

1 **Artificial Intelligence-Enabled Device**
2 **Software Functions: Lifecycle**
3 **Management and Marketing**
4 **Submission Recommendations**

6 **Draft Guidance for Industry and**
7 **Food and Drug Administration Staff**

9 ***DRAFT GUIDANCE***

11 **This draft guidance document is being distributed for comment purposes**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

14 **Document issued on January 7, 2025.**

16 You should submit comments and suggestions regarding this draft document within 90 days of
17 publication in the *Federal Register* of the notice announcing the availability of the draft
18 guidance. Submit electronic comments to <https://www.regulations.gov>. Submit written
19 comments to the Dockets Management Staff, Food and Drug Administration, 5630 Fishers Lane,
20 Room 1061, (HFA-305), Rockville, MD 20852-1740. Identify all comments with the docket
21 number listed in the notice of availability that publishes in the *Federal Register*.

23 For questions about this document regarding CDRH-regulated devices, contact the Digital Health
24 Center of Excellence at digitalhealth@fda.hhs.gov. For questions about this document regarding
25 CBER-regulated devices, contact the Office of Communication, Outreach, and Development
26 (OCOD) at 1-800-835-4709 or 240-402-8010, or by email at ocod@fda.hhs.gov. For questions
27 about this document regarding CDER-regulated products, contact druginfo@fda.hhs.gov. For
28 questions about this document regarding combination products, contact the Office of
29 Combination Products at combination@fda.gov.



33 U.S. Department of Health and Human Services
34 Food and Drug Administration
35 Center for Devices and Radiological Health
36 Center for Biologics Evaluation and Research
Center for Drug Evaluation and Research
Office of Combination Products in the Office of the Commissioner

Preface

37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Additional Copies

CDRH

Additional copies are available from the Internet. You may also send an email request to CDRH-Guidance@fda.hhs.gov to receive a copy of the guidance. Please include the document number GUI00007028 and complete title of the guidance in the request.

CBER

Additional copies are available from the Center for Biologics Evaluation and Research (CBER), Office of Communication, Outreach, and Development (OCOD), 10903 New Hampshire Ave., Bldg. 71, Room 3128, Silver Spring, MD 20993-0002, or by calling 1-800-835-4709 or 240-402-8010, by email, ocod@fda.hhs.gov, or from the Internet at <https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances>.

CDER

Additional copies are available from the Center for Drug Evaluation and Research (CDER), Office of Communication, Division of Drug Information, 10001 New Hampshire Ave., Hillendale Bldg., 4th Floor, Silver Spring, MD 20993-0002, or by calling 855-543-3784 or 301796-3400, or by email, druginfo@fda.hhs.gov, or from the Internet at <https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs>.

Table of Contents

61		
62	I. Introduction.....	1
63	II. Scope.....	2
64	III. TPLC Approach: General Principles	4
65	IV. How to Use this Guidance: Overview of AI-Enabled Device Marketing Submission	
66	Content Recommendations	5
67	A. Quality System Documentation	7
68	V. Device Description.....	8
69	VI. User Interface and Labeling.....	10
70	A. User Interface.....	11
71	B. Labeling	12
72	VII. Risk Assessment	16
73	VIII. Data Management	18
74	IX. Model Description and Development	24
75	X. Validation.....	26
76	A. Performance Validation	27
77	XI. Device Performance Monitoring.....	32
78	XII. Cybersecurity	34
79	XIII. Public Submission Summary	37
80	Appendix A: Table of Recommended Documentation.....	40
81	Appendix B: Transparency Design Considerations.....	41
82	Appendix C: Performance Validation Considerations.....	44
83	Appendix D: Usability Evaluation Considerations.....	48
84	Appendix E: Example Model Card.....	50
85	Appendix F: Example 510(k) Summary with Model Card.....	53

86

87

88

89

Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations

Draft Guidance for Industry and Food and Drug Administration Staff

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

I. Introduction

FDA has long promoted a total product life cycle (TPLC) approach to the oversight of medical devices, including artificial intelligence (AI)-enabled devices, and has committed to developing guidances and resources for such an approach. Some recent efforts include developing guiding principles for good machine learning practice (GMLP)¹ and transparency for machine learning-enabled devices² to help promote safe, effective, and high-quality machine learning models; and a public workshop on fostering a patient-centered approach to AI-enabled devices, including discussions of device transparency for users.³ This guidance intends to continue these efforts, by providing lifecycle management and marketing submission recommendations consistent with a TPLC approach for AI-enabled devices.

This guidance provides recommendations on the contents of marketing submissions for devices that include AI-enabled device software functions including documentation and information that will support FDA's review. To support the development of appropriate documentation for FDA's assessment of devices, this guidance also provides recommendations for the design and

¹ See FDA's website on [Good Machine Learning Practice for Medical Device Development: Guiding Principles](#).

² See FDA's website on [Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles](#).

³ See FDA's website on [Artificial Intelligence and Machine Learning \(AI/ML\) Software as a Medical Device Action Plan](#), the Executive Summary for the "[Patient Engagement Advisory Committee Meeting on Artificial Intelligence \(AI\) and Machine Learning \(ML\) in Medical Devices](#)," and the website on the [Virtual Public Workshop - Transparency of Artificial Intelligence/Machine Learning-enabled Medical Devices](#).

Contains Nonbinding Recommendations

Draft – Not for Implementation

119 development of AI-enabled devices that manufacturers may consider using throughout the
120 TPLC. The recommendations reflect a comprehensive approach to lifecycle management of AI-
121 enabled devices throughout the TPLC. Furthermore, the guidance includes FDA’s current
122 thinking on strategies to address transparency and bias throughout the TPLC of AI-enabled
123 devices, including by collecting evidence to evaluate whether a device benefits all relevant
124 demographic groups (e.g., race, ethnicity, sex, and age) similarly, to help ensure that these
125 devices remain safe and effective for their intended use.

126
127 The emergence of consensus standards related to software has helped to improve the consistency
128 and quality of software development and documentation, particularly with respect to activities
129 such as risk assessment and management. When possible, FDA harmonized the terminology and
130 recommendations in this guidance with software-related consensus standards. The Agency
131 encourages the consideration of such FDA-recognized consensus standards when developing AI-
132 enabled devices and preparing premarket documentation. For the current edition of the FDA-
133 recognized consensus standards referenced in this document, see the [FDA Recognized
134 Consensus Standards Database](#). If submitting a Declaration of Conformity to a recognized
135 standard, we recommend including the appropriate supporting documentation. For more
136 information regarding use of consensus standards in regulatory submissions, refer to the FDA
137 guidance titled “[Appropriate Use of Voluntary Consensus Standards in Premarket Submissions
138 for Medical Devices](#)” and “[Standards Development and the Use of Standards in Regulatory
139 Submissions Reviewed in the Center for Biologics Evaluation and Research](#).”

140
141 In general, FDA’s guidance documents do not establish legally enforceable responsibilities.
142 Instead, guidances describe the Agency’s current thinking on a topic and should be viewed only
143 as recommendations, unless specific regulatory or statutory requirements are cited. The use of
144 the word *should* in Agency guidances means that something is suggested or recommended, but
145 not required.

146

147 **II. Scope**

148 For purposes of this guidance, FDA refers to a software function that meets the definition of a
149 device as a “device software function.” A “device software function” is a software function that
150 meets the device definition in section 201(h) of the Federal Food, Drug, and Cosmetic Act
151 (FD&C Act).⁴ As discussed in other FDA guidance, the term “function” is a distinct purpose of
152 the product, which could be the intended use or a subset of the intended use of the product.⁵

153

154 AI-enabled devices are devices that include one or more AI-enabled device software functions
155 (AI-DSFs). An AI-DSF is a device software function that implements one or more “AI models”
156 (referred to as “models” in this guidance) to achieve its intended purpose. A model is a
157 mathematical construct that generates an inference or prediction based on new input data. In this
158 guidance, when “AI-enabled device” is used, it refers to the whole device, whereas when “AI-
159 DSF” is used, it refers only to the function that uses AI. In this guidance, when “model” is used,

⁴ Device software functions may include Software as a Medical Device (SaMD) and Software in a Medical Device (SiMD). See FDA’s website on [Software as a Medical Device \(SaMD\)](#).

⁵ See FDA’s guidance titled “[Multiple Function Device Products: Policy and Considerations](#).”

Contains Nonbinding Recommendations

Draft – Not for Implementation

160 it refers only to the mathematical construct.

161
162 To continue to support the development of AI enabled devices, this guidance provides
163 recommendations on the documentation and information that should be included in marketing
164 submissions to support FDA’s review of devices that include AI-DSFs. For purposes of this
165 guidance, the term “marketing submission” refers to premarket notification (510(k)) submission,
166 De Novo classification request, Premarket Approval (PMA) application, Humanitarian Device
167 Exemption (HDE), or Biologics License Application (BLA).⁶ Some of the proposed
168 recommendations in this guidance also may apply to Investigational Device Exemption (IDE)
169 submissions. For AI-enabled devices subject to 510(k) requirements, an AI-enabled device can
170 be found substantially equivalent to a non-AI-enabled device with the same intended use
171 provided, among other things, the AI-enabled device does not introduce different questions of
172 safety and effectiveness compared to the non-AI-enabled device and meets other requirements
173 for a determination of substantial equivalence in accordance with section 513(i) of the FD&C
174 Act.

175
176 Generally, the recommendations in this guidance also apply to the device constituent part⁷ of a
177 combination product⁸ when the device constituent part includes an AI-DSF. In developing an AI-
178 DSF, sponsors should consider the impact of the AI-DSF in the context of the combination
179 product as a whole. For a combination product that includes an AI-DSF, we highly encourage
180 early engagement with the FDA lead review division for the combination product.⁹ In
181 accordance with the Inter-Center consult process, the FDA lead review division will consult the
182 appropriate subject matter experts.¹⁰ FDA recommends that sponsors refer to other guidances for
183 recommendations on other aspects of investigational considerations and marketing submissions
184 for combination products.¹¹

185
186 The recommendations proposed within this guidance are based on FDA’s experience with
187 reviewing a variety of AI-enabled devices, as well as current regulatory science research.

188
189 While the proposed recommendations are intended to be broadly applicable to AI-enabled
190 devices, many of these recommendations may be specifically relevant to devices that incorporate
191 the subset of AI known as machine learning, particularly deep learning and neural networks.
192 Additional considerations may apply for other forms of AI.

193
194 In some cases, this guidance highlights recommendations from other guidances in order to assist
195 manufacturers with applying those recommendations to AI-enabled devices. The inclusion of
196 certain recommendations in this guidance does not negate applicable recommendations in other
197 guidances that may not be included. This guidance should be considered in the context of the

⁶ Certain devices are subject to review through a BLA under section 351 of the Public Health Service Act.

⁷ See 21 CFR 4.2.

⁸ See 21 CFR 3.2(e).

⁹ See FDA’s guidance titled “[Principles of Premarket Pathways for Combination Products](#).”

¹⁰ See FDA’s Staff Manual Guide titled “[Combination Products: Inter-Center Consult Request Process](#).”

¹¹ See FDA websites titled “[Combination Products Guidance Documents](#)” and “[Search for FDA Guidance Documents](#).” See also FDA’s website on [Combination Products](#) for additional policy information regarding combination products.

Contains Nonbinding Recommendations

Draft – Not for Implementation

198 FD&C Act, its implementing regulations, and other guidance documents.

199
200 This guidance is not intended to provide a complete description of what may be necessary to
201 include in a marketing submission for an AI-enabled device. In particular, this guidance
202 references sections of the FDA guidance titled “[Content of Premarket Submissions for Device](#)
203 [Software Functions](#)” (hereafter referred to as “[Premarket Software Guidance](#)”), which includes
204 significant additional considerations for AI-enabled devices, but does not include references to
205 every section of that guidance. Additionally, this guidance does not address all of the data and
206 information to be submitted in support of a specific indication for an AI-enabled device. FDA
207 recommends that sponsors also refer to other guidances, as applicable to a particular device, for
208 recommendations on other aspects of a marketing submission.¹² Examples of relevant guidances
209 for specific technologies include the FDA guidances titled “[Technical Performance Assessment](#)
210 [of Quantitative Imaging in Radiological Device Premarket Submissions](#)” and “[Technical](#)
211 [Considerations for Medical Devices with Physiologic Closed-Loop Control Technology](#).” FDA
212 further encourages sponsors to consider other available resources including consensus standards
213 and publicly available information when preparing their marketing submissions. As with all
214 devices, FDA intends to take a risk-based approach to determining specific testing and applicable
215 recommendations to support marketing submissions for AI-enabled devices.

216
217 Early engagement with FDA can help guide product development and submission preparation. In
218 particular, early engagement could be helpful when new and emerging technology is used in the
219 development or design of the device, or when novel methods are used during the validation of
220 the device. FDA encourages sponsors to consider discussing these plans with FDA via the Q-
221 Submission Program.¹³

222

223 **III. TPLC Approach: General Principles**

224 This guidance acknowledges the importance of a TPLC approach to the management of AI-
225 enabled devices. In addition to recommendations regarding the documentation and information
226 that should be included in marketing submissions, which reflect a comprehensive approach to the
227 management of risk throughout the TPLC, the resources provided in this guidance are also
228 intended to assist with the device development and lifecycle management of AI-enabled devices,
229 which should help support the safety and effectiveness of these devices. This guidance provides
230 both specific recommendations on the information and documentation to support a marketing
231 submission for an AI-enabled device, as well as recommendations for the design, development,
232 deployment, and maintenance of AI-enabled devices, including the performance management.¹⁴

¹² For other guidances with digital health content, please see FDA’s website on [Guidances with Digital Health Content](#). For all other guidances, please see FDA’s website on [Guidance Documents \(Medical Devices and Radiation-Emitting Products\)](#).

¹³ See FDA’s guidance titled, “[Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program](#)” (hereafter referred to as the “Q-Submission Program”).

¹⁴ This guidance is not intended to provide recommendations on reporting to FDA when a device has or may have caused or contributed to a death or serious injury as required by section 519 of the FD&C Act, the Medical Device Reporting (MDR) Regulation in 21 CFR Part 803, or the Medical Device Reports of Corrections and Removals Regulation in 21 CFR Part 806. For an explanation of the current reporting and recordkeeping requirements

Contains Nonbinding Recommendations

Draft – Not for Implementation

233
234 This guidance also includes FDA’s current thinking on strategies to address transparency and
235 bias throughout the TPLC of AI-enabled devices. These interconnected considerations are
236 important throughout the TPLC and should be incorporated from the earliest stage of device
237 design through decommission to help design transparency and the control of bias into the device
238 and ensure its safety and effectiveness. Transparency involves ensuring that important
239 information is both accessible and functionally comprehensible and is connected both to the
240 sharing of information, and to the usability of a device. AI bias is a potential tendency to produce
241 incorrect results in a systematic, but sometimes unforeseeable way, which can impact safety and
242 effectiveness of the device within all or a subset of the intended use population (e.g., different
243 healthcare settings, different input devices, sex, age, etc.). A comprehensive approach to
244 transparency and bias is particularly important for AI-enabled devices, which can be hard for
245 users to understand due to the opacity of many models and model reliance on data correlations
246 that may not map directly to biologically plausible mechanisms of action. Recommendations for
247 a design approach to transparency are provided in Appendix B (Transparency Design
248 Considerations). With regards to the control of bias for AI-enabled devices this can include
249 addressing representativeness in data collection for development, testing, and monitoring
250 throughout the product lifecycle, as well as evaluating performance across subgroups of intended
251 use.

252
253 Finally, this guidance includes recommendations that address the performance of AI-enabled
254 devices throughout the TPLC, including in the postmarket setting. For example, AI-enabled
255 devices can be sensitive to differences in input data (also referred to as data drift), such as input
256 data used during development as compared to input data in actual deployments. Further, in
257 addition to data drift, which occurs when systems that produce inputs for AI-enabled devices
258 change over times in ways that may impact the performance of the device but may not be evident
259 to users, AI-enabled devices can also be susceptible to changes in performance due to other
260 factors. Sponsors are also encouraged to consider the use of a predetermined change control plan
261 (PCCP), as discussed in FDA guidance titled “[Marketing Submission Recommendations for a
262 Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software
263 Functions](#),” which describes an approach for manufacturers to prospectively specify and seek
264 premarket authorization for intended modifications to an AI-DSF (e.g., to improve device
265 performance) without needing to submit additional marketing submissions or obtain further FDA
266 authorization before implementing such modification consistent with the PCCP.

267

268 **IV. How to Use this Guidance: Overview of AI-**
269 **Enabled Device Marketing Submission Content**
270 **Recommendations**

271 This guidance provides recommendations on the documentation and information that should be
272 included in marketing submissions to support FDA’s review of devices that include AI-DSFs.
273

applicable to manufacturers of medical devices, please refer to FDA’s guidance titled “[Medical Device Reporting for Manufacturers](#).”

Contains Nonbinding Recommendations

Draft – Not for Implementation

274 There are some differences between the way FDA and the AI community consider the AI-
275 enabled device TPLC and certain terminology. Therefore, this guidance clarifies these
276 differences to facilitate better understanding of the recommendations in this guidance. For
277 example, the AI community often uses the term “validation” to refer to data curation or model
278 tuning that can be combined with the model training phase to optimize the model selection.¹⁵
279 However, validation is defined in 21 CFR 820.3(z)¹⁶ as “...confirmation by examination and
280 provision of objective evidence that the particular requirements for a specific intended use can be
281 consistently fulfilled.” This guidance uses the definition in 21 CFR 820.3(z), specifically when
282 addressing the evaluation of performance of the model for its intended use. For clarity, using the
283 term “validation” to refer to the training and tuning process should be avoided in the context of
284 medical device marketing submissions. Also, the term “development” is used throughout this
285 guidance to refer to training, tuning, and tuning evaluation (often referred to as “internal testing”
286 in the AI community). In this guidance, “test data” is used to refer to data that may be used for
287 verification and validation activities, also known as the testing process, and is not used to
288 describe part of the development process. The [“FDA Digital Health and Artificial Intelligence
289 Glossary – Educational Resource”](#) provides a compilation of commonly used terms in the
290 artificial intelligence and machine learning space and their definitions.

291
292 Sections V through XIII of this guidance describe the marketing submission content
293 recommendations for AI-enabled devices. Specifically, in each section, under, “*Why should it be
294 included in a submission for an AI-enabled device,*” an explanation is provided for why certain
295 information should be included in a marketing submission. An explanation of what
296 documentation and information should be included in a marketing submission can be found
297 under “*What sponsors should include in a submission.*” Finally, recommendations regarding
298 where sponsors should include the information within each section of a marketing submission
299 can be found under “*Where sponsors should provide it in a submission.*” Information regarding
300 recommendations for lifecycle considerations as well as examples of marketing submission
301 materials are provided in the appendices of this guidance.

302
303 The recommendations related to marketing submissions are organized according to how they
304 should appear in the submission (See Appendix A (Table of Recommended Documentation)),
305 which does not always align directly with the order of activities in the TPLC. While all
306 referenced submission sections are provided to FDA during premarket review, they include
307 information about what has already been done to develop and validate the device, as well as what
308 a sponsor plans to do in the future to ensure a device’s ongoing safety and effectiveness. Some

¹⁵ See International Medical Device Regulators Forum Technical Document N67 titled “[Machine Learning-enabled Medical Devices: Key Terms and Definitions.](#)”

¹⁶ On February 2, 2024, FDA issued a final rule amending the device Quality System Regulation (QSR), 21 CFR Part 820, to align more closely with international consensus standards for devices (89 FR 7496). This final rule will take effect on February 2, 2026. Once in effect, this rule will withdraw the majority of the current requirements in Part 820 and instead incorporate by reference the 2016 edition of the International Organization for Standardization (ISO) 13485, Medical devices – Quality management systems – Requirements for regulatory purposes, in Part 820. As stated in the final rule, the requirements in ISO 13485 are, when taken in totality, substantially similar to the requirements of the current Part 820, providing a similar level of assurance in a firm’s quality management system and ability to consistently manufacture devices that are safe and effective and otherwise in compliance with the FD&C Act. When the final rule takes effect, FDA will also update the references to provisions in 21 CFR Part 820 in this guidance to be consistent with that rule.

Contains Nonbinding Recommendations

Draft – Not for Implementation

309 sections of the guidance also describe information relevant to multiple steps in the TPLC. One
310 example of how the sections in this guidance may align with the TPLC is included below:

- 311 • **Development** – Risk Assessment, Data Management, and Model Description and
312 Development
- 313 • **Validation** – Data Management and Validation
- 314 • **Description of the Final Device** – Device Description, Model Description and
315 Development, User Interface and Labeling, Public Submission Summary
- 316 • **Postmarket Management** – Device Performance Monitoring and Cybersecurity

317
318 This guidance generally describes information that would be generated and documented during
319 software development, verification, and validation. However, the information necessary to
320 support market authorization will vary based on the specifics of each AI-enabled device, and
321 during premarket review FDA may request additional information that is needed to evaluate the
322 submission.

323

A. Quality System Documentation

324
325 When considering the recommendations in Sections V through XIII of this guidance, it may be
326 helpful to consider if the documentation and information that should be included in a marketing
327 submission, under “*What sponsors should include in a submission,*” could exist in the Quality
328 System documentation. One source of documentation that may be used as part of demonstrating
329 substantial equivalence or reasonable assurance of safety and effectiveness in the marketing
330 submission for certain AI-enabled devices is documentation related to the ongoing requirements
331 of the Quality System (QS) Regulation.¹⁷ This guidance explains how some documentation that
332 may be relevant for QS regulation compliance for medical devices generally can also be
333 provided premarket to demonstrate how a sponsor or manufacturer is addressing risks associated
334 with AI-enabled devices specifically.

335

336 For example, the QS Regulation requires that manufacturers establish design controls for certain
337 finished devices (see 21 CFR 820.30). Specifically, as part of design controls, a manufacturer
338 must “establish and maintain procedures for validating the device design,” which “shall ensure
339 that devices conform to defined user needs and intended uses and shall include testing of
340 production units under actual or simulated use conditions” (21 CFR 820.30(g)). In addition,
341 under 21 CFR 820.30(i) a manufacturer must establish and maintain procedures to identify,
342 document, validate or where appropriate verify, review, and approve of design changes before
343 their implementation (“design changes”) for all devices, including those automated with

¹⁷ In the postmarket context, design controls may also be important to ensure medical device performance and maintain medical device safety and effectiveness. FDA recommends that device manufacturers implement comprehensive performance risk management programs and documentation consistent with the QS Regulation, including but not limited to management responsibility (21 CFR 820.20), design validation (21 CFR 820.30(g)), design changes (21 CFR 820.30(i)), nonconforming product (21 CFR 820.90), and corrective and preventive action (21 CFR 820.100). While FDA generally does not assess QS Regulation compliance as part of its review of premarket submissions under section 510(k) of the FD&C Act, this guidance is intended to explain how FDA evaluates the performance of the device performance-related outputs of activities that are part and parcel of QS Regulation compliance, and explain how the QS Regulation can be leveraged to demonstrate these performance-related outputs.

Contains Nonbinding Recommendations

Draft – Not for Implementation

344 software. Similarly, as part of the control of nonconforming product, manufacturers must
345 establish and maintain procedures to “control product that does not conform to specified
346 requirements,” including, under some circumstances, user requirements, and to implement
347 corrective and preventative action, including “complaints” and “other sources of quality data” to
348 identify “existing and potential causes of nonconforming product.” (21 CFR 820.90(a) and
349 820.100(a)(1)). Further, manufacturers have ongoing responsibility to manage the quality system
350 and maintain device quality,¹⁸ including by reviewing the “suitability and effectiveness of the
351 quality system at defined intervals and with sufficient frequency according to established
352 procedures” to ensure the quality objectives are being met.¹⁹
353

V. Device Description

354
355 *Why should it be included in a submission for an AI-enabled device:* The following section
356 describes information that sponsors should provide in the device description section of their
357 marketing submission to help FDA understand the general characteristics of the AI-enabled
358 device. The following recommendations supplement device-specific recommendations and
359 recommendations provided in the [Premarket Software Guidance](#), where applicable.
360

361 The device description supports FDA’s understanding of the intended use, expected operational
362 sequence of the device (e.g., clinical workflow of the device), use environment, features of the
363 model, and design of the AI-enabled device. This information is needed for FDA to evaluate the
364 safety and effectiveness of the device. The device description provides important context about
365 what the device does, including how it works, how a user may interact with it, and under what
366 circumstances a device is likely to be used as intended.
367

368 For recommendations related to how to include information in the marketing submission about
369 the technical characteristics of the model, and the method by which the model was developed,
370 see Section IX (Model Description and Development) of this guidance.
371

372 *What sponsors should include in a submission:* In general, sponsors should include the following
373 types of information as part of a device description for an AI-enabled device:

- 374 • A statement that AI is used in the device.
- 375 • A description of the device inputs and device outputs, including whether the inputs are
376 entered manually or automatically, and a list of compatible input devices and acquisition
377 protocols, as applicable.
- 378 • An explanation of how AI is used to achieve the device’s intended use. For devices with
379 multiple functions, this explanation may include how AI-DSFs interact with each other as
380 well as how they interact with non-AI-DSFs.
- 381 • A description of the intended users, their characteristics, and the level and type of training
382 they are expected to have and/or receive. Users include those who will interpret the
383 output. When relevant, list the qualifications or clinical role of the users intended to
384 interpret the output. Users also include all people who interact with the device including

¹⁸ See 21 CFR 820.20.

¹⁹ 21 CFR 820.20(c).

Contains Nonbinding Recommendations

Draft – Not for Implementation

385 during installation, use, and maintenance. For example, users may include technicians,
386 health care providers, patients, and caregivers, as well as administrators and others
387 involved in decisions about how to deploy medical devices, and how the device fits into
388 clinical care.

- 389 • A description of the intended use environment(s) (e.g., clinical setting, home setting).
- 390 • A description of the intended workflow for the use of the device (e.g., intended decision-
391 making role), including:
 - 392 ○ A description of the degree of automation that the device provides in comparison
393 to the workflow for the current standard of care;
 - 394 ○ A description of the clinical circumstances that may lead to use; and
 - 395 ○ An explanation of how the outputs will be used in the clinical workflow.
- 396 • A description of installation and maintenance procedures.
- 397 • A description of any calibration and/or configuration procedures that must be regularly
398 performed by users in order to maintain performance, including when calibration must be
399 performed and how users can identify if calibration is needed again or is incorrect, as
400 applicable.

401

402 Additionally, sponsors should include the following types of information as part of a device
403 description for an AI-enabled device that has elements that can be configured by a user:

- 404 • A description of all configurable elements of the AI-enabled device, for example:
 - 405 ○ Visualizations that the user can turn on/off (e.g., overlays, quality indicators, or
406 heatmaps);
 - 407 ○ Software inputs;
 - 408 ○ Model parameters when they are configured during use; and/or
 - 409 ○ Alert thresholds.
- 410 • A description of how these elements and their settings can be configured, including:
 - 411 ○ A description of the users who make configuration decisions (e.g., clinical user,
412 administrative user, patient) including any necessary qualifications and training
413 needed to make these decisions, as applicable;
 - 414 ○ An explanation of how users know which selections have been made;
 - 415 ○ A description of the level at which the configuration is defined, for example at the
416 patient-, clinical site- or hospital network-level; and
 - 417 ○ A description of customizable pre-defined operating points, their outputs and
418 performance ranges, as applicable. It is also important to specify how the
419 operating points or operating point range(s) were selected based on the indications
420 for use of the device.
- 421 • A description of the potential impact of the configurable elements on user decision
422 making.

423

424 Finally, if a device contains multiple connected applications with separate interfaces, the device
425 description should address all applications in the device. For example, if there is an application
426 for patients, an application for caregivers, and a data portal for healthcare providers, the device
427 description should include details on all functions across the applications and address how they
428 are connected. Sponsors may also wish to consider enhancing the device description with the use
429 of graphics, diagrams, illustrations, screen captured images, or video demonstrations, including

Contains Nonbinding Recommendations

Draft – Not for Implementation

430 screen captured video. For more information on how to share elements of the user interface in
431 the marketing submission, see Section VI.A (User Interface).

432

433 *Where sponsors should provide it in a submission:* The AI-enabled device description
434 information should be included in the “**Device Description**” section of the marketing
435 submission.

436

437 **VI. User Interface and Labeling**

438 The user interface includes all points of interaction between the user and the device, including all
439 elements of the device with which the user interacts (e.g., those parts of the device that users see,
440 hear, touch). It also includes all sources of information transmitted by the device (including
441 packaging and labeling²⁰), training, and all physical controls and display elements (including
442 alarms and the logic of operation of each device component and of the user interface system as a
443 whole), as applicable. A user interface might be used throughout many phases of installation and
444 use, such as while the user sets up the device (e.g., unpacking, set up, calibration), uses the
445 device, or performs maintenance on the device (e.g., cleaning, replacing a battery, repairing
446 parts).²¹ One way to help support the safety and effectiveness of the device for users is to design
447 the user interface such that important information is provided throughout the course of use, to
448 ensure that the device conforms to defined user needs.²² An approach that integrates important
449 information throughout the user interface may help ensure that device users have access to
450 information at the right time and in the right location to support safe and effective use, consistent
451 with the intended use of the device. For software or mobile applications, manufacturers may
452 leverage the user interface elements, such as information on the screen or alerts sent to other
453 products, in addition to device labeling, to communicate risks about the device so that the
454 necessary information is provided at the right time.

455

456 It is important to provide a holistic understanding of the user interface in a marketing submission
457 to support the agency’s understanding of how the device works. If a sponsor references the user
458 interface design in their risk analysis or another section of the submission to control risks,
459 inclusion of the user interface may also support explanations of those risk controls. However, the
460 actual analysis of the efficacy of risk control should be located separately from the description of
461 the user interface. Further information on this topic is described in Section VII (Risk
462 Assessment) and Appendix D (Usability Evaluation Considerations).

463

464 With regard to labeling specifically, a device user interface includes, but is not limited to,
465 labeling. Further, within the user interface, labeling is subject to specific regulations. For
466 example, depending on whether the device is for prescription-use or not, manufacturers are
467 required to provide labeling containing adequate directions for use that would ensure that a
468 layman or, for prescription devices, a practitioner licensed by law to administer the device, “can

²⁰ See section 201(m) of the FD&C Act which defines labeling as “all labels and other written, printed, or graphic matter (1) upon any article or any of its containers or wrappers, or (2) accompanying such article.” 21 U.S.C. § 321(m).

²¹ See FDA’s guidance titled “[Applying Human Factors and Usability Engineering to Medical Devices](#).”

²² See 21 CFR 820.30(g).

Contains Nonbinding Recommendations

Draft – Not for Implementation

469 use a device safely and for the purposes for which it is intended.”²³ One way to satisfy these
470 requirements for AI-enabled devices could be to provide, in the labeling, clear information about
471 the model, its performance characteristics, and how the model is integrated into the device. For
472 example, users may need to know specific information about the model, such as the nature of the
473 data on which the model was trained. These technical characteristics can be critical to the safe
474 and effective use of the device because they can support a user’s understanding of how the
475 device should be expected to perform, and what factors may impact performance.

476
477 The following sections further detail recommended information on the user interface (Section
478 VI.A), and the labeling (Section VI.B), that should be provided in a marketing submission to
479 support FDA’s understanding of what is communicated to users and the elements of the device
480 with which the users interact.

481
482 Appendix B (Transparency Design Considerations) of this guidance outlines a recommended
483 approach to transparency, including examples of types of information, modes of communication,
484 and communication styles that may be helpful to consider when designing the user interface
485 (including labeling) of an AI-enabled medical device. It may also be helpful to integrate a model
486 card in the device labeling to clearly communicate information about an AI-enabled device (see
487 Appendix E (Example Model Card)).

488
489 Note that inclusion of a unique device identifier (UDI) in the labeling is required for devices,
490 including AI-enabled devices, that are subject to UDI requirements.²⁴ A new UDI is required
491 when there is a new version and/or model, and for new device packages.²⁵ See FDA’s website on
492 [Unique Device Identification System](#) for more information.

493

A. User Interface

494
495 *Why should it be included in a submission for an AI-enabled device:* It is important for FDA to
496 understand the device’s user interface, in order to understand how the device is used. The user
497 interface can convey important information about what the device is intended to do, and how
498 users are intended to interact with it. Seeing the user interface can help FDA understand how the
499 device will be operated and how it will fit into the clinical workflow, which can support the
500 review of a device and help the agency determine whether it is safe and effective.

501
502 A representation of the user interface can also serve to support the sponsor’s risk assessment and
503 other documentation when the user interface is referenced as an element of those sections. For
504 example, the user interface can communicate important information to users that supports safe
505 and effective use of the device, and the user interface design may play a crucial role in
506 controlling or eliminating risks associated with not knowing or misunderstanding information
507 that is critical to the safe and effective use of the device. While not required, if a sponsor chooses

²³ See 21 CFR 801.5; 21 CFR 801.109(d); FD&C Act section 502(f), 21 U.S.C. § 352(f). Device labeling must comply with the requirements in 21 CFR part 801 and any device specific labeling requirements such as for hearing aids or in special controls.

²⁴ See 21 CFR 801.20.

²⁵ See 21 CFR 830.50.

Contains Nonbinding Recommendations

Draft – Not for Implementation

508 to use elements of the user interface as part of risk control in the risk assessment, the inclusion of
509 the user interface can help further facilitate review. Further information on this topic is described
510 in Section VII. (Risk Assessment) and Appendix D (Usability Evaluation Considerations).

511
512 While the user interface does include the printed labeling (e.g., packaging and user manuals) and
513 all elements of the user interface should be designed to collectively support the user’s
514 understanding of how to use the device, sponsors should submit labeling separately as described
515 in Section VI.B (Labeling). This section describes how sponsors should provide FDA with an
516 understanding of the remaining elements of the user interface.

517
518 *What sponsors should include in a submission:* Sponsors should provide information about and
519 descriptions of the user interface that makes clear the device workflow, including the information
520 that is provided to users, when the information is provided, and how it is presented. Possible
521 methods to provide this type of information about the user interface include:

- 522 • A graphical representation (e.g., photographs, illustrations, wireframes, line drawings) of
523 the device and its user interface. This may include a depiction of the overall device and
524 all components of the user interface with which the user will interact (e.g., display and
525 function screens, alarm speakers, controls).
- 526 • A written description of the device user interface.
- 527 • An overview of the operational sequence of the device and the user’s expected
528 interactions with the user interface. This may include the sequence of user actions
529 performed to use the device and resulting device responses, when appropriate.
- 530 • Examples of the output format, including example reports representing a range of
531 expected outcomes.
- 532 • A demonstration of the device, for example by providing a recorded video.

533
534 *Where sponsors can provide it in a submission:* The user interface information should be
535 included in the “**Software Description**” in the Software Documentation section of the marketing
536 submission.

537

538 **B. Labeling**

539 *Why should it be included in a submission for an AI-enabled device:* A marketing submission
540 must include labeling information in sufficient detail to help FDA determine that the proposed
541 labeling satisfies applicable requirements for the type of marketing submission.²⁶ Device labeling
542 must satisfy all applicable FDA labeling requirements, including, but not limited to, 21 CFR Part
543 801, as discussed above.²⁷ This section of the guidance includes labeling considerations for AI-
544 enabled devices to support compliance with these requirements.

545
546 *What sponsors should include in a submission:* The labeling for an AI-enabled device should
547 address the following types of information in a format and at a reading level that is appropriate
548 for the intended user (e.g., considering characteristics such as age, education or literacy level,

²⁶ See e.g., 21 CFR 807.87(e) or 21 CFR 814.20(b)(10).

²⁷ Generally, if the device is an in vitro diagnostic device, the labeling must also satisfy the requirements of 21 CFR 809.10.

Contains Nonbinding Recommendations

Draft – Not for Implementation

549 sensory or physical impairments, or occupational specialty) to help ensure users can quickly
550 access important information. Tables and graphics may be used to communicate this information.

551

Inclusion of AI

552

- Statement that AI is used in the device.
- Explanation of how AI is used to achieve the device’s intended use.
 - For devices with multiple functions, this explanation may include how AI-DSFs interact with each other as well as how they interact with non-AI DSFs.

556

Model Input

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

- Description of the model inputs (e.g., signals or patterns acquired from other compatible devices, images from an acquisition system (e.g., MRI), or patient-derived samples, which can be input manually or automatically). Related aspects to consider include:
 - For systems incorporating inputs from an electronic interface, information on the necessary system configuration to ensure the inputs are consistent with the design and validation of the AI-enabled device.²⁸
 - For systems that require input from other medical devices (e.g., an x-ray device), a list of the specific compatible devices or device specification, along with the acceptable acquisition protocols, as applicable.
 - For systems in which the loss of model inputs may prevent the AI-enabled device from generating an output, an explanation of the potential impact of the lost inputs on the performance of the AI-enabled device.
- Instructions on any steps the user is expected to take to prepare input data for processing by the device, including any expected characteristics (e.g., functional capabilities, experience and knowledge levels, and level of training) of those performing these steps. This information should be consistent with the intended use that was studied in the device validation.

577

Model Output

- Explanation of what the model output means and how it is intended to be used.

578

579

580

Automation

- Explanation of the intended degree of automation the device exhibits.

581

582

583

Model Architecture

- High level description of the methods and architecture used to develop the model(s) implemented in the device.

584

585

586

587

Model Development Data

- Description of the development data, including:
 - The source(s) of data;
 - Study sites;

588

589

590

²⁸ For more information, please see FDA guidance titled “[Design Considerations and Premarket Submission Recommendations for Interoperable Medical Devices.](#)”

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 591 ○ Sample size;
- 592 ○ Demographic distributions; and
- 593 ○ Criteria/expertise used for determining clinical reference standard (ground truth).
- 594

Performance Data

- 595 ● Description of the performance validation data, including:
 - 596 ○ The source(s) of data;
 - 597 ○ Study sites;
 - 598 ○ Sample size;
 - 600 ○ Other important study design and data structure information (e.g., randomization
 - 601 schemes, repeated measurements, clinical reference standard);
 - 602 ○ Primary endpoints of the validation study, including pre-specified performance
 - 603 criteria; and
 - 604 ○ Criteria/expertise used for determining clinical reference standard data.
 - 605

Device Performance Metrics

- 606 ● Description of the device performance metrics.
 - 607 ○ An example of performance metrics may include metrics such as the area under
 - 608 the receiver operating characteristic curve (AUROC), sensitivity and specificity,
 - 609 true/false positive and true/false negative counts (e.g., in a confusion matrix),
 - 610 positive/negative predictive values (PPV/NPV), and positive/negative diagnostic
 - 611 likelihood ratios (PLR/NLR). All performance estimates should be provided with
 - 612 confidence intervals.
- 614 ● Explanation of the device performance across important subgroups. Generally, subgroup
- 615 analysis by patient characteristics (e.g., sex,²⁹ gender, age, race, ethnicity,³⁰ disease
- 616 severity), geographic sites, and data collection equipment are appropriate.
- 617 ● Description of the corresponding performance for different operating points, including
- 618 subgroup analysis for each operating point, as applicable.
- 619

Performance Monitoring

- 621 ● Description of any methods or tools to monitor and manage device performance,
- 622 including instructions for the use of such tools, as applicable when ongoing performance
- 623 monitoring and management by the user is considered necessary for the safe and effective
- 624 use of the device.
- 625

Limitations

- 626 ● Description of all known limitations of the AI-enabled device, AI-DSF(s), or model(s).
 - 627 ○ Some limitations of a model may not reach the degree of severity that would
 - 628 warrant a contraindication, warning, or precaution, but they may still be important
 - 629 to include in labeling. For example, the training dataset may have only included a
 - 630 few patients with a rare presentation of a disease or condition; users may benefit
 - 631

²⁹ For more information regarding sex-specific data, please see FDA guidance titled "[Evaluation of Sex-Specific Data in Medical Device Clinical Studies.](#)"

³⁰ For more information regarding the reporting of age, race, and ethnicity related data, please see FDA guidance titled "[Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies.](#)"

Contains Nonbinding Recommendations

Draft – Not for Implementation

632 from knowing the limitations of the data when that rare presentation is suggested
633 by the model as a diagnosis.

634

Installation and Use

- 636 • Information about the installation and implementation instructions, including:
 - 637 ○ Instructions on integrating the AI-enabled device into the site’s data systems and
 - 638 clinical workflow; and
 - 639 ○ Instructions for ensuring that any input data are compatible and appropriate for
 - 640 the device.³¹
 - 641 ▪ Terms may need to be explicitly defined. For example, a healthcare
 - 642 system and a manufacturer may both have data labeled as “sex,” but one
 - 643 may be using sex at birth while the other may be using self-reported sex.
 - 644

644

Customization

- 646 • Description of and instructions on any customizable features, including:
 - 647 ○ When users or healthcare systems can configure the operating points for the
 - 648 device;
 - 649 ○ When it is appropriate to select different configurations; and
 - 650 ○ When operating points are configurable, how end users can discern the operating
 - 651 point the device is currently operating at.
 - 652

652

Metrics and Visualizations

- 654 • Explanation of any additional metrics or visualizations used to add context to the model
- 655 output.
- 656

656

Patient and Caregiver Information

658 For AI-enabled devices intended for use by patients or caregivers, manufacturers should provide
659 labeling material that is designed for patients and caregivers describing the instructions for use,
660 the device’s indication, intended use, risks, and limitations. Patients and caregivers are
661 considered users if they will operate the device, interpret the outcome, or make decisions based
662 on the outcome, even if they are not the only user or the primary operator of the device. This
663 material should be at an appropriate reading level for the intended audience. If patient and
664 caregiver-specific material is not provided, sponsors should provide an explanation of how
665 patients and caregivers will understand how to use the device, including how to make decisions
666 about whether to use the device and how to use the output of the device.

667

668 *Where sponsors should provide it in a submission:* Information regarding the AI-enabled device
669 labeling should be included in the “**Labeling**” section of the marketing submission.

670

³¹ For more information, please see FDA guidance titled, “[Design Considerations and Pre-market Submission Recommendations for Interoperable Medical Devices.](#)”

Contains Nonbinding Recommendations

Draft – Not for Implementation

ADDITIONAL RESOURCES:

- Appendix B (Transparency Design Considerations) outlines a potential approach to understanding a device’s indications for use and a model card, which may aid in the development of the user interface.
- While model cards are not required for presenting information about the labeling or user interface, they may be a helpful tool to organize information. In general, model cards can be adapted to the specific needs and context of each AI-enabled device.
 - Appendix E (Example Model Card) includes an example of a basic model card format intended for users and healthcare providers that conveys information including a summary of the model’s intended use and intended users, and evidence supporting safety and effectiveness.
 - Appendix F (Example 510(k) Submission Summary with Model Card) includes an example of a completed basic model card.
- FDA’s guidance titled “[Device Labeling Guidance #G91-1 \(Blue Book Memo\)](#)” includes suggestions regarding what information should be included within device labeling.

671

672

VII. Risk Assessment

673 *Why should it be included in a submission for an AI-enabled device:* A comprehensive risk
674 assessment helps ensure the device is safe and effective. When included in a marketing
675 submission, a comprehensive risk assessment helps FDA understand whether appropriate risks
676 have been identified and how they are controlled. In Section VI.C of the [Premarket Software](#)
677 [Guidance](#), FDA recommends that marketing submissions that include device software functions
678 include a risk management file composed of a risk management plan, a risk assessment, and a
679 risk management report. Consistent with this, marketing submissions of AI-enabled devices
680 should include a risk management file that takes into account the recommendations of [Premarket](#)
681 [Software Guidance](#) and the recommendations of this guidance, in addition to any other
682 applicable guidance.

683

684 Sponsors should also refer to the FDA-recognized version of ANSI/AAMI/ISO 14971 *Medical*
685 *devices - Applications of risk management to medical devices* for additional information on the
686 development and application of a risk management file, which is also applicable to AI-enabled
687 devices. FDA also recognizes that AI-enabled devices can be associated with new or different
688 risks than device software functions generally. Therefore, FDA also recommends that sponsors
689 incorporate the considerations outlined in the FDA-recognized voluntary consensus standard of
690 AAMI CR34971 *Guidance on the Application of ISO 14971 to Artificial Intelligence and*
691 *Machine Learning*, which is specific to AI-enabled devices.

692

Risks Across the TPLC

694 When conducting a risk analysis, the Medical Devices; Current Good Manufacturing Practice
695 (CGMP), final rule (Oct. 7, 1996, [61 FR 52602](#)) states “manufacturers are expected to identify
696 possible hazards associated with the design in both normal and fault conditions. The risks
697 associated with the hazards, including those resulting from user error, should then be calculated
698 in both normal and fault conditions. If any risk is judged unacceptable, it should be reduced to
699 acceptable levels by the appropriate means.” This risk assessment should take into account all
700 users, as described in Section VI (User Interface and Labeling) of this guidance, across the

Contains Nonbinding Recommendations

Draft – Not for Implementation

701 TPLC. FDA recommends that manufacturers follow this approach for AI-enabled devices across
702 their TPLC.

703

704 Risks Related to Information in AI-Enabled Devices

705 One aspect of risk management that can be particularly important for AI-enabled devices is the
706 management of risks that are related to understanding information that is necessary to use or
707 interpret the device, including risks related to lack of information or unclear information.
708 Misunderstood, misused, or unavailable information can impact the safe and effective use of a
709 device. For example, for devices that utilize complex algorithms, including AI-enabled devices,
710 the performance in different disease subtypes may not be apparent to users, or the logic
711 underlying the output information may not be easily understandable, which can negatively affect
712 user understanding and use of the device. Lack of, or unclear information can also make it
713 difficult for different users to understand whether a device is not performing as expected, or how
714 to correctly follow instructions. FDA recommends that consideration of risks related to
715 understanding information should be one part of a comprehensive approach to risk management
716 for an AI-enabled device.

717

ADDITIONAL RESOURCES:

- ANSI/AAMI HE75 *Human factors engineering - Design of medical devices* includes recommendations on using information in labeling to help control risks.

718

719 *What sponsors should include in a submission:* Sponsors should provide a “Risk Management
720 File” that includes a risk management plan, including a risk assessment. In addition to other
721 considerations, the risk assessment should consider user tasks and knowledge tasks that occur
722 throughout the full continuum of use of the device, including, for example, the process of
723 installing the device, maintaining performance over time, and any risks associated with user
724 interpretation of the results of a device, as appropriate.

725 In addition to the considerations provided in FDA-recognized voluntary consensus standards³²
726 and applicable guidances,³³ FDA recommends that sponsors consider the risks related to
727 understanding information during the risk assessment. As with all identified risks, sponsors
728 should provide an explanation of any risk controls, including elements of the user interface, such
729 as labeling, that address the identified risks. Information that may be helpful to discuss such risks
730 and their controls, as applicable, is provided in Appendix D (Usability Evaluation
731 Considerations).

732

733 *Where sponsors should provide it in a submission:* Much of the information on risk assessment
734 for an AI-enabled device should be included in the “**Risk Management File**” in the Software
735 Documentation section of the marketing submission, as recommended by the [Premarket
736 Software Guidance](#).

³² For more information, see the [FDA Recognized Consensus Standards Database](#).

³³ For more information regarding use of consensus standards in regulatory submissions, refer to the FDA guidances titled “[Appropriate Use of Voluntary Consensus Standards in Premarket Submissions for Medical Devices](#)” and “[Standards Development and the Use of Standards in Regulatory Submissions Reviewed in the Center for Biologics Evaluation and Research](#).”

737

ADDITIONAL RESOURCES:

- Appendix B (Transparency Design Considerations) outlines recommendations for a user-centered design approach to developing a device, which may aid in the identification of risks and development of risk controls.
- Appendix D (Usability Evaluation Considerations) provides recommendations on usability testing, which may help sponsors evaluate the efficacy of proposed controls for information related risks.

738

739 **VIII. Data Management**

740 *Why should it be included in a submission for an AI-enabled device:* For an AI-enabled device,
741 the model is part of the mechanism of action. Therefore, a clear explanation of the data
742 management, including data management practices (i.e., how data has been or will be collected,
743 processed, annotated, stored, controlled, and used) and characterization of data used in the
744 development and validation of the AI-enabled device is critical for FDA to understand how the
745 device was developed and validated. This understanding helps to enable FDA’s evaluation of an
746 AI-enabled device’s safety and effectiveness.

747

748 The performance and behavior of AI systems rely heavily on the quality, diversity, and quantity
749 of data used to train and tune them. The accuracy and usefulness of a validation of an AI-enabled
750 device also depends on the quality, diversity, and quantity of data used to test it. Thus, FDA
751 reviewers evaluate data management in order to understand whether an AI-enabled device is safe
752 and effective. This includes the alignment of the collection and management of training and test
753 data with the intended use and resulting device requirements.

754

755 Data management is also an important means of identifying and mitigating bias. The
756 characterization of sources of bias is necessary to assess the potential for AI bias in the AI-
757 enabled device. AI bias is a potential tendency to produce incorrect results in a systematic, but
758 sometimes unforeseeable, way due to limitations in the training data or erroneous assumptions in
759 the machine learning process. AI bias has been well-documented.³⁴ For example, during training,
760 models can be over-trained to recognize features of images that are unique to specific scanners,
761 patient subpopulations, or clinical sites but have little to do with generalizable patient anatomy,
762 physiology, or condition, which can lead to AI bias in the resulting model. In another example,
763 underrepresentation of certain populations in datasets could lead to overfitting (i.e., data fitting
764 too closely to the potential biases of the training data) based on demographic characteristics,
765 which can impact the AI-enabled device performance in the underrepresented population.

766

767 Using unbiased, representative training data for models promotes generalizability to the intended
768 use population and avoids perpetuating biases or idiosyncrasies from the data itself. For example,
769 in image recognition tasks, confounding may occur when all the diseased cases are imaged with
770 the same instrument, or with a ruler included (e.g., on clinical images of melanoma). Another
771 example of a potential confounding factor is the use of data collected outside the U.S. (OUS) in

³⁴ See Karen Hao, “This is how AI bias really happens—and why it’s so hard to fix,” MIT Technology Review 2019.

Contains Nonbinding Recommendations

Draft – Not for Implementation

772 training, which may bias the model if the OUS population does not reflect the U.S. population
773 due to differences in demographics, practice of medicine, or standard of care. Such confounders
774 in the training data, if not identified and mitigated, can be inadvertently learned by a model,
775 leading to seemingly accurate (but misleading) predictions based on irrelevant characteristics.
776

777 The inclusion of representative data in validation datasets may be important, because
778 underrepresentation may impact the ability to identify any performance problems, including
779 understanding performance in underrepresented populations. Although bias may be difficult to
780 eliminate completely, FDA recommends that manufacturers, as a starting point, ensure that the
781 validation data sufficiently represents the intended use (target) population of a medical device.
782 For more information regarding age-, race-, ethnicity-, and sex-specific data please see the FDA
783 guidances titled, “[Collection of Race and Ethnicity Data in Clinical Trials and Clinical Studies
784 for FDA-Regulated Medical Products](#)”³⁵ “[Evaluation and Reporting of Age-, Race-, and
785 Ethnicity-Specific Data in Medical Device Clinical Studies](#),” and “[Evaluation of Sex-Specific
786 Data in Medical Device Clinical Studies](#).”
787

788 If the same confounders are found in the validation data as the development data, it may be
789 particularly difficult to identify the spurious correlations that appear to be leading to correct
790 predictions. Therefore, information about the representativeness of the datasets used in the
791 development and validation of the AI-enabled device is important to help FDA determine
792 substantial equivalence or if there is a reasonable assurance that the device is safe and effective
793 for its intended use.³⁶ Beyond addressing AI bias, the details of the data management should
794 support the intended use of the device.
795

796 To objectively assess the device performance, it is also important for FDA reviewers to
797 understand whether the test data are independent (e.g., sampled from completely different
798 clinical sites) from the training data and are sequestered from the model developers and the
799 model development stage. Appropriate separation of the development and test datasets can help
800 with evaluating the true performance of an AI-enabled device. Data leakage between the
801 validation and development datasets can create uncertainty regarding the true performance of the
802 AI-enabled device.³⁷
803

804 *What sponsors should include in a submission:* In a submission, a sponsor should provide the
805 following types of information for both the training and testing data, in the appropriate marketing
806 submission sections. It may be helpful to organize data management information by the sections
807 described below. Generally, information on data collection, development and test data
808 independence, reference standards, and representativeness should be provided. Sponsors should
809 also explain any differences in the data management approach and the characteristics of the data

³⁵ When final, this guidance will represent FDA’s current thinking on this topic.

³⁶ For more information, see FDA guidance titled “[Acceptance of Clinical Data to Support Medical Device Applications and Submissions: Frequently Asked Questions](#).”

³⁷ Robert F Wolff, Karel GM Moons, Richard D Riley, et al. “PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies.” *Annals of Internal Medicine* 170, no. 1 (2019): 51–58 <https://doi.org/10.7326/M18-1376>; Altman, Douglas G., and Patrick Royston. “What Do We Mean by Validating a Prognostic Model?” *Statistics in Medicine* 19, no. 4 (2000): 453–73 [https://doi.org/10.1002/\(sici\)1097-0258\(20000229\)19:4<453::aid-sim350>3.0.co;2-5](https://doi.org/10.1002/(sici)1097-0258(20000229)19:4<453::aid-sim350>3.0.co;2-5).

Contains Nonbinding Recommendations

Draft – Not for Implementation

810 between the development and validation phases. The submission should include an explanation
811 for the differences and justification for them.

812

813 Data Collection

- 814 • A description of how data were collected (e.g., clinical study protocols with
815 inclusion/exclusion criteria), including:
 - 816 ○ The names of clinical sites or institutions involved.
 - 817 ■ Sites should be uniquely identified, and they should be referred to
818 consistently throughout the submission.
 - 819 ○ The time period during which the data were acquired.
 - 820 ○ If data were used from a pre-existing database, the appropriateness of the use of
821 this database.
 - 822 ○ If real-world data (RWD) are used, the source and collection of this evidence.
 - 823 ■ If RWD are used, FDA recommends that sponsors provide an assessment
824 of fit-for-purpose data for the selected data source(s) that evaluates both
825 the relevance and reliability of the RWD. FDA encourages sponsors to
826 leverage the [Q-Submission Program](#) for obtaining FDA feedback on
827 proposed uses of RWD. For more information regarding RWD, please see
828 the FDA guidance titled “[Use of Real-World Evidence to Support
829 Regulatory Decision-Making for Medical Devices.](#)”
- 830 • A description of the limitations of the dataset.
- 831 • A description of the quality assurance processes related to the data, including the controls
832 that were put in place to protect from human error during data acquisition, when
833 applicable.
- 834 • A description of the size of each data set.
- 835 • A description of the mechanisms used to improve diversity in enrollment within the
836 scope of the study, and how they ensure the generalizability of study results across
837 patient populations and clinical sites.³⁸ For more information on this topic, please see
838 FDA guidance titled “[Collection of Race and Ethnicity Data in Clinical Trials.](#)”
- 839 • A description of the use of synthetic data.³⁹ Synthetic data used in support of a regulatory
840 submission should be accompanied by a comprehensive explanation of how the data were
841 generated and why they are fit-for-purpose.

842

843 Data Cleaning/Processing

844 To provide optimum training results, it may be important to clean data used for development,
845 such as by removing incorrect, duplicate, or incomplete data. These processing steps should be

³⁸ Sponsors may be required to develop or submit information regarding the enrollment of clinical study participants to help improve the strength and generalizability of the study results. For example, the FD&C Act, as amended by section 3601(b) of the Food and Drug Omnibus Reform Act of 2022 (FDORA), enacted as part of the Consolidated Appropriations Act, 2023 (P.L. 117-328)), requires sponsors to submit to FDA diversity action plans for studies of certain devices. See FD&C Act section 520(g)(9), 21 U.S.C. § 360j(g)(9).

³⁹ For the purposes of this guidance, “synthetic data” is defined as data that have been created artificially (e.g., through statistical modeling, computer simulation) so that new values and/or data elements are generated. Generally, synthetic data are intended to represent the structure, properties and relationships seen in actual patient data, except that they do not contain any real or specific information about individuals. For more information, please see [FDA Digital Health and Artificial Intelligence Glossary – Educational Resource | FDA.](#)

Contains Nonbinding Recommendations

Draft – Not for Implementation

846 described, including data quality factors used, data inclusion/exclusion criteria, treatment of
847 missing data, and whether the steps are internal or external to the AI-DSF.

848
849 Testing data, on the other hand, should only be processed in a manner that is representative of
850 the RWD the model will encounter in its intended use. Any such data processing, data quality
851 factors used, data inclusion/exclusion criteria, and treatment of missing data should be justified
852 as aligned with pre-processing implemented in the final AI-DSF.

853

854 Reference Standard

855 For the purposes of this guidance, a reference standard is the best available representative truth
856 that can be used to define the true condition for each patient/case/record.⁴⁰ It is possible that a
857 reference standard may be used in device training, device validation, or both. A reference
858 standard is validated by evidence from current practice within the medical and regulatory
859 communities for establishing a patient’s true status with respect to a clinical task. The reference
860 standard should reflect the clinical task. Clinical tasks may consist of, for example, classification
861 of a disease or condition, segmentation of contours on medical images, detection by bounding
862 boxes, or localization by markings. The following types of information should be provided
863 regarding the selected reference standard:

- 864 • A description of how the reference standard was established.
- 865 • A description of the uncertainty inherent in the selected reference standard.
- 866 • A description of the strategy for addressing cases where results obtained using a
867 reference standard may be equivocal or missing.
- 868 • If the reference standard is based on evaluations from clinicians, provide:
 - 869 ○ The grading protocol used.
 - 870 ○ What data are provided to these clinicians.
 - 871 ○ How the clinicians’ evaluations are collected/adjudicated for determining the
872 clinical reference standard, including:
 - 873 ■ blinding protocol; and
 - 874 ■ number of participating clinicians and their qualifications.
 - 875 ○ An assessment of the intra- and/or inter-clinician variability for each task, as
876 applicable, as well as an assessment on whether the observed variability is within
877 commonly accepted standards for a particular measurement task.

878 Data Annotation

- 879 • When data annotation is used, the following types of information should be provided
880 regarding the data annotation approach:
 - 881 ○ A description of the expertise of those performing the data annotation.
 - 882 ○ A description of the specific training, instructions or guidelines provided to data
883 annotators to guide their annotation decisions, including whether annotators are
884 blinded to each other.

⁴⁰ For an illustrative example of a reference standard, see FDA guidance titled “[Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data in Premarket Notification \(510\(k\)\) Submissions.](#)” This guidance addresses the reference standard for this device. Other device specific guidances and special controls note the appropriate reference standard to be used. For questions about what the appropriate reference standard may be for a device and proposed intended use, consult the appropriate review division via the Q-Submission Program.

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 885 ○ A description of the methods for evaluating quality/consistency of data
886 annotations and adjudicating disagreements (consensus evaluation, sampling).
887 FDA recommends the use of independent assessments by each annotator, without
888 knowledge of the other annotators' decisions, to ensure objective high-quality
889 data annotations; and
890 ○ A detailed plan for addressing incorrect data annotation.

891 892 Data Storage

893 A description of the data storage of both training and test data. The description should address
894 dataset version control and should ensure the security of the data by addressing the items
895 described in Section XII (Cybersecurity) of this guidance.

896 897 Management and Independence of Data

- 898 ● A description of the development data, including how the development data were split
899 into training, tuning, tuning evaluation, and any additional subsets, and specification of
900 which model development activities were performed using each dataset.
901 ● A description of the controls in place to ensure the data used for testing is sequestered
902 from the development process.
903 ● A justification of why the data used for validation provides a robust external validation.
904 For example, a description of the sites from which test data originates from, because, in
905 general, test data should come from sites different from those used to develop the AI-
906 DSF.

907 908 Representativeness

- 909 ● An explanation of how the data is representative of the intended use population⁴¹ and
910 indications for use, including:
911 ○ A description of the relevant population characteristics, when available,
912 including:
913 ■ Disease conditions (e.g., positive/negative cases, disease severity, disease
914 subtype, comorbidities, distribution of the disease spectrum);
915 ■ Patient population demographics (e.g., sex,⁴² gender, age, race, ethnicity,⁴³
916 height, weight);
917 ■ Data acquisition equipment and conditions (e.g., locations at which data
918 are collected, data acquisition devices/methods, imaging and
919 reconstruction protocols), including any factors that may impact signals
920 analyzed during data acquisition (e.g., patient activities, such as whether a

⁴¹ Sponsors may be required to develop or submit information regarding the representativeness of clinical study participants. For example, the FD&C Act, as amended by section 3601(b) of FDORA, enacted as part of the Consolidated Appropriations Act, 2023 (P.L. 117-328)), requires sponsors to submit to FDA diversity action plans for studies of certain devices. See section 520(g)(9) of the FD&C Act, 21 U.S.C. § 360j(g)(9).

⁴² For more information regarding sex-specific data, please see FDA guidance titled "[Evaluation of Sex-Specific Data in Medical Device Clinical Studies.](#)"

⁴³ For more information regarding age-, race-, and ethnicity-related data, please see FDA guidances titled "[Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies.](#)" and "[Collection of Race and Ethnicity Data in Clinical Trials.](#)"

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 921 patient is ambulatory, resting, standing; or data acquisition environments,
922 such as intensive care unit, MRI); and
- 923 ■ Test data collection sites (e.g., clinical sites, institutions). Generally, while
924 a single data collection site may be a useful starting place during initial
925 data assessment phases, reliance on a single site is generally not
926 appropriate for understanding whether the data are representative of the
927 intended use population and indications for use. The use of multiple data
928 collection sites, such as sites in diverse clinical practice settings (e.g.,
929 large academic hospital vs. community hospital) may assure a more
930 representative sample of the intended use population. For example, the use
931 of at least three geographically diverse US clinical sites (or health care
932 systems) may be appropriate to clinically validate an AI-enabled device.⁴⁴
 - 933 ○ A characterization of the distribution of data along important covariates, including
934 those corresponding to the population characteristics described above.
 - 935 ○ If any of the relevant population characteristics above were not available for the
936 data, an explanation of why, and a justification of the use of the data without this
937 information. FDA understands that, depending on the source of the patients and/or
938 samples used in the training and test data, some relevant patient characteristic
939 information may not be available.
 - 940 ○ A subgroup analysis or analyses stratified by the identified covariates.
 - 941 ○ If OUS data are used during validation, an explanation regarding how the data
942 compares to the U.S. population and U.S. medical practice in terms of general
943 medical practice, disease presentation, prevalence, and progression as well as the
944 demographic characteristics of patients.⁴⁵
 - 945 ■ Due to the data-driven nature of typical models and the obscurity of their
946 algorithms to end users, their generalized performance on the U.S. target
947 population may not be adequately captured in the clinical study if a
948 significant portion of the validation data are OUS data. AI-enabled devices
949 may also be more sensitive than traditional medical devices to the
950 idiosyncratic patterns in the training or test data. For these reasons, they
951 may require higher proportion of U.S. data in the clinical validation. FDA
952 encourages sponsors to leverage the Q-Submission process for obtaining
953 FDA feedback on proposed uses of OUS data.⁴⁶

954 *Where sponsors should provide it in a submission:* The data management information for data
955 used in the development of the model should be included in the “**Software Description**” in the
956 Software Documentation section of the marketing submission, as described in the [Premarket
957 Software Guidance](#).

958
959 The data management information for data used in the performance validation (i.e., clinical

⁴⁴ For more information regarding site selection, please see FDA guidance titled “[Design Considerations for Pivotal Clinical Investigations for Medical Devices](#).”

⁴⁵ For more information on the use of OUS data, please see FDA guidance titled “[Acceptance of Clinical Data to Support Medical Device Applications and Submissions: Frequently Asked Questions](#).”

⁴⁶ For more information on the Q-Submission program, please see FDA guidance titled “[Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program](#).”

Contains Nonbinding Recommendations

Draft – Not for Implementation

960 validation) documentation should be included in the “**Performance Testing**” section of the
961 marketing submission. When the characteristics of data used for model training and validation
962 differ, sponsors should highlight and justify the differences along with the performance
963 validation data management section in the performance testing documentation element.
964

ADDITIONAL RESOURCES: In addition to the considerations in this guidance, to support the TPLC approach to development, FDA recommends that sponsors and investigators consider the unique characteristics of the AI-enabled device during the study design, conduct, and reporting phases for clinical investigations. Researchers should understand how Investigational Device Exemption (IDE), Protection of Human Subjects and Institutional Review Board regulations,⁴⁷ and Good Clinical Practice (GCP) regulations⁴⁸ apply to their devices. Resources include consensus guidelines,⁴⁹ as well as FDA guidances titled:

- “[Significant Risk and Nonsignificant Risk Medical Device Studies](#)”
- “[Informed Consent Guidance for IRBs, Clinical Investigators, and Sponsors](#)”
- “[Acceptance of Clinical Data to Support Medical Device Applications and Submissions: Frequently Asked Questions](#)”

For more information regarding age-, race-, and ethnicity-specific data, and sex-specific data please see the FDA guidances titled:

- “[Collection of Race and Ethnicity Data in Clinical Trials](#)”
- “[Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies](#)”
- “[Evaluation of Sex-Specific Data in Medical Device Clinical Studies](#)”

965

966 IX. Model Description and Development

967 *Why should it be included in a submission for an AI-enabled device:* Information about the
968 model (and device) design, including its biases and limitations, supports FDA’s ability to assess
969 the safety and effectiveness of an AI-enabled device and determine the device’s performance
970 testing specifications.

971
972 Section VI.B of the [Premarket Software Guidance](#) describes information that should be included
973 as part of a software description in a marketing submission, including the model description.
974 Whereas the device description is broader and provides information about the whole device, how
975 users interact with it, and how it fits into the clinical workflow, the model description, as part of
976 the software description, specifically provides detailed information about the technical
977 characteristics of the model(s) themselves and the algorithms and methods that were used in their
978 development. This information helps FDA understand the basis for the functionality of an AI-
979 enabled device. Understanding the methods used to develop the model also helps FDA identify

⁴⁷ See 21 CFR Parts 50 and 56.

⁴⁸ See FDA’s website on [Regulations: Good Clinical Practice and Clinical Trials](#).

⁴⁹ See Liu, Xiaoxuan et al “Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension,” *Natural Medicine* (2020) <https://doi.org/10.1038/s41591-020-1034-x>; Rivera, Samantha C. et al “Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension,” *Lancet* (2020) [https://doi.org/10.1016/S2589-7500\(20\)30219-3](https://doi.org/10.1016/S2589-7500(20)30219-3); Vasey, Baptiste et al Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI,” *BMJ* (2022) <https://doi.org/10.1136/bmj-2022-070904>.

Contains Nonbinding Recommendations

Draft – Not for Implementation

980 potential limitations, sources of AI bias, and considerations for appropriate device labeling.

981

982 *What sponsors should include in a submission:* In a submission, sponsors should include the
983 information described below for each model in the AI-enabled device.

984

985 In situations where multiple models are employed as part of the AI-enabled device, it can be
986 particularly helpful to include a diagram of how model outputs combine to create the device
987 outputs. The description of the algorithms and models should be sufficiently detailed to enable a
988 competent AI practitioner to produce an equivalent model. The use of diagrams in addition to
989 textual descriptions is encouraged to enhance clarity.

990

Model Description

992 • An explanation of each model used as part of the AI-enabled device, including but not
993 limited to:

- 994 ○ Model inputs and outputs;
- 995 ○ A description of model architecture;
- 996 ○ A description of features;
- 997 ○ A description of the feature selection process and any loss function(s) used for
998 model design and optimization, as appropriate; and
- 999 ○ Model parameters.

1000 • In situations where the AI-enabled device has customizable features involving the model,
1001 such as being customizable to operate at multiple pre-defined operating points or with a
1002 variable number of inputs, a description of the technical elements of the model that allow
1003 for and control customization.

1004 • A description of any quality control criteria or algorithms, including AI-based and third-
1005 party ones, for the input data, including how the quality assessment metrics align with the
1006 intended use of the device (e.g., intended patient population and use environment).

1007 • A description of any methods applied to the input and/or output data, including:
1008 ○ Pre-processing of input data (e.g., normalization);
1009 ○ Post-processing of output data; and
1010 ○ Data augmentation or synthesis.

1011

Model Development

1013 • A description of how the model was trained, including but not limited to:

- 1014 ○ Optimization methods;
- 1015 ○ Training paradigms (e.g., supervised, unsupervised or semi-supervised learning,
1016 federated learning, active learning);
- 1017 ○ Regularization techniques employed;
- 1018 ○ Training hyperparameters (e.g., the loss function learning rate) as applicable; and
- 1019 ○ Summary training performance such as the loss function convergence curves for
1020 the different data subsets (such as training, tuning, tuning evaluation).

1021 • If tuning evaluation was conducted, a description of the metrics and results obtained.

1022 • An explanation of any pre-trained models that were used, as applicable.

- 1023 ○ If a pre-trained model was used, specify the dataset that was used for pre-training
1024 and how the pre-trained model was obtained.

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 1025
- 1026
- 1027
- 1028
- A description of the use of ensemble methods (e.g., bagging or boosting), as applicable.
 - An explanation of how any thresholds (e.g., operating points) were determined.
 - An explanation of any calibration of the model output.

1029 *Where sponsors should provide it in a submission:* Information on model development, including
1030 the model description, and the method for model development, should be included as part of the
1031 “**Software Description**” in the Software Documentation section of the marketing submission, as
1032 described in the [Premarket Software Guidance](#).

1033

ADDITIONAL RESOURCES: In situations where manufacturers wish to consider development of models that automatically or continuously update, FDA encourages manufacturers to use the [Q-Submission Program](#) to discuss considerations related to these AI models early in the development process and review the FDA guidance titled “[Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions](#).”

1034

1035 **X. Validation**

1036 For an AI-enabled device, validation includes ensuring that the device, as utilized by users, will
1037 perform its intended use safely and effectively, as well as establishing that the relevant
1038 performance specifications of the device can be consistently met. For AI-enabled devices,
1039 manufacturers should demonstrate users’ ability to interact with and understand the device as
1040 intended in addition to ensuring the device itself meets relevant performance specifications. To
1041 this end, it can be helpful to consider both performance validation (including human factors
1042 validation) and an evaluation of usability. Note that, for the purposes of this guidance (in the
1043 context of risk controls in the absence of human factors validation), usability describes whether
1044 the device can be used safely and effectively by the intended users, including whether users
1045 consistently and correctly receive, understand, interpret, and apply information related to the AI-
1046 enabled device.

1047

1048 The FDA guidance titled “[Applying Human Factors and Usability Engineering to Medical](#)
1049 [Devices](#)” (hereafter referred to as “[Human Factors Guidance](#)”), describes recommendations and
1050 requirements for devices and establishes that human factors validation testing encompasses, “all
1051 testing conducted at the end of the device development process to assess user interactions with a
1052 device user interface to identify use errors that would or could result in serious harm to the
1053 patient or user,” and is also used “to assess the effectiveness of risk management measures.”
1054 While the [Human Factors Guidance](#) outlines specific recommendations and requirements for
1055 human factors validation for devices that have critical tasks, the application of the same or a
1056 similar process can also be helpful to demonstrate the appropriate control of other risks.
1057 Appendix D (Usability Evaluation Considerations) includes recommendations to help sponsors
1058 understand when usability testing may help support the control of risks. The appendix also
1059 includes recommendation to help sponsors develop and describe certain types of usability testing
1060 in addition to human factors validation, or when human factors validation is not required. The
1061 appendix supplements device-specific recommendations and recommendations provided in the
1062 [Human Factors Guidance](#) where applicable.

1063

Contains Nonbinding Recommendations

Draft – Not for Implementation

1064 Together, performance validation and human factors validation (or an evaluation of usability as
1065 appropriate) help provide FDA with information to understand how the device may be used and
1066 perform under real world circumstances. Performance validation may employ a variety of testing
1067 and monitoring methods to evaluate the statistical performance of the model under testing
1068 conditions, and human factors validation testing involves understanding how various users are
1069 likely to use a device in context. In other words, performance validation is meant to provide
1070 confirmation that device specifications conform to user needs and intended uses, and that
1071 performance requirements implemented can be consistently fulfilled, while human factors
1072 validation and an evaluation of usability are meant to specifically address whether all intended
1073 users can achieve specific goals while using the device and whether users will be able to
1074 consistently interact with the device safely and effectively.

1075
1076 Software Version History

1077 Section VI.I of the [Premarket Software Guidance](#) describes information that should be included
1078 as part of a software description in a marketing submission, including information regarding the
1079 software version history. For AI-enabled devices, the software version history includes
1080 consideration of the model version and any differences between the tested version of the model
1081 and the released version, along with an assessment of the potential effect of the differences on
1082 the safety and effectiveness of the device. It is important for FDA to understand what version of
1083 the model was tested in order to ensure that all validation activities will be objective, and the
1084 model has not been adjusted opportunistically in light of the test data (i.e., post-hoc adjustment)
1085 without the Agency’s concurrence.

1086
1087 New unique device identifiers (UDIs) are required for devices that are required to bear a UDI on
1088 its label when there is a new version and/or model, and for new device packages.⁵⁰

1089 **A. Performance Validation**

1090 *Why should it be included in a submission for an AI-enabled device:* The performance validation
1091 for an AI-enabled device provides objective evidence that the device performs predictably and
1092 reliably in the target population according to its intended use. The following recommendations
1093 are intended to supplement device-specific recommendations and recommendations provided in
1094 other FDA guidances where applicable, including “[Design Considerations for Pivotal Clinical](#)
1095 [Investigations for Medical Devices](#),” “[Statistical Guidance on Reporting Results from Studies](#)
1096 [Evaluating Diagnostic Tests](#),” and “[Electronic Source Data in Clinical Investigations](#).”

1097
1098 As part of FDA’s evaluation of safety and effectiveness of the device, it is important for FDA to
1099 understand how the device performs overall in the intended use population, as well as in
1100 subgroups of interest. Acceptable performance in certain subgroups may mask lower
1101 performance in other subgroups when the evaluation is performed only for the total population.
1102 Poor performance in specific subgroups could make the device unsafe for use in those groups,
1103 which may impact the potential scope of the intended use population. Section VIII (Data
1104 Management) outlines why stratification and analyses of subgroups of interest is important to
1105 FDA’s evaluation of safety and effectiveness. An analysis of subgroup performance that supports
1106 safe and effective use across the expected intended use population also helps to ensure that

⁵⁰ See 21 CFR 830.50.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1107 devices can be used for all intended patients.

1108

1109 While differential performance across subgroups is not unique to AI-enabled devices, the
1110 reliance of models on relationships learned from large amounts of data, and the relative opacity
1111 of models to users make AI-enabled devices particularly susceptible to unexpected differences in
1112 performance. Even when the data used to develop the model is representative during training,
1113 models can be over-trained to recognize features of data that are unique to specific characteristics
1114 of the study dataset but may be spurious to the identification or treatment of the disease or
1115 condition. Spurious learnings could impact performance differentially across characteristics of
1116 interest such as disease subtype or patient demographics, especially when data from study
1117 participants from different groups tend to be collected at different sites. For example, models
1118 may erroneously use demographic information, or another variable correlated with demographic
1119 information, as a variable of interest in the model because patients of one demographic tended to
1120 be more likely to have a disease in the training data set. This can be particularly difficult to
1121 identify with complex models in which the variables of interest may not be understandable to
1122 humans. For this reason, the accuracy and usefulness of an evaluation of an AI-enabled device
1123 also depends on the quality, diversity, and quantity of data used to test it.

1124

1125 Subgroup analysis provides the tools to evaluate the performance of the device in specific
1126 populations and can be helpful in identifying scenarios in which the device performs worse than
1127 overall performance. In addition, subgroup analyses are helpful in identifying potential
1128 limitations of the device and can contribute to effective labeling by providing end users with
1129 additional useful information.

1130

1131 Information on the uncertainty of device outputs is also important because it helps reviewers
1132 understand how to interpret device outputs. When not specified for a device type in statute,
1133 regulation, or guidance, repeatability and/or reproducibility studies can still help FDA
1134 understand and quantify the uncertainty associated with device outputs when provided.

1135

1136 Appendix C (Performance Validation Considerations) of this guidance includes additional
1137 recommendations for some common approaches to performance validation. In addition, FDA
1138 encourages sponsors to leverage the [Q-Submission Program](#) for obtaining FDA feedback on
1139 proposed approaches to AI-enabled device development and validation. In particular, early
1140 engagement could be helpful to discuss the use of RWD, the use of new and emerging study
1141 methods, or the validation of new technologies.

1142

Assessing the Performance of the Human-Device Team

1143 It is important for sponsors to consider the interactions between users and the device when
1144 identifying the appropriate methods for performance testing. In the document, “[Good Machine
1146 Learning Practice: Guiding Principles](#),” Principle Seven discusses placing focus on “the
1147 performance of the Human-AI Team.” This principle explains that it is important to understand
1148 the performance of the “Human-AI team, rather than just the performance of the model in
1149 isolation” when a model has a “human in the loop.” The intended use and clinical workflow of
1150 AI-enabled devices span a continuum of decision-making roles from more autonomous systems
1151 to supportive (aid) tools that assist specific users, but rely on the human to interpret the AI
1152 outputs and ultimately make clinical decisions.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1153
1154 As the device moves along this spectrum, the nature of the clinical study or other studies (e.g.,
1155 human factors validation testing) that would be appropriate to support performance evaluation of
1156 an AI-based medical device will vary according to the intended use of the model. For some
1157 devices, more emphasis may be placed on the model’s standalone performance (i.e., Did the
1158 actual output match the expected output?). For others, a focus may be assessing the performance
1159 of the human-AI team, beyond just the performance of the model in isolation (i.e., Did the
1160 intended user working with the new device perform the same or better than the operator alone or
1161 with another device?). Sponsors should consider that, in certain scenarios, both standalone and
1162 human-device team performance evaluations may support the overall performance evaluation of
1163 the AI-enabled device.

1164
1165 Performance evaluation of AI-based medical image analysis systems is an illustrative example of
1166 how the clinical study approaches may change as the intended use of the device moves along the
1167 spectrum of human-device interactions. Standalone assessments measure the model’s
1168 performance independently of human interaction, whereas reader studies compare the
1169 performance of the intended user both with and without the AI-enabled device (i.e., comparing
1170 the human vs. human-device team performance).⁵¹ Reader studies typically serve as the primary
1171 performance evaluation for AI-enabled devices that aid in clinical decision-making in medical
1172 imaging applications, because they allow sponsors to evaluate the tool’s clinical benefit in the
1173 hands of the intended user.⁵²

1174
1175 *What sponsors should include in a submission:* The validation testing should provide objective
1176 information to characterize the model performance with respect to the intended use. A validation
1177 assesses the model’s performance on independent datasets. Assessing the robustness of the
1178 model to anticipate reasonably foreseeable changes in input data and conditions of use should
1179 also be included, as appropriate, based on risk associated with these changes.

1180
1181 Validation methods differ depending on the intended use of a device. For example:
1182 • Devices estimating defined measurements otherwise performed by accepted reference
1183 methods may need a precision study to adequately assess their repeatability and
1184 reproducibility.
1185 • Devices monitoring time-series patient data and needing periodic re-calibrations may need a
1186 stability study and a change tracking study to assess their dynamic responses.
1187 • Devices similar to survey instruments measuring less well-defined patient parameters may
1188 need additional evidence of construct validity (i.e., the extent to which a test measures what it
1189 is proposed to measure).
1190 • Prognostic clinical decision support devices may need longitudinal data with survival
1191 analysis, calibration analysis, and/or discrimination analysis (e.g., risk stratification analysis),
1192 among other methods.

⁵¹ For more information, see FDA’s guidance titled, “[Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data Premarket Notification \[510\(k\)\] Submissions.](#)”

⁵² For more information on computer-assisted detection devices, please see FDA guidance titled, “[Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data in Premarket Notification \(510\(k\)\) Submissions.](#)”

Contains Nonbinding Recommendations

Draft – Not for Implementation

1193
1194 Depending on the specific AI-enabled device, this evidence could come from non-clinical bench
1195 or analytical studies, pre-clinical animal studies, clinical performance studies, clinical outcome
1196 studies, or some combination thereof.

1197 1198 Study Protocols

1199 To support performance validation, sponsors should include information regarding all study
1200 protocols including statistical analysis plans. The statistical analysis plans should include study
1201 design and analysis details. Important aspects for these documents to cover include:

- 1202 • Study design details, including:
 - 1203 ○ A study design description (e.g., prospective, comparative study design with a
1204 sufficient statistical power to demonstrate the key clinical performance metric).
 - 1205 ▪ For a prospective study, procedures and methods that will be followed, a
1206 description of the operators involved in these procedures and methods,
1207 and any tools or equipment to be used.
 - 1208 ▪ For a retrospective study, plans on how to handle, prepare, process, and
1209 select archived data or material.
 - 1210 ○ A description of the data recording mechanisms that will be used to record the
1211 version or state of the AI-enabled device used during the study for a given
1212 patient.
 - 1213 ▪ To ensure accuracy, automated collection of these data implemented in an
1214 electronic case report form (eCRF)⁵³ or electronic data capture (EDC)
1215 system may be appropriate.
 - 1216 ○ A description of the procedures and methods for blinding of the device outputs
1217 from the clinical reference standard determination process, masking of the
1218 clinical reference standard from the users/interpreters of the device outputs, and
1219 masking of the test data from the model developers and clinical team (to avoid
1220 opportunistic tweaking or bias in the study design), as applicable.
 - 1221 ○ A description of the controls in place to address any risks posed to the patient or
1222 user by the AI-enabled device during the study.
 - 1223 ○ If the protocol is altered during the execution of the clinical study, the applicant
1224 should explain the changes, and identify which changes are deemed minor and
1225 major, providing adequate justification for any repeated tests or tests with
1226 deviations from the pre-specified plans. The study protocol should be followed
1227 and all types of protocol deviations, including those deemed minor, should be
1228 minimized.
 - 1229 ○ A full accounting of all enrolled subjects (with an accountability table).
 - 1230 ○ A description of baseline distributions of the study population and other
1231 important factors in the dataset such as data acquisition equipment, device
1232 configurations, and disease status or conditions, and a justification of their
1233 representativeness. For more information on representativeness in AI-enabled
1234 medical devices, refer to Section VIII (Data Management) of this guidance.
- 1235 • Statistical analysis plans, including:

⁵³ For more information, see FDA guidance titled "[Electronic Source Data in Clinical Investigations](#)."

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 1236 ○ A description of the primary endpoint(s) or outcome(s), which should be
- 1237 reflective of the primary objective of the study.
- 1238 ○ Pre-specified study success/failure criteria with respect to each of multiple
- 1239 primary endpoints (e.g., performance goals) that are clinically justified (e.g.,
- 1240 supported by literature or prior investigations).
- 1241 ○ An explanation of the statistical hypotheses, such as null hypothesis, and the
- 1242 alternative (working) hypothesis.
- 1243 ○ A sample size justification that ensures adequate study power.
- 1244 ○ An explanation of the statistical analysis of the primary endpoint(s), including
- 1245 information to justify the sample size calculation.
- 1246 ○ An explanation of the pre-specified, appropriate statistical approaches for
- 1247 handling multiplicity issues and controlling for overall Type I error rates;
- 1248 ○ A description of the appropriate statistical methodology.
- 1249 ○ A subgroup analysis plan.
 - 1250 ▪ The appropriate subgroups are informed by the intended use of the
 - 1251 device, but should generally include patient sex,⁵⁴ gender, age, race,
 - 1252 ethnicity,⁵⁵ disease variables, clinical data site, data acquisition equipment
 - 1253 (e.g., camera brand), and, if applicable, conditions for use (including skill
 - 1254 level of the user when relevant), device configurations, and other relevant
 - 1255 confounding factors that may impact the device performance.
 - 1256 ▪ When a specific performance claim is made with respect to a subgroup,
 - 1257 the subgroup analysis should be statistically significant, including the
 - 1258 inclusion of appropriately powered subgroups. However, when specific
 - 1259 subgroup performance claims are not made, subgroup performance does
 - 1260 not need to be statistically powered for each subgroup, but effort should
 - 1261 be made to include reasonable numbers of patients for each subgroup so
 - 1262 that any reported results have meaning and context.

Study Results

1264 To support performance validation, sponsors should include information regarding the study
1265 results. Important aspects for these documents to cover include:

- 1266 ● An explanation of the pre-specified results for each test, including subgroup analyses.
- 1267 ● An explanation of the results with adequate subgroup analyses for relevant subgroups as
- 1268 described above.
 - 1269 ○ If demographic information is not available for the study data, an explanation of
 - 1270 the reasons it is not available, why performance evaluation can be supported
 - 1271 without demographic subgroup analysis, and how risks associated with the lack
 - 1272 of demographic subgroup analyses have been controlled.
- 1273 ● When feasible, and appropriate, an evaluation of the device repeatability and
- 1274 reproducibility. The specifics of how these studies are conducted will depend on the

⁵⁴ For more information on sex-specific data, please see FDA guidance titled "[Evaluation of Sex-Specific Data in Medical Device Clinical Studies.](#)"

⁵⁵ For more information on age-, race-, and ethnicity-specific data, please see FDA guidance titled "[Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies.](#)"

Contains Nonbinding Recommendations

Draft – Not for Implementation

1275 specific device being evaluated, and may include phantom, simulated, contrived, or
1276 clinical data.

1277
1278 *Where sponsors should provide it in a submission:* Information on the non-clinical or clinical
1279 testing of the device should be included in the appropriate sections of the marketing submission.
1280 For example, clinical study findings should go in the clinical section of the marketing
1281 submission. Information on the software verification and software validation of the model should
1282 be included in the “**Software testing as part of Verification and Validation**” in the Software
1283 Documentation section of the marketing submission, as described in the [Premarket Software](#)
1284 [Guidance](#).

1285

ADDITIONAL RESOURCES: Appendix C (Performance Validation Considerations) includes recommendations to help develop and analyze a performance validation study and its data. Appendix D (Usability Evaluation Considerations) includes information to help sponsors evaluate usability risk controls for AI-enabled device submissions.

FDA encourages sponsors to use the [Q-Submission Program](#) for obtaining FDA feedback on proposed approaches for AI-enabled device development and validation. If real world evidence is used, sponsors may also wish to refer to FDA guidance titled “[Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices](#).”

1286

1287 **XI. Device Performance Monitoring**

1288 *Why should it be included in a submission for an AI-enabled device:* The performance of AI-
1289 enabled devices deployed in a real-world environment (i.e., marketed AI-enabled devices
1290 following approval or clearance) may change or degrade over time, presenting a risk to patients.
1291 In general, as part of the quality system for a medical device, including an AI-enabled device,
1292 manufacturers should have a postmarket performance monitoring plan to help identify and
1293 respond to changes in performance in a postmarket setting. The inclusion of a performance
1294 monitoring plan in the marketing submission may help to reduce uncertainty and support FDA’s
1295 evaluation of risk controls.

1296

1297 As part of their ongoing management of AI-enabled devices manufacturers should proactively
1298 monitor, identify, and address device performance changes, as well as changes to device inputs
1299 and the context in which the device is used that could lead to changes in device performance. In
1300 addition, sponsors must develop and implement plans for comprehensive risk analysis programs
1301 and documentation consistent with the Quality System Regulation (21 CFR Part 820) to manage
1302 risks related to undesirable changes in device performance for AI-enabled devices.⁵⁶ These
1303 regulations include, but are not limited to, management responsibility (21 CFR 820.20), design
1304 validation (21 CFR 820.30(g)), design changes (21 CFR 820.30(i)), nonconforming product (21
1305 CFR 820.90), and corrective and preventive action (21 CFR 820.100). Further, manufacturers
1306 must monitor device performance and report to FDA information about deaths, serious injuries,
1307 and malfunctions in accordance with 21 CFR Parts 803 and 806.

⁵⁶ When the final rule amending the device QSR, 21 CFR Part 820, takes effect on February 2, 2026, the term “risk analysis” will be replaced with “risk management.”

Contains Nonbinding Recommendations

Draft – Not for Implementation

1308

1309 FDA generally does not assess quality system regulation compliance as part of its review of
1310 marketing submissions under section 510(k) of the FD&C Act. However, in some cases, it may
1311 be appropriate for FDA to review details from the sponsor's quality system in the marketing
1312 submission to ensure adequate ongoing performance. Such a review may help support a
1313 determination of substantial equivalence.

1314

1315 Ongoing performance monitoring is important for AI-enabled devices because, as described
1316 above, models are highly dependent on the characteristics of data used to train them, and as such,
1317 their performance can be particularly sensitive to changes in data inputs. Changes in device
1318 performance may originate from many factors, such as changes in patient populations over time,
1319 disease patterns, or data drift from other changes. When performance changes do occur, users
1320 may be less likely to identify them in AI-enabled devices if, for example, the devices are part of
1321 a highly automated process with limited on-going human interaction, or if the output is
1322 prognostic such that different healthcare professionals may be involved in the use of the device
1323 and in confirmatory follow-up interactions with the patient. Because the performance of AI-
1324 enabled devices can change as aspects of the environments in which they are approved or cleared
1325 for use in may change over time, it may not be possible to completely control risks with
1326 development and testing activities performed premarket (prior to device authorization and
1327 deployment).

1328

1329 FDA recognizes that the environments where medical devices are deployed cannot be completely
1330 controlled by the device manufacturer. Further, the presence of factors that may lead to changes
1331 in device performance may not always raise concerns about patient harm. Rather, as part of
1332 ongoing risk management, it is important for device manufacturers to consider the impact of
1333 these factors (e.g., data drift) on the safety and effectiveness of the device. Additional
1334 information about performance management processes may be helpful for FDA to determine
1335 whether risks have been adequately identified, addressed and controlled.

1336

1337 *What sponsors should include in a submission:* Sponsors of AI-enabled devices that elect to
1338 employ proactive performance monitoring as a means of risk control and to provide reasonable
1339 assurance of the device's safety and effectiveness, should include information regarding their
1340 performance monitoring plans as part of the premarket submission. Sponsors are encouraged to
1341 obtain FDA feedback on the plan through the [Q-Submission Program](#). For a 510(k) submission,
1342 FDA generally does not require such plans for devices, absent certain circumstances, for which a
1343 performance monitoring plan is not a special control for the particular device type (i.e., the
1344 applicable classification regulation). For a De Novo classification request, such a plan may be
1345 necessary to control risks posed by the particular device type and so FDA may establish a special
1346 control for the device type going forward. For a PMA, a performance monitoring plan may be a
1347 condition of approval.⁵⁷ However, sponsors may opt to include information regarding the
1348 performance monitoring plan in any submission for an AI-enabled device.

1349

1350 Performance monitoring plans should identify and respond to, in a timely fashion, performance
1351 changes or conditions that may lead to performance change or degradation. A robust

⁵⁷ See 21 CFR 814.44 and 21 CFR 814.82.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1352 performance monitoring plan includes proactive efforts to capture device performance after
1353 deployment. Components of such a plan may include:

- 1354 • A description of the data collection and analysis methods for:
- 1355 ○ Identifying, characterizing, and assessing changes in model performance, including
1356 assessing the results from performance monitoring on safety and effectiveness.
 - 1357 ○ Monitoring potential causes of undesirable changes in performance, such as:
 - 1358 ■ Changes in patient demographics or disease prevalence;
 - 1359 ■ Shifts in input data;
 - 1360 ■ Changes to input data due to corruption in the data pipeline (input data
1361 integrity), such as missing values, duplicate records, data type mismatches;
1362 and
 - 1363 ■ Changes in users' behavior or in user demographics.
 - 1364 • A description of robust software lifecycle processes that include mechanisms for monitoring
1365 in the deployment environment.
 - 1366 • A plan for deploying updates, mitigations, and corrective actions that address changing
1367 performance in a timely manner.
 - 1368 • FDA notes that some actions taken to address performance changes may not require a
1369 marketing submission or authorization (21 CFR 807.81(a)(3) and 21 CFR 814.39(a)) prior to
1370 being taken. Please refer to FDA guidances titled, "[Deciding When to Submit a 510\(k\) for a](#)
1371 [Change to an Existing Device](#)," and "[Deciding When to Submit a 510\(k\) for a Software](#)
1372 [Change to an Existing Device](#)" to help assess whether a particular change may require a
1373 premarket submission to FDA. Sponsors may also wish to consider the use of a PCCP, as
1374 appropriate.⁵⁸
 - 1375 ○ This plan does not replace applicable statutory or regulatory requirements, including
1376 the requirements to report to FDA information about certain adverse events, and
1377 corrections and removals, under 21 CFR Parts 803 and 806.
 - 1378 • A description of the procedures for communicating the results of performance monitoring
1379 and any mitigations to device users.

1380
1381 *Where sponsors should provide it in a submission:* When appropriate, a device performance
1382 monitoring plan should be included in the "**Risk Management File**" in the Software
1383 Documentation section of the marketing submission.
1384

1385 **XII. Cybersecurity**

1386 *Why should it be included in a submission for an AI-enabled device:* As with any digital or
1387 software component integrated into a medical device, AI can present cybersecurity risks. FDA's
1388 general recommendations for designing and maintaining cybersecurity as well as relevant
1389 marketing submission documentation are provided in the guidance document titled
1390 "[Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket](#)
1391 [Submissions](#)" (hereafter referred to as the "[2023 Premarket Cybersecurity Guidance](#)"). The [2023](#)
1392 [Premarket Cybersecurity Guidance](#) identifies security objectives that may be relevant for medical

⁵⁸ See FDA's guidance titled "[Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions](#)."

Contains Nonbinding Recommendations

Draft – Not for Implementation

1393 devices, including AI-enabled devices: authenticity, which includes integrity; authorization;
1394 availability; confidentiality; and secure and timely updatability and patchability.

1395
1396 For AI-enabled devices that meet the definition of a “cyber device” under section 524B(c) of the
1397 FD&C Act, the recommendations in this section of the guidance are intended to help
1398 manufacturers meet their obligations under section 524B of the FD&C Act. Examples of AI risks
1399 which can be impacted by cybersecurity threats include, but are not limited to:

- 1400 • *Data Poisoning*: Cyber threats could lead to data poisoning by deliberately injecting
1401 inauthentic or maliciously modified data, risking outcomes in areas like medical
1402 diagnosis.
- 1403 • *Model inversion/stealing*: Cyber threats could intentionally use forged or altered data to
1404 infer details from or replicate models. These pose risks to continued model performance
1405 as well as intellectual property and privacy breaches.
- 1406 • *Model Evasion*: Cyber attackers could intentionally craft or modify input samples to
1407 deceive models, leading them to incorrect classifications. These pose risks to the
1408 reliability and integrity of model predictions, potentially undermining trust in AI-enabled
1409 devices and exposing them to malicious exploitation.
- 1410 • *Data leakage*: Cyber threats could exploit vulnerabilities to access sensitive training or
1411 inference data in models.
- 1412 • *Overfitting*: Cyber threats could deliberately “overfit” a model, exposing the AI
1413 components to adversarial attacks as these components struggle to adapt effectively to
1414 modified patient data.
- 1415 • *Model Bias*: Cyber threats could lead to manipulation of training data to introduce or
1416 accentuate biases. They could exploit known biases using adversarial examples, embed
1417 backdoors during training to later trigger biased behaviors, or leverage pre-trained models
1418 with inherent biases, amplifying them with skewed fine-tuning data.
- 1419 • *Performance Drift*: Cyber threats could lead to model performance drift by changing the
1420 underlying data distribution, which degrades model performance. Cyber threats could
1421 slightly shift the input data over time or exploit vulnerabilities in dynamic environments,
1422 causing the model to make inaccurate predictions or become more susceptible to
1423 adversarial attacks.

1424
1425 *What sponsors should include in a submission*: Consistent with the submission documentation
1426 recommended in the [2023 Premarket Cybersecurity Guidance](#) regarding the cybersecurity
1427 controls and security risk management relevant to the AI components or features, sponsors
1428 should include the following types of information:

- 1429 • Any additional elements in the cybersecurity risk management report, threat modeling,
1430 cybersecurity risk assessment, labeling, and other deliverables, where there are unique
1431 considerations related to AI cybersecurity.
- 1432 • An explanation regarding how the cybersecurity testing is appropriate to address the risks
1433 associated with the model, including, at minimum the following tests:
 - 1434 ○ Malformed input (fuzz) testing; and
 - 1435 ○ Penetration testing.
- 1436 • A Security Use Case View(s) that covers the AI-enabled considerations for the device.
- 1437 • A description of controls implemented to address data vulnerability and preventing data

Contains Nonbinding Recommendations

Draft – Not for Implementation

1438 leakage, including:

- 1439 ○ Access controls;
- 1440 ○ Any data encryption; and
- 1441 ○ Anonymization or de-identification of sensitive data.

1442

1443 Sponsors should refer to the control recommendations in Appendix 1 of the [2023 Premarket](#)
1444 [Cybersecurity Guidance](#) for how they may wish to address the specific risks above. Example
1445 approaches to controlling cybersecurity risks related to AI-enabled devices include:

1446 ● For data poisoning attacks, consider:

- 1447 ○ Validating, authenticating, and cleansing data.
- 1448 ○ Employing anomaly detection and data integrity checks (e.g., cryptographic
1449 hashes).
- 1450 ○ Applying adversarial training, which is a method used to improve the robustness
1451 and security of models.

1452 ● For cyber threats using forged data to introduce overfitting, model bias, etc., consider:

- 1453 ○ Adopting differential privacy, which is a technique to protect the privacy of
1454 individual data points in a dataset. When utilizing differential privacy, sponsors
1455 should be cognizant of potential trade-offs between privacy and factors such as
1456 model accuracy, utility, and efficiency, and provide information on how the trade-
1457 offs are addressed.
- 1458 ○ Engaging in secure multi-party computation (MPC), which is a technique that can
1459 allow multiple parties to collaboratively train a model without revealing their
1460 local datasets to each other.
- 1461 ○ Employing data authentication and integrity protections.
- 1462 ○ Introducing watermarking, which involves embedding hidden watermarks into AI
1463 models to prove ownership.
- 1464 ○ Applying continuous model performance monitoring.

1465 ● For model evasion, consider adversarial training to enhance model robustness and
1466 implement strict input verification checks to ensure data conforms to expected patterns.
1467 When deploying adversarial training techniques, sponsors should be cognizant of the
1468 trade-offs that may arise between enhanced robustness to attacks and the potential
1469 negative impact on model performance (e.g., accuracy), and provide information on how
1470 the trade-offs are managed.

1471 *Where sponsors should provide it in a submission:* The cybersecurity information should be
1472 included in the “**Cybersecurity/Interoperability**” section of the marketing submission, as
1473 described in the [2023 Premarket Cybersecurity Guidance](#).

1474

ADDITIONAL RESOURCES: Sponsors may also refer to other FDA guidance documents for additional recommendations relevant to cybersecurity:

- [Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions](#)
- [Postmarket Management of Cybersecurity in Medical Devices](#)
- [Cybersecurity for Networked Medical Devices Containing Off-the-Shelf \(OTS\) Software](#)
- [Off-The-Shelf Software Use in Medical Devices](#)

1475

1476 **XIII. Public Submission Summary**

1477 *Why should it be included in a submission for an AI-enabled device:* Transparency is a key
1478 component of premarket authorization and is important to patient care. This is especially
1479 important for AI-enabled devices, which are heavily data driven and incorporate algorithms
1480 exhibiting a degree of opacity. In public workshops and comments, including the October 14,
1481 2021 virtual public workshop on the transparency of AI-enabled devices titled “[Transparency of](#)
1482 [Artificial Intelligence/Machine Learning-enabled Medical Devices](#),” patients noted concerns
1483 with the use of AI in their care. The public has consistently called for additional information
1484 about how FDA makes authorization decisions about AI-enabled devices, as well as more
1485 information about the design and validation of these devices. The public submission summary
1486 should include specific information describing the characteristics of these devices to support
1487 transparency, which can contribute to public health by increasing understanding of AI-enabled
1488 devices and developing public trust.

1489 Public submission summaries are required and available on the FDA website for most marketing
1490 authorization decisions.⁵⁹ These summaries describe the device and the information supporting
1491 regulatory decision-making. Where a public summary is required, details about the AI-enabled
1492 device must be included in sufficient detail in the public-facing documents to support
1493 transparency to users of FDA’s determination of substantial equivalence or reasonable assurance
1494 of safety and effectiveness for the device.^{60,61,62} To ensure public access to important
1495 information on authorized AI-enabled devices, this section describes the types of information
1496 sponsors should include in the public submission summary as well as a possible format for such
1497 information.

1498
1499 For AI-enabled devices submitted through the PMA, HDE, De Novo, BLA, or 510(k) pathways,
1500 FDA recommends that the information discussed in this section be included in the relevant
1501 public submission summary, or the 510(k) Summary (in the section prepared in compliance with
1502 21 CFR 807.92(a)(4)), as applicable. Sponsors should provide the recommended information
1503 excluding any patient identifiers, trade secrets, and confidential commercial information. For
1504 sponsors submitting a 510(k) Statement (21 CFR 807.93), FDA recommends providing the same
1505 information in the submission excluding any patient identifiers, trade secrets, and confidential

⁵⁹ See FDA’s website titled “[CDRH Transparency: Premarket Submissions](#).” See 21 CFR 807.92 for requirements on the form and content of a 510(k) Summary. See 21 CFR 807.93 for requirements on the content and format of a 510(k) Statement. See 21 CFR 814.9(e) for requirements on a PMA decision summary.

⁶⁰ In accordance with 21 CFR 807.92, “a 510(k) summary shall be in sufficient detail to provide an understanding of the basis for a determination of substantial equivalence.” See 21 CFR 807.92 for requirements on the content and format of a 510(k) Summary. If a sponsor chooses to submit a 510(k) Statement rather than 510(k) Summary, the sponsor should provide information that supports FDA’s determination of substantial equivalence. See 21 CFR 807.93 for requirements on the content and format of a 510(k) Statement.

⁶¹ In accordance with 21 CFR 814.9(e), “FDA will make available to the public ... a detailed summary of information submitted to FDA respecting the safety and effectiveness of the device that is the subject of the PMA and that is the basis for the order.” See 21 CFR 814.9(e) for requirements on a PMA decision summary.

⁶² The De Novo decision summary is intended to present an objective and balanced summary of the scientific evidence that served as the basis for the FDA's decision to grant a De Novo request. For more information on De Novo decision summary documents, please see FDA’s website on [De Novo Classification Request](#).

Contains Nonbinding Recommendations

Draft – Not for Implementation

1506 commercial information.⁶³

1507

1508 While not required, the use of a model card may be one way to communicate information about
1509 AI-enabled devices because they are a means to consistently summarize the key aspects of AI-
1510 enabled devices and can be used to concisely describe their characteristics, performance, and
1511 limitations. Appendix E (Example Model Card) provides recommendations for the contents and
1512 formatting of a model card. Research has demonstrated that the use of a model card can increase
1513 user trust and understanding. The use of a model card as part of a public submission summary
1514 specifically is one way to support clear and consistent communication about an AI-enabled
1515 device to the interested parties in the public as well as to users, such as patients, clinicians,
1516 regulators, and researchers. The use of the model card can address the challenges associated with
1517 determining the best approach to communicate important information about the AI-enabled
1518 device.

1519

1520 *What sponsors should include in a submission:* Sponsors must comply with the submission
1521 regulations for their particular submission.⁶⁴ In addition, sponsors should consider FDA
1522 recommendations for the relevant marketing submission type. Sponsors should also provide the
1523 following types of information excluding any patient identifiers, trade secrets, and confidential
1524 commercial information:

- 1525 • A statement that AI is used in the device;
- 1526 • An explanation of how AI is used as part of the device’s intended use. For devices with
1527 multiple functions, this explanation may include how AI-DSFs interact with each other as
1528 well as how they interact with non-AI DSFs;
- 1529 • A description of the class of model (e.g., convolutional neural network, recurrent neural
1530 network, support vector machine, transformers) and limitations of the model within the
1531 device description;
- 1532 • A description of the development and validation datasets (size, source of data), including
1533 information about the demographic characteristics in the training and validation data,
1534 along with information about the demographic characteristics in the population(s) of
1535 intended use. The description should also compare the training dataset to the validation
1536 dataset and model data inputs expected in the intended use. The comparison should
1537 describe how independence of test data from training data was ensured;
- 1538 • A description of the statistical confidence level of predictions, including any other
1539 descriptions or metrics that describe statistical confidence and uncertainty, as applicable;
1540 and
- 1541 • A description of how the model will be updated and maintained over time, if applicable.

1542 Sponsors should consider using a model card to organize information. Appendix E (Example
1543 Model Card) includes recommendations on the elements that may be included within a model
1544 card. While the example model card includes recommended elements and format for a model

⁶³ For more information, see FDA guidance, “[The 510\(k\) Program: Evaluating Substantial Equivalence in Premarket Notifications \[510\(k\)\]](#).”

⁶⁴ For more information regarding the requirements for PMA, see 21 CFR Part 814. For more information regarding the requirements for 510(k), see 21 CFR 807.81 – 807.100. For more information regarding the requirements for De Novo, see 21 CFR 860.200 – 860.260. For more information regarding the requirements for HDE, see 21 CFR 814.100 – 814.126. For more information regarding the requirements for BLA, see 21 CFR Part 600 – 680.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1545 card, sponsors may include additional information and/or follow a different format. In the
1546 absence of the model card structure, sponsors should still consider including the information a
1547 model card contains.

1548 *Where sponsors should provide it in a submission:* The public submission summary should be
1549 included in the “**Administrative Documentation**” section of the marketing submission.

ADDITIONAL RESOURCES: Appendix E (Example Model Card) of this guidance provides one example of the format of a model card. Appendix F (Example 510(k) Summary with Model Card) of this guidance provides an example of a public submission summary for a product, including a completed model card.

1550

1551

1552

1553

1554

Contains Nonbinding Recommendations

Draft – Not for Implementation

1555 **Appendix A: Table of Recommended Documentation**

1556 Sections V-XIII of this guidance provide recommendations regarding the documentation that
1557 may be included within a marketing submission for AI-enabled devices. The table below
1558 summarizes recommended locations within the marketing submission to provide discussed
1559 documentation. One way this documentation may be submitted is through the eSTAR Program.
1560 Specifically, eSTAR is an interactive PDF form that guides applicants through the process of
1561 preparing a comprehensive medical device submission.⁶⁵ eSTAR is free and is required for all
1562 510(k) submissions, unless exempted.

Guidance Section and Recommended Information	Recommended Section in Sponsor’s Marketing Submission
Section V Device Description	Device Description
Section VI.A User Interface	Software Description
Section VI.B Labeling	Labeling
Section VII Risk Assessment	Risk Management File of Software Documentation
Section VIII Data Management	Data for development: Software Description of Software Documentation Data for testing: Performance Testing
Section IX Model Description and Development	Software Description
Section X.A Performance Validation	Clinical and non-clinical testing: Performance Testing Software verification and software validation: Software testing as part of verification and validation of Software Documentation
Section XI Device Performance Monitoring	Risk Management File of Software Documentation
Section XII Cybersecurity	Cybersecurity
Section XIII Public Submission Summary	Administration Information

1563

1564

⁶⁵ For more information on eSTAR, please see FDA’s website on [eSTAR Program](#).

1565 **Appendix B: Transparency Design Considerations**

1566 This appendix contains recommendations for developing a transparent device centered around
1567 users. These recommendations are intended to help sponsors develop safe and effective medical
1568 devices and high-quality marketing submissions. While sponsors may identify alternate
1569 approaches that support FDA’s evaluation of safety and effectiveness, they should integrate
1570 transparency considerations starting at the design phase of the TPLC to ensure the availability of
1571 information to support the marketing submission. It can be difficult to integrate transparency into
1572 a device in later stages of the TPLC when changes to the device might require additional testing.
1573 In this guidance, transparency refers to clearly communicating the contextually relevant
1574 performance and design information of a device to the appropriate stakeholders in a manner that
1575 they can understand and act on. Transparency involves ensuring that important information is
1576 both accessible and functionally comprehensible and is connected both to the sharing of
1577 information, and to the usability of a device. As such, a user-centered approach to transparency
1578 design helps support the safe and effective use of AI-enabled devices. Including appropriate
1579 transparency information has also been shown to more than double willingness to use a device.

1580

1581 Transparency by Design Across the TPLC

1582 Sponsors should take a holistic approach to identifying relevant contextual factors for device use
1583 and how those factors impact device performance when determining what information should be
1584 communicated. Sponsors should consider transparency throughout the full continuum of
1585 implementation through use, maintenance, and decommission of the AI-enabled device, and
1586 should design the device with transparency in mind from the beginning.

1587

1588 The user interface is another area where transparency principles should be used, when
1589 appropriate. The information in other elements of the user interface can complement the printed
1590 labeling (e.g., packaging and user manuals) to support the user’s understanding of how to use the
1591 device by providing timely and contextually relevant information throughout the use process, as
1592 described in Section VI (User Interface and Labeling). Examples of points of interaction include
1593 alerts generated by a device and displayed on the device or pushed to another product,
1594 components of associated hardware, and display screens. Effective transparency planning
1595 identifies the necessary information for the intended user(s) and context of use, as well as the
1596 optimal mediums, timing, and strategies for successful communication of the necessary
1597 information.

1598

1599 Generally, the transparency design process should begin with a holistic approach to obtain an
1600 understanding of the context in which a product is used, followed by identifying user tasks, and
1601 possible risks associated with communication of information during those tasks. This can be
1602 accomplished by determining how and when information is needed, integrating contextually
1603 appropriate risk controls into the design of the product, and finally validating that the intended
1604 users receive and can functionally understand the key information in relevant use contexts. This
1605 process may be iterative and may not flow linearly.

1606

1607 Transparency is contextually dependent, so appropriate information will vary across the range of
1608 AI-enabled devices and depend upon their benefit/risk profiles and the needs of intended users.
1609 The considerations in this appendix are not exhaustive and are intended to help sponsors identify

Contains Nonbinding Recommendations

Draft – Not for Implementation

1610 information about the context in which the device will be used and the needs of the users for the
1611 purpose of developing a consistent approach to understanding the transparency needs for their
1612 AI-enabled device. It is also important to note that while transparency can help to address certain
1613 device risks, particularly those related to misunderstanding or misusing information output by a
1614 device, providing transparency about the existence of a significant clinical risk, including a
1615 significant risk related to performance in subpopulations of intended users, alone may not be an
1616 adequate risk control.

1617

1618 The Right Information at the Right Time

1619 Consider what information the users might need, when they might need it to facilitate decision-
1620 making, and the potential risks if the users do not have the appropriate information at the right
1621 time, at all, or if it is misunderstood. It is important to focus on the tasks that each user has to
1622 perform, and what the user needs to know to perform them in concurrence with the intended use.
1623 To identify what information needs to be gained and is critical for users, consider the intended
1624 use comprehensively with questions, such as:

- 1625 • Who needs the information and what is the most effective method of communication?
- 1626 • When does the user need to understand information to support safe and effective use?
- 1627 • What is the context of use? Examples of questions about the context of use include:
 - 1628 ○ Where will the device be used and what are the conditions in that space?
 - 1629 ○ What else might users be doing at the same time?
 - 1630 ○ How timely is the application of the information?
 - 1631 ○ In what settings will the device output be viewed?
 - 1632 ○ Will users who interpret and apply the output be the same as those who operate
1633 the device?

1634

1635 Information should be communicated at the time that it is needed. Some examples of elements of
1636 the user interface that could be used to communicate transparency information include:

- 1637 • Packaging,
- 1638 • Labeling,
- 1639 • User Training,
- 1640 • Controls,
- 1641 • Display elements
- 1642 • Outputs/ reports,
- 1643 • Alarms/ warnings, and
- 1644 • Logic of operation of each device component and of the user interface system as a whole.

1645

1646 Understanding User Characteristics and Needs

1647 The ability of a user to operate an AI-enabled device depends on their personal characteristics
1648 and the device use environment. The environments in which AI-enabled devices are used may
1649 also influence a user interface design. As part of design inputs, consider the needs of users in the
1650 context of use. Understanding users and their needs and limitations should occur early in the
1651 development process for the AI-enabled device and may be repeated as the design process
1652 continues. Users may include, for example:

- 1653 • Patients,
- 1654 • Purchasers,

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 1655 • Administrators,
- 1656 • Healthcare Professionals,
- 1657 • Caregivers, and
- 1658 • Maintenance Technicians.

1659

1660 It is important to consider the characteristics of each user that may impact the user needs,
1661 including appropriate content and format for communication. Considerations may include:

- 1662 • The user’s functional capabilities, including cognitive, physical and sensory capabilities;
- 1663 • The experience and knowledge levels of the users, including their educational
1664 backgrounds;
- 1665 • The frequency at which the user will interact with the device;
- 1666 • The level of training users are expected to receive; and
- 1667 • The similarities and differences of the new information as compared to information the
1668 users have utilized in the past.

1669

1670 Communication Style and Format

1671 It is also important to consider the format used for communication. The format should be clear
1672 and appropriate for each user at each user task. Factors may include:

- 1673 • The reading level of the user.
- 1674 • The location of information.
- 1675 • Design elements such as:
 - 1676 ○ Hierarchy,
 - 1677 ○ Visualizations, and
 - 1678 ○ Dynamic labeling.

1679

1680 The selection of the timing, mode, and format of communication should be incorporated early to
1681 allow for iterative design.

1682

1683 Explainability Information and Visualizations

1684 It is also important to consider when additional information may detract from understanding,
1685 rather than add to it. For example, explainability tools or visualizations can be valuable in
1686 increasing model transparency and a user’s confidence in a model’s output and could be
1687 developed as part of the user interface. However, if not well designed and validated for the target
1688 user group, explainability tools or visualizations could also significantly mislead users.
1689 Therefore, sponsors should develop and validate explainability metrics and visualizations
1690 through appropriate testing.

1691 **Appendix C: Performance Validation Considerations**

1692 This appendix contains recommendations for some aspects of clinical performance validation in
1693 AI-enabled devices, which are intended to help sponsors develop safe and effective medical
1694 devices. While sponsors may identify alternate approaches that support FDA’s evaluation of
1695 safety and effectiveness, they should rigorously test the device to establish the device’s
1696 performance, and integrate that planning early in the design and development process to ensure
1697 the collection of appropriate data to support the device’s intended use. It can be difficult, for
1698 example, to gather additional supportive data after the completion of the pivotal clinical study.
1699 Sponsors should also follow the recommendations found in other FDA guidances regarding
1700 specific clinical study considerations. For example, additional information on evaluating and
1701 reporting results for AI-enabled devices can be found in the FDA guidances “[Design](#)
1702 [Considerations for Pivotal Clinical Investigations for Medical Devices](#),” “[Statistical Guidance on](#)
1703 [Reporting Results from Studies Evaluating Diagnostic Tests](#),” and “[Electronic Source Data in](#)
1704 [Clinical Investigations](#).” These recommendations may not apply to all device types.

1706 Pre-specification of Study Protocols and Statistical Analysis Plan

1707 Post-hoc analysis may bias the performance assessment. Therefore, to accurately evaluate the
1708 performance of the device, study protocols and statistical analysis plans should be pre-specified.
1709 Regardless of whether data are collected prospectively or retrospectively, study design elements
1710 (such as sample size justification, and plans on how to handle, prepare, process, and select
1711 archived data or material) should be specified prior to beginning the validation study.

1713 Study Reports

1714 All performance and usability assessments should be objective, and the model should not be
1715 tweaked opportunistically in light of the test data results (i.e., no post-hoc adjustment). In
1716 general, proceeding to execute the study protocol only after a sound validation plan (study
1717 protocol and statistical analysis plan) is documented and finalized helps avoid these post-hoc
1718 adjustments. Execution of the plan includes collecting the required data, conducting the pre-
1719 specified analysis, and reporting the study results. Validation study reports should specify the
1720 associated protocol version and adequate justifications should be provided for any repeated tests
1721 or tests with deviations from the pre-specified plans.

1723 Masking Protocol

1724 For diagnostic devices, a masking protocol in the clinical study ensures that the user of the test is
1725 “blinded/masked” to the clinical reference standard result while the provider of the clinical
1726 reference standard result is “blinded/masked” to the test result. The masking protocol also
1727 ensures that model developers and the clinical team are completely masked from the test data
1728 during the model development process.

1730 For therapeutic devices, masking is sometimes implemented through a randomized-controlled
1731 study with two arms (e.g., placebo/sham device arm and subject device arm), when ethically
1732 appropriate such as with non-invasive diagnostic devices. This ensures patients and care
1733 providers are blinded to the actual treatment assignment. The placebo arm may not have any
1734 measurement but only serve as a blinding tool (e.g., so that caretakers will not provide
1735 differentiated care in different arms). When such a two-arms study design is not feasible, there

Contains Nonbinding Recommendations

Draft – Not for Implementation

1736 may be potential bias in the performance assessment due to placebo effects.

1737

1738 Model Precision: Repeatability and Reproducibility

1739 An AI-enabled device may often be intended to measure physiological signals when the device is
1740 placed on a particular anatomical location. It is important to know how robust the device output
1741 is due to potential variations in the measurement system (e.g., whether repeated tests by users
1742 will generate significantly different device output due to operator difference and signal
1743 variation). A precision study gauges the variability of a device output when making repeated
1744 measurements on the same patient, either with the same operator and device (repeatability), or
1745 with different operators and devices (reproducibility). More generally, repeatability is the
1746 closeness of agreement of repeated measurements taken under the same conditions; and
1747 reproducibility is the closeness of agreement of repeated measurements taken under different,
1748 pre-specified conditions.

1749

1750 It is important to note that not every diagnostic device needs a precision study, due to clinical
1751 and feasibility considerations. For example, there is a feasibility concern when a device may be
1752 too harmful on the patients with repeated use (e.g., for radiation or invasive devices). Another
1753 example is a monitoring device that tracks a patient's changing physiological status (e.g.,
1754 hemodynamic parameters) in real-time, where repeated observations of the same truth are not
1755 possible.

1756

1757 Key statistics to summarize the repeatability and reproducibility, based on a variance component
1758 analysis using a model's continuous metric (e.g., a probability score), are the subject-level
1759 standard deviation (SD) and the percent coefficient of variation (%CV). Improving the model to
1760 reduce SD or %CV may provide a low-cost way to improve product quality and the success
1761 likelihood of a future pivotal clinical study. This is, in part, because the clinical reference
1762 standard (i.e., the best available method for establishing the presence or absence of the target
1763 condition) is not measured in a precision study. Depending on the product, additional factors
1764 may be considered in the precision study. In image classification tasks, a model may be sensitive
1765 to data perturbation (e.g., image translation/rotation, light intensity change, random noise). This
1766 phenomenon could be abundant for an AI-enabled medical device software running on a generic
1767 smartphone using its camera to capture measurement data (e.g., skin lesion analyzers).

1768

1769 Study Endpoints and Acceptance Criteria

1770 Primary endpoints are usually assessed using pre-specified acceptance criteria within a statistical
1771 hypothesis testing framework. This approach necessitates an adequate sample size to ensure
1772 sufficient study power (i.e., acceptable type II error rate). Secondary and exploratory endpoints
1773 may also be used to inform the effectiveness of the device and are part of the totality of evidence
1774 that inform regulatory decisions. The evaluations of primary endpoints are typically based on
1775 their 95% two-sided confidence intervals (so that type I error can be protected at 5% for two-
1776 sided testing; and at 2.5% for one-sided testing). The validation of all outputs should be
1777 addressed, appropriately by type (e.g., continuous, categorical, risk scores).

1778

1779 An AI-enabled medical device can produce a variety of outputs, such as diagnostic/prognostic
1780 predictions, or treatment triaging/priority ranking/selection/planning. The validation of these
1781 outputs may involve an analytic study (e.g., precision, bench, simulation study), literature

Contains Nonbinding Recommendations

Draft – Not for Implementation

1782 review, a diagnostic performance study, a reader study (e.g., multi-reader multi-case imaging
1783 study), or a clinical outcome study (e.g., based on a study or randomized-controlled trial design).

1784
1785 When specifically considering an AI-enabled diagnostic device, the key performance assessment
1786 is its diagnostic accuracy, which is evaluated in a pivotal diagnostic performance study. Due to
1787 sampling variation, the uncertainties of the accuracy estimates are typically quantified, usually in
1788 the form of 95% two-sided confidence intervals. The study acceptance criteria can be based on
1789 statistical inferences using hypothesis testing methods (e.g., comparing a lower/upper confidence
1790 limit to a pre-specified performance goal). Note that inferences based on point estimates ignores
1791 the statistical uncertainty of the estimates and is not generally acceptable in the primary analysis.
1792 It is always compared to a comparator that can be tested and evaluated on the same patient/data
1793 as the device. This comparator can be the clinician, another device that is adequately validated
1794 for the same intended use, or standard of care. The evaluation on the same patient/data is key to
1795 mitigate differences in the task difficulty levels and disease spectrum due to sampling variation.

1796
1797 Depending on the nature of the diagnostic output (i.e., binary, polychotomous, or continuous),
1798 different evaluation metrics are possible.

- 1799 • For binary diagnostic output, evaluation may be based on, sensitivity, specificity,
1800 positive/negative predictive values (PPV/NPV), and positive/negative diagnostic
1801 likelihood ratios (LR+/LR-).
- 1802 • For risk stratification output that classifies a patient into one of multiple risk groups and
1803 that may often be found in prognostic models, some evaluation metrics are pre/post-test
1804 risks and likelihood ratios.
- 1805 • For an output that evaluates a patient's disease risk with a continuous score, some risk
1806 evaluation methods are calibration plot, receiver operating characteristic (ROC) curve,
1807 and decision curve analysis. In the context of biomarker evaluation, the predictiveness
1808 curve analysis may be used.
- 1809 • For continuous score, agreement study methods using MAE (mean absolute error),
1810 RMSE (root mean squared error), scatter plot, Deming regression, and Bland-Altman
1811 analysis are often used.

1812
1813 When the test data consists of multiple observations per patient, the within-patient correlations
1814 should be accounted for in the calculation of confidence intervals. Failure to account for the
1815 repeated measurements appropriately in the statistical analysis may lead to biased estimates and
1816 incorrect narrow confidence intervals, which may hinder objective evaluation of the device
1817 performance. Statistical techniques that account for patient-level repeated measurements include
1818 the bootstrap resampling method and analytic methods for clustered data.

1819 1820 Validation of AI-based Pre-processing Steps

1821 Some models may include a quality control algorithm that discards “low” quality cases from
1822 further processing. However, such low-quality cases may actually be truly hard/difficult ones –an
1823 example of missing not at random (MNAR), which may lead to skewed diagnostic performance
1824 in accuracy metrics (e.g., sensitivity and specificity) but also may be biased (e.g., in the sense
1825 that more patients than warranted may not get any results due to declared low-quality events). An

Contains Nonbinding Recommendations

Draft – Not for Implementation

1826 analysis of cases deemed low-quality should be conducted to verify that the quality control
1827 algorithm does not discard challenging cases.

1828
1829 For example, compare two hypothetical AI-enabled diagnostic devices (A and B) using
1830 cellphone cameras for certain skin disease detection. Assume they use the same diagnostic
1831 models, except that A has a more aggressive quality control (QC) algorithm than B in declaring
1832 low-quality cases. After excluding those cases that fail the QC algorithm, it may not be
1833 surprising to observe that A would have a better diagnostic performance than B, because many
1834 low-quality images dropped by A but not by B may in fact be good quality but difficult cases
1835 which are not included in the performance assessment for A.

1836
1837 Thus, a good practice is to examine the influence of a QC algorithm by checking the proportion
1838 of low-quality dropouts and assessing the results of a sensitivity analysis assuming a worst-case
1839 scenario (i.e., assuming the QC failure cases are all difficult ones that the model fails to classify
1840 successfully).

1841
1842
1843

1844 **Appendix D: Usability Evaluation Considerations**

1845
1846 As described in this guidance Section X Validation, sponsors should conduct human factors
1847 evaluations as part of design controls (21 CFR 820.30) for every medical device requiring a
1848 premarket submission. The [Human Factors Guidance](#) outlines analytical approaches to this
1849 evaluation as well as specific requirements for human factors validation for devices when one or
1850 more critical tasks has been identified. Human factors engineering processes typically begin with
1851 preliminary analysis and evaluation of all tasks that identifies critical tasks which, if performed
1852 incorrectly or not performed at all, could cause serious harm.⁶⁶ Sponsors should perform this
1853 analysis to identify whether a device has a critical task. If a critical task is identified, sponsors
1854 should refer to the [Human Factors Guidance](#) and perform human factors validation. While
1855 sponsors of devices that do not have a critical task may not need to submit a human factors
1856 validation testing report, they may choose to use the process outlined in the [Human Factors](#)
1857 [Guidance](#), or another approach of their choosing to evaluate usability,⁶⁷ to test their device
1858 design, and support the efficacy of risk controls. This appendix is focused on the evaluation of
1859 usability to support risk controls when a human factors validation testing report is not required,
1860 where usability addresses whether all intended users can achieve specific goals while using the
1861 device and whether users will be able to consistently interact with the device safely and
1862 effectively. This includes, but is not limited to, whether users can consistently and correctly
1863 receive, understand, interpret, and apply information related to the AI-enabled device.

1864
1865 While FDA’s [Human Factors Guidance](#) outlines recommended analytic approaches for
1866 evaluating usability, sponsors may choose to utilize alternative approaches for the evaluation of
1867 user tasks outside of the scope of that guidance. If this testing is used to support a risk control (as
1868 described in Section VII (Risk Assessment)), sponsors should include a description of the pre-
1869 specified testing protocols and analysis plans, and a justification for the appropriateness of the
1870 assessment method.

1871
1872 For AI-enabled devices, it may be specifically important for sponsors to identify and evaluate
1873 risk controls related to user tasks regarding the interpretation and use of information and
1874 interactions with novel user interfaces. The application of this information is a particular
1875 challenge for users of AI-enabled devices because models developed through AI techniques vary
1876 in explainability and interpretability. For example, some models can be explained using a simple
1877 decision tree which is generally easy for a user to follow and understand the basis of a model’s
1878 recommendations. Other models use complex, deep neural networks, where it may not be
1879 feasible to explain in a way that allows a user to completely understand the basis of
1880 recommendations, even with comprehensive information on its inputs, nodes, and weights. This
1881 means that users may not be able to easily and independently verify whether the
1882 recommendations and decisions made by an AI-enabled device are appropriate. As such, AI-
1883 enabled devices can be prone to errors of device use and information interpretation. The
1884 challenges with interpretability and explainability increase when the intended user has limited

⁶⁶ For more information regarding critical tasks, please see FDA guidance titled “[Applying Human Factors and Usability Engineering to Medical Devices](#).”

⁶⁷ See Section X (Validation) for context regarding “usability” for the purpose of this guidance.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1885 training in interpreting the output of models, when the intended use is in situations that require
1886 urgent action, when the model has no evident biological mechanism of action, and when the
1887 model changes through iterative updates. These errors can cause harm (injury or damage to the
1888 health of including the effects of delayed or incorrect clinical intervention, or damage to property
1889 or the environment)⁶⁸ and impact the safe and effective use of the device.

1890
1891 When sponsors choose to include an evaluation of usability to support the control of risks related
1892 to information, as described in Section VII (Risk Assessment), the evaluation should be
1893 appropriate to demonstrate that the user can both find and apply the information. In such cases,
1894 an impact assessment may be used to determine which user tasks could have an adverse or
1895 positive effect on knowing, understanding, and applying information for the device. As
1896 appropriate for the AI device, this assessment may include, for example, evaluation of the
1897 training program intended for risk control. For more complex AI devices with several sequential
1898 risk controls, an example evaluation approach could include use of the device in a clinical
1899 feasibility study that includes comprehensive assessment of how the user interpreted the AI
1900 outputs and what actions were taken. Ultimately, it is important to evaluate whether the user can
1901 operate and interpret the device, including demonstrating that users can understand and apply
1902 important information about the use of the device and its output in the actual context of clinical
1903 decision-making.

1904
1905 Sponsors may wish to draw on the general structure outlined Section 6.3.1 (Task Analysis) of the
1906 [Human Factors Guidance](#), which provides an example of an analysis technique to systemically
1907 break down device use into discrete user tasks. However, it is important to understand that while
1908 the [Human Factors Guidance](#) focuses on “serious harm,” sponsors may need to provide
1909 documentation evaluating and addressing any potential risk associated with misuse, including
1910 misinterpretation, to ensure that the device is safe and effective for its intended use.

1911
1912 Appendix B (Transparency Design Considerations) of this guidance also outlines
1913 recommendations to user-centered transparency, which may help with the identification of user
1914 tasks and risks related to usability and information interpretation, as well as help sponsors
1915 develop design approaches to control these risks.

1916

⁶⁸ ANSI/AAMI/ISO 14971 *Medical devices—Application of risk management to medical devices*.

1917 **Appendix E: Example Model Card**

1918 A model card is a popular format for communicating information about a device that may align
1919 with the kind of information that FDA may require, for example in the publicly available 510(k)
1920 summaries⁶⁹ and labeling.⁷⁰ The model card format and content discussed below is intended to
1921 serve as an example of possible formatting a sponsor could use to communicate information
1922 about the model and the AI-enabled device in the public submission summary and other
1923 locations where this information may be shared by the sponsor. It is important to note that FDA
1924 does not require the inclusion of a model card or a specific model card format, and this example
1925 should not be considered a template. The example model card below has been designed based on
1926 user-centered research to present data in an order and format that is useful and easy to understand
1927 for non-technical audiences and is provided to sponsors to facilitate the inclusion of a model
1928 card.

1929
1930 In general, model cards can be adapted to the specific needs and context of each AI-enabled
1931 device. However, for the public summary, we encourage sponsors to follow the general
1932 principles for creating model cards outlined in this guidance. Some elements may not be
1933 available for some devices.

1934
1935 When model cards are provided in a digital format, research has demonstrated that a dynamic
1936 approach to formatting that allows users to expand sections individually as needed makes the
1937 information easier to digest. While the public submission summary is provided as a PDF
1938 document and the format is static, sponsors should consider the use of dynamic labeling when
1939 possible.

1940 1941 **DEVICE NAME – Model Card**

1942 1943 **Device Information:**

- 1944 • Name of the Device
- 1945 • Version of the Device
- 1946 • Date when the Device was created (or last updated)
- 1947 • Model Architecture

1948 1949 **Regulatory Status (For model cards used outside of the public submission summary):**

- 1950 • Authorization status
- 1951 • File number

1952 1953 **Description:**

- 1953 • Intended users (e.g., healthcare professionals, caregivers, patients).
- 1954 • Intended use – The general purpose of the device or its function. This includes the
1955 indications for use.
- 1956 • Indications for use – Describes the disease or condition the device will diagnose, treat,

⁶⁹ See 21 CFR 807.92.

⁷⁰ See 21 CFR Part 801.

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 1957 prevent, cure or mitigate, including a description of the target patient population for
1958 which the device is intended and the intended use environment (e.g., intensive care unit,
1959 step-down unit, home).
- 1960 • Instructions for Use – Directions and recommendations for optimal use of the model.
 - 1961 • Clinical benefits (e.g., analyze personalized patient information to improve diagnosis,
1962 treatment assignment, monitoring, or prevention of a medical condition, risk assessment)
1963 and limitations, including whether the device is intended to be used by, or under the
1964 supervision of, a healthcare provider.
 - 1965 • Clinical workflow phase (e.g., patient pre-registration, digitization of forms or clinical
1966 scales, patients' triage, telehealth & virtual rounds, clinical decision support systems,
1967 workflow optimization, evidence-based methods to optimize medical interventions,
1968 feedback from users).
 - 1969 • Inputs and outputs of the model and contribution to healthcare decisions or actions.
 - 1970 • Degree of automation compared to the current standard of care, including whether the
1971 device supports or automates decision making.

Performance and Limitations:

- 1974 • Accuracy (e.g., sensitivity, specificity, positive/negative predictive values, and their 95%
1975 two-sided confidence intervals).
- 1976 • Known biases or failure modes.
- 1977 • Precision (reproducibility) associated with the provided outputs.
- 1978 • Known gaps in the data characterization, such as patient populations that are not well
1979 represented in development (e.g., training) or testing datasets, and therefore, may be at
1980 risk of bias.
- 1981 • Limitations in the model development or performance evaluation.
- 1982 • Known circumstances where the device input will not align with the data used in
1983 development and validation.
- 1984 • Evidence (e.g., clinical trial number or for published results of a supporting study, the
1985 unique reference ID such as Digital Object Identifier, or PubMed Identifier information).
 - 1986 ○ Data Characterization for data used to test the device:
 - 1987 ■ Data sources (e.g., clinical trials, public or proprietary databases)
1988 including details on any devices used to collect data;
 - 1989 ■ Data types used (e.g., structured numerical data, structured categorical
1990 data, unstructured text, images, time-series data, or a combination); and
 - 1991 ■ Relevant details including the sample size, effect size, data quality,
1992 reference standard, diversity, and representativeness.
 - 1993 • Methods used to establish and ensure that the model meets the intended use and user
1994 requirements (e.g., human factors validation/usability evaluation, user acceptance testing,
1995 clinical validation, identification of pre-trained models, other).

Risk Management:

- 1998 • Potential risks associated with the model, the data, and the outputs (e.g.,
1999 contraindications, side effects, data privacy risks, cybersecurity risks, bias risks,
2000 information gaps).
- 2001 • Description of information that could impact risks and patient outcomes, across the

Contains Nonbinding Recommendations

Draft – Not for Implementation

2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024

product lifecycle.

- Interactions, Deployment, and Updates. When appropriate, provide the:
 - Computational Resources Required.
 - Details regarding how the model is deployed and updated, including:
 - How to conduct local site-specific acceptance testing or validation;
 - Ongoing performance monitoring;
 - Transparent reporting of successes and failures;
 - Change management strategies; and
 - Proactive approaches to address vulnerabilities.
 - Communication to parties of as-needed information.
 - Software quality (specify, standards and regulatory compliance issues, intellectual property issues, risk management and safeguards used, other).

Development:

- Data Characterization of data used to develop the device:
 - Data sources (e.g., clinical trials, public or proprietary databases) including details on any devices used to collect data.
 - Data types used (e.g., structured numerical data, structured categorical data, unstructured text, images, time-series data, or a combination).
 - Relevant details including the sample size, effect size, data quality, reference standard, diversity, and representativeness.

2025 **Appendix F: Example 510(k) Summary with Model Card**

2026 In general, publicly available summaries must follow the applicable the requirements for the
2027 specific marketing submission (e.g., 510(k),⁷¹ De Novo,⁷² PMA⁷³). The items below are not an
2028 exhaustive list of topics that a manufacturer may be expected to cover, and all topics may not
2029 apply to all marketing submissions. Likewise, FDA may request additional information to be
2030 included in this summary. This appendix serves as only an example of the types of information
2031 sponsors should generally provide in a 510(k) summary, including an example of a completed
2032 Basic Model Card. Information does not need to be repeated between the model card and other
2033 sections of the public summary, but information can be repeated if the sponsor believes that the
2034 alternate format provides useful context.

2035
2036 **Indications For Use:**

2037
2038 The Disease X screening model is software intended to aid in screening for Disease X on patients
2039 above the age of 22 by analyzing 12-lead electrocardiogram (ECG) recorded from compatible
2040 ECG devices. It is not intended to be a stand-alone diagnostic device for Disease X. However, a
2041 positive result may suggest the need for further clinical evaluation in order to establish a
2042 diagnosis of Disease X. If the patient is at high risk for Disease X, a negative result should not
2043 rule out further non-invasive evaluation. It should not be used to replace the current standard of
2044 care methods for diagnosis of Disease X but applied jointly with clinician judgment.

2045
2046 **Device Description:**

2047
2048 The stand-alone software contains a machine learning model that uses a convolutional neural
2049 network to interpret and analyze 10 seconds of a 12-lead resting electrocardiogram acquired
2050 from 4 compatible ECG devices (A, B, C, and D) and provide an output on the likelihood of
2051 whether a patient has Disease X and further clinical evaluation is required. The software also
2052 contains quality checks that will notify the end user on whether the ECG data provided does or
2053 does not meet the ECG input requirements to generate a model output. If it does not meet the
2054 requirements, an error message will be displayed.

2055 **Summary of Technological Characteristics:**

	Subject Device	Predicate Device	Comparison
Application Number	KXXXXXX	KXXXXXX	-
Product Codes	XXX	XXX	-

⁷¹ See 21 CFR 807.92. For more information, please see FDA guidance titled “[The 510\(k\) Program: Evaluating Substantial Equivalence in Premarket Notifications \[510\(k\)\]](#).”

⁷² See 21 CFR 860.220. For more information, please see FDA guidance titled “[De Novo Classification Process \(Evaluation of Automatic Class III Designation\)](#).”

⁷³ See 21 CFR Part 814.9(e).

Contains Nonbinding Recommendations

Draft – Not for Implementation

	Subject Device	Predicate Device	Comparison
Regulation Number	21 CFR XXXX	21 CFR XXXX	-
Rx/OTC	Rx	Rx	Same
Indication for Use	<p>The Disease X screening model is software intended to aid in screening for Disease X on patients above the age of 22 by analyzing 12-lead <u>electrocardiogram</u> recorded from compatible ECG devices. It is not intended to be a stand-alone diagnostic device for Disease X. However, a positive result may suggest the need for further clinical evaluation in order to establish a diagnosis of Disease X. If the patient is at high risk for Disease X, a negative result should not rule out further evaluation. It should not be used to replace the current standard of care methods for diagnosis of Disease X but applied jointly with clinician judgment.</p>	<p>Software intended to be used as an aid in determining if a patient has Disease X in patients 18 years and above. The software analyzes a 12 lead ECG from compatible devices and should not be used as a stand-alone diagnostic device.</p>	<p>Similar. Both devices are used as aids and screening tools for Disease X. The indications for use for the predicate device is for patients 18 years and above whereas the subject device is for patients 22 and above.</p>
Operational Mode	Spot Check/Not to be used as a diagnostic device	Spot Check/Not to be used as a diagnostic device.	Same

Contains Nonbinding Recommendations

Draft – Not for Implementation

	Subject Device	Predicate Device	Comparison
Hardware Inputs	12 Lead ECG from the following compatible devices: A B C D	12 Lead ECG from the following compatible devices: A B	Similar. While both require inputs from a 12 Lead ECG, the subject device allows for more compatible ECG input devices.
Output	The software provides the following outputs: <ol style="list-style-type: none"> 1. Presence of Disease X. Seek further clinical evaluation to establish a diagnosis of Disease X. 2. Presence of Disease X not likely. However, please use clinical judgment and determine if further evaluation is necessary. 3. Error Message: The 12 lead ECG does not pass the quality checks in place. 	Software provides an output on the possibility of Disease X and if further evaluation is needed.	Similar. Both devices identify if there is presence of Disease X and whether further evaluation is needed. Both identify that it should not be used as a stand-alone and clinical judgment should be used if further evaluation is needed for diagnosis of Disease X.
Ground Truth for Model Training	Echocardiogram	Echocardiogram	Same

Contains Nonbinding Recommendations

Draft – Not for Implementation

	Subject Device	Predicate Device	Comparison
Performance	Sensitivity: 87% (83%, 89%) Specificity: 83% (81%, 85%) Positive Predictive Value (PPV): 56%	Sensitivity: 82% (78%, 85%) Specificity: 81% (79%, 84%) Positive Predictive Value (PPV): 53%	Similar. The subject device has better performance than the predicate device in sensitivity, specificity and PPV.

2056

2057 **Model Training Description:**

2058

2059 The model was trained from a dataset independent from the test dataset. The model was trained
2060 with 30,000 patients that received an ECG and echocardiogram performed within 30 days apart
2061 from one another. The echocardiogram was used to establish clinical reference standard (ground
2062 truth) in patients. The dataset was collected from clinical databases from 2 diverse hospital
2063 networks (Hospital A and Hospital B). Disease X was defined as patients who had a left
2064 ventricular wall thickness ≥ 15 mm based on echocardiographic imaging.

2065 The training dataset contained the following demographic breakdown that was representative of
2066 the disease population:

2067

Race	Percentage (%)
White	75.5
Black or African American	13.6
American Indian or Alaska Native	1.3
Asian	6.3
Native Hawaiian or Pacific Islander	0.3
Two or More Races	3.0

2068

2069 The sample consisted of 49.5% male and 50.5% female participants. The average age was 62
2070 years with the following age breakdown below:

Age (years)	Percentage (%)
Under the age of 40	10
40-49	10
50-59	25
60-69	30
70-79	15
Greater than the age of 79	10

Contains Nonbinding Recommendations

Draft – Not for Implementation

2071 Patients with Disease X were 20% of the overall cohort while patients without Disease X
2072 (control group) consisted of 80% of the overall cohort. Both groups were split into training
2073 (50%), tuning (20%) and tuning evaluation (30%) datasets. The sensitivity and specificity of the
2074 model were calculated from the tuning evaluation datasets. The model was able to achieve the
2075 following:

- 2076 • Sensitivity: 87% (83%, 89%)
- 2077 • Specificity: 83% (81%, 85%)
- 2078 • Positive Predictive Value (PPV): 56%

2079

Summary of Non-Clinical Performance Data

2081 The model was evaluated taking into account applicable requirements of the FD&C Act and
2082 implementing regulations. This included the following testing:

- 2083 • Human Factors and Usability testing was conducted and documentation was provided as
2084 recommended in FDA’s guidance document “[Applying Human Factors and Usability
2085 Engineering to Medical Devices.](#)”
- 2086 • Cybersecurity testing was conducted and documentation was provided as recommended
2087 in FDA’s guidance document “[Cybersecurity in Medical Devices: Quality System
2088 Considerations and Content of Premarket Submissions.](#)”
- 2089 • Software verification and validation testing was conducted and documentation was
2090 provided as recommended in the [Premarket Software Guidance.](#)

2091

Summary of Clinical Validation:

Study Design

2094 The model was validated in a retrospective study of 25,000 patients and their patient records
2095 across 5 different and diverse health systems across the United States. The objective of the study
2096 was to establish the performance of the model on screening for the presence of disease X. The
2097 inclusion criteria for the model were the following:

- 2098 • The patients enrolled in the study were greater than the age of 22 with at least one 12-
2099 lead resting ECG and an echocardiogram within 30 days following the date of the ECG.
2100 The most recent echocardiogram was paired with the most recent ECG for that patient
2101 prior to the echocardiogram.
- 2102 • The following models of ECG devices (A, B, C and D) were used to collect the 12-lead
2103 resting ECG data and used as the inputs to the model.
 - 2104 ○ The 12-lead ECG duration must be 10 seconds long.

2105

2106 The exclusion criteria for the model were the following:

- 2107 • The patients enrolled in the study were less than 22 years old.
- 2108 • Mandatory data were missing (i.e., technical parameters of ECG, age or race
2109 demographic, information regarding the conducted ECG and echocardiogram).
- 2110 • Different device models of 12-lead ECGs were used to collect the ECG data.
- 2111 • The 12-lead ECG duration is not 10 seconds long.
- 2112 • The patient has a pacemaker.

2113

Contains Nonbinding Recommendations

Draft – Not for Implementation

2114 Each of the 5 sites contributed around 5,000 patient-ECG pairs to a final pool of 25,000 patient-
2115 ECG pairs. The study sample had the following demographic breakdown that was representative
2116 of the disease population.

Race	Percentage (%)
White	75.5
Black or African American	13.6
American Indian or Alaska Native	1.3
Asian	6.3
Native Hawaiian or Pacific Islander	0.3
Two or More Races	3.0

2117
2118 The study sample had the following hospital site breakdown:
2119

Hospital Sites	Percentage (%)
A	19.64
B	21.36
C	20.1
D	18.4
E	21.5

2120
2121 The sample consisted of 49.5% male and 50.5% female participants. The average age was 65
2122 years with the following age breakdown below:
2123

Age (years)	Percentage (%)
Under the age of 40	10
40-49	10
50-59	16
60-69	23
70-79	22
Greater than the age of 79	19

2124
2125 The study sample ECG pairs were collected by the following ECG acquisition devices. The
2126 breakdown can be found below:
2127

ECG Acquisition Device	Percentage (%)
A	26.6
B	25.1
C	24.9
D	23.3

2128
2129

Contains Nonbinding Recommendations

Draft – Not for Implementation

2130 Primary Endpoints

2131 The co-primary endpoints regarding this study were to have the lower limits of their 95% two-
2132 sided confidence intervals be:

- 2133 • Sensitivity: 75% or higher
- 2134 • PPV: 50% or higher

2135 2136 Study Results

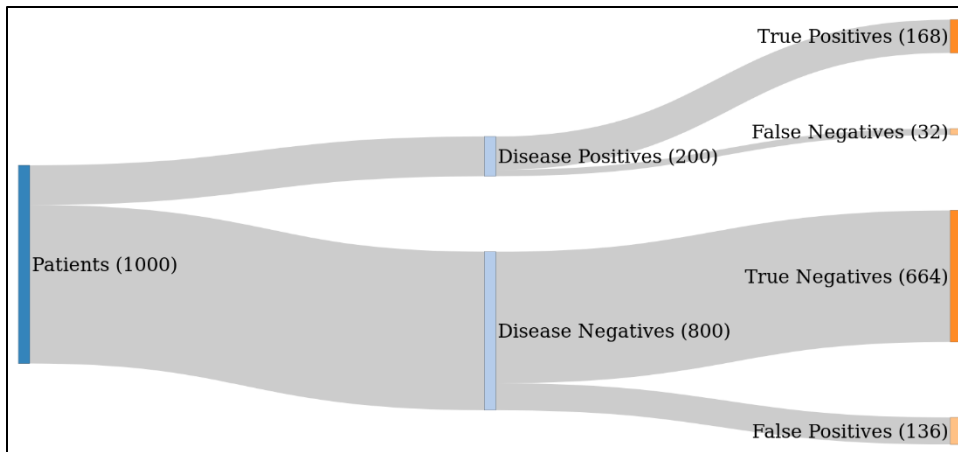
2137 The model achieved a sensitivity of 84%, a specificity of 83%, a PPV of 55%, and a negative
2138 predictive value (NPV) of 95%. Both the point estimates and their 95% two-sided confidence
2139 intervals, along with the confusion matrix, can be reported in a table as shown in the following
2140 example.

	Ref. Pos	Ref. Neg	Sum	Likelihood Ratio	Performance
Test. Pos	4200	3400	7600	4.9 (4.8, 5.1)	PPV= 55.3% (4.9%,654.1%, 56.4%)
Test. Neg	800	16600	17400	0.2 (0.2, 0.2)	NPV=95.4% (95.1% 95.7%)
Sum	5000	20000	25000	1 (1,1)	Prevalence 20% (19.5%, 20.5%)
Performance	Sensitivity = 84% (82.9%, 85%)	Specificity = 83% (82.5%, 83.5%)			

2141
2142 Plain Language to Interpret the Study Results for Benefit Risk Consideration
2143 Assume the prevalence of Disease X in the intended use population of the device is 20%. Among
2144 1000 patients from the target population, about 168 ($1000 \times \text{Prevalence} \times \text{Sensitivity}$) patients
2145 will be correctly classified as having the Disease X (i.e., 168 device true positives out of 200
2146 total reference positive patients), while about 136 ($1000 \times (1 - \text{Prevalence}) \times (1 - \text{Specificity})$)
2147 patients will be wrongly classified as having the Disease X (i.e., 136 device false positives out of
2148 800 total reference negative patients). Thus, each true positive patient comes at the cost of 0.8
2149 ($136/168$) false positive patients (compares to Y from the standard of care, or 4 ($800/200$) from a
2150 worst-case scenario where every patient is called positive). Furthermore, to identify one extra
2151 true positive patient, we need to assess about two patients (considering potential device
2152 positive/negative outcomes) since $\text{NNP (Number Needed to Predict)} = 1 / (\text{PPV} + \text{NPV} - 1) =$
2153 1.97 (compares to the standard of care with NNP of Z, or a perfect device with a NNP of one).
2154

Contains Nonbinding Recommendations

Draft – Not for Implementation



2155
2156
2157
2158
2159
2160
2161
2162
2163

The subgroup analysis for each demographic can be found below.

Please note that while confidence intervals could not be generated for this fictitious example, sponsors should include confidence intervals on all reported results. Placeholders have been included in each cell to represent the confidence interval: (X_{ll}, X_{ul}) , where “ll” stands for lower limit and “ul” stands for upper limit.

Race	Percentage (%)	Sensitivity	PPV
White	75.5	85.3 (X_{ll}, X_{ul})	57.3% (X_{ll}, X_{ul})
Black or African American	13.6	82.9 (X_{ll}, X_{ul})	54.4% (X_{ll}, X_{ul})
American Indian and Alaska Native	1.3	81.6 (X_{ll}, X_{ul})	54.8% (X_{ll}, X_{ul})
Asian	6.3	83.9 (X_{ll}, X_{ul})	56.1% (X_{ll}, X_{ul})
Native Hawaiian and Other Pacific Islander alone	.3	83.6 (X_{ll}, X_{ul})	56.5% (X_{ll}, X_{ul})
Two or More Races	3	84.1 (X_{ll}, X_{ul})	55.4% (X_{ll}, X_{ul})
Age	Percentage (%)	Sensitivity	PPV
Under the age of 40	10	84.9 (X_{ll}, X_{ul})	55% (X_{ll}, X_{ul})
40-49	10	85.1 (X_{ll}, X_{ul})	55.4% (X_{ll}, X_{ul})
50-59	16	84.1 (X_{ll}, X_{ul})	55.4% (X_{ll}, X_{ul})
60-69	23	84.5 (X_{ll}, X_{ul})	56% (X_{ll}, X_{ul})
70-79	22	83.6 (X_{ll}, X_{ul})	55.4% (X_{ll}, X_{ul})
Greater than the age of 79	19	82.1 (X_{ll}, X_{ul})	52.7% (X_{ll}, X_{ul})

Contains Nonbinding Recommendations

Draft – Not for Implementation

2164 The subgroup analysis for each ECG acquisition device can be found below:

ECG Acquisition Device	Percentage (%)	Sensitivity	PPV
A	26.6	84.7 (X_{ll} , X_{ul})	56.5% (X_{ll} , X_{ul})
B	25.1	83.6 (X_{ll} , X_{ul})	54.3% (X_{ll} , X_{ul})
C	24.9	85.4 (X_{ll} , X_{ul})	57.9% (X_{ll} , X_{ul})
D	23.3	84.6 (X_{ll} , X_{ul})	55.1% (X_{ll} , X_{ul})

2165

2166 The subgroup analysis for each hospital site can be found below:

Hospital Sites	Percentage (%)	Sensitivity	PPV
A	19.64	83.6 (X_{ll} , X_{ul})	54.3% (X_{ll} , X_{ul})
B	21.36	85.1 (X_{ll} , X_{ul})	51.4% (X_{ll} , X_{ul})
C	20.1	84.1 (X_{ll} , X_{ul})	55.4% (X_{ll} , X_{ul})
D	18.4	85.4 (X_{ll} , X_{ul})	57.9% (X_{ll} , X_{ul})
E	21.5	84.7 (X_{ll} , X_{ul})	56.5% (X_{ll} , X_{ul})

2167

2168 **Model Card:**

2169

2170 **Device Information**

- 2171 • Model Name: Disease X Screening Model
- 2172 • Model version: version 1.0.1
- 2173 • Model release date: December 2023
- 2174 • Model architecture: Convolutional Neural Network

2175

2176 **Device Description**

- 2177 • Intended User: Healthcare professionals.
- 2178 • Indications for Use: The model is software intended to aid in screening for Disease X on
- 2179 patients above the age of 22 by analyzing recordings of 12-lead ECG made on compatible
- 2180 ECG devices. It is not intended to be a stand-alone diagnostic device for Disease X.
- 2181 However, a positive result may suggest the need for further clinical evaluation in order to
- 2182 establish a diagnosis of Disease X. If the patient is at high risk for Disease X, a negative
- 2183 result should not rule out further non-invasive evaluation. It should not be used to replace
- 2184 the current standard of care methods for diagnosis of Disease X but applied jointly with
- 2185 clinician judgment.

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 2186
- 2187
- 2188
- 2189
- 2190
- Clinical workflow phases: To be used as an aid and screening tool for further clinical follow-up (e.g., echocardiogram) in order to establish a diagnosis of Disease X.
 - Clinical Benefit: To provide point-of-care screening of Disease X where cardiac imaging may not be available.

Performance and Limitations

- 2191
- 2192
- 2193
- 2194
- 2195
- 2196
- 2197
- 2198
- 2199
- 2200
- 2201
- 2202
- 2203
- 2204
- 2205
- 2206
- 2207
- 2208
- 2209
- 2210
- 2211
- 2212
- 2213
- 2214
- 2215
- 2216
- 2217
- 2218
- 2219
- 2220
- 2221
- 2222
- 2223
- 2224
- 2225
- Data type: 12-lead electrocardiogram (ECG)
 - Description: 10 second duration of a 12-lead electrocardiogram (ECG) obtained from the following four compatible ECG devices (A, B, C, and D). The compatible ECG devices have a sampling rate of 500 Hz.
 - Clinical Reference Standard: An echocardiogram obtained within 30 days of the ECG to establish clinical reference standard.
 - Model Validation:
 - Data size and type: A retrospective study of 25,000 patients and their patient records across 5 different and diverse health systems across the United States. Each of the 5 sites contributed 5,000 patient-ECG pairs to a final pool of 25,000 patient-ECG pairs.
 - Exclusion Criteria:
 - The patients enrolled in the study were less than 22 years old.
 - Mandatory data were missing (i.e., technical parameters of ECG, age or race demographic, information regarding the conducted ECG and echocardiogram).
 - ECG data contained either corrupt or missing lead(s).
 - Different models of 12-lead ECGs were used to collect the ECG data.
 - The 12-lead ECG duration is not 10 seconds long.
 - The patient has a pacemaker.
 - Data Results (calculated from test datasets):
 - Sensitivity: 84% (82.9%, 85%)
 - Specificity: 83% (82.5%, 83.5%)
 - PPV: 55.3% (54.1%, 56.4%)
 - NPV: 95.4% (95.1%, 95.7%)
 - Non-Clinical Testing:
 - Human Factors and Usability testing was conducted and documentation was provided as recommended in FDA’s guidance document “[Applying Human Factors and Usability Engineering to Medical Devices](#).”
 - Cybersecurity testing was conducted and documentation was provided as recommended in FDA’s guidance document “[Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions](#).”
 - Software verification and validation testing was conducted and documentation was provided as recommended in the [Premarket Software Guidance](#).

2226

2227

2228

Contains Nonbinding Recommendations

Draft – Not for Implementation

2229 **Risk Management:**

2230

2231 Risk management was conducted, and documentation was provided as recommend in the
2232 [Premarket Software Guidance](#) and in accordance with ANSI/AAMI/ISO 14971 Medical devices
2233 - Applications of risk management to medical devices.

2234 • Potential risks associated with the model, the data, and the outputs (e.g.,
2235 contraindications, side effects, data privacy risks, cybersecurity risks, bias risks,
2236 information gaps): The potential risks associated with the model include incorrect follow-
2237 up due to a false positive or false negative output, which can occur because of (1) model
2238 bias or (2) using the model in an unsupported patient population or with unsupported
2239 input/hardware. Furthermore, information gaps may lead to overreliance on the device
2240 output for follow-up. Controls for identified risks include clinical validation testing,
2241 software verification and validation testing, human factors testing and labeling.

2242 • Description of information that could impact risks and patient outcomes, across the
2243 product lifecycle: Model development and clinical validation included only 10% of
2244 participants under the age of 40, which may mean that the model’s performance on that
2245 subgroup is not fully characterized.

2246 • Interactions, Deployment, and Updates: A comprehensive Device Performance
2247 Monitoring Plan is in place that is consistent with the Quality System Regulation (21
2248 CFR Part 820) which continuously monitors the deployed model to evaluate site-specific
2249 performance, identify vulnerabilities, and ensure transparency of performance and
2250 ongoing maintenance to sites and end users.

2251 ○ Computational resources required.

2252 ○ Details regarding how the model is deployed and updated:

2253 ■ How to conduct local site-specific acceptance testing or validation: Prior
2254 to use of the model in the site’s entire population, the model is deployed,
2255 and data is collected for a one-month period in order to understand any
2256 issues with integration into the sites’ existing systems and measure
2257 performance on a subset of the patient population for that site. Through
2258 this process, issues with deployment can be addressed prior to exposure to
2259 the entire population and can help characterize performance of the model
2260 and the need for additional training and development. Alternatively, sites
2261 may opt to provide historical data that can be used to assess expected
2262 performance at the site.

2263 ■ Ongoing performance monitoring: Automated performance calculation is
2264 deployed along with the model and calculated every 6 months; if the
2265 performance is out of the expected range, an automated e-mail will be sent
2266 to the site administrator and sponsor. This will initiate a process for
2267 understanding performance issues and a mitigation plan will be put in
2268 place to address this.

2269 ■ Transparent reporting of successes and failures: All sites will have access
2270 to anonymized reports that will include successes and failures of deployed
2271 models at various sites, along with site characteristics to contextualize
2272 these successes and failures.

2273 ■ Change management strategies: Change management will be implemented

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 2274 consistent with established Quality System procedures if and when issues
2275 arise that require a change or if features are requested by sites and users.
2276 ▪ Proactive approaches to address vulnerabilities: Sites and users are
2277 encouraged to report any issues within 48 hours of the issue occurring,
2278 which will then follow complaint handling procedures and for which a fix
2279 will be issued according to these procedures.
- 2280 ○ Communication to parties of as-needed information: Automated e-mails will be
2281 generated by the device when performance is out of the expected range, as
2282 described above.
 - 2283 ○ Software quality (specify, standards and regulatory compliance issues, intellectual
2284 property issues, risk management and safeguards used, other):

Development:

- 2285
- 2286
- 2287
- 2288 • Model Training:
 - 2289 ○ Data size: 30,000 patients that received an ECG and echocardiogram performed
2290 within 30 days apart from one another. Dataset collected from clinical databases
2291 from 2 diverse hospital networks (Hospital A and Hospital B).
 - 2292 ○ Patients with Disease X were 20% of the overall cohort while patients without
2293 Disease X (control group) consisted of 80% of the overall cohort. Both groups
2294 were split into training (50%), tuning (20%) and tuning evaluation (30%) datasets.
 - 2295 ○ Data Results (calculated from tuning evaluation datasets):
 - 2296 ▪ Sensitivity: 87% (83%, 89%)
 - 2297 ▪ Specificity: 83% (81%, 85%)

Conclusion:

2300

2301 While there are differences noted in the technological characteristics of the proposed system and
2302 the predicate device, the differences do not raise different questions of safety or effectiveness.
2303 Based on the information provided in this submission, the subject device demonstrates that it is
2304 substantially equivalent to the predicate device through the results of clinical performance and
2305 results of non-clinical verification and validation.

2306