

**Edited Version of Draft  
"Guidance for Industry..."**

---

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

## ***DRAFT GUIDANCE***

**This guidance document is being distributed for comment purposes only.**

Comments and suggestions regarding this draft document should be submitted 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (CDER drugs) Grant Williams at 301-594-5758, (CDER biologics) Patricia Keegan at 301-827-5097, or (CBER biologics) Steven Hirschfeld at 301-827-6536.

**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)**

**April 2005  
Clinical/Medical**

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

*Additional copies are available from:*

*Office of Training and Communications  
Division of Drug Information, HFD-240  
Center for Drug Evaluation and Research  
Food and Drug Administration  
5600 Fishers Lane  
Rockville, MD 20857  
(Tel) 301-827-4573  
<http://www.fda.gov/cder/guidance/index.htm>*

*or*

*Office of Communication, Training, and  
Manufacturers Assistance, HFM-40  
Center for Biologics Evaluation and Research  
Food and Drug Administration  
1401 Rockville Pike, Rockville, MD 20852-1448  
(Tel) 800-835-4709 or 301-827-1800  
<http://www.fda.gov/cber/guidelines.htm>*

**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)**

**April 2005  
Clinical/Medical**

## TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	BACKGROUND .....	2
A.	Regulatory Requirements for Effectiveness.....	2
B.	Endpoints Supporting Past Approvals in Oncology .....	3
III.	GENERAL ENDPOINT CONSIDERATIONS .....	4
A.	Overall Survival.....	65
B.	Endpoints Based on Tumor Assessments.....	76
1.	<i>Disease-Free Survival</i> .....	87
2.	<i>Objective Response Rate</i> .....	98
3.	<i>Time to Progression and Progression-Free Survival</i> .....	108
a.	TTP vs. PFS .....	109
b.	PFS as an endpoint to support drug approval .....	109
c.	PFS trial design issues.....	1140
d.	Analysis of PFS .....	1140
e.	Future methods for assessing progression.....	1211
4.	<i>Time to Treatment Failure</i> .....	1311
C.	Endpoints Involving Symptom Assessment .....	1312
1.	<i>Specific Symptom Endpoints</i> .....	1412
2.	<i>Problems Encountered with Symptom Data</i> .....	1513
D.	Biomarkers.....	1513
IV.	ENDPOINTS AND CLINICAL TRIAL DESIGN; SELECTED ISSUES .....	1614
A.	Single-Arm Studies.....	1614
B.	Studies Designed to Demonstrate Noninferiority .....	1615
C.	No Treatment or Placebo Control .....	1816
D.	Isolating Drug Effect in Combinations.....	1816
E.	Trial Designs for Radiotherapy Protectants and Chemotherapy Protectants.....	1916
V.	SUMMARY AND CONCLUSION .....	1916
	APPENDIX 1: THE COLLECTION OF TUMOR MEASUREMENT DATA.....	2018
	APPENDIX 2: ISSUES TO CONSIDER IN PFS ANALYSIS .....	2119
	APPENDIX 3: EXAMPLE TABLES FOR PFS ANALYSIS.....	2321
	APPENDIX 4: INDEPENDENT REVIEW OF TUMOR ENDPOINTS.....	2623

**Guidance for Industry<sup>1</sup>**  
**Clinical Trial Endpoints**  
**for the Approval of Cancer**  
**Drugs and Biologics**

This draft guidance, when finalized, will represent 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 provides recommendations to sponsors on endpoints for cancer clinical trials submitted to the FDA to support effectiveness claims in new drug applications (NDAs), biologics license applications (BLAs), or supplemental applications.<sup>2</sup>

The FDA is developing guidance on oncology endpoints through a process that includes public workshops of oncology experts and discussions before the FDA's Oncologic Drugs Advisory Committee (ODAC).<sup>3</sup> This guidance is the first in a planned series of cancer endpoint guidances. It provides background information and discusses general regulatory principles. Each subsequent guidance document will focus on endpoints for specific cancer types (e.g., lung cancer, colon cancer) to support drug approval or labeling claims. The endpoints discussed in this guidance document 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.

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

<sup>1</sup> This guidance has been prepared by the Division of Oncology Drug Products and the Division of Therapeutic Biologic Oncology Drug 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.

<sup>2</sup> For the purposes of this guidance, all references to *drugs* include both human drugs and biological products unless otherwise specified.

<sup>3</sup> Transcripts are available at [http://www.fda.gov/cder/drug/cancer\\_endpoints/default.htm](http://www.fda.gov/cder/drug/cancer_endpoints/default.htm).

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

37 cited. The use of the word *should* in Agency guidances means that something is suggested or  
38 recommended, but not required.

39  
40  
41 **II. BACKGROUND**

42  
43 Clinical trial endpoints serve different purposes. In conventional oncology drug development,  
44 early phase clinical trials evaluate safety and identify evidence of biological drug activity, such  
45 as tumor shrinkage. Endpoints for later phase efficacy studies evaluate whether a drug provides  
46 a clinical benefit such as prolongation of survival or an improvement in symptoms. The  
47 following sections discuss the general regulatory requirements for efficacy and how they have  
48 influenced endpoint selection for the approval of cancer drugs. Later sections describe these  
49 endpoints in more detail and discuss whether they might serve as measures of disease activity or  
50 clinical benefit in various clinical settings.

51  
52 **A. Regulatory Requirements for Effectiveness**

53  
54 The requirement that new drugs show effectiveness is based on a 1962 amendment to the Federal  
55 Food, Drug, and Cosmetic Act. This law requires substantial evidence of effectiveness and  
56 specifies that this evidence must be derived from adequate and well-controlled clinical  
57 investigations. Clinical benefits that have supported drug approval have included important  
58 clinical outcomes (e.g., increased survival, symptomatic improvement) but have also included  
59 effects on established surrogate endpoints (e.g., blood pressure or serum cholesterol).

60  
61 In 1992, the accelerated approval regulations (21 CFR part 314, subpart H and 21 CFR part 601,  
62 subpart E) allowed use of additional endpoints for approval of drugs or biological products that  
63 are intended to treat serious or life-threatening diseases and that either demonstrate an  
64 improvement over available therapy or provide therapy where none exists. In this setting, the  
65 FDA may grant approval based on an effect on a surrogate endpoint that is *reasonably likely* to  
66 predict clinical benefit (“based on epidemiologic, therapeutic, pathophysiologic, or other  
67 evidence”). These surrogates are less well-established than surrogates in regular use, such as  
68 blood pressure or cholesterol for cardiovascular disease. A drug is approved under the  
69 accelerated approval regulations on condition that the manufacturer conduct clinical studies to  
70 verify and describe the actual clinical benefit. If the postmarketing studies fail to demonstrate  
71 clinical benefit or if the applicant does not demonstrate due diligence in conducting the required  
72 studies, the drug may be removed from the market under an expedited process. From December  
73 1992 to June 2004, 22 cancer drug applications were approved under the accelerated approval  
74 regulations. In the following discussion, we will use the term *regular approval* to designate the  
75 longstanding route of drug approval based on demonstrating clinical benefit to distinguish it  
76 from *accelerated approval* associated with use of a surrogate endpoint that is reasonably likely to  
77 predict benefit.

78  
79 In 1997 the FDA Modernization Act established that data from one well-controlled clinical trial,  
80 together with confirmatory evidence obtained either before or after that trial, are sufficient to  
81 establish effectiveness. The nature of evidence to support drug approval, including the preferred  
82 number of clinical trials, is discussed in general FDA guidance documents. ~~In most cases, the~~

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

83 ~~FDA has recommended at least two well-controlled clinical trials. In some cases, t~~The FDA has  
84 found that evidence from a single trial was sufficient, but generally only in cases in which a  
85 single multicenter study provided highly reliable and statistically strong evidence of an important  
86 clinical benefit, ~~such as an effect on survival, and or~~ in which confirmation of the result in a  
87 second trial would be practically or ethically impossible.<sup>4</sup> For drugs approved for treatment of  
88 patients with a specific stage of a particular malignancy, evidence from one trial may be  
89 sufficient to support an efficacy supplement for treatment of a different stage of the same  
90 cancer.<sup>5</sup>

91  
92 **B. Endpoints Supporting Past Approvals in Oncology**

93  
94 For regular approval, it is critical that the sponsor show direct evidence of clinical benefit or  
95 improvement in an established surrogate for clinical benefit. In oncology, survival is the gold  
96 standard for clinical benefit, but the FDA has accepted other endpoints for cancer drug approval.  
97 Indeed, in the 1970s the FDA usually approved cancer drugs based on objective response rate  
98 (ORR), determined by tumor assessments from radiologic tests or physical exam. In the early  
99 1980s, after discussion with the ODAC, -the FDA determined that it would be more appropriate  
100 for cancer drug approval to be based on more direct evidence of clinical benefit, such as  
101 improvement in survival or in a patient's quality of life (QOL), improved physical functioning,  
102 or improved tumor-related symptoms — benefits not always predicted by ORR.

103  
104 Over the next decade, several endpoints were used as surrogates for benefit. Improvement in  
105 disease-free survival supported drug approval in selected surgical adjuvant settings (when a large  
106 proportion of patients had cancer symptoms at the time of recurrence). Durable complete  
107 response was considered an acceptable endpoint in testicular cancer and acute leukemia (a de  
108 facto improvement in survival because the untreated conditions were quickly lethal) and in some  
109 chronic leukemias and lymphomas (where it was clear that remission would lead to less  
110 infection, bleeding, and blood product support). The FDA has also considered that a very high  
111 ORR alone might sometimes support regular approval, but that response duration, relief of  
112 tumor-related symptoms, and drug toxicity should also be considered (O'Shaughnessy and  
113 Wittes et al., 1991, Commentary Concerning Demonstration of Safety and Efficacy of  
114 Investigational Anticancer Agents in Clinical Trials, *J Clin Oncol* 9:2225-2232). ORR has been  
115 an especially important endpoint for the less toxic drugs, such as the hormonal drugs for breast  
116 cancer, where improvement in this endpoint has been the basis for regular approval.  
117 Improvement in tumor-related symptoms in conjunction with an improved ORR and an adequate  
118 response duration supported approval in several clinical settings.

119  
120 In the last decade, in addition to its limited role in regular approval, ORR has been the primary  
121 surrogate endpoint used to support cancer drug accelerated approval for several reasons. First,  
122 ORR is directly attributable to drug effect (tumors rarely shrink spontaneously and, therefore,  
123 ORR can be accurately assessed in single-arm studies). Second, tumor response is widely

---

<sup>4</sup> See guidance for industry *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (<http://www.fda.gov/cder/guidance/index.htm>)

<sup>5</sup> See guidance for industry *FDA Approval of New Cancer Treatment Uses for Marketed Drug and Biological Products* (<http://www.fda.gov/cder/guidance/index.htm>)

**Contains Nonbinding Recommendations**  
**Draft — Not for Implementation**

accepted as relevant by oncologists and has a long-accepted role in guiding cancer treatment. Finally, given the remorselessly progressive nature of the disease, if the ORR is likely to represent a high enough and the responses are of sufficient duration, ORR does indeed seem drug effect which is reasonably likely to predict clinical benefit.

Drugs approved under accelerated approval regulations must provide a benefit over available therapy. To satisfy this requirement, many sponsors have designed single-arm studies in patients with refractory tumors where, by definition, no available therapy exists.

### III. GENERAL ENDPOINT CONSIDERATIONS

The following is 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 guidance documents will discuss these issues in more detail with regard to specific treatment indications. Endpoints that will be discussed include overall survival, endpoints based on tumor assessments (e.g., disease-free survival, ORR, time to progression, progression-free survival, time to treatment failure), and endpoints based on symptom assessment. A comparison of important endpoints in cancer drug approval is provided in Table 1. Many of the issues relating to the proper analysis of efficacy endpoints are addressed in general FDA guidance documents.<sup>6</sup> Issues that commonly arise in oncology applications are discussed in this guidance.

**Table 1. A Comparison of Important Cancer Approval Endpoints**

Endpoint	Regulatory Nature of Evidence	Assessment	Some Advantages	Some Disadvantages
Overall Survival	Clinical benefit for <u>regular approval</u>	<ul style="list-style-type: none"> <li>• Randomized studies needed</li> <li>• Blinding <del>not</del> <u>essential preferred</u></li> <li>• <u>May be biased by any imbalances in treatment decisions</u></li> </ul>	<ul style="list-style-type: none"> <li>• Universally accepted direct measure of benefit</li> <li>• Easily measured</li> <li>• Precisely measured</li> </ul>	<ul style="list-style-type: none"> <li>• <del>Requires larger studies</del></li> <li>• Requires longer studies</li> <li>• Potentially affected by crossover <u>and/or sequential</u> therapy</li> <li>• Does not capture symptom benefit</li> <li>• Includes non-cancer deaths</li> </ul>
Disease-Free Survival	Surrogate for accelerated approval or regular approval*	<ul style="list-style-type: none"> <li>• Randomized studies needed</li> <li>• Blinding preferred</li> </ul>	<ul style="list-style-type: none"> <li>• <del>Considered to be</del> <u>Clearly a clinical benefit by some</u></li> <li>• Needs fewer patients and shorter studies than survival</li> </ul>	<ul style="list-style-type: none"> <li>• Not a validated survival surrogate in <del>most</del> <u>all</u> settings</li> <li>• <del>Not precisely measured;</del> <u>Subject to assessment bias in open-label studies</u></li> <li>• <del>Various definitions exist</del></li> </ul>

<sup>6</sup> See ICH guidance for industry *E9 Statistical Principles for Clinical Trials* (<http://www.fda.gov/cder/guidance/index.htm>)

***Contains Nonbinding Recommendations***  
***Draft — Not for Implementation***

148  
149  
150  
151

\*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.

*continued*

**Contains Nonbinding Recommendations**  
Draft — Not for Implementation

Table 1, continued

Endpoint	Regulatory Nature of Evidence	Assessment	Some Advantages	Some Disadvantages
Objective Response Rate (ORR)	Surrogate for accelerated approval or regular approval*	<ul style="list-style-type: none"> <li>• Single-arm or randomized studies can be used</li> <li>• Blinding preferred in comparative studies</li> </ul>	<ul style="list-style-type: none"> <li>• Can be assessed in single-arm studies</li> <li>• <u>In a heterogeneous population, an ORR could capture a treatment effect that other time to event endpoints would miss</u></li> </ul>	<ul style="list-style-type: none"> <li>• <del>Not a direct measure of benefit</del></li> <li>• <del>Usually reflects drug activity in a minority of patients</del></li> <li>• <del>Data are moderately complex compared to survival</del></li> <li>• <del>May not correlate with changes in other endpoints</del></li> </ul>
Complete Response (CR)	Surrogate for accelerated approval or regular approval*	<ul style="list-style-type: none"> <li>• Single-arm or randomized studies can be used</li> <li>• Blinding preferred in comparative studies</li> </ul>	<ul style="list-style-type: none"> <li>• Durable CRs represent obvious benefit in some settings (see text)</li> <li>• Can be assessed in single-arm studies</li> </ul>	<ul style="list-style-type: none"> <li>• Few drugs produce high rates of CR</li> <li>• <del>Data are moderately complex compared to survival</del></li> </ul>
Progression Free Survival (PFS)	<u>Clinical benefit for regular approval and</u> Surrogate for accelerated approval, depending on the setting or regular approval*	<ul style="list-style-type: none"> <li>• Randomized studies needed</li> <li>• Blinding preferred</li> <li>• Blinded review recommended for open-label studies</li> </ul>	<ul style="list-style-type: none"> <li>• Activity measured in responding and stable tumors</li> <li>• Usually assessed prior to change in therapy</li> <li>• <del>Less missing data than for symptom endpoints</del></li> <li>• Assessed earlier and in smaller studies compared with <u>than survival</u></li> </ul>	<ul style="list-style-type: none"> <li>• <del>Various definitions exist</del></li> <li>• <del>Not a direct measure of benefit</del></li> <li>• <del>Not a validated survival surrogate in all settings</del></li> <li>• <del>Time to Progression (TTP) has to be imputed</del></li> <li>• <del>Not precisely measured compared with survival</del></li> <li>• <del>Is May be subject to assessment bias in open-label studies</del></li> <li>• <del>Frequent radiologic studies are needed</del></li> <li>• <del>Data are voluminous and complex compared to survival</del></li> </ul>
Symptom Endpoints	<u>Clinical benefit for regular approval</u>	<ul style="list-style-type: none"> <li>• <u>Blinding recommended</u> Usually needs randomized blinded studies (unless endpoints have an objective component and effects are large — see text)</li> </ul>	<ul style="list-style-type: none"> <li>• Direct measure of benefit</li> </ul>	<ul style="list-style-type: none"> <li>• <del>Blinding is often difficult in oncology trials</del></li> <li>• <del>Missing data are common problematic</del></li> <li>• <del>Requires the use of validated instruments</del> Few instruments are validated for measuring cancer-specific symptoms</li> <li>• <del>Data are voluminous and complex compared to survival</del></li> </ul>

\*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.

Abbreviations: complete response (CR); objective response rate (ORR); progression-free survival (PFS).

**A. Overall Survival**

**Contains Nonbinding Recommendations**  
*Draft — Not for Implementation*

59 Overall survival is defined as the time from randomization until death from any cause, and is  
160 measured in the intent to treat (ITT) population. Survival is the most reliable cancer endpoint,  
161 and when studies can be conducted to adequately assess it, it is usually the preferred endpoint.  
162 In general, An improvement in survival is of unquestioned a clinical benefit. However, the size  
163 of the survival advantage must be weighed against the toxicity of treatment. The endpoint is  
164 precise and easy to measure, documented by the date of death. Bias is not a factor in endpoint  
165 measurement.

166  
167 Overall survival almost always needs to be evaluated in randomized controlled studies.  
168 Historically controlled data are seldom reliable for time-dependent endpoints such as overall  
169 survival unless treatment effects are extreme (e.g., acute leukemia, testicular cancer). Apparent  
170 differences in outcome between historical controls and current treatment groups can arise from  
171 differences other than drug treatment, including patient selection, improved imaging techniques  
172 (which can alter tumor staging and prognosis), or improved supportive care. Randomized  
173 studies minimize the effect of such differences by allowing a comparison of outcomes in patient  
174 groups where such factors should be similar. Demonstration of a statistically significant  
175 improvement in overall survival is usually considered to be clinically significant, and has often  
176 supported new drug approval.

177  
178 Criticisms of survival as an endpoint stem not from doubts about the worth of a proven survival  
179 benefit, but from difficulties in performing studies large enough or long enough to detect a  
180 survival improvement, difficulties in determining a drug's effect on survival because of the  
181 confounding effects of subsequent cancer therapy, or a concern that the drug may be effective in  
182 only a small fraction of those treated, making it difficult to see an effect on survival in the whole  
183 population. It must be noted that there is a potential that increased overall survival may have  
184 arisen from a comparator that has under performed. In this case, improved overall survival may  
185 not always demonstrate a clinical benefit. If the survival increase is small in magnitude, it is not  
186 necessarily indicative of clinical benefit.

187  
188 **B. Endpoints Based on Tumor Assessments**

189  
190 In this section we discuss several endpoints that are based on tumor assessments and are  
191 therefore unique to oncology. These endpoints include disease-free survival, objective response  
192 rate, time to progression, progression-free survival, and time to treatment failure. The data  
193 collection and analysis of all time-dependent endpoints is complex, particularly when the  
194 assessments are indirect and based on calculations and estimates as is the case for tumor  
195 measurements. The discussion of progression-free survival data collection and analysis is  
196 particularly complex and is supplemented by tables in Appendix 3 of this guidance.

197  
198 Selection of tumor-assessment endpoints for efficacy trials should include two judgments. First,  
199 will the endpoint support accelerated approval (is the endpoint a surrogate reasonably likely to  
200 predict clinical benefit and does the drug provide an advantage over available therapy) or regular  
201 approval (is it an established and/or validated surrogate for, or a direct measure of, clinical  
202 benefit)? Second, will the results be reliable, given the potential for uncertainty or bias in tumor  
203 endpoint assessments? ~~Drug applications using studies that rely on tumor measurement based~~  
204 ~~endpoints as sole evidence of efficacy should generally provide confirmatory evidence from a~~

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

205 ~~second trial. Both the precision and the clinical meaning of endpoints based on tumor~~  
206 ~~assessments can vary in different cancer settings. For instance, response rate determinations in~~  
207 ~~malignant mesothelioma and pancreatic cancer are often unreliable because of the difficulty in~~  
208 ~~measuring these tumors with currently available imaging modalities.~~

209  
210 When the primary study endpoint for drug approval is based on tumor measurements (e.g.,  
211 progression-free survival or ORR), the primary endpoint is assessed by the local center. In a  
212 blinded trial, central review is not necessary as there is no opportunity for bias~~it is recommended~~  
213 ~~that tumor endpoint assessments generally be verified by central reviewers blinded to study~~  
214 ~~treatment (see Appendix 4), especially when the study itself cannot be blinded.~~ Although the  
215 FDA will generally not ask that all tumor images be submitted with the marketing application, it  
216 may need to audit a sample of the scans to verify the central review process. In all cases, we  
217 recommend submitting primary electronic data documenting tumor measurements and  
218 assessments.<sup>7</sup> Additional details regarding data collection are listed in Appendix 1.

219  
220 *1. Disease-Free Survival*

221  
222 Disease-free survival (DFS) is usually defined as the time from randomization until recurrence of  
223 tumor or death from any cause. Although DFS can also be an important endpoint when a large  
224 percentage of patients achieve complete responses with chemotherapy, the most frequent use of  
225 this endpoint is in the adjuvant setting after definitive surgery or radiotherapy. In either of these  
226 settings, DFS has special meaning to patients because until a recurrence occurs, a patient can  
227 hope for cure. Whereas overall survival is the standard endpoint for most adjuvant settings, DFS  
228 has been the primary basis of approval for hormonal therapy after initial surgery for breast  
229 cancer. An important consideration is whether prolongation of DFS represents intrinsic benefit  
230 or only a potential surrogate for survival prolongation. In December 2003, the consensus of the  
231 ODAC was that prolongation of DFS represented clinical benefit, but that the magnitude of this  
232 benefit should be carefully weighed against the toxicity of adjuvant treatment, particularly as  
233 measured by effects on patient function. In May 2004, the ODAC recommended that DFS be  
234 considered an acceptable endpoint for colon cancer drugs in the surgical adjuvant setting,  
235 provided certain conditions were met.<sup>8</sup> Additional cancer-specific guidances will address the  
236 acceptability of DFS in other cancer settings.

237  
238 Important considerations in evaluating DFS as a potential endpoint include the estimated size of  
239 the treatment effect, proven benefits of standard therapies, and details of trial design. For  
240 instance, when a new drug is compared to a control drug that is known to improve overall  
241 survival, an important consideration is whether the DFS of the new drug is superior to, or only  
242 noninferior to, the control. Clearly, proof of superiority with regard to a surrogate endpoint is  
243 more persuasive than a demonstration of noninferiority. Furthermore, relying on a conclusion of  
244 noninferiority based on a surrogate endpoint to support a conclusion of noninferiority with  
245 respect to the definitive endpoint is problematic. Another critical issue is whether the duration of  
246 study follow-up is adequate to evaluate the durability of the DFS benefit.

---

<sup>7</sup> See guidance for industry *Cancer Drug and Biological Products — Clinical Data in Marketing Applications*  
(<http://www.fda.gov/cder/guidance/index.htm>)

<sup>8</sup> Transcripts are available at [http://www.fda.gov/cder/drug/cancer\\_endpoints/default.htm](http://www.fda.gov/cder/drug/cancer_endpoints/default.htm).

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

47  
248 We suggest that the protocol carefully detail both the definition of DFS and the schedule for  
249 follow-up studies and visits. Unscheduled assessments can occur for many reasons (including  
250 tumor-related symptoms, drug toxicity, anxiety), and differences between study arms in the  
251 frequency or reason for unscheduled assessments ~~is can be problematic, and may likely to~~  
252 introduce bias. This potential bias can be minimized by blinding patients and investigators to the  
253 treatment assignments if feasible. The potential effects of bias due to unscheduled assessments  
254 can be evaluated by an analysis of the total number of events over the follow-up period  
255 regardless of when the events occurred ~~comparing their frequency between treatment arms and by~~  
256 ~~performing statistical analyses that assign events from unscheduled visits to the time of the next~~  
257 ~~scheduled visit.~~

258  
259 Another issue in defining DFS is whether deaths occurring without prior documentation of tumor  
260 progression should be scored as DFS events (disease recurrences) or should be censored in the  
261 statistical analysis. All methods for statistical analysis of deaths have limitations. The approach  
262 that seems less prone to introducing bias is to consider all deaths as recurrences. Limitations of  
263 this approach are a potential decrease in statistical power of the study (by *diluting* the cancer-  
264 related events with deaths not related to cancer) and a potential to falsely prolong the DFS  
265 estimates in patients who die after a long unobserved period. The latter could introduce bias if  
266 the frequency of long-term follow-up visits is dissimilar on the study arms or if there is  
267 nonrandom dropout due to toxicity. Some analyses count cancer-related deaths as DFS events  
268 and censor noncancer deaths. This method has the potential for bias in the post hoc  
269 determination of the cause of death. Furthermore, any method that censors patients, whether at  
270 death or at the last visit, assumes that the censored patients have the same risk of recurrence as  
271 noncensored patients. This critical assumption needs close examination in any setting where  
272 deaths are to be censored. In settings where deaths due to causes other than cancer are common  
273 (e.g., studies of patients with early metastatic prostate cancer), censoring deaths can be  
274 appropriate.

275  
276 *2. Objective Response Rate*

277  
278 ORR is the proportion of patients with tumor shrinkage of a predefined amount lasting for a  
279 predefined minimum period of time. Response duration is usually measured from the time of  
280 initial response until documented tumor progression. The FDA has generally defined ORR as  
281 the sum of partial responses plus complete responses. When defined in this manner, ORR is a  
282 measure of drug antitumor activity even in a single-arm study. Some sponsors have proposed  
283 including stable disease as a component of ORR; however, evaluating drug effects based on the  
284 stable disease rate generally involves comparison to a randomized concurrent control. Also,  
285 stable disease incorporates components of time to progression or progression-free survival,  
286 which can be captured in a separate measurement. A variety of response criteria have been  
287 considered appropriate, including the RECIST criteria (Therasse and Arbuck et al., 2000, New  
288 Guidelines to Evaluate Response to Treatment in Solid Tumors, J Natl Cancer Inst, 92:205-16).  
289 Important issues for determining the clinical and regulatory significance of ORR include  
290 response duration, the percentage of complete responses, the toxicity of treatment, and associated  
291 improvement in tumor-related symptoms. These issues, in addition to an assessment of benefits  
292 of existing therapies, determine whether ORR will support marketing authorization, either for

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

293 regular approval (as a full surrogate for clinical benefit) or for accelerated approval (as a  
294 *reasonably likely surrogate*).

295  
296 It is important that criteria for response and progression be detailed in the protocol, and data  
297 should be carefully and completely collected at intervals specified in the protocol.

298  
299 3. *Time to Progression and Progression-Free Survival*

300  
301 In the past, time to progression (TTP) (the time from randomization until objective tumor  
302 progression) and progression-free survival (PFS) (the time from randomization until objective  
303 tumor progression or death) have seldom served as primary endpoints for drug approval. Time  
304 to symptomatic progression, which would represent a clear clinical benefit, is infrequently  
305 assessed but would be a credible endpoint of a well-conducted (generally blinded) trial. In  
306 December 2003, the ODAC discussed both potential roles of TTP and PFS in cancer drug  
307 approval and the committee's preference for PFS versus TTP.<sup>9</sup> The ODAC suggested relying on  
308 these endpoints in selected clinical situations, such as diseases with low complete response rates  
309 or when documentation of a survival benefit in clinical trials can be difficult. In settings where  
310 most patients are symptomatic, the ODAC preferred measuring tumor response and symptom  
311 benefit. ~~The definition of tumor progression varies widely; therefore, it is important that it be~~  
312 ~~carefully detailed in the protocol. The precise definition of tumor progression is important and~~  
313 should be carefully detailed in the protocol.

314  
315 a. TTP vs. PFS

316  
317 The ODAC consensus was that PFS is a better predictor of clinical benefit than TTP and thus  
318 preferable as a drug approval endpoint when used as a surrogate for clinical benefit (rather than  
319 just as an indicator of antitumor activity) because PFS includes deaths. Unanticipated effects of  
320 drugs on survival would thus be included in the endpoint. In the analysis of TTP, deaths are  
321 censored, either at the time of death or at an earlier visit. This approach is questionable because  
322 it can represent *informative censoring* (i.e., there may be a nonrandom pattern of loss from the  
323 study). It seems unlikely in most cancer settings that patient deaths are randomly related to  
324 tumor progression (e.g., it is likely that some deaths result from complications of undocumented  
325 cancer progression). Therefore, in most settings PFS is the preferred regulatory endpoint. In  
326 settings where most deaths are due to causes other than cancer, however, TTP can be an  
327 appropriate endpoint.

328  
329 b. PFS as an endpoint to support drug approval

330  
331 Some advantages and disadvantages of using PFS as an endpoint to support cancer drug approval  
332 are listed in Table 1. Conceptually, PFS has desirable qualities of a surrogate endpoint because it  
333 reflects tumor growth (a phenomenon likely to be on the causal pathway for cancer-associated  
334 morbidity and death), can be assessed prior to demonstration of a survival benefit, and is not  
335 subject to the potential confounding impact of subsequent therapy (unless worsening of a blood  
336 marker leads to a change in treatment prior to progression). Moreover, an effect on PFS occurs  
337 earlier than an effect on survival, so that a given advantage, say a median improvement of 3

---

<sup>9</sup> Transcripts are available at [http://www.fda.gov/cder/drug/cancer\\_endpoints/default.htm](http://www.fda.gov/cder/drug/cancer_endpoints/default.htm).

## *Contains Nonbinding Recommendations*

*Draft — Not for Implementation*

338 months, represents a larger (and thus more detectable) hazard ratio improvement than would a 3-  
339 month median survival benefit occurring later. The formal validation of PFS as a surrogate for  
340 survival for the many different malignancies that exist, however, would be difficult. Data are  
341 usually insufficient to allow a robust evaluation of the correlation between effects on survival  
342 and PFS. Oncology trials are often small, and proven survival benefits of existing drugs are  
343 generally modest. The role of PFS as an endpoint to support licensing approval varies in  
344 different cancer settings. In some settings PFS prolongation might be an accepted surrogate  
345 endpoint for clinical benefit to support regular approval, and in others it may be a surrogate  
346 reasonably likely to predict benefit for accelerated approval. Important considerations will be  
347 the magnitude of the effect, the toxicity profile of the treatment, and the clinical benefits and  
348 toxicities of available therapies. These issues will be discussed in future guidance documents for  
349 specific cancer settings.

### c. PFS trial design issues

352  
353 It is important that methodology for assessing, measuring, and analyzing PFS be detailed in the  
354 protocol and statistical analysis plan. It is also important to carefully define tumor progression  
355 criteria in the protocol. There are no standard regulatory criteria for defining progression.  
356 Sponsors have used a variety of different criteria, including the RECIST criteria. The broad  
357 outline presented in most published PFS criteria should be supplemented with additional details  
358 in the protocol and statistical analysis plan. It is important that visits and radiological  
359 assessments be symmetric on the two study arms to prevent systematic bias. When possible,  
360 studies should be blinded. Blinding is particularly important when patient or investigator  
361 assessments are included as components of the progression endpoint. It is important that the  
362 FDA and the sponsor agree prospectively on the protocol, data to be recorded on the case report  
363 form, statistical analysis plan (including analysis of missing data and censoring methods), and, if  
364 applicable, the operating procedures of an independent endpoint review committee (discussed in  
365 Appendix 4). The effect of follow-up visit frequency has been debated. Frequent regular  
366 assessments, depending on the type and stage of cancer, ensure that most progression events will  
367 be detected on radiologic scans rather than as symptomatic events. This approach increases the  
368 expense and difficulty of the study, including an increased data collection burden on the  
369 investigator and an increased number of scans for patients, and may not mirror clinical practice  
370 standards.

### d. Analysis of PFS

371  
372  
373  
374 The analysis of PFS is complicated by missing data. It is important that the protocol specify  
375 what constitutes an adequate assessment visit for each patient (i.e., a visit when all scheduled  
376 tumor assessments have been done). The analysis plan should outline a comparison of the  
377 adequacy of follow-up in each treatment arm and specify how incomplete or missing follow-up  
378 visits will be handled with regard to censoring. For instance, if one or more assessment visits are  
379 missed just prior to the progression event, to what date should the progression event be assigned?  
380 It is important that the analysis plan specify the primary analysis and one or more sensitivity  
381 analyses. For instance, in the previous example, the primary analysis might assign the actual  
382 date of observed progression as the progression date. A sensitivity analysis that ignores the  
383 imputed timing of an event can be conducted. The sensitivity analysis might censor the data at  
384 the last adequate assessment visit. Although both analyses are problematic (the best solution to

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

385 missing data is to have none), the conclusion is probably valid if it is supported by the results of  
386 both the primary and the sensitivity analyses. Other methods could be considered if adequately  
387 supported by the sponsor. The analysis plan should evaluate the number of deaths in patients  
388 who have been lost to follow-up for more than a substantial (prespecified) time. An imbalance  
389 in such deaths could bias the measurement of PFS, artificially prolonging PFS on the arm with  
390 less adequate follow-up.

391  
392 Because progression data can be collected from a variety of sources (including physical exams at  
393 unscheduled visits and radiologic scans of various types) and at a variety of times, it is important  
394 that data collection efforts for each assessment visit be limited to a specified short time interval  
395 prior to the visit. When data are collected over a longer time, the question then arises: What  
396 date should serve as the progression date or the censoring date? A common method is to assign  
397 progression to the earliest observed time when an observation shows progression and to censor at  
398 the date when the last radiologic assessment determined a lack of progression. Because this  
399 method could introduce an assessment bias, especially in unblinded trials, we recommend  
400 assigning the progression and censoring times to the time of the scheduled assessment visits. A  
401 study of time to symptomatic progression, if conducted blindly and with few scheduled  
402 assessments, in contrast, could use the actual time of observed symptom progression. The PFS  
403 date based on a death, however, would be the date of death rather than the assigned visit date  
404 since death ascertainment is not related to visit time and not subject to interpretation.

405  
406 Appendix 3 provides a set of tables for potential analyses of PFS that could be used for primary  
407 or sensitivity analyses. We recommend that plans for PFS data collection and analysis be  
408 discussed with the FDA at end-of-phase 2 meetings and verified in special protocol assessments.

409  
410 e. Future methods for assessing progression

411  
412 ~~In the future, it is important that other methods of progression assessment be evaluated as~~  
413 ~~potential surrogate endpoints for regular approval or accelerated approval. One proposed~~  
414 ~~method (not used to date) is the single time point assessment that could decrease the complexity~~  
415 ~~of progression assessment and eliminate time dependent assessment bias. In the single time~~  
416 ~~point analysis, progression would be assessed at baseline and at one prespecified time after~~  
417 ~~randomization. If patients progress prior to the specified time, radiologic scans could document~~  
418 ~~progression and the patient could go off study. All other patients would have a detailed~~  
419 ~~radiologic evaluation at the prespecified follow up time. The statistical analysis could compare~~  
420 ~~the proportions of patients on each study arm with progression on or before the prespecified time~~  
421 ~~after randomization. Potential problems with this approach are decreased statistical power,~~  
422 ~~potential for missing a small benefit at a time different from the prespecified time, and lack of~~  
423 ~~information regarding the relationship between the single time point analysis and the familiar~~  
424 ~~endpoints of progression free survival and overall survival. Although this approach could~~  
425 ~~provide some advantages and decrease assessment bias, study dropouts prior to progression~~  
426 ~~could present the same difficulty as they do for all progression endpoints. Settings in which~~  
427 ~~further evaluation of this approach seems warranted are those where a significant and durable~~  
428 ~~effect on progression free survival is expected and where complete progression free survival~~  
429 ~~data collection seems impossible or impractical. In the future, it is important that other methods~~  
30 of progression assessment be evaluated as potential surrogate endpoints for regular approval or

**Contains Nonbinding Recommendations**  
*Draft — Not for Implementation*

431 accelerated approval. One proposed method (not used to date) is the ‘single time point’ or, more  
432 accurately, an ‘event count’ analysis which could decrease the complexity of progression  
433 assessment and eliminate time-dependent assessment bias. In this analysis, progression would be  
434 assessed at a minimum of baseline and at one pre-specified time in follow-up after the last  
435 patient had been randomized; typically this would be at the end of the minimum follow-up time  
436 specified in the sample size calculation to achieve a desired number of PFS events. The protocol  
437 would stipulate that, if a patient progressed prior to the specified time, radiologic scans would be  
438 required to document progression. Patients passing through the study without evidence of  
439 progression would be required to have a detailed radiologic evaluation at the pre-specified  
440 follow-up time. The statistical analysis would compare the number of patients on each study arm  
441 with progression on or before the pre-specified time after randomization. In this way the  
442 problems associated with the imputation of progression times are avoided entirely. While there  
443 is some loss of statistical power, this loss has been shown to be minimal if the proportion of  
444 patients with a progression event by the pre-specified follow-up time is not much higher than  
445 75-80%.

446  
447 Although this approach could provide some advantages and decrease assessment bias, study  
448 dropouts prior to progression could present the same difficulty as they do for all progression  
449 endpoints. Further theoretical evaluation of this approach is needed. From a more practical  
450 standpoint, application of this approach to previously reported trials with PFS as an endpoint  
451 would help establish its usefulness and highlight the potential for discrepancy between the  
452 approach and the regular analysis of PFS time.

453  
454 **4. Time to Treatment Failure**

455  
456 Time to treatment failure (TTF) is a composite endpoint measuring time from randomization to  
457 discontinuation of treatment for any reason (including progression of disease, treatment toxicity,  
458 and death). Defined that way, TTF is not recommended as an endpoint for drug approval  
459 because it combines efficacy and toxicity measures. For example, suppose the standard  
460 comparator (Drug A) provides a known survival benefit, but only at the cost of considerable  
461 toxicity with many patients leaving therapy because of that toxicity. A nontoxic investigational  
462 drug (Drug B) could have a significantly longer TTF than Drug A solely because it caused fewer  
463 toxic dropouts. These data alone could not support drug approval because they would not  
464 demonstrate that Drug B is effective. Drug approval would require a demonstration of Drug B  
465 efficacy, such as a survival improvement or other clinical benefit.

466  
467 **C. Endpoints Involving Symptom Assessment**

468  
469 Symptomatic improvement has always been considered a clinical benefit, and many FDA cancer  
470 drug approvals have used patient symptom assessments and/or physical signs thought to  
471 represent symptomatic improvement (e.g., weight gain, decreased effusion) as the primary  
472 evidence of effectiveness. To date, broader measures of health-related quality of life (HRQL  
473 instruments) have not served this role. HRQL is discussed in a separate FDA draft guidance on  
474 patient-reported outcomes (PRO).<sup>10</sup> The FDA has relied on symptom scores, signs, and

---

<sup>10</sup> The draft guidance for industry *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Claims* is currently being developed and is expected to publish in the summer of 2005. When final, this

## *Contains Nonbinding Recommendations*

*Draft — Not for Implementation*

475 symptoms representing obvious benefit (e.g., decreased esophageal obstruction, fewer bone  
476 fractures, reduced size and number of skin lesions, physician actions [need for radiation therapy  
477 in response to painful bone metastases], physician assessments of performance status, and  
478 patient-reported assessments of symptom scales). Relying on such evidence of clinical benefit as  
479 the basis for approval has allowed the FDA to approve cancer drugs earlier than if demonstration  
480 of a survival benefit had been required. It seems self-evident that cancer patients will be in most  
481 cases the best source for determining effects on patient symptoms, so that PRO instruments seem  
482 most appropriate. Formal PRO instruments can be designed that focus on specific symptoms  
483 (e.g., a pain scale) or on a broader array of physical, emotional, and activity measures.

484  
485 The use of improvement of signs and symptoms or QOL assessments as primary endpoints to  
486 support cancer drug approval requires discrimination between tumor symptoms and drug  
487 toxicity, especially when evidence is based on comparison to a toxic active control. This poses  
488 particular problems for general HRQL scales, which, by definition, are multidimensional scales  
489 including elements other than physical problems. An apparent effectiveness advantage of one  
490 drug over another measured on a global HRQL instrument might simply indicate less toxicity of  
491 one product or regimen versus the other, a matter of interest but not an effectiveness measure.  
492 Morbidity endpoints used to date for cancer drug approvals have possessed *face validity* (value  
493 obvious to patients and physicians, for example, an endpoint based on functional measures such  
494 as the ability to swallow solids, liquids, or nothing) and have not measured benefit and toxicity  
495 on the same scale.

### 496 497 1. *Specific Symptom Endpoints*

498  
499 One endpoint the FDA has suggested to sponsors is *time to progression of cancer symptoms*, an  
500 endpoint similar to time to progression. This endpoint would be a direct measure of clinical  
501 benefit rather than a potential surrogate. Sponsors have cited several problems with this  
502 approach. First, because few cancer trials are blinded, assessments can be biased and therefore  
503 unreliable. Another problem is the usual delay between tumor progression and the onset of  
504 cancer symptoms. Often alternative treatments are begun before reaching the symptom endpoint,  
505 which can confound the results. Many cancer trials are performed in patients with little prior  
506 exposure to chemotherapy and who usually have minimal cancer symptoms. Finally, it can  
507 sometimes be difficult to differentiate tumor symptoms from drug toxicity, a problem noted in  
508 discussions of time to treatment failure and HRQL. *Time to progression of symptoms* and *time to*  
509 *onset of symptoms* can be reasonable endpoints in cancer settings where treatment can be  
510 blinded, most progressing patients are symptomatic, no effective therapy exists, and less frequent  
511 radiologic follow-up is appropriate. Symptom data should be carefully collected using a  
512 validated instrument according to a schedule detailed in the protocol.

513  
514 A *composite symptom endpoint* can be appropriate when the benefit of a drug is multifaceted. It  
515 is important that the components of the endpoint be related and generally of similar clinical  
516 importance. Drugs have been approved for treatment of patients with cancer metastases to the  
517 skeleton based on a composite benefit endpoint consisting of one or more skeletal-related event

---

guidance will represent the FDA's current thinking on this topic. For the most recent version of a CDER or CBER guidance, check the CDER guidance Web page at <http://www.fda.gov/cder/guidance/index.htm> and the CBER Web page at <http://www.fda.gov/cber/guidance/index.htm>.

## *Contains Nonbinding Recommendations*

*Draft — Not for Implementation*

518 (SRE) that would be anticipated to be associated with pain and other distress. SREs are defined  
519 as pathologic fractures, radiation therapy to bone, surgery to bone, and spinal cord compression.  
520 Clinical Benefit Response, a composite endpoint of pain and analgesic consumption reported by  
521 the patient, and performance status assessed by a physician, in part supported approval of a drug  
522 to treat pancreatic cancer.

523  
524 Selection of the appropriate population for study can be critical for documenting symptom  
525 benefit. Patients symptomatic at study baseline can be evaluated with a categorical symptom  
526 response analysis. This approach can be appropriate for diseases such as lung cancer, when most  
527 patients have symptoms at diagnosis. Studies of asymptomatic patients could use a time-to-first-  
528 symptom analysis. Even if the patient discontinues the study drug or begins a new drug,  
529 symptomatic progression could still be assessed if follow-up is continued until documentation of  
530 the first symptom. This approach is worth considering but has been infrequently attempted.

### 531 532 2. *Problems Encountered with Symptom Data*

533  
534 Many problems have been encountered in the analysis of symptom data submitted to the FDA.  
535 The most important problem in oncology is that few trials are blinded so that the possibility of  
536 observer bias is difficult to exclude. Missing data are common and often cast doubt on study  
537 conclusions. It is critically important to have frequent assessments to minimize long unobserved  
538 gaps. In addition, symptom severity should be addressed, rather than providing only a binary  
539 present or absent. Withdrawing treatment because of drug toxicity or tumor progression is one  
540 cause of missing symptom data. Ideally, when patients stop treatment, data collection forms  
541 should continue to gather information to inform the analysis. Symptom data could lead to a large  
542 number of different endpoints, and prospectively defined statistical plans need to correct for  
543 multiplicity if each symptom is treated as a separate endpoint.

### 544 545 **D. Biomarkers**

546  
547 To date, evidence from biomarkers assayed from blood or body fluids has not served as primary  
548 endpoints for cancer drug approval, although paraprotein levels measured in blood and urine  
549 have contributed to response endpoints for myeloma. Further research is needed to establish the  
550 validity of the available tests and determine whether improvements in such biomarkers are  
551 reasonably likely to predict clinical benefit (accelerated approval) or are established surrogates  
552 for clinical benefit (regular approval).

553  
554 Although tumor markers are not yet used alone as a basis for marketing approval, the FDA has  
555 sometimes accepted their inclusion as elements in composite endpoints. For instance, women  
556 with ovarian cancer often show clinical deterioration from progression of unmeasured tumor. In  
557 blinded randomized controlled trials in advanced refractory ovarian cancer, the FDA has  
558 accepted use of a composite endpoint that included CA-125. The occurrence of certain clinical  
559 events (a significant decrease in performance status, or bowel obstruction) coupled with marked  
560 increases in CA-125 was considered progression in these patients. The use of prostate specific  
561 antigen (PSA) was discussed at a recent workshop on prostate cancer endpoints. Different  
562 methods of evaluating PSA as an endpoint were discussed, including PSA response, PSA slope,  
563 and PSA velocity. Although the FDA has not yet accepted a PSA endpoint to support drug

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

564 approval, evaluation of additional data and further discussions of PSA endpoints are planned in  
565 future workshops and ODAC meetings.<sup>11</sup>  
566

567  
568 **IV. ENDPOINTS AND CLINICAL TRIAL DESIGN; SELECTED ISSUES**  
569

570 By law, the FDA must base new drug approval decisions on substantial evidence of efficacy  
571 from “adequate and well-controlled investigations.” Regulations describe the meaning of  
572 “adequate and well-controlled investigations.” Studies must allow a valid comparison to a  
573 control and must provide a quantitative assessment of the drug’s effect. (See 21 CFR 314.126.)  
574 Below we discuss several issues related to the design of cancer trials intended to support drug  
575 approval.  
576

577 **A. Single-Arm Studies**  
578

579 The most reliable method for demonstrating efficacy is to show a statistically significant  
580 improvement in a clinically meaningful endpoint in blinded randomized controlled trials. Other  
581 approaches have also been successful in certain settings. In settings where there is no effective  
582 therapy and where major tumor regressions can be presumed to occur infrequently in the absence  
583 of treatment (a historical control), the FDA has sometimes accepted ORR and response duration  
584 observed in single-arm studies as substantial evidence supporting accelerated approval or even  
585 regular approval (e.g., when many complete responses were observed or when toxicity was  
586 minimal or modest). In contrast to the success of this approach, evidence from historically  
587 controlled trials attempting to show improvement in time-to-event endpoints such as survival,  
588 time to progression, or progression-free survival have seldom been persuasive support for drug  
589 approval, except when treatment provides survival outcomes that contrast markedly with  
590 historical experience (e.g., testicular cancer, acute leukemias). In most cases, however, these  
591 outcomes vary among study populations in ways that cannot always be predicted; for example,  
592 changes in concomitant supportive care or frequency and method of tumor assessment can differ  
593 by location or change over time. Consequently, comparisons involving these time-to-event  
594 endpoints generally need a concurrent control (preferably in a randomized trial), unless, as noted,  
595 the effect is very large.  
596

597  
598 **B. Studies Designed to Demonstrate Noninferiority**  
599

600 ~~The goal of noninferiority (NI) trials is to demonstrate the effectiveness of a new drug showing~~  
601 ~~that it is not less effective, by a predefined amount, than a standard regimen known to have the~~  
602 ~~effect being investigated (Temple and Ellenberg, 2000, Placebo Controlled Trials and Active-~~  
603 ~~Control Trials in the Evaluation of New Treatments, Part 1: Ethical and Scientific Issues, Ann~~  
604 ~~Intern Med, 2000 Sep 19; 133(6):455-63).<sup>12</sup> ~~The difference to be ruled out, the noninferiority~~  
605 ~~margin, cannot be larger than the effect of the control drug in the new study. As that effect is not~~~~

---

<sup>11</sup> Transcripts are available at [http://www.fda.gov/cder/drug/cancer\\_endpoints/default.htm](http://www.fda.gov/cder/drug/cancer_endpoints/default.htm).

<sup>12</sup> See ICH guidance for industry *E10 Choice of Control Group and Related Issues in Clinical Trials* (<http://www.fda.gov/cder/guidance/index.htm>)

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

606 measured (the new study does not have a no-treatment arm), the effect must be assumed based  
607 on the previous studies of the control drug that documented its effect. If the new drug is inferior  
608 by more than the noninferiority margin, it would have no effect at all. In most cases the NI  
609 margin is not set at the control drug's full effect, but at some fraction of it (e.g., 50 percent), so  
610 that the study seeks to show that at least 50 percent of the control drug effect is preserved.

611  
612 There are multiple difficulties with NI trials. NI trials rely on historical data to establish the  
613 expected size of treatment effect of the active control. In many situations adequate historical  
614 data for the control do not exist. Moreover, a critical assumption is that the treatment effect of  
615 the active control that was observed historically will also be observed in the current population in  
616 the new study. This assumption is difficult to support, as results of trials are almost never  
617 identical (although one can evaluate control regimen response rates in the historical and NI trial  
618 populations as some measure of comparability). Optimally, the estimated size of the treatment  
619 effect of the active control would be based on a comprehensive meta-analysis of historical  
620 studies that reproducibly demonstrate the effectiveness, compared to no treatment, of the control  
621 agent. In the oncology setting, however, information is often lacking on effects compared to a  
622 no-treatment control. The variability in the meta-analysis will be reflected in the choice of the  
623 noninferiority margin. But there may be little data from randomized controlled trials available to  
624 estimate the treatment effect and thus no basis for estimating the control treatment effect.  
625 Furthermore, subsequent events in the trial, especially crossover from the control, can invalidate  
626 NI survival analyses (producing a bias toward a showing of no difference). NI designs generally  
627 require many patients in order to provide meaningful results. Given the complex issues  
628 involved, we strongly recommend that sponsors designing noninferiority trials consult early with  
629 the FDA. Because of the difficulties with the design, conduct, and analysis of NI trials, a single  
630 NI trial seldom provides sufficient evidence of efficacy to support drug approval.

631  
632 When the new treatment has a different toxicity profile from available treatments, it may be  
633 possible to design around the NI study problem by conducting an add-on study, adding new drug  
634 or placebo/no-treatment to the standard therapy. This will not be possible if the goal is to show a  
635 new treatment to be less toxic than existing therapy (but still effective). In this case the NI  
636 design is unavoidable in order to demonstrate that the survival benefit of the standard drug is  
637 retained by the experimental drug. If the standard drug is associated with only a small proven  
638 survival benefit, however, interpretation of an NI study is difficult or impossible. Moreover, the  
639 size of such NI trials can be prohibitively large. A randomized trial comparing a new drug to  
640 placebo is the most direct and effective way of establishing efficacy and safety of the new drug.  
641 However, in oncology, placebo controlled trials are often impossible due to the availability of  
642 either approved agents or the use of unapproved, but nevertheless commonly accepted agents. In  
643 such circumstances, active controlled, non-inferiority (NI) trials are necessary. The goal of such  
644 trials is to demonstrate, indirectly, the absolute effectiveness of a new drug by showing that it  
645 would most likely have beaten placebo if placebo controlled trial could have been conducted. A  
646 secondary objective of these trials is to examine how well the new drug compares in terms of  
647 efficacy and safety, to the active control (ref Wang, Fisher, Carroll). This latter objective is  
648 commonly achieved by defining in advance a difference between the new drug and the active  
649 control that is to be ruled out statistically. This difference is referred to as the NI margin, and is  
650 determined from historical studies of the active control that documented its effect. If the new  
651 drug is inferior by more than the non-inferiority margin, then non-inferiority to the degree

***Contains Nonbinding Recommendations***  
***Draft — Not for Implementation***

652 captured by the margin cannot be established. Previously in oncology drug applications (e.g.,  
653 Xeloda vs. 5-FU, Cisplatin + taxotere vs. cisplatin + vinorelbine), the NI margin has been  
654 arbitrarily set to be 50% of the active control effect, so that, in these examples, NI was defined as  
655 showing at least 50 percent of the active control effect is preserved.

656 There are several challenges in the design of active-control, NI trials. NI trials necessarily rely  
657 on historical data to establish the expected size of treatment effect of the active control. In some  
658 situations, the effect of the control drug may not have been established with narrow confidence  
659 limits. However, methods do exist that compensate for the level of precision in the active control  
660 effect albeit at the expense of the size of the new active control trial, which may need to be  
661 extremely large (ref Rothmann). Also, a critical assumption is that the treatment effect of the  
662 active control that was observed historically will also be observed in the current population in the  
663 new study. This assumption is often difficult to demonstrate unequivocally. Informal  
664 comparison of response and death rates on the control arm, of the new active control NI trial with  
665 the response and death rates based on historical data may provide some reassurance that this  
666 assumption has, or has not, been met. However, it is important to recognize that the performance  
667 of the active control is just as much an issue for superiority trials as NI trials; superiority of a new  
668 drug to an active control that has grossly under performed can pose difficulties in interpreting  
669 whether the new drug has had an true effect, or at least a clinically relevant one. A further  
670 problem in NI trials is crossover from the new drug to the control drug, which can bias overall  
671 survival toward a showing of no difference. Given the complex issues involved, we strongly  
672 recommend that sponsors designing non-inferiority trials consult early with the FDA.

673  
674 **C. No Treatment or Placebo Control**

675  
676 Giving no anticancer drug treatment to patients in the control arm of a cancer study is often  
677 considered unethical, but, in some settings, it can be acceptable. For instance, in early stage  
678 cancer when standard practice is to give no treatment, comparison of a new agent to a no-  
679 treatment control would be acceptable. This approach would not be an ethical problem in the so-  
680 called *add-on* design, when all patients receive standard treatment plus either no additional  
681 treatment or the experimental drug. Using a control group that receives only best supportive care  
682 is acceptable in an advanced refractory setting where there is no effective therapy. Placebos  
683 (identical appearing inactive controls) are generally preferred to no-treatment controls because  
684 they permit blinding. With many cytotoxic cancer drugs, blinding may not be feasible because  
685 of a relatively high rate of recognizable toxicities, but newer interventions, many of them much  
686 less toxic, are increasingly being studied in blinded trials.

687  
688 **D. Isolating Drug Effect in Combinations**

689  
690 Because marketing approval is usually for a single drug product rather than for a drug  
691 combination, clinical trials supporting regulatory approval need to isolate the effectiveness of the  
692 proposed agent. Evidence is needed showing not only the effectiveness of the regimen but also  
693 establishing the contribution of the new drug to that regimen. One way to demonstrate the  
694 individual contribution of a new drug in a regimen is using the *add-on* design previously  
695 discussed. Sometimes the clinical effects seen in early phases of development can be used to  
696 establish the contribution of a drug to a drug regimen, particularly if the combination is more

***Contains Nonbinding Recommendations***  
***Draft — Not for Implementation***

697 effective than any of the individual components. We recommend discussing these issues with  
698 the FDA at end-of-phase 1 or end-of-phase 2 meetings.  
699

700 **E. Trial Designs for Radiotherapy Protectants and Chemotherapy Protectants**  
701

702 Radiotherapy protectants and chemotherapy protectants are drugs designed to ameliorate the  
703 toxicities of radiotherapy or chemotherapy. Trials to evaluate these agents usually have two  
704 objectives. The first is to assess whether the protecting drug achieves its intended purpose of  
705 ameliorating the cancer treatment toxicity. Unless the mechanism of protection is clearly  
706 unrelated to the mechanism of antitumor activity (e.g., antiemetic agents which ameliorate  
707 nausea via central nervous system receptors), a second trial objective is to determine whether  
708 anticancer efficacy is compromised by the protectant. Because the comparison of antitumor  
709 activity between the two arms of the trial is a noninferiority comparison, a large number of  
710 patients may be required to achieve this objective. Generally, a second study is needed to  
711 confirm the findings. A critical question for the future is whether, in such cases where the same  
712 drug is studied in both arms, ORR should be considered a sufficient endpoint for comparing drug  
713 activity and benefit.  
714

715  
716 **V. SUMMARY AND CONCLUSION**  
717

718 Although general principles outlined in this guidance should help sponsors select endpoints for  
719 marketing applications, we recommend that sponsors meet with the FDA before submitting  
720 protocols intended to support NDA or BLA marketing applications. The FDA will ensure that  
721 these meetings include a multidisciplinary FDA team of oncologists, statisticians, clinical  
722 pharmacologists, and often external expert consultants. Sponsors may submit protocols after  
723 these meetings and request a special protocol assessment that provides the acceptability of  
724 endpoints and protocol design to support drug marketing applications.<sup>13</sup> Ultimately, of course,  
725 marketing approval will depend not only on the design of a single trial, but on FDA review of the  
726 results and data from all studies in the drug marketing application.  
727

---

<sup>13</sup> See guidance for industry *Special Protocol Assessment* (<http://www.fda.gov/cder/guidance/index.htm>)

**APPENDIX 1:**  
**THE COLLECTION OF TUMOR MEASUREMENT DATA<sup>14</sup>**

The following are important considerations for tumor measurement data. The Agency recommends that:

- The case report form (CRF) and electronic data document the target lesions identified during the baseline visit prior to treatment. Retrospective identification of such lesions would rarely be considered reliable.
- Tumor lesions are assigned a unique identifying letter or number. This allows differentiating among multiple tumors occurring at one anatomic site and matching of tumors measured at baseline and tumors measured during follow-up.
- A mechanism ensures complete collection of data at critical times during follow-up. It is important that the CRF ensures that all target lesions are assessed at each follow-up visit and that all required follow-up tests are done with the same imaging/measuring method.
- The CRF contains data fields that indicate whether scans were performed at each visit.
- A zero is recorded when a lesion has completely resolved. Otherwise, disappearance of a lesion cannot be differentiated from a missing value.
- Follow-up tests allow timely detection of new lesions both at initial and new sites of disease. It is important that the occurrence of and location of new lesions be recorded in the CRF and the submitted electronic data.

---

<sup>14</sup> *Tumor data* in this section refers to data in SAS transport files, not images. Images are not generally submitted to the NDA/BLA, but may be audited by the FDA during the review process.

**APPENDIX 2:  
ISSUES TO CONSIDER IN PFS ANALYSIS**

751  
752  
753  
754 The protocol and statistical analysis plan (SAP) of a study should detail the primary analysis of  
755 progression-free survival (PFS). This includes a detailed description of the endpoint, acceptable  
756 modalities for evaluating tumors, and procedures for minimizing bias when determining  
757 progression status, such as procedures for an independent endpoints review committee. It is  
758 important that one or two secondary analyses be specified to evaluate anticipated problems in  
759 trial conduct and to assess whether results are robust. The following are several important  
760 factors to consider.

761  
762 • **Definition of progression date.** Survival analyses use the exact date of death. In analyses  
763 of PFS, however, the exact progression date is unknown. The following are two methods for  
764 defining the *recorded progression date (PDate)* used for PFS analysis.

765  
766 1. One approach assigns PDate to the first time at which progression can be declared:

- 767 • For progression based on a new lesion, the PDate is the date of the first observation  
768 that detects the new lesion.
- 769 • For progression based on the sum of target lesion measurements, PDate is the date of  
770 the last observation or radiologic assessment of target lesions (if multiple assessments  
771 are done at different times).

772 This approach can introduce between-arm bias if radiologic assessments are done earlier  
773 or more frequently in one treatment arm.

774  
775 2. A second approach assigns the PDate to the date of the scheduled clinic visit immediately  
776 after all radiologic assessments (which collectively document progression) have been  
777 done. Although this approach provides a less accurate estimate of the true date of  
778 progression, the error should be symmetrically distributed between arms, and between-  
779 arm bias is minimized.

780  
781 • **Definition of censoring date.** Censoring dates are defined in patients with no documented  
782 progression prior to data cutoff or dropout. In these patients, the censoring date is often  
783 defined as the last date on which progression status was adequately assessed. One acceptable  
784 approach uses the date of the last assessment performed. However, multiple radiologic tests  
785 can be evaluated in the determination of progression. A second acceptable approach uses the  
786 date of the clinic visit corresponding to these radiologic assessments.

787  
788 • **Definition of an adequate PFS evaluation.** In patients with no evidence of progression,  
789 censoring for PFS often relies on the date of the last *adequate tumor assessment*. A careful  
790 definition of what constitutes an adequate tumor assessment includes adequacy of target  
791 lesion assessments and adequacy of radiologic tests both to evaluate nontarget lesions and to  
792 search for new lesions.

793  
794 • **Analysis of partially missing tumor data.** Analysis plans should describe the method for  
795 calculating progression status when data are partially missing from *adequate tumor*

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

assessment visits. For instance, are the values for missing target lesions to be carried forward?

- 799 • **Completely missing tumor data.** Assessment visits where no data are collected are  
800 sometimes followed by death or by assessment visits showing progression; in other cases the  
801 subsequent assessment shows no progression. In the latter case, at first glance, it might seem  
802 acceptable to continue the patient on study and continue monitoring for evidence of  
803 progression. This approach, however, treats missing data differently depending upon  
804 subsequent events and could represent informative censoring. Therefore, another possibility  
805 is for the primary analysis to include data from subsequent PFS assessments when only a  
806 single follow up visit is missed but censor data when there are two or more missed visits.  
807 While missed visits for progression assessment are problematic, all efforts should be made to  
808 keep following patients for disease progression irrespective of the number of visits missed.  
809 In order to avoid over estimating the true progression time, consideration should be given in  
810 the protocol to simple algorithms for handling a series of missing visits. For example,  
811 patients dying without progression, say, 3 months after their last assessment for progression  
812 status, might be censored at the time of their last assessment plus 3 months, whereas patients  
813 dying without progression within 3 months after their last assessment for progression status  
814 would be included with their date of death as the time of progression. It is important that the  
815 SAP detail primary and secondary PFS analyses to evaluate the potential effect of missing  
816 data. Reasons for dropouts should be incorporated into procedures for determining censoring  
817 and progression status. For instance, for the primary analysis, patients going off study for  
818 undocumented clinical progression, change of cancer treatment, or decreasing performance  
819 status could be censored at the last adequate tumor assessment. The secondary sensitivity  
820 analysis would include these dropouts as progression events. Patients lost to follow-up  
821 should be handled in the same way as patients with missing visits. Patients without  
822 progression who stop randomized therapy for any reason, for example due to undocumented  
823 clinical progression, change of cancer treatment, decreasing performance status or  
824 unacceptable toxicity should continue to be followed in so far as is possible for disease  
825 progression. Due to the informative nature of events that lead to the cessation of randomized  
826 therapy, analyses that censor patients who stop treatment without progression at the last  
827 adequate tumor assessment can be biased and misleading and hence can only be considered  
828 exploratory in nature.
- 830 • **Progression of nonmeasurable disease.** When appropriate, progression criteria should be  
831 described for each assessment modality (e.g., CT scan, bone scan). It is important that scans  
832 documenting progression based on nonmeasurable disease be verified by a blinded review  
833 committee and be available for verification by the FDA if needed.
- 835 • **Suspicious lesions.** Sometimes new lesions are identified as suspicious. An algorithm  
836 should be provided for following up these lesions and for assignment of progression status at  
837 the time of analysis. For example, a radiological finding identified as suspicious at visit one  
838 might be verified as being a new tumor at visit three. It is important that the protocol or  
839 analytical plan clarify whether the progression time would be visit one or visit three.

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

**APPENDIX 3:  
EXAMPLE TABLES FOR PFS ANALYSIS**

As discussed in Section III.B., sensitivity analyses may be helpful in determining whether the PFS analysis is robust. Different sensitivity analyses can be described in tables that specify how to assign dates of progression events and dates for censoring of progression data. 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 radiologic assessments verified by an independent review committee (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. Deaths occurring after two or more missed visits, however, are censored at last visit.

**Table A. PFS 1 (includes documented progression only)**

<b>Situation</b>	<b>Date of Progression or Censoring</b>	<b>Outcome</b>
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Earliest of: <ul style="list-style-type: none"> <li>• Date of radiologic assessment showing new lesion (if progression is based on new lesion); or</li> <li>• Date of last radiologic assessment of measured lesions (if progression is based on increase in sum of measured lesions)</li> </ul>	Progressed
No progression	Date of last radiologic assessment of measured lesions	Censored
Treatment discontinuation for undocumented progression	Date of last scan of measured lesions	Censored
Treatment discontinuation for toxicity or other reason	Date of last radiologic assessment of measured lesions	Censored
New anticancer treatment started	Date of last radiologic 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 radiologic assessment of measured lesions	Censored

In line with the intent-to-treat principle underpinning a valid and meaningful analysis of survival, all patients should be followed for disease progression irrespective of any interruption to their

***Contains Nonbinding Recommendations***  
***Draft — Not for Implementation***

864 randomized therapy. Hence, patients who stop randomized therapy for any reason without  
865 progression (i.e., due to undocumented clinical progression, change of cancer treatment,  
866 decreasing performance status or unacceptable toxicity) should continue to be followed in so far  
867 as is possible for disease progression. Patients who experience a progression event would be  
868 included as such in the analysis and those who continue without progression would be censored  
869 at their last adequate visit for progression assessment.

870 The primary analysis, as defined in Table A and incorporating an intent-to-follow patients for  
871 progression irrespective of any interruption to their randomized therapy, will therefore compare  
872 treatment policies in exactly the same fashion as is standard and common place for overall  
873 survival.

874 Due to the informative nature of events that lead to the cessation of randomized therapy, it is  
875 important to recognize that analyses that censor patients who stop treatment without progression  
876 at the last adequate tumor assessment can be biased and misleading and hence can only be  
877 considered exploratory in nature.

**Contains Nonbinding Recommendations**  
*Draft — Not for Implementation*

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.

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

<b>Situation</b>	<b>Date of Progression or Censoring</b>	<b>Outcome</b>
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.

**Table C. PFS 3 (includes investigator claims)**

<b>Situation</b>	<b>Date of Progression or Censoring</b>	<b>Outcome</b>
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

*Contains Nonbinding Recommendations*  
*Draft — Not for Implementation*

**APPENDIX 4:**  
**INDEPENDENT REVIEW OF TUMOR ENDPOINTS**

890  
891  
892  
893 Sponsors and the FDA need to be able to verify clinical trial results that support drug approval,  
894 including ORR and progression-free survival. ORR determined in single-arm studies can be  
895 verified by scrutiny of a limited number of images. However, when drug approval is based on  
896 measurement of progression-free survival in a randomized study, careful planning is needed to  
897 minimize bias and to allow the sponsor and the FDA to verify results. This is especially true  
898 when investigators and patients cannot be blinded to treatment assignment because of drug  
899 toxicities or manner of administration. An independent endpoints review committee (IRC)  
900 provides a mechanism to minimize bias in interpretation of the radiologic findings and  
901 independent adjudication of endpoints. We recommend that a clearly described written plan  
902 outlining the IRC function and process, sometimes called an independent review charter, be  
903 agreed upon with the FDA prior to study conduct. It is important that the plan describe how the  
904 independence of the committee will be assured; how images will be collected, stored,  
905 transported, and reviewed; how differences in image interpretation will be resolved; how clinical  
906 data will be used in final endpoint interpretation; and how, if needed, images and IRC results will  
907 be made available to the FDA for audit. The use of an IRC is discussed further in a draft  
908 guidance for the development of medical imaging products.<sup>15</sup>  
909

---

<sup>15</sup> See draft guidance for industry *Developing Medical Imaging Drug and Biological Products, Part 3: Design, Analysis, and Interpretation of Clinical Studies*. When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a CBER guidance, check the CBER guidance Web page at <http://www.fda.gov/cber/guidelines.htm>.