BROOKE & ASSOCIATES

April 7, 2025

BY ELECTRONIC SUBMISSION

Dockets Management Staff (HFA–305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

RE: Comments Regarding *Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations—Draft Guidance for Industry and Food and Drug Administration Staff* (Docket # FDA-2024-D-4488)

To Whom It May Concern:

Thank you for the opportunity to submit to the FDA these comments on the recent draft guidance document titled *Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations—Draft Guidance for Industry and Food and Drug Administration Staff*[1] on behalf of Brooke & Associates.[2]

Brooke & Associates is a boutique law and advisory firm that provides legal and consulting services to innovative medical device, digital health, and digital therapeutics companies across the globe and at various stages of commercialization. We advise clients of all sizes and clinical, therapeutic, and technological domains to address issues across a broad spectrum—from counseling on compliance with legal and regulatory requirements to providing practical guidance on product-specific market authorization strategies to supporting post-market surveillance activities. Our team represents a diverse set of stakeholders reflective of the evolving device industry and includes attorneys, scientists, engineers, clinical, and regulatory affairs professionals, government affairs professionals, and former FDA personnel. Our firm has a particular focus on hardware and software medical devices incorporating wearable sensors, artificial intelligence/machine learning, and virtual, augmented, and mixed reality technologies. In this domain, we have served as an advocate for development of smart regulatory policy for more than 15 years.

[1] U.S. FOOD & DRUG ADMIN., ARTIFICIAL INTELLIGENCE-ENABLED DEVICE SOFTWARE FUNCTIONS: LIFECYCLE MANAGEMENT AND MARKETING SUBMISSION RECOMMENDATIONS—DRAFT GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2025), *available at* https://www.fda.gov/media/184856/download [hereinafter AI LIFECYCLE MANAGEMENT DRAFT GUIDANCE].

[2] These comments solely represent the opinions of Brooke & Associates and do not represent the views of any of the firm's clients.

To that end, while we applaud the FDA for preparing a well-written, cogent set of expectations for lifecycle management of and premarket submissions involving AI-enabled device software functions ("AI-DSFs"), we respectfully submit the comments in this letter and the attached Table 1 for consideration by the Agency during its effort to finalize the guidance. To summarize, our comments reflect concerns regarding the FDA's expectations related to 1) the reference standard required to evaluate the performance of the AI model, 2) demonstrating generalizability of the AI model, 3) subgroup analyses and the use of unpowered subgroup analyses to make decisions regarding safety and effectiveness of the AI-DSF, and 4) labeling to achieve transparency.

* * *

Thank you in advance for your time and consideration of these comments. If you have any questions or require additional information, please do not hesitate to contact me via phone or email at the contact information provided below.

Sincerely yours,

Signed by:
M. Jason Brooke
F43BACF6CEBF454...

M. Jason Brooke, MSE, JD, CSQE
Attorney & Managing Member
Brooke & Associates
+1.202.258.1422
jbrooke@devicecounsel.com

Table 1: Specific Line-Item Comments on the AI Lifecycle Management Draft Guidance

#	Page/Line #	Draft Guidance Language	Comments
1	Page 2, Lines 121-125	Furthermore, the guidance includes FDA's current thinking on strategies to address transparency and bias throughout the TPLC of AI-enabled devices, including by collecting evidence to evaluate whether a device benefits all relevant demographic groups (e.g., race, ethnicity, sex, and age) similarly, to help ensure that these devices remain safe and effective for their intended use.	It is not clear why the FDA expects that "all relevant demographic groups" are benefited "similarly". The FDA should explain the basis for this expectation, including providing a definition of "similarly" and provide guidance on how a Sponsor can determine whether their device meets this "similarly" standard. Specifically, the Agency should explain what degree of similarity is acceptable across demographic groups. Does the FDA expect to see statistical analyses across demographic groups? If so, what degree of statistical significance is required? If not, how is acceptable similarity determined?
2	Page 3, Lines 169-174	For AI-enabled devices subject to 510(k) requirements, an AI-enabled device can be found substantially equivalent to a non-AI-enabled device with the same intended use provided, among other things, the AI-enabled device does not introduce different questions of safety and effectiveness compared to the non-AI-enabled device and meets other requirements for a determination of substantial equivalence in accordance with section 513(i) of the FD&C Act.	The FDA should provide examples of when the use of AI technology itself introduces different questions of safety or effectiveness and when it does not. Sponsors need to be able to independently assess whether and to what extent the use of AI will impact the premarket submission pathway (i.e., 510(k), De Novo, or PMA) and whether and to what extent modifications to a device that contains AI requires a new premarket submission. As such, in addition to adding clarity in this guidance, we recommend updating existing final guidance to ensure considerations related to AI-enabled device have been incorporated, including the following: • *The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)]* (2014), • *Deciding When to Submit a 510(k) for a Change to an Existing Device* (2017), and • *Deciding When to Submit a 510(k) for a Software Change to an Existing Device* (2017).
3	Page 7, Lines 336-337	For example, the QS Regulation requires that manufacturers establish design controls for certain finished devices (see 21 CFR 820.30).	By stating that design controls are required "for certain finished devices", the FDA creates an impression that in some cases an AI-enabled device would not require compliance with the Quality System Regulation. Given that 21 C.F.R. § 820.30(a)(2) indicates that all Class III, Class II, and Class I devices that "are automated with computer software" must be developed in compliance with design controls, which suggests that all AI-DSFs would require design controls (unless a device-type-specific exemption exists), the Agency should clarify when, if ever, an AI-DSF would <u>not</u> require compliance with the Quality System Regulation and provide examples.

#	Page/Line #	Draft Guidance Language	Comments
4	Page 8, Lines 349-352	Further, manufacturers have ongoing responsibility to manage the quality system and maintain device quality, including by reviewing the "suitability and effectiveness of the quality system at defined intervals and with sufficient frequency according to established procedures" to ensure the quality objectives are being met.	To be comprehensive, the FDA should include a discussion of the consequences for manufacturers (and other regulated entities) who do not comply with the Quality System Regulation, including those who do not manage the Quality System and maintain device quality. Alternatively, the Agency should reference relevant statutes or regulations that establish such consequences.
5	Page 8, Lines 364-366	The device description provides important context about what the device does, including how it works, how a user may interact with it, and under what circumstances a device is likely to be used as intended.	The FDA should add "foreseeable misuse" to the end of this sentence such that the sentence reads as follows: *The device description provides important context about what the device does, including how it works, how a user may interact with it, under what circumstances a device is likely to be used as intended, and any foreseeable misuse associated with the device.*
6	Page 9, Lines 402-422	Additionally, sponsors should include the following types of information as part of a device description for an AI-enabled device that has elements that can be configured by a user: • A description of all configurable elements of the Al-enabled device . . . • A description of how these elements and their settings can be configured . . . • A description of the potential impact of the configurable elements on user decision making.	While the FDA mentions the potential for AI-enabled devices to contain user-configurable elements, the guidance only discusses this aspect in the context of including information about those user-configurable elements in the device description. The Agency should further describe its expectations regarding user-configurable elements and discuss the impact of such elements of the device on the regulatory requirements throughout the product lifecycle. For example, the FDA should explain its expectations regarding testing of user-configurable elements and whether and to what extent clinical validation testing is required for such elements.
7	Page 9-10, Lines 428-430	Sponsors may also wish to consider enhancing the device description with the use of graphics, diagrams, illustrations, screen captured images, or video demonstrations, including screen captured video.	The FDA should clarify the preferred method for incorporating videos into the device description. For example, does the FDA expect to receive video files (e.g., AVI or MPEG4 files) or links to YouTube videos?

#	Page/Line #	Draft Guidance Language	Comments
8	Pages 12-14, Lines 546-593	The labeling for an AI-enabled device should address the following types of information in a format and at a reading level that is appropriate for the intended user (e.g., considering characteristics such as age, education or literacy level, sensory or physical impairments, or occupational specialty) to help ensure users can quickly access important information. . . . Model Development Data • Description of the development data, including: o The source(s) of data; o Study sites; o Sample size; o Demographic distributions; and o Criteria/expertise used for determining clinical reference standard (ground truth).	The FDA should explain why inclusion of model development data are critical to include in public-facing labeling. In some cases, information related to model development are proprietary/trade secret information and manufacturers of AI-enabled medical devices should not be required to disclose such information to the public. While it is reasonable for the Agency to want to review model development data in a confidential manner, requiring disclosure of such information in the labeling is unreasonable and overly burdensome. For example, in the context of a physical medical device, the FDA does not require the disclosure of confidential information about material properties or trade secret manufacturing processes that are employed to build the device. Similarly, the FDA does not require the disclosure of early feasibility testing that is performed as part of the product development process for traditional medical devices. Yet, the Agency is requiring such analogous information to be included in public labeling for AI-enabled medical devices. To demand such information, the FDA should justify why the information is uniquely required for AI-enabled medical devices.

9	Page 14, Lines 614-616	Explanation of the device performance across important subgroups. Generally, subgroup analysis by patient characteristics (e.g., sex, gender, age, race, ethnicity, disease severity), geographic sites, and data collection equipment are appropriate.	The FDA's expectations for the reporting of subgroup analyses in labeling is unreasonably burdensome and will result in misleading conclusions drawn by patients, clinicians, the general public, and the Agency itself. The FDA should justify why subgroup analyses must be reported in labeling and the scientific, statistical, and clinical basis for reporting such information in the labeling. It is well understood that *power* in a statistical context is, simply put, the probability that a device will produce a correct output (e.g., detect a disease if the disease is present). When an analysis lacks statistical power, the results of that analysis cannot be relied upon to make any meaningful conclusions, meaning that if the results of the analysis indicate that the device is reasonably good at detecting a disease, the reality might be that the device is not actually that good at detecting the disease. Likewise, if the results of the analysis indicate that the device is bad at detecting a disease, the lack of power means that the reality might be that the device is actually quite good at detecting the disease. In other words, the results of the analysis cannot be relied upon to draw a conclusion as to the quality of the device output. As the FDA is aware, in most cases the subgroup analyses are not designed to be statistically powered (see Page 31, Lines 1258-1260). Yet the Agency requires such subgroup analyses to evaluate whether the results of a powered analysis on a population level differ from the results of the unpowered analysis on a subgroup level. For example, take a hypothetical where the powered analysis shows that the device is quite good at detecting a condition across a patient population aged 18-75 years of age such that the sensitivity of the device is 92% with a 95% confidence interval ranging from 88% to 95%. The FDA would expect to see unpowered subgroup analyses for patients ranging from, for example, 18-25 yo, 26-35 yo, and so on. In any one of the numerous unpowered analyses for age subgroups, the result might indicate that the device is not so great at detecting the condition. To be illustrative, the subgroup analysis might indicate that for 18-25 year olds, the sensitivity is 75% with a 95% confidence interval ranging from 60% to 85%. The FDA may draw the conclusion that the device performance is diminished in this subgroup relative to the overall patient population but such a conclusion would be unfounded and unsupported by basic statistical principles. To be sure, the FDA in drawing such a conclusion might be correct, but the Agency might be equally incorrect. Requiring such unpowered subgroup analyses to be included in the product labeling will likely cause patients, clinicians, and the general public to mistakenly come to similar incorrect conclusions, leading to confusion and misjudgments as to whether or not to rely on the device in specific subgroups. As such, the Agency should justify why such information must be included in the labeling. Furthermore, to the extent subgroup analysis information must be presented in labeling, the

#	Page/Line #	Draft Guidance Language	Comments
			FDA should provide guidance on how a Sponsor can determine what is considered to be an "important subgroup".
10	Page 14, Lines 617-618	Description of the corresponding performance for different operating points, including subgroup analysis for each operating point, as applicable.	The FDA should provide specific examples of the expected subgroup analyses for each operating point and explain why such detailed analyses are necessary to determine the safety and effectiveness of the subject device. As noted elsewhere in these comments, subgroup analyses (particularly those that are not statistically powered) provide no meaningful information upon which to draw conclusions about the safety or effectiveness of the device. Requiring analyses across numerous subgroups for *each* operating point is an overly burdensome demand that will result in massive amounts of data analysis that are not necessary to make a safety and effectiveness determination.
11	Pages 14-15, Lines 628-633	Some limitations of a model may not reach the degree of severity that would warrant a contraindication, warning, or precaution, but they may still be important to include in labeling. For example, the training dataset may have only included a few patients with a rare presentation of a disease or condition; users may benefit from knowing the limitations of the data when that rare presentation is suggested by the model as a diagnosis.	If the FDA is requiring the inclusion of limitations in the labeling that do not reach the degree of severity that would warrant a contraindication, warning, or precaution, the Agency must explain its legal authority for requiring such information. While the FDA has long required the inclusion of "indications for use and appropriate contraindications, warnings, precautions and adverse reaction information" in the labeling as part of the directions for use (see *Device Labeling Guidance #G91-1*), demanding additional information that does not fall within this scope is a new labeling requirement. In addition to explaining the legal basis for such a new requirement, the FDA should provide guidance on how a device manufacturer would determine what is considered to be important enough to be included in the labeling but not important enough to be considered a contraindication, warning, or precaution. To that end, the FDA should provide guidance on how much benefit to the users triggers the requirement to include such information in the labeling and how to quantify the benefit in such a what that such a determination can be made. Without such guidance, the device manufacturer will be subject to the whim of the FDA reviewers during the premarket review process meaning that device manufacturers will effectively be held hostage unless the labeling includes whatever information the FDA reviewer requires. Such an approach will lead to arbitrary and inconsistent requirements where some manufacturers include certain limitations while other manufacturers do not. Such inconsistencies will lead to confusion in the marketplace and unfairly and capriciously advantage one manufacturer over another.

#	Page/Line #	Draft Guidance Language	Comments
12	Page 18, Lines 767-770	Using unbiased, representative training data for models promotes generalizability to the intended use population and avoids perpetuating biases or idiosyncrasies from the data itself. For example, in image recognition tasks, confounding may occur when all the diseased cases are imaged with the same instrument, or with a ruler included (e.g., on clinical images of melanoma).	The FDA should clarify that the concern about perpetuating bias as a result of using a single or limited set of instruments for generating training (and validation) data is irrelevant in the situation where the AI model is meant to be used only with the single or limited set of instruments that was used data generation. For example, the FDA should clarify that, if a manufacturer of a wearable device sensor creates an AI-enabled software device that analyzes only data from that sensor, the fact that the training (and validation) data are generated from that wearable sensor only does not indicate that the AI model has an inappropriate, inherent bias such that the results of the model would not be adequately generalizable.
13	Pages 18-19, Lines 770-773	Another example of a potential confounding factor is the use of data collected outside the U.S. (OUS) in training, which may bias the model if the OUS population does not reflect the U.S. population due to differences in demographics, practice of medicine, or standard of care.	For many companies, it is unduly burdensome for the FDA to require that model training rely primarily on US data. It is often cheaper, faster, and logistically more efficient to conduct training based on OUS data (especially for foreign manufacturers). If the model has inherently detrimental bias (whether from the use of primarily OUS data in training or for some other reason), the validation testing (if designed properly) will demonstrate such bias. If the model is not detrimentally biased by the training data, then the performance metrics at the validation testing stage will satisfy the pre-specified acceptance criteria, which should be sufficient for the FDA's determination of safety and effectiveness. The FDA should explain (with real-world evidence) why the use of OUS data in training is likely to result in bias that cannot be accounted for at the validation testing stage and why the Agency is not able to determine safety and effectiveness if the pre-specified acceptance criteria are satisfied at the validation testing stage.
14	Page 19, Lines 777-779	The inclusion of representative data in validation datasets may be important, because underrepresentation may impact the ability to identify any performance problems, including understanding performance in underrepresented populations.	The FDA should explain what it considers to be underrepresentation in the context of validation datasets. Providing examples or more detailed guidance will help manufacturers better understand the threshold for underrepresentation so as to avoid it in the testing protocols.

#	Page/Line #	Draft Guidance Language	Comments
15	Page 19, Lines 796-799	To objectively assess the device performance, it is also important for FDA reviewers to understand whether the test data are independent (e.g., sampled from completely different clinical sites) from the training data and are sequestered from the model developers and the model development stage.	The FDA should explain why "completely different clinical sites" are necessary to establish independence in the datasets. Such a requirement creates a major bar (especially for small manufacturers) to obtaining market authorization and is not absolutely necessary to evaluating safety and effectiveness. Importantly, the FDA does not require non-AI-enabled devices to conduct pivotal studies at "completely different clinical sites" from those sites involved in pilot-stage studies. To be sure, Section 6.6 of *Design Considerations for Pivotal Clinical Investigations for Medical Devices* speaks to various considerations when selecting sites for a pivotal study (e.g., the use of multiple sites as well as the need for diversity in the study subjects and clinical study investigators) but does not indicate that the pivotal study sites must be "completely different" from the pilot study sites. The FDA should explain why this requirement exists for AI-enabled medical devices but not other medical devices.

#	Page/Line #	Draft Guidance Language	Comments
16	Page 21, Lines 855-856	For the purposes of this guidance, a reference standard is the best available representative truth that can be used to define the true condition for each patient/case/record.	The FDA should explain how the requirement that the reference standards be the "best available" meets the least burdensome requirement. As the FDA is aware, Section 513(i)(1)(D)(i) of the Federal Food, Drug & Cosmetic Act states that, for example, in the context of a 510(k) premarket notification review, the FDA "shall only request information that is necessary to making substantial equivalence determinations. In making such [information] requests, the [FDA] shall consider the least burdensome means of demonstrating substantial equivalence and request information accordingly." Similar least burdensome provisions exist for other premarket reviews. See, e.g., Sections 513(a)(3)(D)(ii)-(iii) and 515(c)(5)(A) of the Federal Food, Drug & Cosmetic Act. As such, the FDA is prohibited from requiring reference standards that are not "necessary to making substantial equivalence determinations" or from requiring a study design that relies on anything but the minimum reference standard required to demonstrate substantial equivalence (in the context of a 510(k) submission). By requiring the best available reference standard, the FDA is inherently requiring more than the bare minimum. For example, if the validation testing of the AI model involves a clinical investigation wherein a panel of readers is used to determine the ground truth, the "best available" reference standard may involve several readers (e.g., 5 or more). However, the minimum number of readers on the panel to establish an appropriate level of adjudication may only require 3 readers. In such a situation, the FDA's requirement for the "best available" reference standard does not meet the least burdensome requirement. Similarly, in a situation where a panel of 3 readers is used, the "best available" reference standard from a qualification's standpoint might be board certified, actively practicing physicians; however, the minimum qualifications "necessary" may be a retired physician or a physician's assistant or some other healthcare professional with less training than the "best available". Again, requiring the "best available" qualifications in establishing the reference standard does not satisfy the least burdensome requirement. As such, the FDA should explain how it's "best available" reference standard requirement satisfies the least burdensome requirement. To the extent that the "best available" reference standard requirement meets the least burdensome requirements, the FDA should provide clarity as to how a manufacturer can determine whether the best available reference standard is being used, providing guidance on the selection of reference standards and characteristics to consider when selecting a reference standard.

#	Page/Line #	Draft Guidance Language	Comments
17	Page 21, Lines 857-859	A reference standard is validated by evidence from current practice within the medical and regulatory communities for establishing a patient's true status with respect to a clinical task.	The FDA should clarify how it makes a determination that a reference standard is validated. Importantly, the FDA should provide examples and discuss whether (and under what circumstances) the Agency would (or would not) accept a legally marketed medical device to be a reference standard. For example, if a manufacturer of a heart rate monitor is seeking 510(k) clearance and wishes to rely on an FDA-cleared heart rate monitor from another manufacturer, it is not clear whether the Agency would accept the FDA-cleared heart rate monitor to serve as the validated reference standard. If the Agency would not allow such a legally marketed device to serve as the validated reference standard, the FDA should provide a justification and how requiring another reference standard meets the least burdensome requirements.
18	Page 22-23, Lines 909-923	An explanation of how the data is representative of the intended use population and indications for use, including: . . . Test data collection sites (e.g., clinical sites, institutions).	The FDA should clarify how the expectations for representativeness in the dataset applies to the use of retrospective data for validation testing. Does the FDA have different expectations for data collection sites if the validation testing is performed using retrospective data versus prospectively collected data?
19	Page 23, Lines 945-948	Due to the data-driven nature of typical models and the obscurity of their algorithms to end users, their generalized performance on the U.S. target population may not be adequately captured in the clinical study if a significant portion of the validation data are OUS data.	The FDA should clarify what constitutes a "significant portion" of the validation data. Does the majority (i.e., at least 51%) of the validation data need to come from the US?
20	Page 25, Lines 1022-1024	An explanation of any pre-trained models that were used, as applicable. If a pre-trained model was used, specify the dataset that was used for pre-training and how the pre-trained model was obtained.	The FDA should clarify what it considers to be a "pre-trained" model and what information must be provided if the pre-trained model was developed by a third party. Likewise, the FDA should clarify its expectations if the third party (e.g., a developer of a general-purpose AI model) does not provide information about the dataset used for pre-training or the process used to develop the pre-trained model.
21	Page 27, Lines 1087-1088	New unique device identifiers (UDIs) are required for devices that are required to bear a UDI on its label when there is a new version and/or model, and for new device packages.	The FDA should clarify whether the use of the term "model" in this sentence refers to a "model number" that is used to catalog the device itself or to the "AI model" specifically.

#	Page/Line #	Draft Guidance Language	Comments
22	Page 27, Lines 1098-1100	As part of FDA's evaluation of safety and effectiveness of the device, it is important for FDA to understand how the device performs overall in the intended use population, as well as in subgroups of interest.	The FDA should clarify what it considers to be a "subgroup of interest". The FDA should consider and discuss how the "subgroups of interest" are linked to the intended use (e.g., generic versus specific claims) and risk profile of the device.
23	Page 28, Lines 1125-1129	Subgroup analysis provides the tools to evaluate the performance of the device in specific populations and can be helpful in identifying scenarios in which the device performs worse than overall performance. In addition, subgroup analyses are helpful in identifying potential limitations of the device and can contribute to effective labeling by providing end users with additional useful information.	As discussed elsewhere in these comments, subgroup analyses that are not statistically powered cannot provide evidence of better or worse performance as compared to overall population-based performance. The FDA should justify why the Agency believes poorly powered subgroup analyses can be used to identify scenarios in which the device performs worse than overall performance. Similarly, as discussed elsewhere in these comments, the FDA should explain why it believes that providing misleading conclusions about subgroup analyses should be included in labeling.
24	Page 31, Lines 1258-1262	[W]hen specific subgroup performance claims are not made, subgroup performance does not need to be statistically powered for each subgroup, but effort should be made to include reasonable numbers of patients for each subgroup so that any reported results have meaning and context.	The FDA should explain what it considers to be "reasonable numbers of patients for each subgroup" and what "meaning and context" can be gleaned from subgroup analyses that are not statistically powered. As noted elsewhere in these comments, the reporting of unpowered subgroup analyses as a means of drawing conclusions about comparisons to the overall population are misleading and can cause confusion that risks patient safety. The FDA should explain why such misleading and confusing information is necessary to establish substantial equivalence or a reasonable assurance of safety and effectiveness.
25	Page 31, Lines 1264-1268	To support performance validation, sponsors should include information regarding the study results. Important aspects for these documents to cover include: • An explanation of the pre-specified results for each test, including subgroup analyses. • An explanation of the results with adequate subgroup analyses for relevant subgroups as described above.	The FDA should clarify the difference between "an explanation of the pre-specified results for each test" and "an explanation of the results with adequate subgroup analyses".

#	Page/Line #	Draft Guidance Language	Comments
26	Page 32, Lines 1291-1295	In general, as part of the quality system for a medical device, including an AI-enabled device, manufacturers should have a postmarket performance monitoring plan to help identify and respond to changes in performance in a postmarket setting. The inclusion of a performance monitoring plan in the marketing submission may help to reduce uncertainty and support FDA's evaluation of risk controls.	The FDA should clarify under what circumstances and on what legal basis is a post-market performance monitoring plan required. In cases where the AI-enabled device is locked and non-adapting, a post-market performance monitoring plan is not necessary to establish substantial equivalence or a reasonable assurance of safety and effectiveness.
27	Page 33, Lines 1309-1313	FDA generally does not assess quality system regulation compliance as part of its review of marketing submissions under section 510(k) of the FD&C Act. However, in some cases, it may be appropriate for FDA to review details from the sponsor's quality system in the marketing submission to ensure adequate ongoing performance. Such a review may help support a determination of substantial equivalence.	The FDA should explain under what circumstances it is necessary for the Agency to review details about the manufacturer's Quality System during the premarket review process in order to support a substantial equivalence determination. Furthermore, the FDA should justify on what legal basis the Agency has the authority to request or review Quality System documentation during a 510(k) premarket notification review process.
28	Page 33, Lines 1337-1340	Sponsors of AI-enabled devices that elect to employ proactive performance monitoring as a means of risk control and to provide reasonable assurance of the device's safety and effectiveness, should include information regarding their performance monitoring plans as part of the premarket submission.	Again, the FDA should explain under what circumstances a manufacturer must submit a performance monitoring plan as part of the premarket submission. Specifically, the Agency should clarify whether manufacturers of AI-enabled devices who do not elect to employ proactive performance monitoring as a means of risk control or to provide reasonable assurance of the device's safety and effectiveness are required to include information regarding their performance monitoring plans in the premarket submission.

#	Page/Line #	Draft Guidance Language	Comments
29	Page 43, Lines 1684-1690	It is also important to consider when additional information may detract from understanding, rather than add to it. For example, explainability tools or visualizations can be valuable in increasing model transparency and a user's confidence in a model's output and could be developed as part of the user interface. However, if not well designed and validated for the target user group, explainability tools or visualizations could also significantly mislead users. Therefore, sponsors should develop and validate explainability metrics and visualizations through appropriate testing.	The FDA should clarify its expectations as to the need to develop and validate explainability metrics and visualizations. It is not clear what exactly the FDA expects manufacturers to develop on top of their actual medical device. Outside of developing a Quality System and tools to support maintenance of the Quality System (including post-market monitoring of the product quality), the FDA has no established precedent for requiring a manufacturer to develop additional tools. The lack of clarity, and the expectation itself, creates a significant burden (particularly on small device manufacturers) that is not required in other contexts. The FDA should clarify under what circumstances a manufacturer must develop and validate explainability metrics and visualizations in order to establish substantial equivalence or a reasonable assurance of safety and effectiveness. Furthermore, the FDA should explain the legal basis for requiring such information as part of a premarket review.

#	Page/Line #	Draft Guidance Language	Comments
30	Page 46, Lines 1785-1795	When specifically considering an AI-enabled diagnostic device, the key performance assessment is its diagnostic accuracy, which is evaluated in a pivotal diagnostic performance study. Due to sampling variation, the uncertainties of the accuracy estimates are typically quantified, usually in the form of 95% two-sided confidence intervals. The study acceptance criteria can be based on statistical inferences using hypothesis testing methods (e.g., comparing a lower/upper confidence limit to a pre-specified performance goal). Note that inferences based on point estimates ignores the statistical uncertainty of the estimates and is not generally acceptable in the primary analysis. It is always compared to a comparator that can be tested and evaluated on the same patient/data as the device. This comparator can be the clinician, another device that is adequately validated for the same intended use, or standard of care. The evaluation on the same patient/data is key to mitigate differences in the task difficulty levels and disease spectrum due to sampling variation.	The FDA should clarify what it considers to be "adequately validated" and whether the use of "another device that is adequately validated for the same intended use" would include an FDA-authorized device. For example, if a manufacturer of a wearable vital sign monitoring patch compares its performance to a bedside vital sign monitor that has been authorized by the FDA for the same intended use (i.e., vital sign monitoring), would the fact that the bedside monitor is legally marketed make it an "adequately validated" comparator? The FDA should clarify what manufacturers should consider when deciding what comparator is appropriate, including providing examples of characteristics in specific situations.
31	Page 47, Lines 1837-1840	[A] good practice is to examine the influence of a QC algorithm by checking the proportion of low-quality dropouts and assessing the results of a sensitivity analysis assuming a worst-case scenario (i.e., assuming the QC failure cases are all difficult ones that the model fails to classify successfully).	The FDA should clarify that sensitivity analyses validating the Quality Control process under worst-case scenarios should be limited to scenarios that remain within the scope of the intended use of the device. In the example provided by the Agency in the draft guidance, the FDA assumes that "low-quality images" that were discarded were potentially within the scope of the intended use of the product. If, though, the discarded cases were outside the scope of the intended use of the device, the manufacturer and the AI-model performance should not be detrimented by analysis of such cases.

#	Page/Line #	Draft Guidance Language	Comments
32	Page 49, Lines 1907-1910	[I]t is important to understand that while the Human Factors Guidance focuses on "serious harm," sponsors may need to provide documentation evaluating and addressing any potential risk associated with misuse, including misinterpretation, to ensure that the device is safe and effective for its intended use.	The FDA should explain why it is necessary to conduct human factors engineering analysis on *any* potential risk associated with misuse. This requirement significantly raises the bar from analyzing those risks that are associated with "serious harm" to any and all possible risks. Not only is it inappropriate to establish such a new requirement via guidance, the FDA must justify why information related any and all possible risks associated with misuse is necessary (particularly for AI-enabled devices but not other medical devices) to establish substantial equivalence or a reasonable assurance of safety and effectiveness.
33	Page 52, Lines 2016-2022	Data Characterization of data used to develop the device: o Data sources (e.g., clinical trials, public or proprietary databases) including details on any devices used to collect data. o Data types used (e.g., structured numerical data, structured categorical data, unstructured text, images, time-series data, or a combination). o Relevant details including the sample size, effect size, data quality, reference standard, diversity, and representativeness.	The FDA should explain why information related to early stage product development is necessary to disclose. Historically, prototyping and early feasibility testing for medical devices has not been required to be disclosed in any meaningful detail. While some early development related information is important for the FDA to review, public disclosure (e.g., in a model card) of such information is inconsistent with existing regulatory principles.

#	Page/Line #	Draft Guidance Language	Comments
34	Page 53, Lines 2026-2034	In general, publicly available summaries must follow the applicable the requirements for the specific marketing submission (e.g., 510(k), De Novo, PMA). The items below are not an exhaustive list of topics that a manufacturer may be expected to cover, and all topics may not apply to all marketing submissions. Likewise, FDA may request additional information to be included in this summary. This appendix serves as only an example of the types of information sponsors should generally provide in a 510(k) summary, including an example of a completed Basic Model Card. Information does not need to be repeated between the model card and other sections of the public summary, but information can be repeated if the sponsor believes that the alternate format provides useful context.	The FDA should justify why significantly more information than has historically been required to disclose in a 510(k) Summary must now be disclosed for AI-enabled medical devices. For example, the information presented in the *Model Training Description* section as well as the detailed subgroup analyses is substantially more information than what is typically included in publicly available documentation. In fact, it is not uncommon for the Agency to redact as trade secrets information related to clinical validation testing (even where such testing is published in peer-reviewed journals). As discussed elsewhere in these comments, early feasibility testing data and results (e.g., model training information) is often trade secret information and, thus, should not be disclosed publicly. Similarly, disclosure of unpowered subgroup analyses can mislead and confuse the public if disclosed in a 510(k) Summary. Hence, it is not clear why the FDA believes such information is necessary to disclose. In addition, the FDA should clarify that the inclusion of a Basic Model Card is voluntary and that choosing to exclude a Basic Model Card will not in any way detriment the manufacturer or the premarket submission review.
35	Page 56, Line 2056	Performance Subject Device Sensitivity: 87% (83%, 89%) Specificity: 83% (81%, 85%) Positive Predictive Value (PPV): 56% Predicate Device Sensitivity: 82% (78%, 85%) Specificity: 81% (79%, 84%) Positive Predictive Value (PPV): 53% Comparison Similar. The subject device has better performance than the predicate device in sensitivity, specificity, and PPV.	The draft guidance presents a substantial equivalence table that includes a line item for performance of the hypothetical subject and predicate device. The values presented in the table for the subject device reflect the performance observed during model development (as described on Page 57, Lines 2074-2078) as opposed to the performance observed during clinical validation (as described on Page 59, Lines 2140-2141). The FDA should explain the reasoning for presenting the model development performance in the substantial equivalence table or update the table to reflect the clinical validation performance presented for this hypothetical. In addition, the FDA should provide guidance on how similar is sufficiently similar when it comes to the comparison of performance values or provide references other guidance document(s) that inform how to determine whether the performance values are sufficiently similar.

#	Page/Line #	Draft Guidance Language	Comments
36	Page 60, Lines 2159-2160	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.*	As noted elsewhere in these comments, we strongly disagree with the notion of publicly disclosing unpowered subgroup analyses as doing so is misleading and will cause user confusion. Requiring the disclosure of such misleading information is unduly prejudicial to the manufacturer and would likely violate the First Amendment's prohibition on government compelled speech.
37	Page 61, Lines 2168-2226	Model Card	The FDA should clarify whether the Model Card portion of the example 510(k) Summary is presented as an alternative format for satisfying the 510(k) Summary requirements or if it is presented as an addition to the other information in the 510(k) Summary. If the latter, the FDA should explain why the model card is required given that the vast majority of the information presented in the model card is already presented elsewhere in the 510(k) Summary. For example, the device description, the non-clinical testing, and the clinical performance information are presented in the 510(k) Summary and the Model Card.
38	Page 63, Lines 2242-2245	Description of information that could impact risks and patient outcomes, across the product lifecycle: Model development and clinical validation included only 10% of participants under the age of 40, which may mean that the model's performance on that subgroup is not fully characterized.	The FDA should clarify what it means for a subgroup to be "fully characterized". Specifically, the Agency should clarify whether this concept applies to subgroup analyses that are powered as well as unpowered.

#	Page/Line #	Draft Guidance Language	Comments
39	Page 63, Lines 2253-2262	How to conduct local site-specific acceptance testing or validation: Prior to use of the model in the site's entire population, the model is deployed, and data is collected for a one-month period in order to understand any issues with integration into the sites' existing systems and measure performance on a subset of the patient population for that site. Through this process, issues with deployment can be addressed prior to exposure to the entire population and can help characterize performance of the model and the need for additional training and development. Alternatively, sites may opt to provide historical data that can be used to assess expected performance at the site.	The FDA should clarify whether the hypothetical deployment method is simply an example or is an approach that the Agency expects manufacturers to utilize.
40	Page 64, Lines 2283-2284	Software quality (specify, standards and regulatory compliance issues, intellectual property issues, risk management and safeguards used, other):	This sentence appears to be incomplete. The FDA should update the guidance to fully convey the expectations for risk management related to software quality.