

---

# **Guidance for Industry**

## **Developing Medical Imaging Drug and Biological Products**

### **Part 3: Design, Analysis, and Interpretation of Clinical Studies**

**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)**

**June 2004  
Clinical Medical**

---

# Guidance for Industry

## Developing Medical Imaging Drug and Biological Products

### Part 3: Design, Analysis, and Interpretation of Clinical Studies

*Additional copies of this Guidance are available from:*

*Division of Drug Information HFD-240  
Center for Drug Evaluation and Research  
Food and Drug Administration  
5600 Fishers Lane, Rockville, MD 20857  
(Phone 301-827-4573)  
Internet: <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  
Internet: <http://www.fda.gov/cber/guidelines.htm>.  
Mail: the Voice Information System at 800-835-4709 or 301-827-1800.*

**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)**

**June 2004  
Clinical Medical**

## Table of Contents

<b>I. INTRODUCTION.....</b>	<b>3</b>
<b>II. SCOPE — TYPES OF MEDICAL IMAGING AGENTS .....</b>	<b>4</b>
<b>A. Contrast Agents.....</b>	<b>4</b>
<b>B. Diagnostic Radiopharmaceuticals .....</b>	<b>4</b>
<b>III. GENERAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF MEDICAL IMAGING AGENTS .....</b>	<b>5</b>
<b>A. Phase 1 Studies .....</b>	<b>5</b>
<b>B. Phase 2 Studies .....</b>	<b>6</b>
<b>C. Phase 3 Studies .....</b>	<b>7</b>
<b>IV. ADDITIONAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF EFFICACY .....</b>	<b>7</b>
<b>A. Selecting Subjects.....</b>	<b>7</b>
<b>B. Imaging Conditions and Image Evaluations .....</b>	<b>8</b>
1. <i>Imaging Conditions.....</i>	8
2. <i>Methods and Considerations for Image Evaluation .....</i>	8
3. <i>Steps in Image Evaluation .....</i>	9
a. Assessing objective image features.....	9
b. Image interpretation.....	10
4. <i>Endpoints in Trials.....</i>	10
a. Image interpretations as endpoints.....	10
b. Objective image features as endpoints.....	11
c. Subjective image assessments as endpoints.....	11
d. Clinical outcomes as endpoints.....	11
5. <i>Case Report Forms .....</i>	12
6. <i>CRFs for Image Evaluation .....</i>	12
7. <i>Blinded Imaging Evaluations.....</i>	12
a. Fully blinded image evaluation.....	13
b. Image evaluation blinded to outcome .....	14
c. Sequential Unblinding .....	15
d. Unblinded image evaluations .....	15
8. <i>Independent Image Evaluations.....</i>	16
a. Consensus image evaluations.....	16
b. Repeated image evaluations by the same reader.....	16
9. <i>Offsite and Onsite Image Evaluations .....</i>	17
10. <i>Assessment of Interreader and Intrareader Variability.....</i>	17
11. <i>Protocol and Nonprotocol Images .....</i>	18
a. Protocol images.....	18
b. Nonprotocol images.....	19
12. <i>Separate or Combined Image Evaluations.....</i>	19
a. Separate image evaluations.....	19
b. Combined image evaluations.....	20

<b>C. Truth Standards (Gold Standards)</b> .....	<b>22</b>
<b>D. Comparison Groups</b> .....	<b>23</b>
1. <i>Comparison to an Agent or Modality Approved for a Similar Indication</i> .....	23
a. Noninferiority studies .....	24
b. Agreement studies .....	25
2. <i>Comparison to Placebo</i> .....	26
<b>V. STATISTICAL ANALYSIS</b> .....	<b>26</b>
<b>A. Statistical Methods</b> .....	<b>27</b>
<b>B. Diagnostic Performance</b> .....	<b>28</b>
<b>GLOSSARY</b> .....	<b>29</b>

1  
2  
3 **Guidance for Industry<sup>1</sup>**  
4 **Developing Medical Imaging Drug and Biological Products**  
5 **Part 3: Design, Analysis and Interpretation of Clinical Studies**  
6  
7

8  
9 This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It  
10 does not create or confer any rights for or on any person and does not operate to bind FDA or the public.  
11 An alternative approach may be used if such approach satisfies the requirements of the applicable statutes  
12 and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for  
13 implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate  
14 number listed on the title page of this guidance.  
15

16  
17  
18  
19 **I. INTRODUCTION**  
20

21 This guidance is one of three guidances intended to assist developers of medical imaging drug  
22 and biological products (*medical imaging agents*) in planning and coordinating their clinical  
23 investigations and preparing and submitting investigational new drug applications (INDs), new  
24 drug applications (NDAs), biologics license applications (BLAs), abbreviated NDAs (ANDAs),  
25 and supplements to NDAs or BLAs. The three guidances are: *Part 1: Conducting Safety*  
26 *Assessments; Part 2: Clinical Indications; and Part 3: Design, Analysis, and Interpretation of*  
27 *Clinical Studies.*  
28

29 Medical imaging agents generally are governed by the same regulations as other drug and  
30 biological products. However, because medical imaging agents are used solely to diagnose and  
31 monitor diseases or conditions as opposed to treat them, development programs for medical  
32 imaging agents can be tailored to reflect these particular uses. Specifically, this guidance  
33 discusses our recommendations on how to design a clinical development program for a medical  
34 imaging agent including selecting subjects and acquiring, analyzing, and interpreting medical  
35 imaging data.  
36

37 FDA's guidance documents, including this guidance, do not establish legally enforceable  
38 responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should  
39 be viewed only as recommendations, unless specific regulatory or statutory requirements are

---

<sup>1</sup> This guidance has been prepared by the Division of Medical Imaging and Radiopharmaceutical Drug Products and the Office of Therapeutics Research and Review in the Center for Drug Evaluation and Research (CDER) at the Food and Drug Administration.

## *Contains Nonbinding Recommendations*

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

42  
43 A glossary of common terms used in diagnostic medical imaging is provided at the end of this  
44 document.

45  
46

## 47 **II. SCOPE — TYPES OF MEDICAL IMAGING AGENTS**

48  
49 This guidance discusses medical imaging agents that are administered in vivo and are used for  
50 diagnosis or monitoring with a variety of modalities, such as radiography, computed tomography  
51 (CT), ultrasonography, magnetic resonance imaging (MRI), and radionuclide imaging. The  
52 guidance is not intended to apply to the development of in vitro diagnostic or therapeutic uses of  
53 these agents.<sup>2</sup>

54  
55

Medical imaging agents can be classified into at least two general categories:

56  
57

### 57 **A. Contrast Agents**

58  
59

As used in this guidance, a contrast agent is a medical imaging agent used to improve the  
60 visualization of tissues, organs, and physiologic processes by increasing the relative difference of  
61 imaging signal intensities in adjacent regions of the body. Types of contrast agents include  
62 (1) iodinated compounds used in radiography and CT; (2) paramagnetic metallic ions (such  
63 as ions of gadolinium, iron, and manganese) linked to a variety of molecules and microparticles  
64 (such as superparamagnetic iron oxide) used in MRI; and (3) microbubbles, microaerosomes,  
65 and related microparticles used in diagnostic ultrasonography.

66  
67

### 67 **B. Diagnostic Radiopharmaceuticals**

68  
69

As used in this guidance, a *diagnostic radiopharmaceutical* is (1) an article intended for use in  
70 the diagnosis or monitoring of a disease or a manifestation in humans and that exhibits  
71 spontaneous disintegration of unstable nuclei with the emission of nuclear particles or photons or

---

<sup>2</sup> The guidance is not intended to apply to the development of research drugs that do not provide direct patient benefit with respect to diagnosis, therapy, prevention, or prognosis, or other clinically useful information. These include radioactive drugs for research that are used in accordance with 21 CFR 361.1. Section 361.1 states that radioactive drugs (defined in 21 CFR 310.3(n)) are generally recognized as safe and effective when administered under specified conditions to human research subjects in the course of a project intended to obtain basic information about the metabolism of a radioactively labeled drug or about human physiology, pathophysiology, or biochemistry. However, if a radioactive drug is used for immediate therapeutic, diagnostic, or similar purpose or to determine the safety and effectiveness of the drug in humans, or if the radioactive drug has a pharmacological effect in the body, an IND is required. FDA is developing a guidance on determining when research with radioactive drugs may be conducted under § 361.1.

The Agency recognizes the potential of imaging agents as research tools for aiding the development of therapeutic drugs, and some of the principles of the guidance may be applicable to such research.. Sponsors of such imaging research agents are urged to contact the Division of Medical Imaging and Radiopharmaceutical Drug Products for advice on development of the imaging research agent.

## *Contains Nonbinding Recommendations*

72 (2) any nonradioactive reagent kit or nuclide generator that is intended to be used in the  
73 preparation of such an article.<sup>3</sup> As stated in the preamble to FDA's proposed rule on Regulations  
74 for In Vivo Radiopharmaceuticals Used for Diagnosis and Monitoring, the Agency interprets this  
75 definition to include articles that exhibit spontaneous disintegration leading to the reconstruction  
76 of unstable nuclei and the subsequent emission of nuclear particles or photons (63 FR 28301 at  
77 28303; May 22, 1998).

78  
79 Diagnostic radiopharmaceuticals are generally radioactive drugs or biological products that  
80 contain a radionuclide that typically is linked to a ligand or carrier.<sup>4</sup> These products are used in  
81 planar imaging, single photon emission computed tomography (SPECT), positron emission  
82 tomography (PET), or with other radiation detection probes.

83  
84 Diagnostic radiopharmaceuticals used for imaging typically have two distinct components.

- 85  
86 • A radionuclide that can be detected in vivo (e.g., technetium-99m, iodine-123,  
87 indium-111).

88 The radionuclide typically is a radioactive atom with a relatively short physical half-life  
89 that emits radioactive decay photons having sufficient energy to penetrate the tissue mass  
90 of the patient. These photons can then be detected with imaging devices or other  
91 detectors.

- 92 • A nonradioactive component to which the radionuclide is bound that delivers the  
93 radionuclide to specific areas within the body.

94 This nonradionuclidic portion of the diagnostic radiopharmaceutical often is an organic  
95 molecule such as a carbohydrate, lipid, nucleic acid, peptide, small protein, or antibody.

96 As technology advances, new products may emerge that do not fit into these traditional  
97 categories (e.g., agents for optical imaging, magnetic resonance spectroscopy, combined contrast  
98 and functional imaging). It is anticipated, however, that the general principles discussed here  
99 could apply to these new diagnostic products. Developers of these products are encouraged to  
100 contact the appropriate reviewing division for advice on product development.

### 101 102 103 **III. GENERAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF** 104 **MEDICAL IMAGING AGENTS**

#### 105 106 **A. Phase 1 Studies** 107

---

<sup>3</sup> 21 CFR 315.2 and 601.31.

<sup>4</sup> In this guidance, the terms *ligand* and *carrier* refer to the entire nonradionuclidic portion of the diagnostic radiopharmaceutical.

## *Contains Nonbinding Recommendations*

108 The general goal of phase 1 studies<sup>5</sup> of medical imaging agents is to obtain pharmacokinetic and  
109 human safety assessments of a single mass dose and increasing mass doses of a drug or  
110 biological product. We recommend that evaluation of a medical imaging agent that targets a  
111 specific metabolic process or receptor include assessments of its potential effects on these  
112 processes or receptors.

113  
114 We recommend that, for diagnostic radiopharmaceuticals, organ and tissue distribution data over  
115 time be collected to optimize subsequent imaging protocols and calculate radiation dosimetry  
116 (see Part I, section IV.D). We also recommend that, as appropriate, pharmacokinetic and  
117 pharmacodynamic evaluations be made of the intact diagnostic radiopharmaceutical, the carrier  
118 or ligand, and other vial contents, especially when large amounts of cold components are present  
119 as determined by absolute measurement or by relative concentration of labeled to unlabeled  
120 carrier or ligand. This can be achieved by administering large mass doses of a medical imaging  
121 agent with low specific activity, administering the contents of an entire vial of a medical imaging  
122 agent (assuming that this approximates a worst-case scenario in clinical practice), or both.  
123 Because of potential toxicities, this approach may not be appropriate for some drugs nor for most  
124 biological products. In such cases, we recommend you contact the review division.

125

### **B. Phase 2 Studies**

126

127  
128 The general goals of phase 2 studies of medical imaging agents include (1) refining the agent's  
129 clinically useful mass dose and radiation dose ranges or dosage regimen (e.g., bolus  
130 administration or infusion) in preparation for phase 3 studies, (2) answering outstanding  
131 pharmacokinetic and pharmacodynamic questions, (3) providing preliminary evidence of  
132 efficacy and expanding the safety database, (4) optimizing the techniques and timing of image  
133 acquisition, (5) developing methods and criteria by which images will be evaluated, and  
134 (6) evaluating other critical questions about the medical imaging agent. With the  
135 accomplishment of these elements, phase 3 development should proceed smoothly.

136

137 We recommend that sponsors explore the consequences of both mass dose and radiation dose (or  
138 dosage regimen) adjustment on image acquisition and on the safety or effectiveness of the  
139 administered product. We recommend that additional exploration include adjusting the  
140 following if relevant:

141

- 142 • Character and amount of active and inactive ingredients
- 143 • Amount of radioactivity
- 144 • Amount of nonradioactive ligand or carrier
- 145 • Specific activity
- 146 • Radionuclide that is used

147

---

<sup>5</sup> See also the guidance *Content and Format of Investigational New Drug Applications (INDs) for Phase-1 Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-Derived Products*. This and all other guidances cited in this document are available at FDA's Web site at <http://www.fda.gov/cder/guidance/index.htm>.



## *Contains Nonbinding Recommendations*

148 We recommend that methods used to determine the comparability, superiority, or inferiority of  
149 different mass and radiation doses or regimens be discussed with the Agency. To the extent  
150 possible, the formulation that will be used for marketing should be used during phase 2 studies.  
151 When a different formulation is used, we recommend that bioequivalence and/or other bridging  
152 studies be used to document the relevance of data collected with the original formulation.  
153

154 We recommend that phase 2 studies be designed to define the appropriate patient populations  
155 and clinical settings for phase 3 studies. To gather preliminary evidence of efficacy, however,  
156 both subjects with known disease (or patients with known structural or functional abnormalities)  
157 and subjects known to be normal for these conditions may be included in clinical studies.  
158 However, for products that are immunogenic or exhibit other toxicities, use of healthy subjects  
159 may not be appropriate. We recommend that methods, endpoints, and items on the case report  
160 form (CRF) that will be used in critical phase 3 studies be tested and refined.  
161

### **C. Phase 3 Studies**

162  
163  
164 The general goals of phase 3 efficacy studies for medical imaging agents include confirming the  
165 principal hypotheses developed in earlier studies, demonstrating the efficacy and continued  
166 safety of the medical imaging agent, and validating instructions for use and for imaging in the  
167 population for which the agent is intended. We recommend that the design of phase 3 studies  
168 (e.g., dosage, imaging techniques and times, patient population, and endpoints) be based on the  
169 findings in phase 2 studies. We recommend that the formulation intended for marketing be used,  
170 or bridging studies be performed.  
171

172 When multiple efficacy studies are performed, the studies can be of different designs.<sup>6</sup> To  
173 increase the extent to which the results can be generalized, we recommend the studies be  
174 independent of one another and use different investigators, clinical centers, and readers that  
175 perform the blinded image evaluations (see section IV.B).  
176

## **IV. ADDITIONAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF EFFICACY**

177  
178  
179  
180  
181 The following sections describe special considerations for the evaluation of efficacy in clinical  
182 trials for medical imaging agents (see *Part 2: Clinical Indications*, section IV, for  
183 recommendations on general considerations for establishing effectiveness, clinical usefulness,  
184 and clinical setting).  
185

### **A. Selecting Subjects**

186  
187  
188 We recommend that subjects included in phase 3 clinical efficacy studies be representative of the  
189 population in which the medical imaging agent is intended to be used. We also recommend that  
190 the protocol and study reports specify the method by which patients were selected for

---

<sup>6</sup> See the guidance *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*.

## *Contains Nonbinding Recommendations*

191 participation in the study (e.g., consecutive subjects enrolled, random selection) to facilitate  
192 assessments of potential selection bias (e.g., using a comparator test result to pre-select subjects  
193 most likely to have the desired image finding).<sup>7</sup>

194

### **B. Imaging Conditions and Image Evaluations**

196

197 The following guidance may be customized to the specific medical imaging drug, biological  
198 product, or imaging modality under development. (The term *images* is nonspecific and may refer  
199 to an individual image or to a set of images acquired from different views, different sequences  
200 and timing.)

201

#### *1. Imaging Conditions*

202

203  
204 We recommend that the effects of changes in relevant imaging conditions (e.g., timing of  
205 imaging after product administration, views, instrument settings, patient positioning) on  
206 image quality and reproducibility, including any limitations imposed by changes in such  
207 conditions, be evaluated in early product development. We recommend that subsequent,  
208 phase 3 efficacy trials substantiate and possibly refine these conditions for use.  
209 Appropriate imaging conditions, including limitations, can be described in the product  
210 labeling.

211

#### *2. Methods and Considerations for Image Evaluation*

212

213  
214 We recommend that methods and criteria for image evaluation (including criteria for  
215 image interpretation) be evaluated in early product development. Subsequently, we  
216 recommend that the methods and criteria that are anticipated for clinical use be employed  
217 and substantiated in the phase 3 efficacy trials. For example, early clinical trials might  
218 compare ways in which regions of interest on images are selected or ways in which an  
219 organ will be subdivided on images for purposes of analysis. Similarly, early clinical  
220 trials might evaluate which objective image features (e.g., lesion conspicuity, relative  
221 count rate density) appear to be most affected by the medical imaging agent and which of  
222 these are most useful in image interpretation, such as making a determination of whether  
223 a mass is benign or malignant (see section IV.B.3).

224

225 We recommend that the most appropriate of these methods and criteria for image  
226 evaluation be incorporated into the protocols of the phase 3 efficacy trials.

---

<sup>7</sup> To aid in the subsequent use of this information in clinical trial design, the pretest odds or pretest probabilities of disease can be used as part of the selection criteria as a method of ensuring enrollment of the population of intended use and/or as part of the patient stratification or subsetting criteria for analysis. We recommend that the range of pretest probabilities enrolled be determined by the type of clinical setting that will support the labeling (e.g., a screening setting, a case finding setting, a pivotal decision setting). We recommend that the pretest odds or probabilities be estimated for all subjects after enrollment, but before any trial results are made available. We also recommend that these odds and probabilities be derived from prespecified criteria for disease (e.g., history, physical findings, results of other diagnostic evaluations) according to prespecified algorithms. We recommend that the estimated pretest odds and probabilities of disease should be compared with the pretest odds and probabilities actually observed in the studies. (See the glossary for the definition of terms relating to pretest odds and probabilities for study analysis.)

## *Contains Nonbinding Recommendations*

227  
228 A description of the appropriate methods and criteria for image evaluation, including  
229 limitations, should be described in the product labeling.

230  
231 We recommend that sponsors seek FDA comment on the designs and analysis plans for  
232 the principal efficacy trials before they are finalized. In some cases, special protocol  
233 assessments may be appropriate (see guidance for industry *Special Protocol Assessment*).  
234 In addition, we recommend that the following elements be completed and submitted to  
235 the IND before the phase 3 efficacy studies enroll subjects:

- 236
- 237 • Proposed indications for use
  - 238 • Protocols for the phase 3 efficacy trials
  - 239 • Investigators' brochure
  - 240 • CRFs to be used by on-site investigators
  - 241 • Plan for blinded image evaluations<sup>8</sup>
  - 242 • CRFs to be used by the blinded readers
  - 243 • Statistical analysis plan
  - 244 • Plan for on-site image evaluation and intended use of such evaluation in patient  
245 management, if any
- 246

247 We recommend that sponsors submit a single comprehensive statistical analysis plan for  
248 each principal efficacy study. We recommend that this statistical analysis plan be part of  
249 the study protocol, include the plan for blinded image evaluations, and be submitted to  
250 the protocol before images have been collected.

### 251 252 3. *Steps in Image Evaluation*

253  
254 The evaluation of medical images generally consists of two distinct steps: assessing  
255 objective image features and interpreting findings on the image.

#### 256 257 a. Assessing objective image features

258  
259 As used in this guidance, *objective image features* are attributes on the image that  
260 are either visually perceptible or that can be detected with instrumentation.  
261 Examples of objective image features include signal-to-noise ratios; degree of  
262 delineation; extent of opacification; and the size, number, or density of lesions.

263  
264 Objective image features can be captured on scales that are continuous (e.g., the  
265 diameter of a mass), ordinal (e.g., a feature can be classified as definitely  
266 increased, probably increased, neither increased nor decreased, probably  
267 decreased, definitely decreased), or dichotomous (e.g., a feature can be classified  
268 as present or absent).

269

---

<sup>8</sup> *Blinded* image evaluations may also be referred to as *masked* or as *uninformed* image evaluations.

## *Contains Nonbinding Recommendations*

270 Medical imaging agents have their intended effects by altering objective image  
271 features. We recommend that both the nature and location of such changes on the  
272 image be documented fully during image evaluations in clinical trials intended to  
273 demonstrate efficacy. We also recommend that such documentation also include  
274 changes that are unintended or undesirable. For example, a diagnostic  
275 radiopharmaceutical intended for cardiac imaging also might localize in the liver,  
276 thereby obscuring visualization of parts of the heart.

277  
278 When possible, it is often desirable to perform both a qualitative visual evaluation  
279 of images as well as a quantitative analysis of images with instrumentation.  
280 However, a quantitative image analysis with instrumentation by itself may not be  
281 sufficient to establish efficacy of the medical imaging agent, such as in cases  
282 where images are not intended (or not likely) to be evaluated quantitatively with  
283 instrumentation in clinical practice.

### b. Image interpretation

284  
285  
286 As used in this guidance, an *image interpretation* is the explanation or meaning  
287 that is attributed to objective image features. We recommend that interpretations  
288 of image features be supported by objective, quantitative, and/or qualitative  
289 information derived from the images. For example, the interpretation that cardiac  
290 tissue seen on an image is infarcted, ischemic, or normal might be supported by  
291 objective image features such as the extent and distribution of localization of the  
292 medical imaging agent in the heart (e.g., increased, normal, decreased, or absent),  
293 the time course of such localization, and how these features are affected by  
294 exercise or pharmacologic stress.

## 4. *Endpoints in Trials*

295  
296  
297  
298  
299 Medical imaging agents could be developed for structural delineation; functional,  
300 physiological, or biochemical assessment; disease or pathology detection or assessment;  
301 diagnostic or therapeutic patient management; or multiple or other indications. The  
302 primary endpoints (response variables) relate to the indication's clinical usefulness (see  
303 Part 2: Clinical Indications, section IV.B).

### a. Image interpretations as endpoints

304  
305  
306  
307 Image interpretations that are clinically useful can be incorporated into the  
308 primary endpoint in phase 3 clinical trials. For example, the primary analysis  
309 endpoints of a trial for a medical imaging agent intended for the indication  
310 *disease or pathology detection or assessment* might be the proportions of subjects  
311 with and without the disease who are properly classified against an appropriate  
312 truth standard. In this example, the interpretation that a pulmonary lesion seen on  
313 an image is benign or malignant has direct clinical meaning and can be  
314 incorporated into the primary endpoint.

315

## *Contains Nonbinding Recommendations*

### 316 b. Objective image features as endpoints

317  
318 When the clinical usefulness of particular objective image features is obvious and  
319 apparent, the objective imaging features can be incorporated into the primary  
320 endpoint. For example, in a study of a medical imaging agent intended for brain  
321 imaging, the ability to delineate anatomy that indicates the presence or absence of  
322 cranial masses on images has direct clinical usefulness. The primary endpoint  
323 (e.g., cranial mass detection) serves as the primary basis for the indication for the  
324 product (e.g., the medical imaging agent is indicated for detecting cranial masses  
325 in patients in a particular defined clinical setting).

326  
327 However, in some cases the clinical usefulness of particular objective image  
328 features may not be readily apparent without additional interpretation. In these  
329 cases, we recommend that the objective image features serve as secondary  
330 imaging endpoints. For example, the finding that a medical imaging agent alters  
331 the conspicuity of masses differentially could lead to the interpretation that  
332 specific masses are benign or malignant; acute or chronic; inflammatory,  
333 neoplastic, or hemorrhagic; or lead to some other clinically useful interpretations.  
334 The interpretations can be incorporated into the primary endpoint and can serve as  
335 the primary basis for the indication for the product. However, the objective image  
336 feature of lesion conspicuity might be designated more appropriately as a  
337 secondary imaging endpoint.

### 338 c. Subjective image assessments as endpoints

339  
340 As used in this guidance, *subjective image assessments* are perceptions or  
341 inferences made by the reader. Such assessments are tangible and cannot be  
342 measured objectively. For example, a conclusion that use of a medical imaging  
343 agent alters *diagnostic confidence* is a subjective assessment as is the conclusion  
344 that a medical imaging agent provides *more diagnostic information*.

345  
346 We recommend that subjective image assessments be linked to objective image  
347 features so that the objective basis for such assessments can be understood.  
348 Subjective image assessments can be difficult to validate and replicate. They may  
349 introduce bias as well. Therefore, subjective image assessments should not be  
350 used as primary imaging endpoints.

### 351 d. Clinical outcomes as endpoints

352  
353 Clinical outcomes, such as measurement of symptoms, functioning, or survival,  
354 are among the most direct ways to measure clinical usefulness. Clinical outcomes  
355 can serve as primary endpoints in trials of medical imaging agents. For example,  
356 the primary endpoint of a trial of a medical imaging agent intended for the  
357 indication *therapeutic patient management* in patients with colon cancer might be  
358 a response variable that measures changes in symptoms, functioning, or survival.  
359  
360  
361

## *Contains Nonbinding Recommendations*

### 5. *Case Report Forms*

We recommend that case report forms (CRFs) in trials of medical imaging agents prospectively define the types of observations and evaluations for investigators to record. In addition to data that are usually recorded in CRFs (e.g., inclusion/exclusion criteria, safety findings, efficacy findings), we recommend that the onsite investigator's CRF for a medical imaging agent capture the following information:

- The technical performance of the diagnostic radiopharmaceutical used in the study, if any (e.g., specific activity, percent bound, percent free, percent active, percent inactive)
- The technical characteristics and technical performance of the imaging equipment (e.g., background flood, quality control analysis of the imaging device, pulse height analyzer)
- Methods of image acquisition, output processing, display, reconstruction, and archiving of the imaging study

The collection and availability of the data on the CRF may be important for labeling how the imaging agent is intended to be administered and the appropriate device settings for optimal imaging.

### 6. *CRFs for Image Evaluation*

We recommend that imaging CRFs be designed to capture imaging endpoints, including objective features of the images as well as the location and interpretation of any findings. We recommend that interpretations of image features be supported by objective quantitative or qualitative information derived from the images. We recommend that image interpretations be recorded as distinct items from the assessments of the objective image features. We also recommend that items on the CRFs for image evaluation be carefully constructed to gather information without introducing a bias that suggests the answer that is being sought. We recommend that the proposed labeled indication be clearly derived from specific items in the CRF and from endpoints and hypotheses that have been prospectively stated in the protocol.

### 7. *Blinded Imaging Evaluations*

We recommend that image evaluations be designed to demonstrate that the specific effects of the medical imaging agent, as manifested in the images, provide such information reproducibly and apart from other possible confounding influences or biases. We recommend that blinded image evaluations by multiple independent readers be performed in the phase 3 efficacy studies.

We recommend that either a *fully blinded image evaluation* or an *image evaluation blinded to outcome* by independent readers serve as the principal image evaluation for

### *Contains Nonbinding Recommendations*

408 demonstration of efficacy.<sup>9</sup> Alternatively, both types of image evaluations can be used; if  
409 so, the evaluations can be performed through sequential unblinding. Both primary and  
410 secondary imaging endpoints should be evaluated in this manner. We recommend that  
411 the nature and type of information available to the readers be discussed with FDA before  
412 the trials are initiated.

413

414 In addition to the items outlined in the sections below, we recommend that plans for  
415 blinded image evaluations include the following elements:

416

- 417 • We recommend that the protocol clearly specify the elements to which readers are  
418 blinded.
- 419
- 420 • We recommend that meanings of all endpoints be clearly understood for consistency.  
421 We recommend that terms to be used in image evaluation and classification be  
422 defined explicitly in the image evaluation plan, including such terms as *technically*  
423 *inadequate*, *uninterpretable*, *indeterminate*, or *intermediate*. Blinded readers can be  
424 trained in scoring procedures using sample images from phase 1 and phase 2 studies.  
425
- 426 • We recommend that images be masked for all patient identifiers.  
427
- 428 • We recommend that blinded readers evaluate images in a random sequence.  
429 *Randomization* of images refers to merging the images obtained in the study (to the  
430 fullest degree that is practical) and then presenting images in this merged set to the  
431 readers in a random sequence.

432

433 For example, when images of several diagnostic radiopharmaceuticals read by the  
434 same criteria are being compared to establish relative efficacy (e.g., a comparison of a  
435 test drug or biological product to an established drug or biological product), we  
436 recommend the readers evaluate individual images from the merged set of images in a  
437 random sequence.

438

439 a. Fully blinded image evaluation

440

441 During a *fully blinded image evaluation*, we recommend that readers not have any  
442 knowledge of the following types of information:

443

- 444 • Results of evaluation with the truth standard, of the final diagnosis, or of  
445 patient outcome
- 446
- 447 • Any patient-specific information (e.g., history, physical exam, laboratory  
448 results, results of other imaging studies)

449

---

<sup>9</sup> See section IV.B.8 for a definition of *independent readers*.

## *Contains Nonbinding Recommendations*

450 We recommend that general inclusion and exclusion criteria for patient  
451 enrollment, other details of the protocol, or anatomic orientation to the images not  
452 be provided to the readers.

453  
454 During a *fully blinded image evaluation* in studies where images obtained by  
455 different treatments are being evaluated, we recommend that readers not have  
456 knowledge of treatment identity, to the greatest extent to which that is possible.<sup>10</sup>  
457 For example, in a comparative study of two or more medical imaging agents (or  
458 of two or more doses or regimens of a particular medical imaging agent), we  
459 suggest the blinded readers not know which agent (or which dose or regimen) was  
460 used to obtain a given image.

461  
462 For contrast agents, we suggest this also can include lack of knowledge about  
463 which images were obtained before product administration and which were  
464 obtained after product administration, although sometimes this is apparent upon  
465 viewing the images.

466  
467 In cases where the instructions for image evaluation differ according to treatment  
468 (e.g., as might be the case when images are obtained using different imaging  
469 modalities), blinding the readers to treatment identity may be infeasible.

470  
471 b. Image evaluation blinded to outcome

472  
473 As in a *fully blinded image evaluation*, we recommend that readers performing an  
474 *image evaluation blinded to outcome* not have any knowledge of the results of  
475 evaluation with the truth standard, of the final diagnosis, or of patient outcome.

476  
477 However, in an *image evaluation blinded to outcome*, the readers might have  
478 knowledge of particular elements of patient-specific information (e.g., history,  
479 physical exam, laboratory results, or results of other imaging studies). In some  
480 cases, the readers also might be aware of general inclusion and exclusion criteria  
481 for patient enrollment, other details of the protocol, or anatomic orientation to the  
482 images. We recommend that the particular elements about which the reader will  
483 have information be standardized for all patients and defined prospectively in the  
484 clinical trial protocol, statistical plan, and the blinded image evaluation plan.

485  
486 In studies where images obtained by different treatments are being evaluated  
487 (including *no treatment*, such as in unenhanced image evaluation of a contrast  
488 agent), we recommend that the readers not have knowledge of treatment identity,  
489 to the greatest extent to which that is possible (see section IV.B.7.a).

490

---

<sup>10</sup> This is the common meaning of *blinding* in therapeutic clinical trials. See the ICH guidelines *E8 General Considerations for Clinical Trials* and *E9 Statistical Principles for Clinical Trials*.



## *Contains Nonbinding Recommendations*

### 491 c. Sequential Unblinding

492

493 As used in this guidance, *sequential unblinding* is an assessment where readers  
494 typically evaluate images with progressively more information (e.g., clinical  
495 information) on each read. Sequential unblinding might be used to provide  
496 incremental information under a variety of conditions that may occur in routine  
497 clinical practice (e.g., when no clinical information is available, when limited  
498 clinical information is available, and when a substantial amount of information is  
499 available). This can be used to determine when or how the test agent should be  
500 used in a diagnostic algorithm. We recommend that a typical *sequential*  
501 *unblinding* image evaluation be a three-step process.

502

503 • We recommend that a fully blinded image evaluation be performed. We  
504 recommend that this evaluation be recorded and locked in a dataset by  
505 methods that can be validated. In a *locked* dataset, we recommend that it not  
506 be possible to alter the evaluation later when additional information is  
507 available, or if input is received from the clinical investigators, other readers,  
508 or the sponsor.

509 • We recommend that an image evaluation blinded to outcome be performed.  
510 We recommend this evaluation be recorded and locked in the dataset.

511 • To determine diagnostic performance of the imaging agent, we recommend  
512 that the result of the above two blinded evaluations be compared to the results  
513 of evaluation with the truth standard (or of the final diagnosis, or of patient  
514 outcome).

515

516 Such sequential unblinding can be expanded to include other types of image  
517 evaluations where additional clinical information is provided to the readers. If  
518 sequential unblinding is used, we recommend that the protocol specify the  
519 hypothesis that is to be evaluated at each step. Also, we recommend that the  
520 protocol specify which image evaluation will be the primary one for determining  
521 efficacy.<sup>11</sup>

522

### 523 d. Unblinded image evaluations

524

525 In an *unblinded image evaluation*, readers are aware of the results of patient  
526 evaluation with the truth standard, of the final diagnosis, or of patient outcome.  
527 Unblinded readers also typically are aware of patient-specific information  
528 (e.g., history, physical exam, laboratory results, results of other imaging studies),  
529 of treatment identity where images obtained by different treatments (including no  
530 treatment) are being evaluated, of inclusion and exclusion criteria for patient

---

<sup>11</sup> The labeling should reflect the image methods (blinded, sequentially unblinded, or unblinded, as appropriate) that provided substantial evidence that the Agency used to reach an approval decision and to develop appropriate labeling recommendations for use.

## *Contains Nonbinding Recommendations*

531 enrollment, other details of the protocol, and of anatomic orientation to the  
532 images.

533  
534 Unblinded image evaluations can be used to show consistency with the results of  
535 fully blinded image evaluations or image evaluations blinded to outcome. We  
536 recommend that these blinded and unblinded image evaluations use the same  
537 endpoints so that the results can be compared. However, we recommend that  
538 unblinded image evaluations not be used as the principal image evaluation for  
539 demonstration of efficacy. The unblinded readers may have access to additional  
540 information that may alter the readers' diagnostic assessments and may confound  
541 or bias the image evaluation by these readers.

### 542 543 8. *Independent Image Evaluations*

544  
545 Two events are independent if knowing the outcome of one event says nothing about the  
546 outcome of the other. Therefore, as used in this guidance, *independent readers* are  
547 readers that are completely unaware of findings of other readers (including findings of  
548 other blinded readers and onsite investigators) and are readers who are not otherwise  
549 influenced by the findings of other readers. To ensure that blinded reader's evaluations  
550 remain independent, we recommend that each blinded reader's evaluation be locked in  
551 the dataset shortly after it is obtained and before additional types of image evaluations are  
552 performed (see section IV.B.7.c).

#### 553 554 a. Consensus image evaluations

555  
556 As used in this guidance, *consensus image evaluations (consensus reads)* are  
557 image evaluations during which readers convene to evaluate images together.  
558 Consensus image evaluations can be performed after the individual readings are  
559 completed and locked. However, readers are not considered independent during  
560 consensus reads and therefore we recommend that such reads not serve as the  
561 primary image evaluation used to demonstrate the efficacy of medical imaging  
562 agents. Although a consensus read is performed by several readers, it is actually a  
563 single image-evaluation and is unlikely to fulfill our interest in image evaluations  
564 by multiple blinded readers. As with the individual blinded evaluations, we  
565 recommend that the consensus reads be locked once obtained and before  
566 additional types of blinded readings are performed.

#### 567 568 b. Repeated image evaluations by the same reader

569  
570 In studies where readers evaluate the same image multiple times (e.g., as in  
571 sequential unblinding, or in readings designed to assess *intrareader* variability),  
572 we recommend that the readings be performed independently of one another to  
573 the fullest extent practical. The goal is to minimize *recall bias*. We further  
574 recommend that readers be unaware, to the fullest extent practical, of their own  
575 previous image findings and not be otherwise influenced by those previous  
576 findings.

## *Contains Nonbinding Recommendations*

577  
578 We recommend that different pages in the CRF be used for the two image  
579 evaluations and that each image evaluation be performed with sufficient time  
580 between readings to decrease recall and without reference to prior results.  
581

### 9. *Offsite and Onsite Image Evaluations*

582  
583  
584 As used in this guidance, *offsite image evaluations* are image evaluations performed at  
585 sites that have not otherwise been involved in the conduct of the study and by readers  
586 who have not had contact with patients, investigators, or other individuals involved in the  
587 study. We recommend that Phase 3 trials include offsite image evaluations that are  
588 performed at a limited number of sites (or preferably at a centralized site). In such offsite  
589 evaluations, it is usually easier to control factors that can compromise the integrity of the  
590 blinded image evaluations and to ensure that the blinded readers perform their image  
591 evaluations independently of other image evaluations.  
592

593 As used in this guidance, *onsite image evaluations* are image evaluations performed by  
594 investigators involved in the conduct of the protocol or in the care of the patient. The  
595 term also can refer to blinded image evaluations performed at sites involved with the  
596 conduct of the study. Onsite investigators may have additional information about the  
597 patients that was not predefined in the clinical trial protocol. Such additional information  
598 may alter the investigators' diagnostic assessments and may confound or bias the image  
599 evaluation by the investigators. Therefore, we recommend that onsite image evaluations  
600 usually not be used as the principal image evaluation for demonstration of efficacy, but  
601 be regarded as supportive of the blinded image evaluations.  
602

603 However, we suggest onsite investigators who are blinded to *truth* (e.g., blinded to any  
604 test result that makes up the truth standard, to the final diagnosis, and to patient final  
605 outcome as in an image evaluation blinded to outcome see (section IV.B.7.b)) can be  
606 used for principal image evaluation. In such instances, we recommend that all clinical  
607 information available to the investigator at the time of the image evaluation be clearly  
608 specified and fully documented. We also recommend that a critical assessment of how  
609 such information might have influenced the readings be performed. In addition, we  
610 recommend that an independent blinded evaluation that is supportive of the finding of  
611 efficacy be performed.  
612

### 10. *Assessment of Interreader and Intrareader Variability*

613  
614  
615 We recommend that at least two blinded readers (and preferably three or more) evaluate  
616 images for each study that is intended to demonstrate efficacy. (The truth standard,  
617 however, may be read by a single blinded reader.) The use of multiple readers allows for  
618 an evaluation of the reproducibility of the readings (i.e., interreader variability) and  
619 provides a better basis for subsequent generalization of any findings. Ideally, we  
620 recommend that each reader view all of the images intended to demonstrate efficacy,  
621 both for the investigational imaging agent and the truth standard, so that interreader  
622 agreement can be measured. In large studies, where it may be impractical to have every

## *Contains Nonbinding Recommendations*

623 image read by each reader, a properly chosen subset of images can be selected for such  
624 duplicate image evaluations. We recommend that consistency among readers be  
625 measured quantitatively (e.g., with the kappa statistic).

626  
627 We recommend that *intrareader* variability be assessed during the development of  
628 medical imaging agents. This can be accomplished by having individual blinded readers  
629 perform repeated image evaluations on some or all images (see section IV.B.8.b).

### 630 631 11. *Protocol and Nonprotocol Images*

632  
633 Images obtained in a clinical trial of a medical imaging agent can generally be considered  
634 either protocol or nonprotocol images.

#### 635 636 a. Protocol images

637  
638 As used in this guidance, *protocol images* are images obtained under protocol-  
639 specified conditions and at protocol-specified time points with the goal of  
640 demonstrating or supporting efficacy. We recommend that efficacy evaluations  
641 be based on the evaluations of such protocol images. We also recommend that all  
642 protocol images (e.g., not just those images determined to be evaluable) be  
643 evaluated by the blinded readers, including images of test patients, control  
644 patients, and normal subjects. In addition, we recommend that evaluation of the  
645 protocol images be completed before other images, such as nonprotocol images,  
646 are reviewed by the readers (see section IV.B.11.b).

647  
648 In some cases where large numbers of images are obtained or where image tapes  
649 are obtained (e.g., cardiac echocardiography), sponsors have used image selection  
650 procedures. This is discouraged because the selection of images can introduce the  
651 bias of the selector.

652  
653 We recommend that sponsors specify prospectively in protocols of efficacy  
654 studies how missing images (and images that are technically inadequate,  
655 uninterpretable or show results that are indeterminate or intermediate) will be  
656 handled in the data analysis. Sponsors are encouraged to incorporate analyses in  
657 the statistical analysis plan that incorporate the principle of *intention-to-treat*, but  
658 that are adapted to a diagnostic setting (e.g., *intention-to-diagnose* considers all  
659 subjects enrolled in a diagnostic study regardless of whether they were imaged  
660 with the test drug and regardless of the image quality).<sup>12</sup> Images (including truth  
661 standard images) may be missing from analysis for many reasons, including  
662 patient withdrawal from the study, technical problems with imaging, protocol

---

<sup>12</sup> The *intention-to-treat principle* is defined as the principle that asserts that the effect of a treatment policy can be best assessed by evaluating on the basis of the intention to treat a subject (i.e., the planned treatment regimen) rather than the actual treatment given. As a consequence, we recommend that subjects allocated to a treatment group be followed up, assessed, and analyzed as members of that group irrespective of their compliance with the planned course of treatment (see *E9 Statistical Principles for Clinical Trials*, p. 28).

## *Contains Nonbinding Recommendations*

663 violations, and image selection procedures. We suggest that appropriate methods  
664 be prospectively developed to deal with missing values in the primary response  
665 variable analysis.<sup>13</sup>

666  
667 b. Nonprotocol images

668  
669 As used in this guidance, *nonprotocol image* refers to an image that is not a  
670 protocol image, as defined above (see section IV.B.11.a). These are sometimes  
671 obtained for exploratory purposes and are excluded from the locked phase 3  
672 datasets.

673  
674 12. *Separate or Combined Image Evaluations*

675  
676 Performance of a separate image evaluation does not preclude performance of a  
677 combined image evaluation, and vice versa. If multiple image evaluations are performed,  
678 however, we recommend that the protocol specify which image evaluation will serve as  
679 the primary evaluation and which image evaluations are secondary.

680  
681 a. Separate image evaluations

682  
683 As used in this guidance, a *separate* image evaluation has a reader evaluate test  
684 images obtained from a patient independently of other test images obtained from  
685 that patient, to the fullest degree practical.<sup>14</sup> A reader evaluates each test image  
686 for a patient on its own merits without reference to, or recall of, any other test  
687 images obtained from that patient, to the fullest degree practical.

688  
689 A separate image evaluation often can be performed by combining test images  
690 obtained under different conditions (or at different times) into an intermixed set.  
691 Images in this intermixed set can then be evaluated individually in random order  
692 so that multiple images are not viewed simultaneously, and so that images are not  
693 evaluated sequentially within patients. Alternatively, test images obtained under  
694 one condition (or at a particular time) can be evaluated individually in a random  
695 order, followed by an evaluation in random order of the individual test images  
696 obtained under different conditions (or at different times).

697  
698 As described in the first example below, we recommend that an appropriately  
699 designed separate image evaluation be performed when a goal of a study is to  
700 make comparative inferences about product performance (e.g., to compare the  
701 diagnostic performance of one medical imaging agent with another). As  
702 described in the second example, an appropriately designed separate image  
703 evaluation also can be used to demonstrate that a contrast agent contributes  
704 additional information to images obtained with the device alone.

---

<sup>13</sup> See *E9 Statistical Principles for Clinical Trials*, p. 31.

<sup>14</sup> In the special case where only two test images are being evaluated, a *separate* image evaluation may also be referred to as an *unpaired* image evaluation.

## *Contains Nonbinding Recommendations*

705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750

### Example 1: Comparative inferences of product performance

In a comparative study designed to show that the diagnostic performance of a new medical imaging agent is superior to that of an approved agent and that the new agent can replace the approved agent (see section IV.D.1), we recommend that an appropriate separate image evaluation of test images be performed as the principal image analysis. The *test images* in this case are the images obtained with the new and the approved medical imaging agents. The two agents are not intended to be used together in actual clinical practice, and we therefore recommend that the goal of such an *unpaired* image evaluation be to show that the information obtained with the new agent is clinically and statistically superior to the information obtained with the approved agent. For any given patient, we recommend that images obtained with the new agent be evaluated independently of the evaluation of the images obtained with the approved agent, to the fullest degree practical.

If desired, a side-by-side (*paired*) comparison of images obtained with the new agent and the approved agent can be performed as a secondary image analysis. However, such a side-by-side comparison may yield estimates of diagnostic performance that are biased. The blinded reader may tend to *overread* the presence of masses on the image obtained with the new agent in such a paired comparison. Similarly, the blinded reader may tend to *underread* the image obtained with the new agent in a paired evaluation where a mass is not seen clearly on the image obtained with the approved agent.

In general, these procedures for image evaluation also are applicable to studies designed to show noninferiority. We recommend that sponsors seek Agency comment on proposed study designs and analytical plans before enrolling patients in such studies (see also section IV.D.1 for additional discussion).

### Example 2: Contribution of additional information by a contrast agent

In a study intended to demonstrate that a contrast agent contributes additional information to images obtained with the device alone, it is often highly desirable to perform an appropriate separate image evaluation of test images as the principal image analysis (see the next section for an alternative approach). The *test images*, in this case, include both the images obtained before administration of contrast (the *unenhanced* images) and those obtained after administration of contrast (the *enhanced* images). We recommend that the goal of such an unpaired image evaluation be to show that the information obtained from the enhanced image is clinically and statistically superior to the information obtained from the unenhanced image.

- b. Combined image evaluations

## *Contains Nonbinding Recommendations*

751 As used in this guidance, a *combined* image evaluation has a reader  
752 simultaneously evaluate two or more test images that were obtained under  
753 different conditions or at different times with respect to agent administration.<sup>15</sup> A  
754 combined image evaluation may resemble the conditions under which the product  
755 will be used clinically. For example, in some clinical situations both unenhanced  
756 and enhanced imaging studies are typically performed in patients.<sup>16</sup> If so, such  
757 images often are evaluated concurrently in a comparative fashion.<sup>17</sup> However, as  
758 noted above, such combined image evaluations may increase the likelihood that  
759 bias will be introduced into the image evaluations (e.g., by systematic overreading  
760 or underreading particular findings on images).

761  
762 A combined image evaluation can be performed by creating a set of combined  
763 images for each patient. These sets can then be presented to the blinded readers  
764 in random sequence.

765  
766 When this type of reading is performed, however, we recommend that an  
767 additional independent *separate* image evaluation be completed on at least one of  
768 the members of the combination. We recommend that the member chosen be the  
769 member that usually is obtained under the current standard of practice (e.g., the  
770 unenhanced image). In this way, differences in the evaluations of the combined  
771 reading with those of the separate reading can be assessed. When the goal is to  
772 show that the medical imaging agent adds information to images, we suggest that  
773 these differences demonstrate that the information from the combined images is  
774 clinically and statistically superior to information obtained from the separate  
775 image alone. The results of the combined and separate image evaluations can be  
776 analyzed statistically using paired comparisons.

777  
778 For example, when a two-dimensional ultrasound study of blood vessels is  
779 performed with a microbubble contrast agent, a combined image evaluation could  
780 be performed by evaluating for each patient the unenhanced and enhanced images  
781 side-by-side (or in close temporal proximity). A separate independent evaluation  
782 of the unenhanced image of the blood vessel (i.e., images obtained with the  
783 device alone) for each patient could also be performed. Assessing the differences  
784 for each patient between the results of the combined reading with those of the  
785 separate readings could allow the effects of the microbubble on the images to be  
786 determined.

---

<sup>15</sup> In the special case where only two test images are being evaluated, a *combined* image evaluation can also be referred to as a *paired* image evaluation.

<sup>16</sup> Also, combined images may refer to results from the test drug and modality plus images from a different modality.

<sup>17</sup> Under sections 505 and 502 of the Act, if images are evaluated only in a combined fashion, the approved labeling of the medical imaging agent likely will have to specify that combined evaluations should be performed in clinical practice. If such labeling restrictions are not desired, we recommend that additional separate image evaluations be performed.

## *Contains Nonbinding Recommendations*

787  
788 As noted above, we recommend that combined and separate image evaluations be  
789 performed independently of one another to decrease recall bias (see section  
790 IV.B.8.b). We recommend that different pages in the CRF be used for the  
791 combined and separate evaluations and that the combined and separate image  
792 evaluations be performed at different times without reference to prior results.

793  
794 We recommend that when differences between the combined and separate images  
795 are to be assessed, the combined CRF and separate CRF contain items or  
796 questions that are identical so that differences can be calculated and biases can be  
797 reduced by avoiding questions asking for comparative judgment.

### 799 **C. Truth Standards (Gold Standards)**

800  
801 A truth standard provides an independent way of evaluating the same variable being assessed by  
802 the investigational medical imaging agent. A truth standard is known or believed to give the true  
803 state of a patient or true value of a measurement. Truth standards are used to demonstrate that  
804 the results obtained with the medical imaging agent are valid and reliable and to define summary  
805 test statistics (e.g., sensitivity, specificity, positive and negative predictive value). We  
806 recommend that the following general principles be incorporated prospectively into the design,  
807 conduct, and analysis of the phase 3 efficacy trials for medical imaging agents:

- 808  
809 1. We recommend that the test results obtained with the medical imaging agent be  
810 evaluated without knowledge of the results obtained with the truth standard and without  
811 knowledge of outcome (see section IV.B.7).  
812
- 813 2. We recommend that the true state of the subjects (e.g., diseased or nondiseased)  
814 be determined with a truth standard without knowledge of the test results obtained with  
815 the medical imaging agent.  
816
- 817 3. We recommend that truth standards not include as a component any test results  
818 obtained with the test medical imaging agent (i.e., to avoid *incorporation bias*). This is  
819 because the features of the test image obtained with the test agent (e.g., the *enhanced*  
820 *image*) are likely to be correlated to the features of the image obtained with the device  
821 alone (e.g., the *unenanced image*). For example, in the case of a CT contrast agent  
822 intended to visualize abdominal masses, unenhanced abdominal CT images should not be  
823 included in the truth standard. However, components of the truth standard might include  
824 results from other imaging modalities (e.g., MRI, ultrasonography).  
825
- 826 4. We recommend that evaluation with the truth standard be planned for all enrolled  
827 subjects, and the decision to evaluate a subject with the truth standard not be affected by  
828 the test results with the medical imaging agent under study. For example, if patients with  
829 positive results with the test agent are evaluated preferentially with the truth standard (as  
830 compared to patients with negative test results), the results of the study may be affected  
831 by *partial verification bias*. Similarly, if patients with positive results with the test agent  
832 are evaluated preferentially with the truth standard and those with negative test results are



## *Contains Nonbinding Recommendations*

833 evaluated preferentially with a less rigorous standard, the results of the study may be  
834 affected by *differential verification bias*.<sup>18</sup>

835  
836 We encourage sponsors to seek FDA comment when it is anticipated that a meaningful  
837 proportion of enrolled subjects might not be evaluated with the truth standard or might be  
838 evaluated with a less rigorous standard. In such situations, it may be appropriate to  
839 evaluate clinical outcomes for the enrolled subjects (see section IV.D.4).

840  
841 From a practical perspective, diagnostic standards are derived from procedures that are  
842 considered more definitive in approximating the truth than the test agent. For  
843 example, histopathology or long-term clinical outcomes may be acceptable diagnostic standards  
844 for determining whether a mass is malignant. Diagnostic standards may not be error free, but for  
845 purposes of the clinical trial, they generally are regarded as definitive. However,  
846 misclassification of disease by the truth standard can lead to positive or negative biases in  
847 diagnostic performance measures (*misclassification bias*). Thus, we recommend that the choice  
848 of the truth standard be discussed with the Agency during design of the clinical trials to ensure  
849 that it is appropriate.

850  
851 After the truth standard has been selected, we recommend that the hypothesis for the summary  
852 test statistic in reference to the truth standard be determined and prospectively incorporated into  
853 the study protocol. We recommend that the hypothesis and expected summary statistics reflect  
854 the intended clinical setting for use of the imaging agent (e.g., screening test, sequential  
855 evaluation, alternative to or replacement of another imaging study (see section V)).

### 856 **D. Comparison Groups**

857  
858 Before selecting comparison groups, discussions with the Agency are recommended. General  
859 principles relating to the choice of control groups in clinical trials are set forth in the ICH  
860 guideline *E10 Choice of Control Group and Related Issues in Clinical Trials* (ICH *E10*), and  
861 these principles are applicable to diagnostic trials.

#### 862 863 *1. Comparison to an Agent or Modality Approved for a Similar Indication*

864  
865 If the test agent is being developed as an advance over an approved drug, biological  
866 product, or other diagnostic modality, we recommend that a direct, concurrent  
867 comparison to the approved comparator(s) be performed. We recommend that the  
868 comparison include an evaluation of both the safety and the efficacy data for the  
869 comparator(s) and the test agent. Because of disease variability, typically such  
870 comparisons are performed in the same patient. We recommend that the image  
871 evaluation for the test product or modality be done without knowledge of the imaging  
872 results obtained from the approved products or modalities (see section IV.B.7).

873  
874

---

<sup>18</sup> Partial verification bias and differential verification bias are forms of *diagnostic work-up bias*.

## *Contains Nonbinding Recommendations*

875 We recommend that information from both the test and comparator images (i.e., using the  
876 new and old methods) be compared not only to one another but also to an independent  
877 truth standard. This will facilitate an assessment of possible differences between the  
878 medical imaging agent and the comparator and will enable comparative assessments of  
879 diagnostic performance. Such assessments could be obtained, for example, by comparing  
880 estimates of sensitivity, specificity, positive and negative predictive values, likelihood  
881 ratios, related measures, or receiver operating characteristic (ROC) curves for the  
882 different diagnostic agents. Note that two medical imaging agents could have similar  
883 values for sensitivity and specificity in the same set of patients, yet have poor agreement  
884 rates with each other. Similarly, two medical imaging agents could have good agreement  
885 rates, yet both have poor sensitivity and specificity values. In ROC analysis, overall  
886 areas under the curves obtained with different agents may be comparable, but areas under  
887 partial spans of the curves may be dissimilar. Likewise, one diagnostic agent may have  
888 superior diagnostic performance characteristics over another at one point on the ROC  
889 curve, but may have inferior diagnostic performance characteristics at a different point  
890 (see section V.B).

891  
892 When a medical imaging drug or biological product is being developed for an indication  
893 for which other drugs, biological products, or diagnostic modalities have already been  
894 approved, a direct, concurrent comparison to the approved drug, biological product, or  
895 diagnostic modality is encouraged. However, prior approval of a medical imaging agent  
896 for use in a particular indication does not necessarily mean that the results of a test with  
897 that agent alone can be used as a truth standard. For example, if a medical imaging agent  
898 has been approved on the basis of sufficient concordance of findings with truth as  
899 determined by histopathology, we recommend that assessment of the proposed medical  
900 imaging agent also include determination of truth by histopathology. In this case, the  
901 direct and concurrent comparison of the proposed medical imaging agent to the approved  
902 agent with histopathology serving as the truth standard best measures the performance  
903 difference between the two agents.

904  
905 In studies that compare the effects of a test agent with another drug, biological product,  
906 or imaging modality, we recommend that any images obtained using a nontest agent that  
907 are taken before enrollment be used only as enrollment criteria. We recommend that  
908 these images not be part of the database used to determine test agent performance. Such  
909 baseline enrollment images have inherent selection bias because they are unblinded and  
910 based on referral and management preferences. We recommend that test agent  
911 administration be within a time frame when the disease process is expected not to have  
912 changed significantly. This provides for a fair, balanced comparison between the test and  
913 the comparator agent.

### a. Noninferiority studies

914  
915  
916  
917 Trials can be designed to show that a new test agent is not inferior to a reference  
918 product. In general, the requirements for such studies are more stringent than the  
919 requirements for studies designed to show superiority. Imaging studies, in  
920 particular, can lack assay sensitivity for several reasons, including inappropriate

### *Contains Nonbinding Recommendations*

921 study population, lack of objective imaging endpoints, and inaccuracy in the truth  
922 standard. Moreover, assay sensitivity is difficult to validate because imaging  
923 studies often lack historical evidence of sensitivity to drug effects, and it is not  
924 always clear that the conduct of the imaging procedures and the subsequent image  
925 evaluations did not undermine the trial's ability to distinguish effective treatments  
926 from less effective ones. ICH *E10* provides further guidance on these matters.

927  
928 We recommend that noninferiority studies be based on a concurrent comparison  
929 of the test agent and a reference product and that such studies use objectively  
930 defined endpoints validated by an acceptable truth standard. Such designs allow  
931 comparative assessment of the diagnostic (or functional) performance of the new  
932 and reference tests. For example, if the study endpoint is the presence or absence  
933 of disease, the sensitivities and specificities of the test product and the reference  
934 product can each be compared. The statistical hypotheses may be superiority,  
935 noninferiority, or both. If the test agent is to be used primarily to rule out disease,  
936 high negative predictive value and thus high sensitivity might be more important  
937 than specificity. The objective then would be to show that the new agent, when  
938 compared to the reference test, is superior with regard to sensitivity but not  
939 inferior with regard to specificity.

940  
941 When the study design includes a truth standard but no comparison to a reference  
942 product, the performance levels of the new test agent can only be compared to  
943 some fixed threshold (e.g., prespecified levels of sensitivity and specificity). The  
944 statistical objective should then be to show superiority to the threshold values.  
945 Such values should be based on substantial clinical evidence supporting the  
946 assertion that exceeding the thresholds clearly demonstrates product efficacy.

947  
948 To obtain a noninferiority claim against a reference product, a sponsor should  
949 show that its test agent has been shown to have similar performance  
950 characteristics as the reference product and can be used as an alternative modality  
951 in a precisely defined clinical setting. In other situations, the noninferiority  
952 comparison might only serve as a demonstration of efficacy of the test product.  
953 Generally, non-inferiority trials are designed to show that new and comparator  
954 test performance differ at most by a clinically acceptable margin that has been  
955 agreed to by the Agency. We recommend that noninferiority trials be carefully  
956 planned and that discussions with the Agency begin early in the development  
957 program.

#### 958 b. Agreement studies

959  
960 Similarity between a new test agent and a reference product can also be shown by  
961 demonstrating that both agents consistently give identical results. In this case, the  
962 use of a truth standard is not possible, and the objective is to show agreement  
963 between test and comparator outcomes even though the validity (accuracy) of the  
964 outcomes cannot be verified. High agreement between a new test product and a  
965

## *Contains Nonbinding Recommendations*

966 reference product can support a claim that the new test is an acceptable alternative  
967 to the reference product.

968  
969 In agreement studies, assay sensitivity is critical. In particular, outcomes should  
970 be objectively defined and the two agents should be compared in subjects who  
971 represent an appropriate spectrum of disease conditions. For example, showing  
972 that two diagnostic tests give the same positive diagnosis for a large percentage of  
973 the trial subjects might not be sufficient. We recommend that the sponsor also  
974 demonstrate that the test agent and the reference product respond similarly when a  
975 negative diagnosis prevails and that the probability of discordant outcomes is  
976 negligible. When outcomes are multivalued as opposed to dichotomous,  
977 agreement should be shown across the entire range of test values.

978  
979 An agreement hypothesis should not imply that the agreement between test and  
980 comparator outcomes exceeds agreement among comparator outcomes. Thus, an  
981 understanding of intra-test and intra-reader variability should be taken into  
982 account. For example, consider a new pharmacological stress agent used with  
983 myocardial perfusion imaging to assess perfusion defects. One possible design  
984 would be to apply the comparator procedure to all subjects for a first evaluation  
985 and, for a second evaluation, randomize subjects to receive either the comparator  
986 procedure or the new test agent. This would allow the inter-test agreement to be  
987 directly compared with the intra-test agreement of the comparator using a  
988 noninferiority hypothesis.

989  
990 Because agreement studies do not provide direct evidence of new test validity,  
991 they are difficult to design and execute effectively. Therefore, we recommend  
992 that sponsors pursue agreement studies in limited circumstances and consider  
993 alternative designs that employ an acceptable truth standard.

### 994 995 2. *Comparison to Placebo*

996  
997 Whether the use of a placebo is appropriate in the evaluation of a medical imaging agent  
998 depends on the specific imaging agent, proposed indication, and imaging modality. In  
999 some cases, the use of placebos can help reduce potential bias in the conduct of the study  
1000 and can facilitate unambiguous interpretation of efficacy or safety data. However, in  
1001 some diagnostic studies (such as ultrasonography), products that are considered to be  
1002 placebos (e.g., water, saline, or vehicle) can have some diagnostic effects. We  
1003 recommend that these be used as controls to demonstrate that the medical imaging agent  
1004 has an effect above and beyond that of its vehicle.

## 1005 1006 1007 **V. STATISTICAL ANALYSIS**

1008  
1009 We recommend that statistical methods and the methods by which diagnostic performance will  
1010 be assessed be incorporated prospectively into the statistical analysis plan for each study (see  
1011 section IV.B.2). In addition, we recommend that each study protocol clearly state the hypotheses

## *Contains Nonbinding Recommendations*

1012 to be tested, present sample size assumptions and calculations, and describe the planned  
1013 statistical methods and other data analysis considerations. The ICH guideline *E9 Statistical*  
1014 *Principles for Clinical Trials* provides guidance on these matters.

1015

### 1016 **A. Statistical Methods**

1017

1018 One part of imaging evaluation is the determination of how well the test measures what it is  
1019 intended to measure (validity). The overall diagnostic performance of the product can be  
1020 measured by factors such as sensitivity, specificity, positive and negative predictive values, and  
1021 likelihood ratios. Outcome validity can be demonstrated by a showing that use of the test  
1022 enhances a clinical result.

1023

1024 The reliability of an imaging agent reflects the reproducibility of the result (i.e., the value of a  
1025 measure repeated in the same individual, repeated evaluations of the same image by different  
1026 readers, or repeated evaluations of the same image by the same reader). (See the glossary for  
1027 other related definitions.)

1028

1029 Many studies of imaging agents are designed to provide dichotomous, ordered, or categorical  
1030 outcomes. We think it important that appropriate assumptions and statistical methods be applied  
1031 in their analysis. Statistical tests for proportions and rates are commonly used for dichotomous  
1032 outcomes, and methods based on ranks are often applied to ordinal data. We recommend that  
1033 study outcomes be stratified in a natural way, such as by center or other subgroup category, and  
1034 the Mantel-Haenszel<sup>19</sup> procedures provide effective ways to examine both binomial and ordinal  
1035 data. We recommend that exact methods of analysis, based on conditional inference, be  
1036 employed when necessary. We recommend that the use of model-based methods also be  
1037 encouraged. These models include logistic regression models for binomial data and proportional  
1038 odds models for ordinal data. Log-linear models can be used to evaluate nominal outcome  
1039 variables.

1040

1041 In studies that compare images obtained after the administration of the test agent to images  
1042 obtained before administration, dichotomous outcomes are often analyzed as matched pairs,  
1043 where differences in treatment effects can be assessed using methods for correlated binomial  
1044 outcomes. These studies, however, may be problematic because they often do not employ  
1045 blinding and randomization. For active- and placebo-control studies, including dose-response  
1046 studies, crossover designs can often be used to gain efficiency. We recommend that subjects be  
1047 randomized to order of treatment. If subjects are not randomized to order of treatment, we  
1048 otherwise recommend that the order in which images are evaluated be appropriately randomized.

1049 We recommend that study results from a crossover trial always be analyzed according to  
1050 methods specifically designed for such trials.

1051

---

<sup>19</sup> For more on this topic, see Fleiss, Joseph, L., *Statistical Methods for Rates and Proportions*, 2nd ed., 1981, John Wiley and Sons, New York; and Woolson, Robert, *Statistical Methods for the Analysis of Biomedical Data*, 1987, John Wiley and Sons, New York.

1052 **B. Diagnostic Performance**  
1053

1054 Diagnostic validity can be assessed in a number of ways. For example, both the unenhanced and  
1055 enhanced images could be compared to the truth standard, and the sensitivity and specificity of  
1056 the unenhanced image could be compared to that of the enhanced image. Two different active  
1057 agents can be compared in the same manner. Diagnostic comparisons can also be made when  
1058 there are more than two outcomes to the diagnostic test results. Common methods used to test  
1059 for differences in diagnosis include the McNemar test and the Stuart Maxwell test.<sup>20</sup> In addition,  
1060 we recommend that confidence intervals for sensitivity, specificity, and other measures be  
1061 provided in the analyses. ROC analysis also may be useful in assessing the diagnostic  
1062 performance of medical imaging agents over a range of threshold values.<sup>21</sup> For example, ROC  
1063 analysis can be used to describe the relative diagnostic performance of two medical imaging  
1064 agents if each test can be interpreted using several thresholds to define a positive (or negative)  
1065 test result (see section IV.D.1). For all planned statistical analyses, we recommend that details  
1066 of the analysis methods and specific hypotheses to be tested be stated prospectively in the  
1067 protocol as part of the statistical analysis plan. We recommend that sponsors seek Agency  
1068 comment on the design of and statistical approach to analyses before the protocols are finalized.  
1069

---

<sup>20</sup> Ibid.

<sup>21</sup> For an introduction to this topic, see Metz, Charles E., *Basic Principles of ROC Analysis*, Seminars in Nuclear Medicine 1978;VIII(4):283-298. For a current treatment of statistical issues in diagnostic trials, see Zhou, Xiao-Hua, et al., *Statistical Methods in Diagnostic Medicine*, 2002, John Wiley and Sons, New York.

**Contains Nonbinding Recommendations**

**GLOSSARY**

1070  
1071  
1072  
1073  
1074  
1075  
1076

*Note:* Subjects in trials of medical imaging agents are often classified into one of four groups depending on (1) whether disease is present (often determined with a truth standard or *gold standard*) and (2) the results of the diagnostic test of interest (positive or negative). The following table identifies the variables that are used to estimate the parameters defined below.

<b>Test Result:</b>	<b>Disease:</b>		
	<b>Present (+)</b>	<b>Absent (-)</b>	
<b>Positive (+)</b>	<b>TP (a)</b> true positive=TP	<b>FP (b)</b> false positive=FP	<b>m1 = a+b = TP+FP</b> total with positive test
<b>Negative (-)</b>	<b>FN (c)</b> false negative=FN	<b>TN (d)</b> true negative=TN	<b>m2 = c+d = FN+TN</b> total with negative test
	<b>n1 = a+c = TP+FN</b> total with disease	<b>n2 = b+d = FP+TN</b> total without disease	<b>N = a+b+c+d</b> = TP+FP+FN+TN total in study

1077  
1078

**Accuracy:** (1) In common usage, *accuracy* is the quality of being true or correct. (2) As a measure of diagnostic performance, *accuracy* is a measure of how faithfully the information obtained using a medical imaging agent reflects reality or *truth* as measured by a truth standard or *gold standard*. Accuracy is the proportion of cases, considering both positive and negative test results, for which the test results are correct (i.e., concordant with the truth standard or *gold standard*). Accuracy = (a+d)/N = (TP+TN)/(TP+FP+FN+TN).

1085  
1086  
1087  
1088  
1089

**Comparator:** An established test against which a proposed test is compared to evaluate the effectiveness of the proposed test. A comparator usually means an agent or modality approved for a similar indication. (See also the definition of *reference product*.)

1090  
1091  
1092  
1093  
1094

**Likelihood ratio:** A measure that can be interpreted either as (a) the relative *odds* of a diagnosis, such as being diseased or nondiseased, for a given test result, or (b) the relative *probabilities* of a given test result in subjects with and without the disease. This latter interpretation is analogous to a relative risk or risk ratio.

1095  
1096  
1097  
1098

1. For tests with dichotomous results (e.g., positive or negative test results), the likelihood ratio of a positive test result can be expressed as LR(+), and the likelihood of a negative test result can be expressed as LR(-). See the equations below:

$$LR(+) = \frac{\frac{a}{n1}}{\frac{b}{n2}} = \frac{\text{sensitivity}}{1 - \text{specificity}} = \frac{\text{TruePositiveRate}}{\text{FalsePositiveRate}} = \frac{\frac{a}{b}}{\frac{n1}{n2}} = \frac{\text{PostTestOdds}(+)}{\text{PreTestOdds}}$$

1100

### Contains Nonbinding Recommendations

1101 
$$LR(-) = \frac{\frac{c}{n1}}{\frac{d}{n2}} = \frac{1 - \text{sensitivity}}{\text{specificity}} = \frac{\text{FalseNegativeRate}}{\text{TrueNegativeRate}} = \frac{\frac{c}{d}}{\frac{n1}{n2}} = \frac{\text{PostTestOdds}(-)}{\text{PreTestOdds}}$$

1102

1103 LR(+): *Interpreted as relative odds:* LR(+) is the post-test odds of the disease  
1104 (among those with a positive test result) compared to the pretest odds of  
1105 the disease.

1106

1107 *Interpreted as relative probabilities:* LR(+) is the probability of a positive  
1108 test result in subjects with the disease compared to the probability of a  
1109 positive test result in subjects without the disease.

1110

1111 LR(-): *Interpreted as relative odds:* LR(-) is the post-test odds of the disease  
1112 (among those with a negative test result) compared to the pretest odds of  
1113 the disease.

1114

1115 *Interpreted as relative probabilities:* LR(-) is the probability of a negative  
1116 test result in subjects with the disease compared to the probability of a  
1117 negative test result in subjects without the disease.

1118

1119 2. For tests with several levels of results, such as tests with results expressed on ordinal or  
1120 continuous scales, the likelihood ratio can be used to compare the proportions of subjects  
1121 with and without the disease at different levels of the test result. Alternatively, the  
1122 likelihood ratio can be used to compare the post-test odds of disease at a particular level  
1123 of test result compared with the pretest odds of disease. Thus, the generalized likelihood  
1124 ratio can reflect diagnostic information at any level of the test result.

1125

1126 **Negative predictive value:** The probability that a subject does not have the disease when the  
1127 test result is negative. Synonyms include *predictive value negative*. Negative predictive value =  
1128  $d/m2 = TN/(TN+FN)$ .

1129

1130 By application of Bayes' Rule, the negative predictive value also can be defined as a function of  
1131 pretest probability of disease (p), sensitivity, and specificity:

1132

1133 Negative predictive value =  $[(1-p) \cdot \text{specificity}] / [(1-p) \cdot \text{specificity} + p \cdot (1-\text{sensitivity})]$

1134

1135 **Odds:** The probability that an event will occur compared to the probability that the event will  
1136 not occur. Odds =  $(\text{probability of the event}) / (1 - \text{probability of the event})$ .

1137

1138 **Positive predictive value:** The probability that a subject has disease when the test result is  
1139 positive. Synonyms include *predictive value positive*. Positive predictive value =  $a/m1 =$   
1140  $TP/(TP+FP)$ .

1141

1142 By application of Bayes' Rule, the positive predictive value also can be defined as a function of  
1143 pretest probability of disease (p), sensitivity, and specificity:



## Contains Nonbinding Recommendations

1144

1145 Positive predictive value =  $(p \cdot \text{sensitivity}) / [p \cdot \text{sensitivity} + (1-p) \cdot (1-\text{specificity})]$

1146

1147 **Post-test odds of disease:** The odds of disease in a subject after the diagnostic test results are  
1148 known. Synonyms include *posterior odds of disease*. For subjects with a positive test result, the  
1149 post-test odds of disease =  $a/b = TP/FP$ . For subjects with a negative test result, the post-test  
1150 odds of disease =  $c/d = FN/TN$ . The following expression shows the general relationship  
1151 between the post-test odds and the likelihood ratio: Post-test odds of disease = Pretest odds of  
1152 disease x Likelihood ratio.

1153

1154 **Post-test probability of disease:** The probability of disease in a subject after the diagnostic test  
1155 results are known. Synonyms include *posterior probability of disease*. For subjects with a  
1156 positive test result, the post-test probability of disease =  $a/m1 = TP/(TP+FP)$ . For subjects with a  
1157 negative test result, the post-test probability of disease =  $c/m2 = FN/(TN+FN)$ .

1158

1159 **Precision:** A measure of the reproducibility of a test, including reproducibility within and  
1160 across doses, rates of administration, routes of administration, timings of imaging after product  
1161 administration, instruments, instrument operators, patients, and image interpreters, and possibly  
1162 other variables. Precision is usually expressed in terms of variability, using such measures as  
1163 confidence intervals and/or standard deviations. Precise tests have relatively narrow confidence  
1164 intervals (or relatively small standard deviations).

1165

1166 **Pretest odds of disease:** The odds of disease in a subject before doing a diagnostic test.  
1167 Synonyms include *prior odds of disease*. Pretest odds of disease =  $n1/n2 = (TP+FN)/(TN+FP)$ .

1168

1169 **Pretest probability of disease:** The probability of disease in a subject before doing a diagnostic  
1170 test. Synonyms include *prevalence of disease* and *prior probability of disease*. Pretest  
1171 probability of disease =  $n1/N = (TP+FN)/(TP+FP+FN+TN)$ .

1172

1173 **Probability:** The likelihood of occurrence of an event, expressed as a number between 0 and 1  
1174 (inclusive).

1175

1176 **Receiver operating characteristic (ROC) curve:** A graphical representation of pairs of values  
1177 for *true positive rate* (or sensitivity) and the corresponding *false positive rate* (or 1-specificity)  
1178 for a diagnostic test. Each pair is established by classifying the test result as *positive* when the  
1179 test outcome equals or exceeds the value set by a given threshold, and *negative* when the test  
1180 outcome is less than this threshold value. For example, if a five-point ordinal scale is used to  
1181 rate the likelihood of malignancy for a tumor (e.g., definitely benign, probably benign,  
1182 equivocal, probably malignant, definitely malignant), setting the threshold at *equivocal* will  
1183 classify tumors as malignant (i.e., a *positive* test result) when the test outcome is at this level or  
1184 higher and will classify tumors as nonmalignant (i.e., a *negative* test result) when the test  
1185 outcome is less than this level. To generate an ROC curve, the sensitivity and specificity of the  
1186 diagnostic test are calculated and graphed for several thresholds (e.g., all values of the rating  
1187 scale). In a typical ROC curve, values for *true positive rate* (or sensitivity) are plotted on the  
1188 vertical axis, and the corresponding values for *false positive rate* (or 1-specificity) are plotted on  
1189 the horizontal axis.

### *Contains Nonbinding Recommendations*

- 1190
- 1191 **Reference product:** An FDA-approved drug product having an indication similar to that of an  
1192 investigational drug or biological product to which it is being compared for the purpose of  
1193 evaluating the effectiveness of the investigational drug or biological product.
- 1194
- 1195 **Sensitivity:** The probability that a test result is positive when the subject has the disease.  
1196 Synonyms include *true positive rate*. Sensitivity =  $a/n1 = TP/(TP+FN)$ .
- 1197
- 1198 **Specificity:** The probability that a test result is negative when the subject does not have the  
1199 disease. Synonyms include *true negative rate*. Specificity =  $d/n2 = TN/(TN+FP)$ .
- 1200
- 1201 **Truth standard (gold standard):** An independent method of measuring the same variable  
1202 being measured by the investigational drug or biological product that is known or believed to  
1203 give the *true* value of a measurement.