendations

Table 2. Advantages and Disadvantages of Important Cancer Approval Endpoints

Endpoint	Advantages	Disadvantages
Overall Survival	<ul style="list-style-type: none"> • Easily and precisely measured • Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> • May be affected by switch-over of control to treatment or subsequent therapies • Needs longer follow-up • Includes noncancer deaths
Symptom Endpoints (patient-reported outcomes)	<ul style="list-style-type: none"> • Generally assessed earlier and with smaller sample size compared with survival studies 	<ul style="list-style-type: none"> • Blinding is important for assessing the endpoint • Potentially subject to assessment bias, particularly in open-label studies • Lack of validated instruments in many disease areas • Definitions vary among studies • Balanced timing of assessments among treatment arms is critical
Disease-Free Survival or Event-Free Survival	<ul style="list-style-type: none"> • Generally assessed earlier and with smaller sample size compared with survival studies • Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> • Potentially subject to assessment bias, particularly in open-label studies • Definitions vary among studies • Balanced timing of assessments among treatment arms is critical • Includes noncancer deaths
Objective Response Rate	<ul style="list-style-type: none"> • Generally assessed earlier and with smaller sample size compared with survival studies • Effect on tumor attributable to drug(s), not natural history • Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> • Definitions vary among studies • Frequent radiological or other assessments • May not always correlate with survival
Complete Response	<ul style="list-style-type: none"> • Generally assessed earlier and with smaller sample size compared with survival studies • Effect on tumor attributable to drug(s), not natural history • Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> • Definitions vary among studies • Frequent radiological or other assessments • May not always correlate with survival
Progression-Free Survival or Time to Progression	<ul style="list-style-type: none"> • Generally assessed earlier and with smaller sample size compared with survival studies • Measurement of stable disease included • Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> • Potentially subject to assessment bias, particularly in open-label studies • Definitions vary among studies • Frequent radiological or other assessments • Balanced timing of assessments among treatment arms is critical • May not always correlate with survival

Contains Nonbinding Recommendations

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 externally controlled trials are seldom reliable for time-to-event endpoints, including overall survival. Apparent differences in outcome between external 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 known and unknown 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. In many cancer types, radiographic tumor assessments directly measure components of the disease, and tumor measures commonly trigger treatment decisions in clinical practice. Therefore, tumor measure-based endpoints are considered more clinically relevant than other biomarkers. Tumor measure-based endpoints may support either traditional (as a clinical endpoint that represents clinical benefit or a surrogate endpoint for traditional approval) or accelerated approval and include DFS (and EFS), ORR, CR, TTP, and PFS.

Tumor assessment endpoint selection should include two judgments. First, a determination of whether the endpoint may support either accelerated approval or traditional approval should be ascertained. Second, the endpoint should be evaluated for the potential of bias or uncertainty. Drug applications using studies that rely on tumor assessment 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 ORR determinations can be imprecise in locations where there is a lack of demarcated margins (e.g., pleural or peritoneal mesothelioma, pancreatic cancer, brain tumors).

When the primary study endpoint is based on tumor measurements (e.g., PFS or ORR), tumor assessments generally should be verified by central reviewers blinded to study treatments (see Appendix) to ascertain lack of assessment bias. Additional details regarding data collection are listed in the guidance for industry *Clinical Trial Endpoints for the Approval of Non-Small Cell*

Contains Nonbinding Recommendations

Lung Cancer Drugs and Biologics, Appendix A.¹⁶ Alternatively, a random sample-based blinded central review auditing approach could be used with a detailed auditing plan prespecified including a strategy to detect potential assessment bias. If bias cannot be excluded based upon the audit, a blinded central review of all radiographic images will be necessary.¹⁷ Sponsors should seek FDA input prior to conducting an audit based central review. Centralized independent verification of tumor assessment endpoints (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 investigator bias.

1. Disease-Free Survival (and Event-Free Survival)

Generally, DFS is defined as the time from randomization until disease recurrence 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 CRs 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 an overall survival endpoint impractical. An endpoint that is similar to DFS but is differentiated from it in that randomization takes place before definitive surgery or radiotherapy in the adjuvant setting is EFS. For the purpose of this guidance, EFS is defined as time from randomization to any of the following events: progression of disease that precludes surgery, local or distant recurrence, or death due to any cause. DFS has been the primary basis of approval for adjuvant breast cancer hormonal therapy, adjuvant colon cancer, and adjuvant cytotoxic breast cancer therapy, whereas EFS is an appropriate endpoint for the evaluation of neoadjuvant breast cancer therapy given prior to definitive surgery. Compared with standard cytotoxic therapies, hormonal therapies carry minimal side effects and thus a favorable risk-benefit relationship. Treatment effect measured by DFS or EFS can be a surrogate endpoint to support accelerated approval, a surrogate endpoint to support traditional approval, or it can represent direct clinical benefit based on the specific disease, context of use, magnitude of the effect, the disease setting, available therapy, and the risk-benefit relationship. 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 a clinical endpoint for colon cancer drugs in the surgical adjuvant setting.¹⁸ DFS has been used for traditional approval for diseases such as breast cancer, colorectal cancer, gastrointestinal stromal tumors, melanoma, and renal cell carcinoma.

Important considerations in evaluating DFS or EFS as a potential endpoint include the estimated size of the treatment effect and proven benefits of standard therapies. The protocol should

¹⁶ Appendices A-D in the guidance for industry *Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics*, regarding tumor measurement data collection and PFS analysis also apply more broadly to oncology endpoints and should be used in conjunction with this guidance, as needed.

¹⁷ Discussed during July 24, 2012 ODAC meeting. Transcripts are available at <https://wayback.archive-it.org/7993/20170403224012/https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/default.htm>.

¹⁸ Transcripts are available at <https://wayback.archive-it.org/7993/20170403224012/https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/default.htm>.

Contains Nonbinding Recommendations

carefully delineate both the definition of DFS or EFS and the schedule for follow-up assessments 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, as appropriate.

Application of the definition of DFS or EFS in a study 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 deaths from all causes as recurrences can minimize bias. DFS or EFS 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 or EFS events and censor noncancer deaths. This method can introduce bias in the attribution of the cause of death. Furthermore, any method that censors observations on patients, whether at death or at the last visit, assumes that the patients with censored observations have the same risk of recurrence as patients with noncensored observations who have not yet experienced the event.

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 CRs. When defined in this manner, ORR is a direct measure of a 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, whereas tumor reduction is a direct therapeutic effect. Also, stable disease can be more accurately assessed by TTP or PFS analysis (see section III.B.4). If available, standardized criteria should be used to ascertain response. A variety of response criteria have been considered appropriate (e.g., revised RECIST guideline (version 1.1))¹⁹ 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 CRs. Treatment effect measured by ORR can be a surrogate endpoint to support accelerated approval, a surrogate endpoint to support traditional approval, or it can represent direct clinical benefit based on the specific disease, context of use, magnitude of the effect, the number of CRs, the durability of response, the disease setting, the location of the tumors, available therapy, and the risk-benefit relationship.

3. Complete Response

CR is defined as no detectable evidence of tumor. CR is generally measured through imaging studies (e.g., CT scans) or through histopathologic assessment (e.g., bone marrow biopsy or breast cancer resection specimens). Treatment effect measured by CR can be a surrogate endpoint to support accelerated approval, a surrogate endpoint to support traditional approval, or it can represent direct clinical benefit based on the specific disease, context of use, magnitude of the

¹⁹ Eisenhauer EA, P Therasse, J Bogaerts, LH Schwartz, D Sargent, R Ford, et al., 2009, New Response Evaluation Criteria in Solid Tumours: Revised RECIST guideline (version 1.1), *Eur J Cancer*, 45(2):228-247.

Contains Nonbinding Recommendations

effect, effect duration, disease setting, location of disease, available therapy, and the risk-benefit relationship. For hematologic malignancies such as acute leukemia, CR has been used as a clinical endpoint for traditional approval. For early high-risk breast cancer, pathologic CR (pCR) has been used as a surrogate endpoint for accelerated approval.

4. 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, whichever occurs first. The precise definition of tumor progression is important and should be carefully detailed in the protocol.

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, death events 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 that death events are randomly related to tumor progression.

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. Data are usually insufficient to allow a robust evaluation of the correlation between effects on overall survival and PFS. Cancer trials are often small, and proven survival benefits of existing drugs are generally modest. Treatment effect measured by PFS can be a surrogate endpoint to support accelerated approval, a surrogate endpoint to support traditional approval, or it can represent direct clinical benefit based on the specific disease, context of use, magnitude of the effect, the disease setting, location of metastatic sites, available therapy, the risk-benefit relationship, and the clinical consequences of delaying or preventing progression in key disease sites (e.g., delay of new lesions in the brain or spine) or delaying administration of more toxic therapies.

It is 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 statistical analysis plan (SAP).

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. The guidance for industry *Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics*, Appendices C and D 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 FDA at end-of-phase 2 meetings.

Contains Nonbinding Recommendations

5. *Time to Treatment Failure*

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 generally not recommended as a regulatory endpoint for new molecular entity drug approval.

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 an important efficacy endpoint. For the improvement of signs and symptoms or other clinical outcome assessments (COA) to be used as efficacy endpoints to support cancer drug approval, symptoms should be assessed that are due to cancer rather than drug toxicity to the extent possible.

1. *Specific Symptom Endpoints*

Symptom improvement/palliation is a direct measure of clinical benefit rather than a surrogate endpoint. A decrease in the severity of cancer symptoms has been used to support traditional approval of anti-cancer agents where anti-tumor activity has also been demonstrated. The use of a symptom palliation endpoint requires that the population be symptomatic at baseline, which can be problematic in many cancer trials where patients can often be asymptomatic at baseline. This endpoint can also be subject to open label response bias, the magnitude of which is not well described.

Time to progression of cancer symptoms, an endpoint similar to TTP, is also a direct measure of clinical benefit rather than a potential surrogate endpoint. 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, symptom assessments can also be subject to response bias. 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. In addition, tumor symptoms can be difficult to differentiate from drug toxicity.

A composite symptom endpoint or symptom scale should have components of similar clinical importance and an analysis of the contribution of the components should be submitted with the primary analysis of the overall composite endpoint. An example is the composite symptom scale that includes several important symptoms of myelofibrosis, the myelofibrosis symptom assessment form.²⁰ An example of a composite endpoint for clinical events exists for patients with cancer metastases to the skeleton. Drugs have been approved based on delaying the time to skeletal-related events defined as pathological fractures, radiation therapy to bone, surgery to bone, and spinal cord compression.

²⁰ Deisseroth A, E Kamiskas, J Grillo, W Chen, H Saber, HL Lu, et. al., 2012, U.S. Food and Drug Administration Approval: Ruxolitinib for the Treatment of Patients with Intermediate and High-Risk Myelofibrosis, Clin Cancer Res, 18(12): 3212-7.

Contains Nonbinding Recommendations

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, although this approach can be limited by intermittent missing data and discontinuation of the study drug prior to symptomatic progression.

2. Problems Encountered with Symptom Data

Missing data and infrequent assessments can complicate the evaluation of symptom data particularly for time to deterioration analyses. Procedures should be put in place to maximize completion rate during trial conduct and the statistical analysis plan should outline how missing data will be handled. Data collection to support multiple symptom endpoints should be addressed prospectively regarding multiple hypotheses testing and the necessary statistical adjustments should be specified in the SAP. Additional information on the use of patient-reported outcomes can be found in the guidance for industry *Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*.

D. Blood or Body Fluid-Based 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, and durable major molecular response is a surrogate endpoint used for traditional approval for therapies in chronic myelogenous leukemia.

The FDA has accepted blood-based 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. In addition, blood-based biomarkers can be useful in identifying prognostic factors and in selection of patients and stratification factors to be considered in study designs.

E. Emerging Endpoints

In addition to the endpoints discussed in this section, FDA recognizes that advances in science are facilitating development of oncology products, which may also result in the identification of additional endpoints that may be used to support approval of oncology products. As examples, minimal residual disease²¹ has been used as a surrogate endpoint for accelerated approval for a therapy in acute lymphoblastic leukemia and metastasis-free survival has been used as a clinical endpoint for traditional approval for a therapy in non-metastatic castration-resistant prostate cancer. If a sponsor is planning to use an emerging endpoint in its clinical development program, we recommend discussing such use with the applicable FDA Division or Office prior to initiating a trial (see section V).

²¹ See draft guidance for industry *Hematologic Malignancies: Regulatory Considerations for Use in Minimal Residual Disease in Development of Drug and Biological Products for Treatment* for information to assist sponsors planning to use minimal residual disease to support marketing approval of drugs and biological products for the treatment of specific hematologic malignancies. When finalized this guidance will represent FDA current thinking on the topics it addresses.

IV. CLINICAL TRIAL DESIGN AND ANALYSIS 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.²² The most reliable method for demonstrating efficacy is to show a statistically significant improvement in a clinically meaningful endpoint in randomized controlled trials (i.e., superiority). The following sections discuss other approaches that may be indicated when a randomized controlled trial demonstrating superiority is not feasible or ethical. If a sponsor is planning to use one of these approaches, we recommend discussing such use with the applicable FDA Division or Office prior to initiating a trial (see section V).

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 accepted 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 traditional approval where CRs 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 overall survival, DFS (and EFS), 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. Randomized 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*).²³ 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 externally controlled (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 externally controlled study and the current study. This assumes constancy of patient population characteristics, supportive care measures, and evaluation techniques between the current trial and the externally controlled 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 externally controlled studies. These studies should reliably reproduce the active-control effect compared with placebo arm.

²² See 21 CFR 314.126(b)(2) and (7).

²³ See guidance for industry *Non-Inferiority Clinical Trials to Establish Effectiveness*. See also the ICH guidance for industry *E10 Choice of Control Group and Related Issues*.

Contains Nonbinding Recommendations

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 overall survival and ORR 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 earlier endpoints; for example, ORR or PFS, rather than overall survival.

D. Clinical Trial Design Considerations

The methodology for assessing, measuring, and analyzing the endpoint(s) should be detailed in the protocol and SAP.

Visits and radiological assessments, when applicable, should be symmetric between the two study arms to prevent systematic bias. The FDA and the applicant should agree prospectively on the following items and the applicant should finalize the items before the initiation of the study, to the extent possible:

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

E. Clinical Trial Analysis Issues

Missing data can complicate endpoint analysis. For endpoints based on tumor assessments, 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. When applicable, the evaluation should include the number of deaths in patients who have been lost to follow-up for a prolonged time period. For example, an imbalance in such deaths could bias a PFS measurement by overestimating PFS in the treatment arm with less follow-up.

V. CONCLUSION

Although the 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 external expert consultants as needed. 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.²⁴ 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.

²⁴ See the guidance for industry *Special Protocol Assessment*.

APPENDIX: INDEPENDENT REVIEW OF TUMOR ENDPOINTS

Clinical trial results that support drug approval should be verifiable by applicants and the Food and Drug Administration (FDA). Objective response rate (ORR) determined in single-arm studies can be evaluated by reviewing a limited number of images. When drug approval is based on measurement of progression-free survival (PFS), careful planning can minimize bias and enable the applicant and the FDA to verify results. An independent endpoint review committee (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 in place 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*.