

---

# Guidance for Industry

## Medical Imaging Drug and Biological Products

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

**This guidance document is being distributed for comment purposes only.**

Comments and suggestions regarding this draft document should be submitted within 30 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane., rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability published in the *Federal Register*.

For questions regarding this draft document contact (CDER) Kyong Kang, 301-827-7510 or (CBER) George Mills, 301-827-5097.

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

**May 2003  
Clinical Medical**

# Guidance for Industry

## 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)**

**May 2003  
Clinical Medical**

## Table of Contents

<b>I. INTRODUCTION.....</b>	<b>2</b>
<b>II. SCOPE — TYPES OF MEDICAL IMAGING AGENTS .....</b>	<b>3</b>
<b>A. Contrast Agents.....</b>	<b>3</b>
<b>B. Diagnostic Radiopharmaceuticals .....</b>	<b>3</b>
<b>III. GENERAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF MEDICAL IMAGING AGENTS .....</b>	<b>4</b>
<b>A. Phase 1 Studies .....</b>	<b>4</b>
<b>B. Phase 2 Studies .....</b>	<b>5</b>
<b>C. Phase 3 Studies .....</b>	<b>5</b>
<b>IV. ADDITIONAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF EFFICACY .....</b>	<b>6</b>
<b>A. Selecting Subjects.....</b>	<b>6</b>
<b>B. Imaging Conditions and Image Evaluations .....</b>	<b>6</b>
1. <i>Imaging Conditions.....</i>	<i>7</i>
2. <i>Methods and Considerations for Image Evaluation .....</i>	<i>7</i>
3. <i>Steps in Image Evaluation .....</i>	<i>8</i>
4. <i>Endpoints in Trials.....</i>	<i>9</i>
5. <i>Case Report Forms .....</i>	<i>10</i>
6. <i>CRFs for Image Evaluation .....</i>	<i>11</i>
7. <i>Blinded Imaging Evaluations.....</i>	<i>11</i>
8. <i>Independent Image Evaluations.....</i>	<i>15</i>
9. <i>Offsite and Onsite Image Evaluations .....</i>	<i>16</i>
10. <i>Assessment of Interreader and Intrareader Variability.....</i>	<i>16</i>
11. <i>Protocol and Nonprotocol Images .....</i>	<i>17</i>
12. <i>Separate or Combined Image Evaluations.....</i>	<i>18</i>
<b>C. Truth Standards (Gold Standards) .....</b>	<b>21</b>
<b>D. Comparison Groups.....</b>	<b>22</b>
1. <i>Comparison to an Agent or Modality Approved for a Similar Indication .....</i>	<i>22</i>
2. <i>Comparison to Placebo .....</i>	<i>23</i>
<b>V. STATISTICAL ANALYSIS.....</b>	<b>24</b>
<b>A. Statistical Methods.....</b>	<b>24</b>
<b>B. Diagnostic Performance .....</b>	<b>25</b>
<b>GLOSSARY.....</b>	<b>26</b>

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

8  
9 This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current  
10 thinking on this topic. It does not create or confer any rights for or on any person and does not operate to  
11 bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements  
12 of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA  
13 staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call  
14 the appropriate 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 of Medical Imaging Agents*; *Part 2: Clinical Indications*; and *Part 3: Design,*  
27 *Analysis, and Interpretation of Clinical Studies*.  
28

29 Medical imaging agents generally are governed by the same regulations as other drug and  
30 biological products.<sup>2</sup> However, because medical imaging agents are used to diagnose and  
31 monitor diseases or conditions, development programs for medical imaging agents can be  
32 tailored to reflect how these products are used. Specifically, this guidance discusses our  
33 recommendations on how to design a clinical development program for a medical imaging agent  
34 including selecting subjects, and acquiring, analyzing and interpreting medical imaging data.  
35

---

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

<sup>2</sup> Sponsors developing medical imaging agents should be familiar with Agency regulations and guidances pertaining to the development of these products.

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

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

41

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

44

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

45

46  
47 This guidance discusses medical imaging agents that are administered in vivo and are used for  
48 diagnosis or monitoring. Included are medical imaging agents used with medical imaging  
49 techniques such as radiography, computed tomography (CT), ultrasonography, magnetic  
50 resonance imaging (MRI), and radionuclide imaging. It is not intended to cover the development  
51 of in vitro diagnostic uses, or to therapeutic uses of these agents.<sup>3</sup>

52

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

54

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

55

56  
57 Contrast agents improve the visualization of tissues, organs, and physiologic processes by  
58 increasing the relative difference of imaging signal intensities in adjacent regions of the body.  
59 Products include: (1) iodinated compounds used in radiography and CT; (2) paramagnetic  
60 metallic ions (such as ions of gadolinium, iron, and manganese) linked to a variety of molecules  
61 and used in MRI; and (3) microbubbles, microaerosomes, and related microparticles used in  
62 diagnostic ultrasonography.

63

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

64

65  
66 As used in this guidance, a *diagnostic radiopharmaceutical* is (1) an article intended for use in  
67 the diagnosis or monitoring of a disease or a manifestation in humans and that exhibits  
68 spontaneous disintegration of unstable nuclei with the emission of nuclear particles or photons or  
69 (2) any nonradioactive reagent kit or nuclide generator that is intended to be used in the  
70 preparation of such an article.<sup>4</sup> As stated in the preamble to FDA's proposed rule on Regulations  
71 for In Vivo Radiopharmaceuticals Used for Diagnosis and Monitoring, the Agency interprets this  
72 definition to include articles that exhibit spontaneous disintegration leading to the reconstruction

---

<sup>3</sup> The guidance is not intended to apply to the development of research drugs that do not have clinical usefulness. The Agency recognizes the potential of imaging as a research tool and some of the principles of the guidance may be applicable. Sponsors of such products are urged to contact the appropriate review division for advice on product development.

<sup>4</sup> 21 CFR 315.2 and 601.31.

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

73 of unstable nuclei and the subsequent emission of nuclear particles or photons (63 FR 28301 at  
74 28303; May 22, 1998).

75  
76 Diagnostic radiopharmaceuticals are generally radioactive drugs or biological products that  
77 contain a radionuclide that may be linked to a ligand or carrier.<sup>5</sup> These products are used in  
78 planar imaging, single photon emission computed tomography (SPECT), positron emission  
79 tomography (PET), or with other radiation detection probes.

80  
81 Diagnostic radiopharmaceuticals used for imaging typically have two distinct components.

- 82  
83 • A radionuclide that can be detected in vivo (e.g., technetium-99m, iodine-123, indium-111).  
84 The radionuclide typically is a radioactive molecule with a relatively short physical half-life  
85 that emits radioactive decay photons having sufficient energy to penetrate the tissue mass of  
86 the patient. These photons can then be detected with imaging devices or other detectors.
- 87 • A nonradioactive component that delivers the molecule to specific areas within the body.  
88 This nonradionuclidic portion of the diagnostic radiopharmaceutical often is an organic  
89 molecule such as a carbohydrate, lipid, nucleic acid, peptide, small protein, or antibody. In  
90 general, the purpose of the nonradioactive component is to direct the radionuclide to a  
91 specific body location or process.

92  
93 As technology advances, new products may emerge that do not fit into these traditional  
94 categories (e.g., agents for optical imaging, magnetic resonance spectroscopy, combined contrast  
95 and functional imaging). In such instances, developers of these products are encouraged to  
96 contact the appropriate reviewing division for advice on product development.

### 97 98 99 **III. GENERAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF** 100 **MEDICAL IMAGING AGENTS**

#### 101 102 **A. Phase 1 Studies**

103  
104 The general goal of phase 1 studies<sup>6</sup> is to obtain pharmacokinetic and human toxicology  
105 assessments of the safety of a single dose and increasing doses of a drug or biological product.  
106 We recommend that evaluation of a medical imaging agent that targets a specific metabolic  
107 process or receptor include assessments of its potential effects on these processes or receptors.

108  
109 We recommend that, for diagnostic radiopharmaceuticals, organ and tissue distribution data over  
110 time be collected to optimize subsequent imaging protocols and calculate radiation dosimetry

---

<sup>5</sup> In this guidance, the terms *ligand* and *carrier* refer to the entire nonradionuclidic portion of the diagnostic radiopharmaceutical.

<sup>6</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*.

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

111 (see Part I, Section IV.D). We also recommend that, as appropriate, pharmacokinetic and  
112 pharmacodynamic evaluations be made of the intact diagnostic radiopharmaceutical, the carrier  
113 or ligand, and other vial contents, especially when large amounts of cold components are  
114 present. This can be achieved by administering large doses of a medical imaging agent with low  
115 specific activity, administering the contents of an entire vial of a medical imaging agent  
116 (assuming that this approximates a worst-case scenario in clinical practice), or both. Because of  
117 potential toxicities, this approach may not be appropriate for some drugs nor for most biological  
118 products. In such cases, we recommend you contact the review division.

119

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

120

121  
122 The general goals of phase 2 studies of medical imaging agents include (1) refining the agent's  
123 clinically useful dose range or dosage regimen (e.g., bolus administration or infusion), (2)  
124 answering outstanding pharmacokinetic and pharmacodynamic questions, (3) providing  
125 preliminary evidence of efficacy and expanding the safety database, (4) optimizing the  
126 techniques and timing of image acquisition, (5) developing methods and criteria by which  
127 images will be evaluated, and (6) evaluating other critical questions about the medical imaging  
128 agent.

129

130 We recommend that sponsors explore the consequences of dose (or dosage regimen) adjustment  
131 on image acquisition and on the safety or effectiveness of the administered product. We  
132 recommend that additional exploration include adjusting the following if relevant:

133

- 134 • character and amount of active and inactive ingredients
- 135 • amount of radioactivity
- 136 • amount of nonradioactive ligand or carrier
- 137 • specific activity
- 138 • radionuclide that is used

139

140 We recommend that methods used to determine the comparability, superiority, or inferiority of  
141 different doses or regimens be discussed with the Agency. To the extent possible, the  
142 formulation that will be used for marketing should be used during phase 2 studies. When a  
143 different formulation is used, we recommend that bioequivalence and/or other bridging studies  
144 be used to document the relevance of data collected with the original formulation.

145

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

153

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

154

155

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

156 The general goals of phase 3 efficacy studies include confirming the principal hypotheses  
157 developed in earlier studies, demonstrating the efficacy and continued safety of the medical  
158 imaging agent, and validating instructions for use and for imaging in the population for which  
159 the agent is intended. We recommend that the design of phase 3 studies (e.g., dosage, imaging  
160 techniques and times, patient population, and endpoints) be based on the findings in phase 2  
161 studies. We recommend that the formulation intended for marketing be used, or bridging studies  
162 be performed.

163  
164 When multiple efficacy studies are performed, the studies can be of different designs.<sup>7</sup> To  
165 increase the extent to which the results can be generalized, we recommend the studies be  
166 independent of one another and use different investigators, clinical centers, and readers that  
167 perform the blinded image evaluations (see Section IV.B).

### **168 IV. ADDITIONAL CONSIDERATIONS IN THE CLINICAL EVALUATION OF 169 EFFICACY 170**

171  
172 The following sections describe special considerations for the evaluation of efficacy in clinical  
173 trials for medical imaging agents (see *Part 2: Clinical Indications*, Section IV, for  
174 recommendations on general considerations for establishing effectiveness, clinical usefulness,  
175 and clinical setting).

#### **176 A. Selecting Subjects 177**

178  
179 We recommend that subjects included in phase 3 clinical efficacy studies be representative of the  
180 population in which the medical imaging agent is intended to be used. We also recommend that  
181 the protocol and study reports specify the method by which patients were selected for  
182 participation in the study (e.g., consecutive subjects enrolled, random selection) to facilitate  
183 assessments of potential selection bias (e.g., selecting patients most likely to have the desired  
184 image finding).<sup>8</sup>

#### **185 B. Imaging Conditions and Image Evaluations 186 187**

---

<sup>7</sup> See the guidance *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*.

<sup>8</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 Part 2, Section IV.C and the glossary for use of pretest odds or probabilities in the analysis of a study.)

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

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

192

### 193 1. *Imaging Conditions*

194

195 We recommend that the effects of changes in relevant imaging conditions (e.g., timing of  
196 imaging after product administration, views, instrument settings, patient positioning) on  
197 image quality and reproducibility, including any limitations imposed by changes in such  
198 conditions, be evaluated in early product development. We recommend that subsequent,  
199 phase 3 efficacy trials substantiate and may refine these conditions for use. Appropriate  
200 imaging conditions, including limitations, can be described in the product labeling.

201

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

203

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

214

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

217

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

220

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

226

- 227 • Proposed indications for use
- 228 • Protocols for the phase 3 efficacy trials
- 229 • Investigators=brochure
- 230 • CRFs to be used by on-site investigators

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

- 231
- 232
- 233
- 234
- 235
- Plan for blinded image evaluations<sup>9</sup>
  - CRFs to be used by the blinded readers
  - Statistical analysis plan
  - Plan for on-site image evaluation and intended use of such evaluation in patient management, if any

236

237 We recommend that sponsors submit a single comprehensive statistical analysis plan for

238 each principal efficacy study. We recommend that this statistical analysis plan be part of

239 the study protocol, include the plan for blinded image evaluations, and be submitted to

240 the protocol before images have been collected.

241

242 3. *Steps in Image Evaluation*

243

244 The evaluation of medical images generally consists of two distinct steps: assessing

245 objective image features and interpreting findings on the image.

246

247 a. Assessing objective image features

248

249 As used in this guidance, *objective image features* are attributes on the image that

250 are either visually perceptible or that can be detected with instrumentation.

251 Examples of objective image features include signal-to-noise ratios; degree of

252 delineation; extent of opacification; and the size, number, or density of lesions.

253

254 Objective image features can be captured on scales that are continuous (e.g., the

255 diameter of a mass), ordinal (e.g., a feature can be classified as definitely

256 increased, probably increased, neither increased nor decreased, probably

257 decreased, definitely decreased), or dichotomous (e.g., a feature can be classified

258 as present or absent).

259

260 Medical imaging agents have their intended effects by altering objective image

261 features. We recommend that both the nature and location of such changes on the

262 image be documented fully during image evaluations in clinical trials intended to

263 demonstrate efficacy. We also recommend that such documentation also include

264 changes that are unintended or undesirable. For example, a diagnostic

265 radiopharmaceutical intended for cardiac imaging also might localize in the liver,

266 thereby obscuring visualization of parts of the heart.

267

268 When possible, it is often desirable to perform both a qualitative visual evaluation

269 of images as well as a quantitative analysis of images with instrumentation.

270 However, a quantitative image analysis with instrumentation by itself may not be

271 sufficient to establish efficacy of the medical imaging agent, such as in cases

---

<sup>9</sup> *Blinded* image evaluations may also be referred to as *masked* or as *uninformed* image evaluations.

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

272 where images are not intended (or not likely) to be evaluated quantitatively with  
273 instrumentation in clinical practice.

274  
275 b. Image interpretation

276  
277 As used in this guidance, an *image interpretation* is the explanation or meaning  
278 that is attributed to objective image features. We recommend that interpretations  
279 of image features be supported by objective, quantitative, and/or qualitative  
280 information derived from the images. For example, the interpretation that cardiac  
281 tissue seen on an image is infarcted, ischemic, or normal might be supported by  
282 objective image features such as the extent and distribution of localization of the  
283 medical imaging agent in the heart (e.g., increased, normal, decreased, or absent),  
284 the time course of such localization, and how these features are affected by  
285 exercise or pharmacologic stress.

286  
287 4. *Endpoints in Trials*

288  
289 Medical imaging agents could be developed for structural delineation; functional,  
290 physiological, or biochemical assessment; disease or pathology detection or assessment;  
291 diagnostic or therapeutic patient management; or multiple or other indications. The  
292 primary endpoints (response variables) relate to the indication's clinical usefulness (see  
293 Part II: Clinical Indications, section IV.B).

294  
295 a. Image interpretations as endpoints

296  
297 Image interpretations that are clinically useful can be incorporated into the  
298 primary endpoint in phase 3 clinical trials. For example, the primary endpoint  
299 (response variable) of a trial for a medical imaging agent intended for the  
300 indication *disease or pathology detection or assessment*, might be the proportion  
301 of subjects with and without the disease who are properly classified against an  
302 appropriate truth standard. In this example, the interpretation that a pulmonary  
303 lesion seen on an image is benign or malignant has direct clinical meaning and  
304 can be incorporated into the primary endpoint.

305  
306 b. Objective image features as endpoints

307  
308 When the clinical usefulness of particular objective image features are obvious  
309 and apparent, the objective imaging features can be incorporated into the primary  
310 endpoint. For example, in a study of a medical imaging agent intended for brain  
311 imaging, the ability to delineate anatomy that indicates the presence or absence of  
312 cranial masses on images has direct clinical usefulness. The primary endpoint  
313 (e.g., cranial mass detection) serves as the primary basis for the indication for the  
314 product (e.g., the medical imaging agent is indicated for detecting cranial masses  
315 in patients in a particular defined clinical setting).

316

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

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

### c. Subjective image assessments as endpoints

330  
331 As used in this guidance, *subjective image assessments* are assessments that are  
332 perceptible only to the reader. Such assessments are not visually perceptible and  
333 cannot be detected with instrumentation. For example, a conclusion that use of a  
334 medical imaging agent alters *diagnostic confidence* is a subjective assessment as  
335 is the conclusion that a medical imaging agent provides *more diagnostic*  
336 *information*.

337  
338 We recommend that subjective image assessments be linked to objective image  
339 features so that the objective basis for such assessments can be understood.  
340 Subjective image assessments can be difficult to validate and replicate. They may  
341 introduce bias as well. Therefore, subjective image assessments should not be  
342 used as primary imaging endpoints.

### d. Clinical outcomes as endpoints

343  
344  
345  
346 Clinical outcomes, such as measurement of symptoms, functioning, or survival,  
347 are among the most direct ways to measure clinical usefulness. Clinical outcomes  
348 can serve as primary endpoints in trials of medical imaging agents. For example,  
349 the primary endpoint of a trial of a medical imaging agent intended for the  
350 indication *therapeutic patient management* in patients with colon cancer might be  
351 a response variable that measures changes in symptoms, functioning, or survival.

## 5. *Case Report Forms*

352  
353  
354  
355 We recommend that case report forms (CRFs) in trials of medical imaging agents  
356 prospectively define the types of observations and evaluations for investigators to record.  
357 In addition to data that are usually recorded in CRFs (e.g., inclusion/exclusion criteria,  
358 safety findings, efficacy findings), we recommend that the onsite investigator's CRF for a  
359 medical imaging agent capture the following information:  
360

## Contains Nonbinding Recommendations

Draft — Not for Implementation

- 361 • The technical performance of the diagnostic radiopharmaceutical used in the  
362 study, if any (e.g., specific activity, percent bound, percent free, percent  
363 active, percent inactive)  
364
- 365 • The technical characteristics and technical performance of the imaging  
366 equipment (e.g., background flood, quality control analysis of the imaging  
367 device, pulse height analyzer)  
368
- 369 • Methods of image acquisition, output processing, display, reconstruction, and  
370 archiving of the imaging study  
371

372 The collection and availability of the data on the CRF may be important for labeling how  
373 the imaging agent is intended to be administered and the appropriate device settings for  
374 optimal imaging.  
375

### 376 6. CRFs for Image Evaluation 377

378 We recommend that imaging CRFs be designed to capture imaging endpoints, including  
379 objective features of the images as well as the location and interpretation of any findings.

380 We recommend that interpretations of image features be supported by objective  
381 quantitative or qualitative information derived from the images. We recommend that  
382 image interpretations be recorded as distinct items from the assessments of the objective  
383 image features. We also recommend that items on the CRFs for image evaluation be  
384 carefully constructed to gather information without introducing a bias that indicates the  
385 answer that is being sought. We recommend that the proposed labeled indication be  
386 clearly derived from specific items in the CRF and from endpoints and hypotheses that  
387 have been prospectively stated in the protocol.  
388

### 389 7. Blinded Imaging Evaluations 390

391 We recommend that image evaluations be designed to demonstrate that the specific  
392 effects of the medical imaging agent, as manifested in the images, provide such  
393 information reproducibly and apart from other possible confounding influences or biases.

394 We recommend that blinded image evaluations by multiple independent readers be  
395 performed in the phase 3 efficacy studies.  
396

397 We recommend that a *fully blinded image evaluation* or an *image evaluation blinded to*  
398 *outcome* by independent readers serve as the principal image evaluation for  
399 demonstration of efficacy.<sup>10</sup> Such image evaluations can be performed through  
400 *sequential unblinding*. Both primary and secondary imaging endpoints should be  
401 evaluated in this manner. We recommend that the nature and type of information  
402 available to the readers be discussed with FDA before the trials are initiated.  
403

---

<sup>10</sup> See Section IV.B.8 for a definition of *independent readers*.

## *Contains Nonbinding Recommendations*

*Draft — Not for Implementation*

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

- 406
- 407 • We recommend that the protocol clearly specify the elements to which readers are  
408 blinded.
  - 409
  - 410 • We recommend that meanings of all endpoints be clearly understood for consistency.  
411 We recommend that terms to be used in image evaluation and classification be  
412 defined explicitly in the image evaluation plan, including such terms as *technically*  
413 *inadequate*, *uninterpretable*, *indeterminate*, or *intermediate*. Blinded readers can be  
414 trained in scoring procedures using sample images from phase 1 and phase 2 studies.  
415
  - 416 • We recommend that images be masked for all patient identifiers.
  - 417
  - 418 • We recommend that blinded readers evaluate images in a random sequence.  
419 *Randomization* of images refers to merging the images obtained in the study (to the  
420 fullest degree that is practical) and then presenting images in this merged set to the  
421 readers in a random sequence.
  - 422

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

428

429 a. Fully blinded image evaluation

430

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

- 433
- 434 • Results of evaluation with the truth standard, of the final diagnosis, or of  
435 patient outcome
  - 436
  - 437 • Any patient-specific information (e.g., history, physical exam, laboratory  
438 results, results of other imaging studies)
  - 439

440 In some cases, we recommend that general inclusion and exclusion criteria for  
441 patient enrollment, other details of the protocol, or anatomic orientation to the  
442 images not be provided to the readers.

443

444 During a *fully blinded image evaluation* in studies where images obtained by  
445 different treatments are being evaluated, we recommend that readers not have

## Contains Nonbinding Recommendations

Draft — Not for Implementation

446 knowledge of treatment identity, to the greatest extent to which that is possible.<sup>11</sup>  
447 For example, in a comparative study of two or more medical imaging agents (or  
448 of two or more doses or regimens of a particular medical imaging agent), we  
449 suggest the blinded readers not know which agent (or which dose or regimen) was  
450 used to obtain a given image.

451  
452 For contrast agents, we suggest this also can include lack of knowledge about  
453 which images were obtained before product administration and which were  
454 obtained after product administration, although sometimes this is apparent upon  
455 viewing the images.

456  
457 In cases where the instructions for image evaluation differ according to treatment  
458 (e.g., as might be the case when images are obtained using different imaging  
459 modalities), blinding the readers to treatment identity may not be feasible.

460  
461 b. Image evaluation blinded to outcome

462  
463 As in a *fully blinded image evaluation*, we recommend that readers performing an  
464 *image evaluation blinded to outcome* not have any knowledge of the results of  
465 evaluation with the truth standard, of the final diagnosis, or of patient outcome.

466  
467 However, in an *image evaluation blinded to outcome*, the readers may have  
468 knowledge of particular elements of patient-specific information (e.g., history,  
469 physical exam, laboratory results, or results of other imaging studies). In some  
470 cases, the readers also may be aware of general inclusion and exclusion criteria  
471 for patient enrollment, other details of the protocol, or anatomic orientation to the  
472 images. We recommend that the particular elements about which the reader will  
473 have information be standardized for all patients and defined prospectively in the  
474 clinical trial protocol, statistical plan, and the blinded image evaluation plan.

475  
476 In studies where images obtained by different treatments are being evaluated  
477 (including *no treatment*, such as in unenhanced image evaluation of a contrast  
478 agent), we recommend that the readers not have knowledge of treatment identity,  
479 to the greatest extent to which that is possible (see Section IV.B.7.a).

480  
481 c. Sequential Unblinding

482  
483 As used in this guidance, in *sequential unblinding* readers typically evaluate  
484 images with progressively more information (e.g., clinical information) on each  
485 read. Sequential unblinding might be used to provide incremental information  
486 under a variety of conditions that may occur in routine clinical practice (e.g.,

---

<sup>11</sup> This is the common meaning of *blinding* in therapeutic clinical trials. See *E8 General Considerations for Clinical Trials*, and *E9 Statistical Principles for Clinical Trials*.

## Contains Nonbinding Recommendations

Draft — Not for Implementation

487 when no clinical information is available, when limited clinical information is  
488 available, and when a substantial amount of information is available). This can be  
489 used to determine when or how the test agent should be used in a diagnostic  
490 algorithm. We recommend that a typical *sequential unblinding* image evaluation  
491 be a three-step process.  
492

- 493 • We recommend that a fully blinded image evaluation be performed. We  
494 recommend that this evaluation be recorded and locked in a dataset by  
495 methods that can be validated. In a *locked* dataset, we recommend that it not  
496 be possible to alter the evaluation later when additional information is  
497 available, or if input is received from the clinical investigators, other readers,  
498 or the sponsor.
- 499 • We recommend that an image evaluation blinded to outcome be performed.  
500 We recommend this evaluation be recorded and locked in the dataset.
- 501 • To determine diagnostic performance of the imaging agent, we recommend  
502 that the result of the above two blinded evaluations be compared to the results  
503 of evaluation with the truth standard (or of the final diagnosis, or of patient  
504 outcome).

505 Such sequential unblinding can be expanded to include other types of image  
506 evaluations where additional clinical information is provided to the readers. If  
507 sequential unblinding is used, we recommend that the protocol specify the  
508 hypothesis that is to be evaluated at each step. Also, we recommend that the  
509 protocol specify which image evaluation will be the primary one for determining  
510 efficacy.<sup>12</sup>  
511

### 512 d. Unblinded image evaluations

513  
514 In an *unblinded image evaluation*, readers are aware of the results of patient  
515 evaluation with the truth standard, of the final diagnosis, or of patient outcome.  
516 Unblinded readers also typically are aware of patient-specific information  
517 (e.g., history, physical exam, laboratory results, results of other imaging studies),  
518 of treatment identity where images obtained by different treatments (including no  
519 treatment) are being evaluated, of inclusion and exclusion criteria for patient  
520 enrollment, other details of the protocol, and of anatomic orientation to the  
521 images.  
522

523 Unblinded image evaluations can be used to show consistency with the results of  
524 fully blinded image evaluations or image evaluations blinded to outcome. We

---

<sup>12</sup> The labeling would 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***

*Draft — Not for Implementation*

525 recommend that these blinded and unblinded image evaluations use the same  
526 endpoints so that the results can be compared. However, we recommend that  
527 unblinded image evaluations not be used as the principal image evaluation for  
528 demonstration of efficacy. The unblinded readers may have access to additional  
529 information that may alter the readers' diagnostic assessments and may confound  
530 or bias the image evaluation by these readers.

### 531 532 8. *Independent Image Evaluations*

533  
534 Two events are independent if knowing the outcome of one event says nothing about the  
535 outcome of the other. Therefore, as used in this guidance, *independent readers* are  
536 readers that are completely unaware of findings of other readers (including findings of  
537 other blinded readers and onsite investigators) and are readers who are not otherwise  
538 influenced by the findings of other readers. To ensure that blinded reader's evaluations  
539 remain independent, we recommend that each blinded reader's evaluation be locked in  
540 the dataset shortly after it is obtained and before additional types of image evaluations are  
541 performed (see Section IV.B.7.c).

#### 542 543 a. Consensus image evaluations

544  
545 As used in this guidance, *consensus image evaluations (consensus reads)* are  
546 image evaluations during which readers convene to evaluate images together.  
547 Consensus image evaluations can be performed after the individual readings are  
548 completed and locked. However, readers are not independent during consensus  
549 reads and therefore we recommend that such reads not serve as the primary image  
550 evaluation used to demonstrate efficacy of medical imaging agents. Although a  
551 consensus read is performed by several readers, it is actually a single image-  
552 evaluation and is unlikely to fulfill our interest in image evaluations by multiple  
553 blinded readers. As with the individual blinded evaluations, we recommend that  
554 the consensus reads be locked once obtained and before additional types of  
555 blinded readings are performed.

#### 556 557 b. Repeated image evaluations by the same reader

558  
559 In studies where readers evaluate the same image multiple times (e.g., as in  
560 sequential unblinding, or in readings designed to assess *intrareader variability*),  
561 we recommend that the readings be performed independently of one another to  
562 the fullest extent practical. The goal is to minimize *recall bias*. We further  
563 recommend that readers be unaware, to the fullest extent practical, of their own  
564 previous image findings and not be otherwise influenced by those previous  
565 findings.

566  
567 We recommend that different pages in the CRF be used for the two image  
568 evaluations, and each image evaluation usually be performed with sufficient time  
569 between readings to decrease recall and without reference to prior results.

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613

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

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

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

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

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

We recommend that at least two blinded readers (and preferably three or more) evaluate images for each study that is intended to demonstrate efficacy. This allows for an evaluation of the reproducibility of the readings (i.e., interreader variability) and provides a better basis for subsequent generalization of any findings. Ideally, we recommend that each reader view all of the images intended to demonstrate efficacy so that interreader agreement can be measured. In large studies, where it may be impractical to have every image read by each reader, a properly chosen subset of images can be selected for such duplicate image evaluations. We recommend that consistency among readers be measured quantitatively (e.g., with the kappa statistic).

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

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

617  
618 *11. Protocol and Nonprotocol Images*

619  
620 Images obtained in a clinical trial of a medical imaging agent can generally be considered  
621 either as protocol or nonprotocol images.

622  
623 a. Protocol images

624  
625 As used in this guidance, *protocol images* are images obtained under protocol-  
626 specified conditions and at protocol-specified time points with the goal of  
627 demonstrating or supporting efficacy. We recommend that efficacy evaluations  
628 be based primarily upon the evaluations of such protocol images. Ideally, we  
629 recommend that all protocol images (e.g., not just those images determined to be  
630 evaluable) be evaluated by the blinded readers, including images of test patients,  
631 control patients, and normal subjects. We recommend that evaluation of the  
632 protocol images be completed before other images, such as nonprotocol images,  
633 are reviewed by the readers (see Section IV.B.11.b).

634  
635 In some cases where large numbers of images are obtained or where image tapes  
636 are obtained (e.g., cardiac echocardiography), sponsors have used image selection  
637 procedures. This is discouraged because the selection of images can introduce the  
638 bias of the selector. In cases where preselection is thought to be needed, the  
639 sponsor is encouraged to clearly identify and discuss the selection procedures  
640 with the appropriate Agency division before their implementation.

641  
642 We recommend that sponsors specify prospectively in protocols of efficacy  
643 studies how missing images (and images that are technically inadequate,  
644 uninterpretable or show results that are indeterminate or intermediate) will be  
645 handled in the data analysis. Sponsors are encouraged to incorporate analyses in  
646 the statistical analysis plan that incorporate the principle of *intention-to-treat*, but  
647 that are adapted to a diagnostic setting (e.g., *intention-to-diagnose* considers all  
648 subjects enrolled in a diagnostic study regardless of whether they were imaged  
649 with the test drug and regardless of the image quality).<sup>13</sup> Images may be missing  
650 from analysis for many reasons, including patient withdrawal from the study,  
651 technical problems with imaging, protocol violations, and image selection

---

<sup>13</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*, p. 49597).

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

652 procedures. We suggest these missing results be included in the primary response  
653 variable analysis, assigned to the arm the subjects were randomized.

654  
655 b. Nonprotocol images

656  
657 As used in this guidance, *nonprotocol image* refers to an image that is not a  
658 protocol image, as defined above (see Section IV.B.11.a). These are sometimes  
659 obtained for exploratory purposes and are excluded from the locked phase 3  
660 datasets.

661  
662 12. *Separate or Combined Image Evaluations*

663  
664 Performance of a separate image evaluation does not preclude performance of a  
665 combined image evaluation, and vice versa. If multiple image evaluations are performed,  
666 however, we recommend that the protocol specify which image evaluation will serve as  
667 the primary evaluation and which image evaluations are secondary.

668  
669 a. Separate image evaluations

670  
671 As used in this guidance, a *separate* image evaluation has a reader evaluate test  
672 images obtained from a patient independently of other test images obtained from  
673 that patient, to the fullest degree practical.<sup>14</sup> A reader evaluates each test image  
674 for a patient on its own merits without reference to, or recall of, any other test  
675 images obtained from that patient, to the fullest degree practical.

676  
677 A separate image evaluation often can be performed by combining test images  
678 obtained under different conditions (or at different times) into an intermixed set.  
679 Images in this intermixed set can then be evaluated individually in random order  
680 so that multiple images are not viewed simultaneously, and so that images are not  
681 evaluated sequentially within patients. Alternatively, test images obtained under  
682 one condition (or at a particular time) can be evaluated individually in a random  
683 order, followed by an evaluation in random order of the individual test images  
684 obtained under different conditions (or at different times).

685  
686 As described in the first example below, we recommend that an appropriately  
687 designed separate image evaluation be performed when a goal of a study is to  
688 make comparative inferences about product performance (e.g., to compare the  
689 diagnostic performance of one medical imaging agent with another). As  
690 described in the second example, an appropriately designed separate image  
691 evaluation also can be used to demonstrate that a contrast agent contributes  
692 additional information to images obtained with the device alone.

---

<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***

*Draft — Not for Implementation*

### 694 Example 1: Comparative inferences of product performance

695  
696 In a comparative study designed to show that the diagnostic performance of a new  
697 medical imaging agent is superior to that of an approved agent and that the new  
698 agent can replace the approved agent (see Section IV.D.1), we recommend that an  
699 appropriate separate image evaluation of test images be performed as the principal  
700 image analysis. The *test images* in this case are the images obtained with the new  
701 and the approved medical imaging agents. The two agents are not intended to be  
702 used together in actual clinical practice, and we therefore recommend that the  
703 goal of such an *unpaired* image evaluation be to show that the information  
704 obtained with the new agent is clinically and statistically superior to the  
705 information obtained with the approved agent. For any given patient, we  
706 recommend that images obtained with the new agent be evaluated independently  
707 of the evaluation of the images obtained with the approved agent, to the fullest  
708 degree practical.

709  
710 If desired, a side-by-side (*paired*) comparison of images obtained with the new  
711 agent and the approved agent can be performed as a secondary image analysis.  
712 However, such a side-by-side comparison may yield estimates of diagnostic  
713 performance that are biased. The blinded reader may tend to *overread* the  
714 presence of masses on the image obtained with the new agent in such a paired  
715 comparison. Similarly, the blinded reader may tend to *underread* the image  
716 obtained with the new agent in a paired evaluation where a mass is not seen  
717 clearly on the image obtained with the approved agent.

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

719  
720  
721 In a study intended to demonstrate that a contrast agent contributes additional  
722 information to images obtained with the device alone, it is often highly desirable  
723 to perform an appropriate separate image evaluation of test images as the  
724 principal image analysis (see the next section for an alternative approach). The  
725 *test images*, in this case, include both the images obtained before administration  
726 of contrast (the *unenhanced* images) and those obtained after administration of  
727 contrast (the *enhanced* images). We recommend that the goal of such an unpaired  
728 image evaluation be to show that the information obtained from the enhanced  
729 image is clinically and statistically superior to the information obtained from the  
730 unenhanced image.

#### 731 732 b. Combined image evaluations

733  
734 As used in this guidance, a *combined* image evaluation has a reader  
735 simultaneously evaluate two or more test images that were obtained under

## Contains Nonbinding Recommendations

Draft — Not for Implementation

736 different conditions or at different times with respect to agent administration.<sup>15</sup> A  
737 combined image evaluation may resemble the conditions under which the product  
738 will be used clinically. For example, in some clinical situations both unenhanced  
739 and enhanced imaging studies are typically performed in patients.<sup>16</sup> If so, such  
740 images often are evaluated concurrently in a comparative fashion.<sup>17</sup> However, as  
741 noted above, such combined image evaluations may increase the likelihood that  
742 bias will be introduced into the image evaluations (e.g., by systematic overreading  
743 or underreading particular findings on images).

744  
745 A combined image evaluation can be performed by creating a set of combined  
746 images for each patient. These sets can then be presented to the blinded readers  
747 in random sequence.

748  
749 When this type of reading is performed, however, we recommend that an  
750 additional independent *separate* image evaluation be completed on at least one of  
751 the members of the combination. We recommend that the member chosen be the  
752 member that usually is obtained under the current standard of practice (e.g., the  
753 unenhanced image). In this way, differences in the evaluations of the combined  
754 reading with those of the separate reading can be assessed. When the goal is to  
755 show that the medical imaging agent adds information to images, we suggest that  
756 these differences demonstrate that the information from the combined images is  
757 clinically and statistically superior to information obtained from the separate  
758 image alone. The results of the combined and separate image evaluations can be  
759 analyzed statistically using paired comparisons.

760  
761 For example, when a two-dimensional ultrasound study of blood vessels is  
762 performed with a microbubble contrast agent, a combined image evaluation could  
763 be performed by evaluating for each patient the unenhanced and enhanced images  
764 side-by-side (or in close temporal proximity). A separate independent evaluation  
765 of the unenhanced image of the blood vessel (i.e., images obtained with the  
766 device alone) for each patient could be performed. Assessing the differences for  
767 each patient between the results of the combined reading with those of the  
768 separate readings could allow the effects of the microbubble on the images to be  
769 determined.

770

---

<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> If images are evaluated only in a combined fashion, labeling of the medical imaging agent likely will 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***

*Draft — Not for Implementation*

771 As noted above, we recommend that combined and separate image evaluations be  
772 performed independently of one another to decrease recall bias (see Section  
773 IV.B.8.b). We recommend that different pages in the CRF be used for the  
774 combined and separate evaluations, and the combined and separate image  
775 evaluations usually be performed at different times without reference to prior  
776 results.

777  
778 We recommend that when differences between the combined and separate images  
779 are to be assessed, the combined CRF and separate CRF contain items or  
780 questions that are identical so that differences can be calculated and to reduce  
781 biases by avoiding questions asking for comparative judgment.  
782

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

783  
784 A truth standard provides an independent way of evaluating the same variable being assessed by  
785 the investigational medical imaging agent. A truth standard is known or believed to give the true  
786 state of a patient or true value of a measurement. Truth standards are used to demonstrate that  
787 the results obtained with the medical imaging agent are valid and reliable and to define summary  
788 test statistics (e.g., sensitivity, specificity, positive and negative predictive value). We  
789 recommend that the following general principles be incorporated prospectively into the design,  
790 conduct, and analysis of the phase 3 efficacy trials for medical imaging agents:  
791

792  
793 1. We recommend that the test results obtained with the medical imaging agent be  
794 evaluated without knowledge of the results obtained with the truth standard and without  
795 knowledge of outcome (see Section IV.B.7).  
796

797 2. We recommend that the true state of the subjects (e.g., diseased or nondiseased)  
798 be determined with a truth standard without knowledge of the test results obtained with  
799 the medical imaging agent.  
800

801 3. We recommend that truth standards not include as a component any test results  
802 obtained with the test medical imaging agent (i.e., to avoid *incorporation bias*). This is  
803 because the features of the test image obtained with the medical imaging agent (e.g., the  
804 *enhanced image*) are likely to be correlated to the features of the image obtained with the  
805 device alone (e.g., the *unenanced image*). For example, in the case of a CT contrast  
806 agent intended to visualize abdominal masses, unenhanced abdominal CT images should  
807 not be included in the truth standard. However, components of the truth standard might  
808 include results from other imaging modalities (e.g., MRI, ultrasonography).  
809

810 4. We recommend that evaluation with the truth standard be planned for all enrolled  
811 subjects, and the decision to evaluate a subject with the truth standard not be affected by  
812 the test results with the medical imaging agent under study. For example, if patients with  
813 positive results with the test agent are evaluated preferentially with the truth standard (as  
814 compared to patients with negative test results), the results of the study may be affected  
815 by *partial verification bias*. Similarly, if patients with positive results with the test agent

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

816 are evaluated preferentially with the truth standard and those with negative test results are  
817 evaluated preferentially with a less rigorous standard, the results of the study may be  
818 affected by *differential verification bias*.<sup>18</sup>  
819

820 We encourage sponsors to seek FDA comment when it is anticipated that a meaningful  
821 proportion of enrolled subjects might not be evaluated with the truth standard or might be  
822 evaluated with a less rigorous standard. In such situations, it may be appropriate to  
823 evaluate clinical outcomes for the enrolled subjects (see Section IV.D.4).  
824

825 From a practical perspective, diagnostic standards are derived from procedures that are  
826 considered more definitive in approximating the truth than the test agent. For  
827 example, histopathology or long-term clinical outcomes may be acceptable diagnostic standards  
828 for determining whether a mass is malignant. Diagnostic standards may not be error free, but for  
829 purposes of the clinical trial, they generally are regarded as definitive. Please note however, that  
830 misclassification of disease by the truth standard can lead to positive or negative biases in  
831 diagnostic performance measures (*misclassification bias*). Thus, we recommend that the choice  
832 of the truth standard be discussed with the Agency during design of the clinical trials to ensure  
833 that it is appropriate.  
834

835 After the truth standard has been selected, we recommend that the hypothesis for the summary  
836 test statistic in reference to the truth standard be determined and prospectively incorporated into  
837 the study protocol. We recommend that the hypothesis and expected summary statistics reflect  
838 the intended clinical setting for use of the imaging agent (e.g., screening test, sequential  
839 evaluation, alternative or replacement of another imaging study (see Section V.))  
840

### **D. Comparison Groups**

841 Before selecting comparison groups, discussions with the Agency are recommended.  
842

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

843 If the test agent is being developed as an advance over an approved drug, biological  
844 product, or other diagnostic modality, we recommend that a direct, concurrent  
845 comparison to the approved comparator(s) be performed. We recommend that the  
846 comparison include an evaluation of both the safety and the efficacy data for the  
847 comparator(s) and the test agent. Because of disease variability, typically such  
848 comparisons are performed in the same patient. We recommend that the image  
849 evaluation for the test product or modality be done without knowledge of the imaging  
850 results obtained from the approved products or modalities (see Section IV.B.7).  
851

852 We recommend that information from both test and comparator images be compared not  
853 only to one another but also to an independent truth standard. This will facilitate an  
854  
855  
856  
857

---

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

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

858 assessment of possible differences between the medical imaging agent and the  
859 comparator(s) and will enable comparative assessments of diagnostic performance. Such  
860 assessments could be obtained, for example, by comparing estimates of sensitivity,  
861 specificity, positive and negative predictive values, likelihood ratios, related measures, or  
862 receiver operating characteristic (ROC) curves for the different diagnostic agents. Note  
863 that two medical imaging agents could have similar values for sensitivity and specificity  
864 in the same set of patients, yet have poor agreement rates with each other. Similarly, two  
865 medical imaging agents could have good agreement rates, yet both have poor sensitivity  
866 and specificity values. In ROC analysis, overall areas under the curves obtained with  
867 different agents may be comparable, but areas under partial spans of the curves may be  
868 dissimilar. Likewise, one diagnostic agent may have superior diagnostic performance  
869 characteristics over another at one point on the ROC curve, but may have inferior  
870 diagnostic performance characteristics at a different point (see Section V.B).

871  
872 When a medical imaging drug or biological product is being developed for an indication  
873 for which other drugs, biological products, or diagnostic modalities have already been  
874 approved, a direct, concurrent comparison to the approved drug, biological product, or  
875 diagnostic modality is encouraged. However, prior approval of a medical imaging agent  
876 for use in a particular indication does not necessarily mean that the results of a test with  
877 that agent can be used as a truth standard. For example, if a medical imaging agent has  
878 been approved on the basis of sufficient concordance of findings with truth as determined  
879 by histopathology, we recommend that assessment of the new medical imaging agent also  
880 usually include determination of truth by histopathology.

881  
882 In studies that compare the effects of a test agent with another drug, biological product,  
883 or imaging modality, we recommend that any images obtained using nontest agent that  
884 are taken before enrollment be used only as enrollment criteria. We recommend that  
885 these images not be part of the database used to determine test agent performance. Such  
886 baseline enrollment images have inherent selection bias because they are unblinded and  
887 based on referral and management preferences. We recommend that test agent  
888 administration be within a time frame when the disease process is expected not to have  
889 changed significantly. This provides for a fair, balanced comparison between the test and  
890 the comparator agent.

891  
892 *2. Comparison to Placebo*

893  
894 Whether the use of a placebo is appropriate in the evaluation of a medical imaging agent  
895 depends on the specific imaging agent, proposed indication, and imaging modality. In  
896 some cases, the use of placebos can help reduce potential bias in the conduct of the study  
897 and can facilitate unambiguous interpretation of efficacy or safety data. However, in  
898 some diagnostic studies (such as ultrasonography), products that are considered to be  
899 placebos (e.g., water, saline, or vehicle) can have some diagnostic effects. We  
900 recommend that these be used as controls to demonstrate that the medical imaging agent  
901 has an effect above and beyond that of the vehicle.

902

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941

### **V. STATISTICAL ANALYSIS**

We recommend that statistical methods and the methods by which diagnostic performance will be assessed be incorporated prospectively into the statistical analysis plan (see Section IV.B.2).

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

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

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

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

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

---

<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.

## *Contains Nonbinding Recommendations*

*Draft — Not for Implementation*

942 We recommend that study results from a crossover trial always be analyzed with methods  
943 specifically designed for such trials.

944

### **B. Diagnostic Performance**

946

947 Diagnostic validity can be assessed in a number of ways. For example, both with unenhanced  
948 and enhanced images, each could be compared to the truth standard, and the sensitivity and  
949 specificity of the unenhanced image could be compared to that of the enhanced image. Two  
950 different active agents can be compared in the same manner. Diagnostic comparisons can also  
951 be made when there are more than two outcomes to the diagnostic test results. Common  
952 methods used to test for differences in diagnosis include the McNemar test and the Stuart  
953 Maxwell test.<sup>20</sup> In addition, we recommend that confidence intervals for sensitivity, specificity,  
954 and other measures be provided in the analyses. Receiver operating characteristic analysis also  
955 may be useful in assessing the diagnostic performance of medical imaging agents over a range of  
956 threshold values.<sup>21</sup> For example, receiver operating characteristic analysis can be used to  
957 describe the relative diagnostic performance of two medical imaging agents if each test can be  
958 interpreted using several thresholds to define a positive (or negative) test result (see Section  
959 IV.D.1). For all planned statistical analyses, we recommend that details of the analysis methods  
960 and specific hypotheses to be tested be stated prospectively in the protocol as part of the  
961 statistical analysis plan. We recommend that sponsors seek Agency comment on the design and  
962 statistical approach to analyses before the protocols are finalized.

963

---

<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.

**Contains Nonbinding Recommendations**

Draft — Not for Implementation

964  
965  
966  
967  
968  
969  
970

**GLOSSARY**

*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.

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

971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988

**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 =  $(a+d)/N = (TP+TN)/(TP+FP+FN+TN)$ .

**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.

- 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:

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

990

## Contains Nonbinding Recommendations

Draft — Not for Implementation

991 
$$LR(-) = \frac{\frac{c}{n1}}{\frac{d}{n2}} = \frac{1 - \text{sensitivity}}{\text{specificity}} = \frac{\text{FalseNegativeRate}}{\text{TrueNegativeRate}} = \frac{\frac{c}{d}}{\frac{n1}{n2}} = \frac{\text{PostTestOdds}(-)}{\text{PreTestOdds}}$$

992

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

996

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

1000

1001 LR(-): *Interpreted as relative odds:* LR(-) is the post-test odds of the disease  
1002 (among those with a negative test result) compared to the pretest odds of  
1003 the disease.

1004

1005 *Interpreted as relative probabilities:* LR(-) is the probability of a negative  
1006 test result in subjects with the disease compared to the probability of a  
1007 negative test result in subjects without the disease.

1008

1009 2. For tests with several levels of results, such as tests with results expressed on ordinal or  
1010 continuous scales, the likelihood ratio can be used to compare the proportions of subjects  
1011 with and without the disease at different levels of the test result. Alternatively, the  
1012 likelihood ratio can be used to compare the post-test odds of disease at a particular level  
1013 of test result compared with the pretest odds of disease. Thus, the generalized likelihood  
1014 ratio can reflect diagnostic information at any level of the test result.

1015

1016 **Negative predictive value:** The probability that a subject does not have the disease given that  
1017 the test result is negative. Synonyms include *predictive value negative*. Negative predictive  
1018 value =  $d/m2 = \text{TN}/(\text{TN}+\text{FN})$ .

1019

1020 By application of Bayes' Rule, the negative predictive value also can be defined as a function of  
1021 pretest probability of disease (p), sensitivity, and specificity:

1022

1023 Negative predictive value =  $[(1-p) \cdot \text{specificity}]/[(1-p) \cdot \text{specificity} + p \cdot (1-\text{sensitivity})]$

1024

1025 **Odds:** The probability that an event will occur compared to the probability that the event will  
1026 not occur. Odds = (probability of the event)/(1 - probability of the event).

1027

1028 **Positive predictive value:** The probability that a subject has disease given that the test result is  
1029 positive. Synonyms include *predictive value positive*. Positive predictive value =  $a/m1 =$   
1030  $\text{TP}/(\text{TP}+\text{FP})$

1031

## Contains Nonbinding Recommendations

Draft — Not for Implementation

1032 By application of Bayes' Rule, the positive predictive value also can be defined as a function of  
1033 pretest probability of disease ( $p$ ), sensitivity, and specificity:

1034  
1035 Positive predictive value =  $(p \cdot \text{sensitivity}) / [p \cdot \text{sensitivity} + (1-p) \cdot (1-\text{specificity})]$   
1036

1037 **Post-test odds of disease:** The odds of disease in a subject after the diagnostic test results are  
1038 known. Synonyms include *posterior odds of disease*. For subjects with a positive test result, the  
1039 post-test odds of disease =  $a/b = TP/FP$ . For subjects with a negative test result, the post-test  
1040 odds of disease =  $c/d = FN/TN$ . The following expression shows the general relationship  
1041 between the post-test odds and the likelihood ratio: Post-test odds of disease = Pretest odds of  
1042 disease x Likelihood ratio.  
1043

1044 **Post-test probability of disease:** The probability of disease in a subject after the diagnostic test  
1045 results are known. Synonyms include *posterior probability of disease*. For subjects with a  
1046 positive test result, the post-test probability of disease =  $a/m1 = TP/(TP+FP)$ . For subjects with a  
1047 negative test result, the post-test probability of disease =  $c/m2 = FN/(TN+FN)$ .  
1048

1049 **Precision:** A measure of the reproducibility of a test, including reproducibility within and  
1050 across doses, rates of administration, routes of administration, timings of imaging after product  
1051 administration, instruments, instrument operators, patients, and image interpreters, and possibly  
1052 other variables. Precision is usually expressed in terms of variability, using such measures as  
1053 confidence intervals and/or standard deviations. Precise tests have relatively narrow confidence  
1054 intervals (or relatively small standard deviations).  
1055

1056 **Pretest odds of disease:** The odds of disease in a subject before doing a diagnostic test.  
1057 Synonyms include *prior odds of disease*. Pretest odds of disease =  $n1/n2 = (TP+FN)/(TN+FP)$ .  
1058

1059 **Pretest probability of disease:** The probability of disease in a subject before doing a diagnostic  
1060 test. Synonyms include *prevalence of disease* and *prior probability of disease*. Pretest  
1061 probability of disease =  $n1/N = (TP+FN)/(TP+FP+FN+TN)$ .  
1062

1063 **Probability:** The likelihood of occurrence of an event, expressed as a number between 0 and 1  
1064 (inclusive).  
1065

1066 **Receiver operating characteristic (ROC) curve:** A graphical representation of pairs of values  
1067 for *true positive rate* (or sensitivity) and the corresponding *false positive rate* (or 1-specificity)  
1068 for a diagnostic test. Each pair is established by classifying the test result as *positive* when the  
1069 test outcome equals or exceeds the value set by a given threshold, and *negative* when the test  
1070 outcome is less than this threshold value. For example, if a five-point ordinal scale is used to  
1071 rate the likelihood of malignancy for a tumor (e.g., definitely benign, probably benign,  
1072 equivocal, probably malignant, definitely malignant), setting the threshold at *equivocal* will  
1073 classify tumors as malignant (i.e., a *positive* test result) when the test outcome is at this level or  
1074 higher and will classify tumors as nonmalignant (i.e., a *negative* test result) when the test  
1075 outcome is less than this level. To generate an ROC curve, the sensitivity and specificity of the  
1076 diagnostic test are calculated and graphed for several thresholds (e.g., all values of the rating

## ***Contains Nonbinding Recommendations***

*Draft — Not for Implementation*

1077 scale). In a typical ROC curve, values for *true positive rate* (or sensitivity) are plotted on the  
1078 vertical axis, and the corresponding values for *false positive rate* (or 1-specificity) are plotted on  
1079 the horizontal axis.

1080  
1081 **Sensitivity:** The probability that a test result is positive given the subject has the disease.  
1082 Synonyms include *true positive rate*.  $\text{Sensitivity} = a/n1 = \text{TP}/(\text{TP}+\text{FN})$ .

1083  
1084 **Specificity:** The probability that a test result is negative given that the subject does not have the  
1085 disease. Synonyms include *true negative rate*.  $\text{Specificity} = d/n2 = \text{TN}/(\text{TN}+\text{FP})$ .

1086  
1087 **Truth standard (gold standard):** An independent method of measuring the same variable  
1088 being measured by the investigational drug or biological product that is known or believed to  
1089 give the *true* value measurement