U.S. FOOD & DRUG ADMINISTRATION

Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)

Discussion Paper and Request for Feedback

Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - *Discussion Paper and Request for Feedback*

I. Introduction

Artificial intelligence (AI)- and machine learning (ML)-based technologies have the potential to transform healthcare by deriving new and important insights from the vast amount of data generated during the delivery of healthcare every day. Example high-value applications include earlier disease detection, more accurate diagnosis, identification of new observations or patterns on human physiology, and development of personalized diagnostics and therapeutics. One of the greatest benefits of AI/ML in software resides in its ability to learn from real-world use and experience, and its capability to improve its performance. The ability for AI/ML software to learn from real-world feedback (training) and improve its performance (adaptation) makes these technologies uniquely situated among software as a medical device (SaMD)[1] and a rapidly expanding area of research and development. Our vision is that with appropriately tailored regulatory oversight, AI/ML-based SaMD will deliver safe and effective software functionality that improves the quality of care that patients receive.

FDA has made significant strides in developing policies[2, 3] that are appropriately tailored for SaMD to ensure that safe and effective technology reaches users, including patients and healthcare professionals. Manufacturers submit a marketing application to FDA prior to initial distribution of their medical device, with the submission type and data requirements based on the risk of the SaMD (510(k) notification, De Novo, or premarket approval application (PMA) pathway). For changes in design that are specific to software that has been reviewed and cleared under a 510(k) notification, FDA's Center for Devices and Radiological Health (CDRH) has published guidance (Deciding When to Submit a 510(k) for a Software Change to an Existing Device,[4] also referred to herein as the software modifications guidance) that describes a risk-based approach to assist in determining when a premarket submission is required.[5]

The International Medical Device Regulators Forum (IMDRF) defines 'Software as a Medical Device (SaMD)' as software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.[1] FDA, under the Federal Food, Drug, and Cosmetic Act (FD&C Act) considers medical purpose as those purposes that are intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions.

[1] Software as a Medical Device (SaMD): Key Definitions: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf.
[2] Pre-Cert Program Version 1.0 Working Model: https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM629276.pdf.
[3] Software as a Medical Device (SaMD): Clinical Evaluation: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm524904.pdf.
[4] Deciding When to Submit a 510(k) for a Software Change to an Existing Device: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf.
[5] 21 CFR 807.81(a)(3). Modifications to a device approved through a PMA are governed by the criteria in 21 CFR 814.39(a). Modifications to Devices Subject to Premarket Approval (PMA) - The PMA Supplement Decision-Making Process: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM089360.pdf.

The 510(k) software modifications guidance focuses on the risk to users/patients resulting from the software change. Categories of software modifications that may require a premarket submission include:

- A change that introduces a new risk or modifies an existing risk that could result in significant harm;
- A change to risk controls to prevent significant harm; and
- A change that significantly affects clinical functionality or performance specifications of the device.

When applied to AI/ML-based SaMD, the above approach would require a premarket submission to the FDA when the AI/ML software modification significantly affects device performance, or safety and effectiveness[6]; the modification is to the device's intended use; or the modification introduces a major change to the SaMD algorithm. For a PMA-approved SaMD, a supplemental application would be required for changes that affect safety or effectiveness, such as new indications for use, new clinical effects, or significant technology modifications that affect performance characteristics.

To address the critical question of when a continuously learning AI/ML SaMD may require a premarket submission for an algorithm change, we were prompted to reimagine an approach to premarket review for AI/ML-driven software modifications. Such an approach would need to maintain reasonable assurance of safety and effectiveness of AI/ML-based SaMD, while allowing the software to continue to learn and evolve over time to improve patient care.

To date, FDA has cleared or approved several AI/ML-based SaMD. Typically, these have only included algorithms that are "locked[7]" prior to marketing, where algorithm changes likely require FDA premarket review for changes beyond the original market authorization. However, not all AI/ML-based SaMD are locked; some algorithms can adapt over time. The power of these AI/ML-based SaMD lies within the ability to continuously learn, where the adaptation or change to the algorithm is realized after the SaMD is distributed for use and has "learned" from real-world experience. Following distribution, these types of continuously learning and adaptive AI/ML algorithms may provide a different output in comparison to the output initially cleared for a given set of inputs.

The traditional paradigm of medical device regulation was not designed for adaptive AI/ML technologies, which have the potential to adapt and optimize device performance in real-time to continuously improve healthcare for patients. The highly iterative, autonomous, and adaptive nature of these tools requires a new, total product lifecycle (TPLC) regulatory approach that facilitates a rapid cycle of product improvement and allows these devices to continually improve while providing effective safeguards.

This discussion paper proposes a framework for modifications to AI/ML-based SaMD that is based on the internationally harmonized International Medical Device Regulators Forum (IMDRF) risk categorization principles, FDA's benefit-risk framework, risk management principles in the software

[6] 21 CFR 807.81(a)(3).
[7] We define a "locked" algorithm as an algorithm that provides the same result each time the same input is applied to it and does not change with use. Examples of locked algorithms are static look-up tables, decision trees, and complex classifiers.

modifications guidance[8], and the organization-based TPLC approach as envisioned in the Digital Health Software Precertification (Pre-Cert) Program.[9] It also leverages practices from our current premarket programs, including the 510(k), De Novo, and PMA pathways.

This discussion paper describes an innovative approach that may require additional statutory authority to implement fully. The proposed framework is being issued for discussion purposes only and is not a draft guidance. This document is not intended to communicate FDA's proposed (or final) regulatory expectations but is instead meant to seek early input from groups and individuals outside the Agency prior to development of a draft guidance.

This proposed TPLC approach allows FDA's regulatory oversight to embrace the iterative improvement power of AI/ML SaMD while assuring that patient safety is maintained. It also assures that ongoing algorithm changes are implemented according to pre-specified performance objectives, follow defined algorithm change protocols, utilize a validation process that is committed to improving the performance, safety, and effectiveness of AI/ML software, and include real-world monitoring of performance. This proposed TPLC regulatory framework aims to promote a mechanism for manufacturers to be continually vigilant in maintaining the safety and effectiveness of their SaMD, that ultimately, supports both FDA and manufacturers in providing increased benefits to patients and providers.

II. Background: AI/ML-Based Software as a Medical Device

In this paper, we use John McCarthy's definition of AI as the science and engineering of making intelligent machines, especially intelligent computer programs.[10] AI can use different techniques, such as ML, to produce intelligent behavior, including models based on statistical analysis of data, and expert systems that primarily rely on if-then statements. In this paper, we refer to an ML system as a system that has the capacity to learn based on training on a specific task by tracking performance measure(s). AI, and specifically ML, are techniques used to design and train software algorithms to learn from and act on data. These AI/ML-based software, when intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions, are medical devices under the FD&C Act, and called "Software as a Medical Device" (SaMD) by FDA and IMDRF. The intended use of AI/ML-based SaMD, similar to other SaMDs, may exist on a spectrum of impact to patients as categorized by IMDRF SaMD risk categorization framework.[11]

Non-device software functions are not subject to FDA device regulation and are not within the scope of this paper. In addition, as detailed in section 502(o) of the FD&C Act, software functions intended (1) for administrative support of a health care facility, (2) for maintaining or encouraging a healthy lifestyle, (3) to serve as electronic patient records, (4) for transferring, storing, converting formats, or displaying data, or (5) to provide certain, limited clinical decision support are not medical devices and are not subject to FDA regulation.

[8] Deciding When to Submit a 510(k) for a Software Change to an Existing Device: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf.
[9] Developing a Software Precertification Program: A Working Model; v1.0 – January 2019: https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM629276.pdf.
[10] http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.
[11] Software as a Medical Device (SaMD): Possible Framework for Risk Categorization and Corresponding Considerations: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf.

The IMDRF SaMD risk categorization framework takes a risk-based approach to categorize SaMD based on intended use, similar to traditional risk-based approaches used by the FDA. The IMDRF risk framework identifies the following two major factors as providing a description of the intended use[12] of the SaMD:

1) **Significance of information provided by the SaMD to the healthcare decision**, which identifies the intended use of the information provided by the SaMD – i.e., to treat or diagnose; to drive clinical management; or to inform clinical management; and
2) **State of healthcare situation or condition**, which identifies the intended user, disease or condition, and the population for the SaMD – i.e., critical; serious; or non-serious healthcare situations or conditions.

Taken together, these factors describing the intended use can be used to place the AI/ML-based SaMD into one of four categories, from lowest (I) to highest risk (IV) to reflect the risk associated with the clinical situation and device use.

State of healthcare situation or condition	Significance of information provided by SaMD to healthcare decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

Figure 1: SaMD IMDRF risk categorization

While AI/ML-based SaMD exist on a spectrum categorized by risk to patients, they also exist on a spectrum from locked to continuously learning. "Locked" algorithms are those that provide the same result each time the same input is provided. As such, a locked algorithm applies a fixed function (e.g., a static look-up table, decision tree, or complex classifier) to a given set of inputs. These algorithms may use manual processes for updates and validation. In contrast to a locked algorithm, an adaptive algorithm (e.g., a continuous learning algorithm) changes its behavior using a defined learning process. The algorithm adaptation or changes are implemented such that for a given set of inputs, the output may be different before and after the changes are implemented. These algorithm changes are typically implemented and validated through a well-defined and possibly fully automated process that aims at improving performance based on analysis of new or additional data.

The adaptation process can be intended to address several different clinical aspects, such as optimizing performance within a specific environment (e.g., based on the local patient population), optimizing performance based on how the device is being used (e.g., based on preferences of a specific physician), improving performance as more data are collected, and/or changing the intended use of the device. The adaptation process follows two stages: learning and updating. The algorithm "learns" how to change its behavior, for example, from the addition of new input types or adding new cases to an already existing training database. The "update" then occurs when the new version of the algorithm is deployed. As a

[12] Information that may be used to describe intended use for FDA purposes is set forth in 21 CFR 807.92(a)(5), 814.20(b)(3), and 860.7(b), and could be written using terminology as described in the IMDRF risk categorization framework.

result, given the same set of inputs at time A (before update) and time B (after update), the output of the algorithm may differ.

Although AI/ML-based SaMD exists on a spectrum from locked to continuously adaptive algorithms, a common set of considerations for data management, re-training, and performance evaluation can be applied to the entire spectrum of SaMD. For example, the rigor of performance evaluation for both locked and continuously adaptive algorithms depend on the test methods, quality and applicability of dataset used for testing, and the algorithm's training methods. Robust algorithms typically require the availability of large, high-quality, and well-labeled training data sets. Likewise, a common set of principles can be applied to considerations about how to provide confidence in function and performance to users through appropriate validation, transparency, and claims after the modification.

III. Types of AI/ML-based SaMD Modifications

There are many possible modifications to an AI/ML-based SaMD. Some modifications may not require a review based on guidance provided in "Deciding When to Submit a 510(k) for a Software Change to an Existing Device."[13] This paper anticipates that many modifications to AI/ML-based SaMD involve algorithm architecture modifications and re-training with new data sets, which under the software modifications guidance would be subject to premarket review. The types of modifications generally fall into three broad categories:

- **Performance** – clinical and analytical performance[14];
- **Inputs** used by the algorithm and their clinical association to the SaMD output; and/or
- **Intended use**[15] – The intended use of the SaMD, as outlined above and in the IMDRF risk categorization framework, described through the significance of information provided by the SaMD for the state of the healthcare situation or condition.

The changes described may not be mutually exclusive – one software modification may impact, for example, both a change in input and change in performance; or, a performance change may increase a device's clinical performance that in turn impacts the intended use. These software changes in AI/ML-based SaMD, grouped by the types of changes as described above, have different impact on users, which may include either patients, healthcare professionals, or others:

i. **Modifications related to performance, with no change to the intended use or new input type:** This type of modification includes improvements to analytical and clinical performance that can result from a number of changes. This may include re-training with new data sets within the intended use population from the same type of input signal, a change in the AI/ML architecture, or other means. For this type of modification, the manufacturer commonly aims to update users on the performance, without changing any of the explicit use claims about their product (e.g., increased sensitivity of the SaMD at detecting breast lesions suspicious for cancer in digital mammograms).

[13] Deciding When to Submit a 510(k) for a Software Change to an Existing Device: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf.
[14] Software as a Medical Device (SaMD): Clinical Evaluation: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm524904.pdf.
[15] In this document, "modifications related to intended use" refers to changes within the parameters of the cleared/approved intended use as defined in the classification regulation or FDA approval or authorization.

ii. **Modifications related to inputs, with no change to the intended use:** These types of modifications are those that change the inputs used by the AI/ML algorithm. These modifications may involve changes to the algorithm for use with new types of input signals, but do not change the product use claims. Examples of these changes could be:

a. Expanding the SaMD's compatibility with other source(s) of the same input data type (e.g., SaMD modification to support compatibility with CT scanners from additional manufacturers); or

b. Adding different input data type(s) (e.g., expanding the inputs for a SaMD that diagnoses atrial fibrillation to include oximetry data, for example, in addition to heart rate data).

iii. **Modifications related to the SaMD's intended use:** These types of modifications include those that result in a change in the significance of information provided by the SaMD (e.g., from a confidence score that is 'an aid in diagnosis' (drive clinical management) to a 'definitive diagnosis' (diagnose)). These types of modifications also include those that result in a change in the state of the healthcare situation or condition and are explicitly claimed by the manufacturer, such as an expanded intended patient population (e.g., inclusion of pediatric population where the SaMD was initially intended for adults ages 18 years or older); or the intended disease or condition (e.g., expansion to use a SaMD algorithm for lesion detection from one type of cancer to another). Changes related to either the significance of the information provided by the SaMD or the healthcare situation or condition may be limited in scope by the pre-specified performance objectives and algorithm change protocols.

Questions / Feedback on the types of AI/ML-SaMD modifications:

- *Do these categories of AI/ML-SaMD modifications align with the modifications that would typically be encountered in software development that could require premarket submission?*
- *What additional categories, if any, of AI/ML-SaMD modifications should be considered in this proposed approach?*
- *Would the proposed framework for addressing modifications and modification types assist the development AI/ML software?*

IV. A Total Product Lifecycle Regulatory Approach for AI/ML-Based SaMD

As envisioned in the Software Pre-Cert Program,[16] applying a TPLC approach to the regulation of software products is particularly important for AI/ML-based SaMD due to its ability to adapt and improve from real-world use. In the Pre-Cert TPLC approach, FDA will assess the culture of quality and organizational excellence of a particular company and have reasonable assurance of the high quality of their software development, testing, and performance monitoring of their products. This approach

[16] Developing a Software Precertification Program: A Working Model; v1.0 – January 2019: https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM629276.pdf.

would provide reasonable assurance of safety and effectiveness throughout the lifecycle of the organization and products so that patients, caregivers, healthcare professionals, and other users have assurance of the safety and quality of those products. This TPLC approach enables the evaluation and monitoring of a software product from its premarket development to postmarket performance, along with continued demonstration of the organization's excellence (Figure 2).

This proposed regulatory approach would apply to only those AI/ML based-SaMD that require premarket submission and not those that are exempt from requiring premarket review (i.e., Class I exempt and Class II exempt).

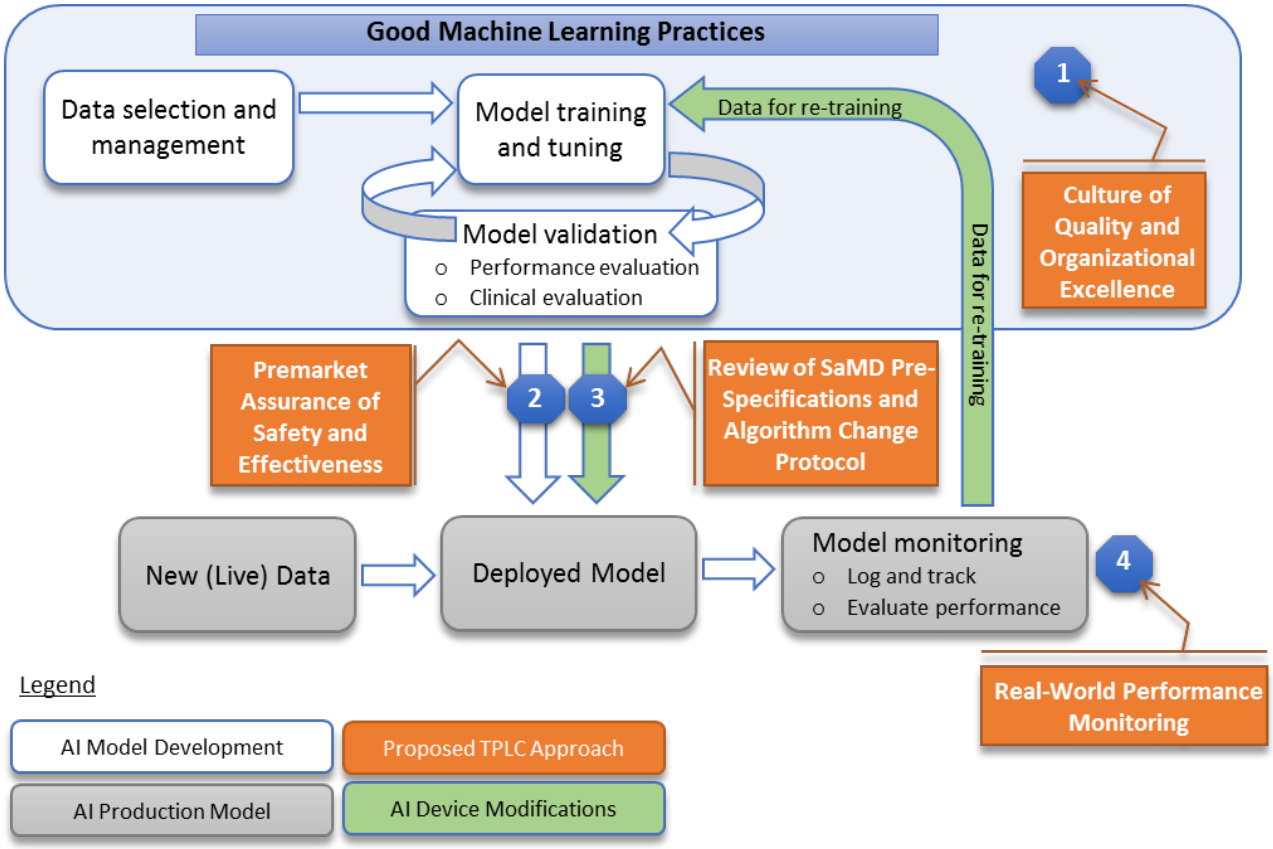

Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

To fully realize the power of AI/ML learning algorithms while enabling continuous improvement of their performance and limiting degradations, the FDA's proposed TPLC approach is based on the following general principles that balance the benefits and risks, and provide access to safe and effective AI/ML-based SaMD:

1. Establish clear expectations on quality systems and good ML practices (GMLP);

2. Conduct premarket review for those SaMD that require premarket submission[17] to demonstrate reasonable assurance of safety and effectiveness and establish clear expectations for manufacturers of AI/ML-based SaMD to continually manage patient risks throughout the lifecycle;
3. Expect manufacturers to monitor the AI/ML device and incorporate a risk management approach and other approaches outlined in "Deciding When to Submit a 510(k) for a Software Change to an Existing Device" Guidance[18] in development, validation, and execution of the algorithm changes (SaMD Pre-Specifications and Algorithm Change Protocol); and
4. Enable increased transparency to users and FDA using postmarket real-world performance reporting for maintaining continued assurance of safety and effectiveness.

1. Quality Systems and Good Machine Learning Practices (GMLP):

The FDA expects every medical device manufacturer to have an established quality system that is geared towards developing, delivering, and maintaining high-quality products throughout the lifecycle that conforms to the appropriate standards and regulations.[19] Similarly, for AI/ML-based SaMD, we expect that SaMD developers embrace the excellence principles of culture of quality and organizational excellence.[20]

As is the case for all SaMD, devices that rely on AI/ML are expected to demonstrate analytical and clinical validation, as described in the SaMD: Clinical Evaluation guidance (Figure 3).[21] The specific types of data necessary to assure safety and effectiveness during the premarket review, including study design, will depend on the function of the AI/ML, the risk it poses to users, and its intended use.

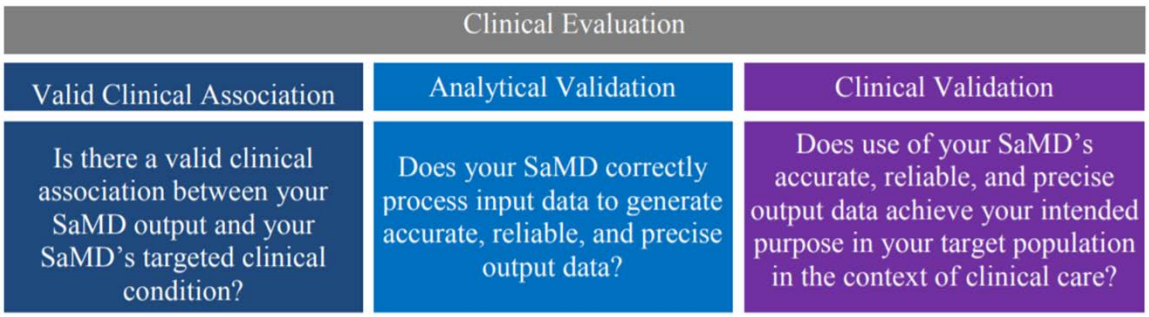

Figure 3: IMDRF description of Clinical Evaluation components

AI/ML algorithm development involves learning from data and hence prompts unique considerations that embody GMLP. In this paper, GMLP are those AI/ML best practices (e.g., data management, feature extraction, training, and evaluation) that are akin to good software engineering practices or quality system practices. Examples of GMLP considerations as applied for SaMD include:

[17] 21 CFR Part 807 Subpart E or 21 CFR Part 814 Subpart B.
[18] Deciding When to Submit a 510(k) for a Software Change to an Existing Device: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf.
[19] 21 CFR Part 820.
[20] See the discussion in Developing a Software Precertification Program: A Working Model; v1.0 – January 2019: https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM629276.pdf.
[21] Software as a Medical Device (SaMD): Clinical Evaluation: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm524904.pdf.

- Relevance of available data to the clinical problem and current clinical practice;
- Data acquired in a consistent, clinically relevant and generalizable manner that aligns with the SaMD's intended use and modification plans;
- Appropriate separation between training, tuning, and test datasets; and
- Appropriate level of transparency (clarity) of the output and the algorithm aimed at users.

Questions / Feedback on GMLP:

- *What additional considerations exist for GMLP?*
- *How can FDA support development of GMLP?*
- *How do manufacturers and software developers incorporate GMLP in their organization?*

2. Initial Premarket Assurance of Safety and Effectiveness:

This framework gives manufacturers the option to submit a plan for modifications during the initial premarket review of an AI/ML-based SaMD. FDA's premarket review and determination regarding the acceptability of such plans would provide reasonable assurance of safety and effectiveness and would include review of the SaMD's performance, the manufacturer's plan for modifications, and the ability of the manufacturer to manage and control resultant risks of the modifications. FDA has successfully explored this voluntary approach to review device modification plans in certain recent De Novo classifications regarding several in-vitro diagnostic next generation sequencing products.[22] This paper proposes a framework for modifications to AI/ML-based SaMD that relies on the principle of a "**predetermined change control plan**." Using this proposed regulatory approach, we believe that our oversight will enable responsible performance enhancements in AI/ML-based technologies.

The predetermined change control plan would include the types of anticipated modifications – **SaMD Pre-Specifications** – based on the retraining and model update strategy, and the associated methodology – **Algorithm Change Protocol** – being used to implement those changes in a controlled manner that manages risks to patients.

SaMD Pre-Specifications (SPS): A SaMD manufacturer's anticipated modifications to "performance" or "inputs," or changes related to the "intended use" of AI/ML-based SaMD. These are the types of changes the manufacturer plans to achieve when the SaMD is in use. The SPS draws a "region of potential changes" around the initial specifications and labeling of the original device. This is "what" the manufacturer intends the algorithm to become as it learns.

Algorithm Change Protocol (ACP): Specific methods that a manufacturer has in place to achieve and appropriately control the risks of the anticipated types of modifications delineated in the SPS. The ACP is a step-by-step delineation of the data and procedures to be followed so that the modification achieves its goals and the device remains safe and effective after the modification. Figure 4 below provides a

[22] CDRH's Approach to Tumor Profiling Next Generation Sequencing Tests: https://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM584603.pdf.

general overview of the components of an ACP. This is "how" the algorithm will learn and change while remaining safe and effective.

Data Management	➢ For new training & test data: • Collection protocols • Quality assurance • Reference standard determination ➢ Auditing and sequestration of training and test sets
Re-training	➢ Re-training objectives ➢ Changes related to: • ML methods, including architecture and parameters • Data pre-processing ➢ Criteria to initiate performance evaluation
Performance Evaluation	➢ Assessment metrics ➢ Statistical analysis plans ➢ Frequency and triggers for evaluation ➢ Performance targets ➢ Methods for testing with "clinicians in the loop" when necessary
Update Procedures	➢ Software verification and validation ➢ When and how updates will be implemented ➢ Plans for global and local updates ➢ Communication and transparency to users

Figure 4: Algorithm Change Protocol components

Scope and limitations for establishing SPS and ACP: The FDA acknowledges that the types of changes that could be pre-specified in a SPS and managed through an ACP may necessitate individual consideration during premarket review of benefits and risks to patients of that particular SaMD. The extent to which pre-approval of a SPS and an ACP can be relied on to support future modifications depends on various factors. The following are example scenarios that illustrate the general concept of establishing an appropriate SPS and its corresponding ACP:

- Changes that involve improvements in performance, or changes in input, without affecting the intended use of the SaMD, may be accomplished with an appropriate level of pre-specification and an appropriate ACP that provides reasonable assurance that performance will be improved or maintained. The ACP may include the basis of validation and methods to adequately monitor and control for significant degradation in performance or introduce risks to patients.

- Certain changes related to the intended use, in particular, an increase in the significance of the information provided to the user for the same healthcare situation or condition. Using the IMDRF risk framework as the basis for an example, a SPS may include a modification related to the intended use within "drive clinical management," which may shift the intended use from "identify early signs of a disease or conditions" to "aid in making a definitive diagnosis" for the same healthcare situation or condition. An appropriate ACP might be developed, reviewed, and

agreed by FDA and the manufacturer to adequately improve the performance to a level that increases the confidence in its ability to be used as an aid in making a definitive diagnosis.

- Certain changes related to the intended use, in particular, the "indications for use." For example, a manufacturer may intend to expand the use of their SaMD to a new patient population for which there had been insufficient evidence available to initially support that indication for use. In some cases, an appropriate reference standard may initially not be available for the new patient population; a manufacturer's ACP may include a characterization plan for the reference standard in the disease population to assure it provides a meaningful representation of the disease. In other cases, an input data type used by the AI/ML-based SaMD may not normally be available for the patient population; a developer's ACP may include a demonstration of the clinical association between the disease and input data type in the new patient population, as well as a plan for data collection and algorithm testing in the patient population.

There are many scenarios for which an appropriate SPS and ACP could be crafted, however, we also anticipate that in certain cases, the SaMD's risk or the intended use may significantly change after learning. In these cases, it may not be appropriate for a proposed SPS and ACP to manage the risks to patients or align with the initial authorized intended use. For example, it would not be appropriate for a SPS and ACP initially indicated for a "low risk" (non-serious) healthcare situation or condition, such as using skin images to manage the healing of scars, to be leveraged for the same SaMD in diagnosing melanoma, which would be considered a "critical healthcare situation or condition."

Questions / Feedback on SPS and ACP:

- *What are the appropriate elements for the SPS?*
- *What are the appropriate elements for the ACP to support the SPS?*
- *What potential formats do you suggest for appropriately describing a SPS and an ACP in the premarket review submission or application?*

3. Approach for modifications after initial review with an established SPS and ACP:

Learning, adaptation, and optimization are inherent to AI/ML-based SaMD. These capabilities of AI/ML would be considered modifications to SaMD after they have received market authorization from FDA. This paper proposes an approach to appropriately manage risks to patients from these modifications, while enabling manufacturers to improve performance and potentially advance patient care.

As outlined in Figure 5, manufacturers are expected to evaluate the modifications based on risk to patients as outlined in the software modifications guidance.[23] The software modifications guidance uses a risk-based approach and expects a manufacturer to perform a risk assessment and evaluate that the risks are reasonably mitigated. Depending on the type of modification, the current software

[23] Deciding When to Submit a 510(k) for a Software Change to an Existing Device: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf.

modifications guidance results in either 1) submission of a new 510(k) for premarket review or 2) documentation of the modification and the analysis in the risk management and 510(k) files. If, for AI/ML SaMD with an approved SPS and ACP, modifications are within the bounds of the SPS and the ACP, this proposed framework suggests that manufacturers would document the change in their change history and other appropriate records, and file for reference, similar to the "document" approach outlined in the software modifications guidance.

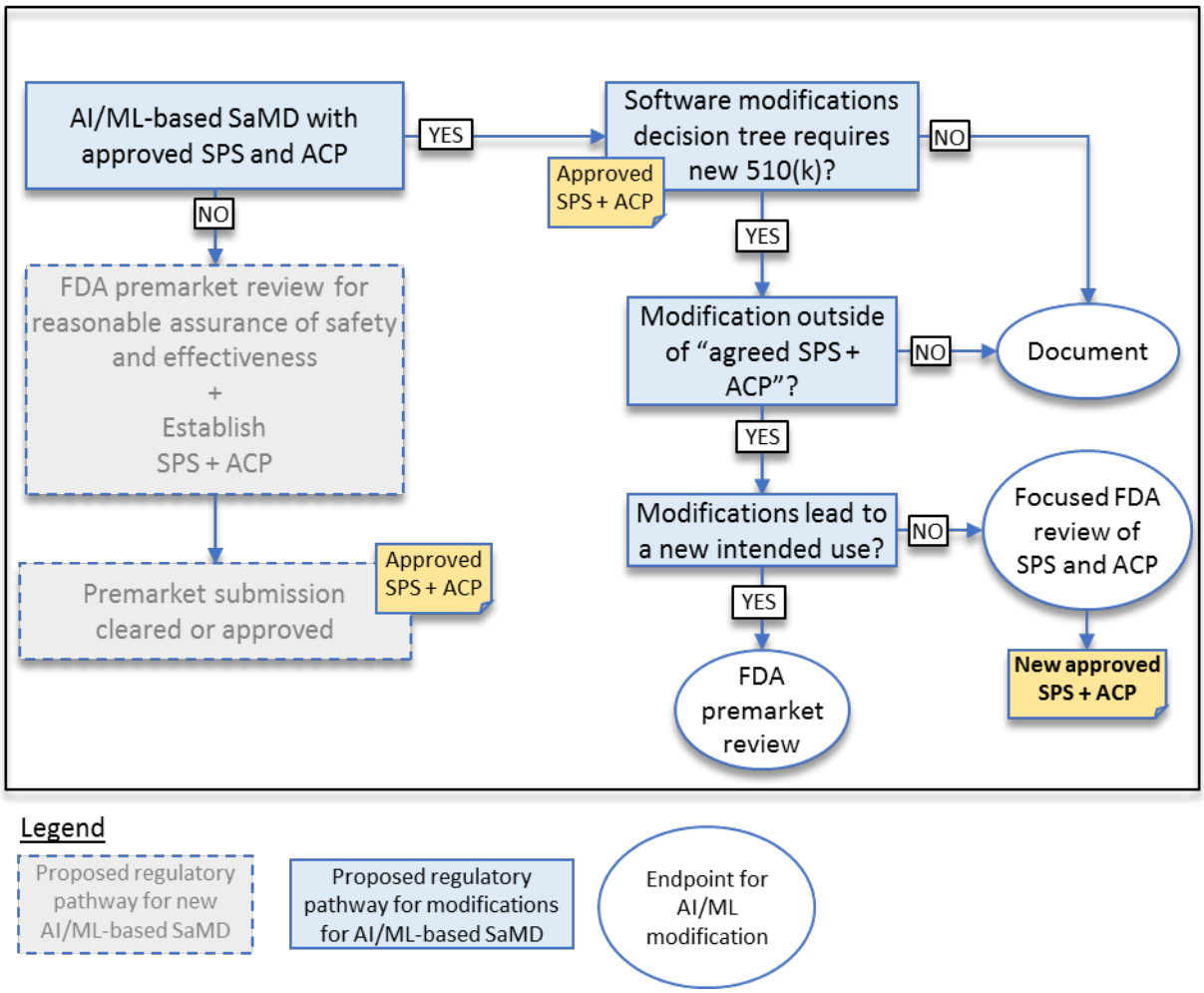

Figure 5: Approach to modifications to previously approved SaMD with SPS and ACP. This flowchart should only be considered in conjunction with the accompanying text in this white paper.

In the software modifications guidance, depending on the type of change, if the modification is beyond the intended use for which the SaMD was previously authorized, manufacturers are expected to submit a new premarket submission.[24] For this proposed approach, we anticipate that there may be cases where the SPS or ACP can be refined based on the real-world learning and training for the same intended use of AI/ML SaMD model. In those scenarios, FDA may conduct a "focused review" of the proposed SPS and ACP for a particular SaMD. Manufacturers may leverage some of the following options to engage with FDA on the SPS and ACP for a particular SaMD:

[24] 21 CFR 807.81(a)(3) or 21 CFR 814.39(a).

a. Contact the appropriate review division to obtain concurrence that the modification fits under current SPS and ACP; or
b. Submit a pre-submission[25] for a discussion on the modification and how it is within the bounds of the current SPS and ACP; or
c. Submit a premarket submission or application of the modification to SPS and ACP.

Questions / Feedback on premarket review:

- *How should FDA handle changes outside of the "agreed upon SPS and ACP"?*
- *What additional mechanisms could achieve a "focused review" of an SPS and ACP?*
- *What content should be included in a "focused review"?*

4. Transparency and real-world performance monitoring of AI/ML-based SaMD:

To fully adopt a TPLC approach in the regulation of AI/ML-based SaMD, manufacturers can work to assure the safety and effectiveness of their software products by implementing appropriate mechanisms that support transparency and real-world performance monitoring. Transparency about the function and modifications of medical devices is a key aspect of their safety. This is especially important for devices, like SaMD that incorporate AI/ML, which change over time. Further, many of the modifications to AI/ML-based SaMD may be supported by collection and monitoring of real-world data. Gathering performance data on the real-world use of the SaMD may allow manufacturers to understand how their products are being used, identify opportunities for improvements, and respond proactively to safety or usability concerns. Real-world data collection and monitoring is an important mechanism that manufacturers can leverage to mitigate the risk involved with AI/ML-based SaMD modifications, in support of the benefit-risk profile in the assessment of a particular AI/ML-based SaMD.

Through this framework, manufacturers would be expected to commit to the principles of transparency and real-world performance monitoring for AI/ML-based SaMD. FDA would also expect the manufacturer to provide periodic reporting to FDA on updates that were implemented as part of the approved SPS and ACP, as well as performance metrics for those SaMD. This commitment could be achieved through a variety of mechanisms.

Transparency may include updates to FDA, device companies and collaborators of the manufacturer, and the public, such as clinicians, patients, and general users. For modifications in the SPS and ACP, manufacturers would ensure that labeling changes accurately and completely describe the modification, including its rationale, any change in inputs, and the updated performance of the SaMD. Manufacturers may also need to update the specifications or compatibility of any impacted supporting devices, accessories, or non-device components. Finally, manufacturers may consider unique mechanisms for how to be transparent – they may wish to establish communication procedures that could describe how users will be notified of updates (e.g., letters, email, software notifications) and what information could be provided (e.g., how to appropriately describe performance changes between the current and previous version).

[25] Requests for Feedback on Medical Device Submissions: The Pre-Submission Program and Meetings with Food and Drug Administration Staff: https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM311176.

Real-world performance monitoring may also be achieved in a variety of suggested mechanisms that are currently employed or under pilot at FDA, such as adding to file or an annual report, Case for Quality activities,[26] or real-world performance analytics via the Pre-Cert Program.[27] Reporting type and frequency may be tailored based on the risk of the device, number and types of modifications, and maturity of the algorithm (i.e., quarterly reports are unlikely to be useful if the algorithm is at a mature stage with minimal changes in performance over the quarter).[28] Involvement in pilot programs, such as Case for Quality and the Pre-Cert Program, may also impact the reporting type and frequency given the insight into the manufacturer's TPLC and organization. Participation in these programs could provide another avenue to support continued assurance of safety and effectiveness in development and modifications of AI/ML-based SaMD.

Questions / Feedback on the transparency and real-world performance monitoring:

- *In what ways can a manufacturer demonstrate transparency about AI/ML-SaMD algorithm updates, performance improvements, or labeling changes, to name a few?*
- *What role can real-world evidence play in supporting transparency for AI/ML-SaMD?*
- *What additional mechanisms exist for real-world performance monitoring of AI/ML-SaMD?*
- *What additional mechanisms might be needed for real-world performance monitoring of AI/ML-SaMD?*

V. Appendix A: Examples

The following are hypothetical examples of AI/ML-based SaMD modifications that may or may not be permitted under this proposed framework. We welcome comment on the proposed hypothetical examples, in addition to the questions and feedback on specific topic areas and framework.

Note that these generalized examples do not contain the detail necessary for the SaMD device description, SPS, and ACP in a submission. The reader should reference the above sections and Appendix B (ACP) for more information on the appropriate content.

Please also note that the manufacturer may submit a new premarket submission for the algorithm modification, as appropriate with current policy.[29] The scenarios are not exhaustive or definitive; they are only intended to assist in illustrating the concept and framework for types of SaMD, and potential modifications that may or may not be permitted through this proposed framework.

[26] Case for Quality Pilot Activities: https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/MedicalDeviceQualityandCompliance/ucm590419.htm.
[27] Developing a Software Precertification Program: A Working Model; v1.0 – January 2019: https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM629276.pdf.
[28] New reporting mechanisms for this approach may require additional statutory authority to implement fully.
[29] Deciding When to Submit a 510(k) for a Software Change to an Existing Device: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf.

1. **Intensive Care Unit (ICU) SaMD**

Description of SaMD: An AI/ML application intended for ICU patients receives electrocardiogram, blood pressure, and pulse-oximetry signals from a primary patient monitor. The physiologic signals are processed and analyzed to detect patterns that occur at the onset of physiologic instability. When physiologic instability is detected, an audible alarm signal is generated to indicate that prompt clinical action is needed to prevent potential harm to the patient. This SaMD AI/ML application will 'drive clinical management' in a 'critical healthcare situation or condition.'

SPS: The manufacturer proposes two potential modifications for ICU SaMD:

- Modify the algorithm to ensure consistent performance across sub-populations, especially in cases where real-world monitoring suggests the algorithm underperforms; and
- Reduce false-alarm rates while maintaining or increasing sensitivity to the onset of physiologic instability.

ACP: For these modifications, the ACP details the methods for database generation, reference standard labeling, and comparative analysis along with the performance requirements and statistical analysis plan. The manufacturer follows GMLP.

Modification Scenario 1A: Increase in performance (type i modification), consistent with SPS and ACP

In accordance with the ACP, data was collected and used to modify the algorithm in a way that the manufacturer believes will lower the false-alarm rate while maintaining the sensitivity. A separate independent validation data set was collected. The manufacturer used the independent data set to perform analytical validation and found that the false-alarm rate was reduced while the sensitivity remained the same. Labeling was updated in accordance with the modified SaMD performance, and communication was provided to SaMD users. The algorithm modification may be made without additional FDA review.

Modification Scenario 1B: Increase in performance and change related to intended use (type iii modification), inconsistent with SPS and ACP

In accordance with the ACP, the manufacturer re-trained their algorithm using additional data to improve the sensitivity, however, analytical validation demonstrated that the revised algorithm has the same sensitivity and false-alarm rate as the previous version. The manufacturer noticed that the modified algorithm can maintain that same sensitivity 15 minutes prior to the onset of physiologic instability, which the previous version of the algorithm could not do.

The manufacturer would like to update their algorithm, labeling, and intended use to indicate that the alarm condition now reflects prediction of a physiologic instability within the next 15 minutes, which was not previously included in the SPS and ACP. Because the methods required for analysis, performance, and statistics are not consistent with predicting a future state, and the significance of information provided by the SaMD has changed, FDA may review a new SPS and ACP that includes this information for the algorithm modification before the manufacturer is permitted to make the change.

2. Skin Lesion Mobile Medical App (MMA)

Description of SaMD: An AI/ML MMA uses images taken by a smartphone camera to provide detailed information to a dermatologist on the physical characteristics of a skin lesion in order for the dermatologist to label the skin lesion as benign or malignant. The MMA will 'drive clinical management' in a 'serious healthcare situation or condition.'

SPS: The manufacturer proposes two potential modifications for the Skin Lesion MMA:

- Improve sensitivity and specificity in analyzing physical characteristics of benign or malignant skin lesions using real-world data; and
- Extend the MMA to be used with similar smartphone image acquisition systems, with pre-specified acceptance criteria for the image acquisition characteristics and a real-world performance plan to monitor performance across image acquisition systems.

ACP: For these modifications, the ACP includes detailed methods for database generation, reference standard labeling, and comparative analysis along the performance requirements, including sensitivity and specificity. The manufacturer incorporates the acceptance criteria for image acquisition systems intended for future compatibility with the MMA, and the validation study to demonstrate MMA performance requirements using new input. The manufacturer also includes the real-world performance plan. The manufacturer follows GMLP.

Modification Scenario 2A: Increase in performance (type i modification), consistent with SPS and ACP

The manufacturer collected real-world data from use of the MMA on various smartphone platforms. The actively 'learning' (but not distributed) MMA demonstrated improved sensitivity and specificity in assessment of skin lesion physiological characteristics following analytical validation, which was performed as described in the ACP. Labeling was updated in accordance with the updated MMA performance, and communication was provided to users on the improved performance characteristics. The modified algorithm that 'learned' on real-world data can be marketed without additional FDA review.

Modification Scenario 2B: Change in input (type ii modification), consistent with SPS and ACP

The manufacturers analytical validation demonstrated the MMA can be deployed on two additional smartphones that have image acquisition criteria consistent with what was provided in the SPS and ACP. The analytical performance using the new image acquisition systems was consistent with the initial performance. Labeling was updated to reflect the new MMA compatibility with additional smartphones, which may increase access of the MMA in the healthcare community. Communication updates on device compatibility were also provided. The algorithm modification may be made without additional FDA review.

Modification Scenario 2C: Change related to intended use (type iii modification), inconsistent with SPS and ACP

The manufacturer would like to distribute a new version of the MMA on smartphone platforms that is patient-facing. The MMA would provide an analysis of the physiological characteristics of skin lesions, as it does currently, and direct patients to follow-up with a dermatologist based on the preliminary analysis

of the malignancy of the skin lesion. The modification also introduces many new, unconsidered risks that were not yet mitigated in the current SPS or ACP given that the new MMA will be patient-facing. FDA may require a new premarket submission or application and updated SPS and ACP for this algorithm modification.

3. X-Ray Feeding Tube Misplacement SaMD

Description of SaMD: A SaMD analyzes chest x-rays taken of hospitalized inpatients after they had undergone placement of a feeding tube, in order to evaluate the tube placement, detect incorrectly placed tube, and triage radiologists review of those films among the queue of similar images. This SaMD will 'drive clinical management' in a 'serious healthcare situation or condition.'

SPS: The manufacturer proposes two potential modifications for the X-Ray Feeding Tube Misplacement SaMD:

- Improve accuracy of performance in identification of incorrect tube placements using real-world data; and
- Allow the algorithm to notify nursing staff to check on the patient, in parallel with its triaging of the film in the radiologist's queue, based on achieving a pre-specified performance threshold.

ACP: For these modifications, the ACP details methods for real-world data collection, including inclusion and exclusion criteria, reference standard information, and comparative and statistical analysis for performance testing. The ACP also details the analytical validation for performance improvement, as well as the clinical validation for determining high-confidence cases. The manufacturer follows GMLP.

Modification Scenario 3A: Increase in performance and change related to intended use (type iii modification), consistent with SPS and ACP

The manufacturer re-trained and re-validated the algorithm on real-world data, as described in the ACP, which improved the SaMD accuracy in identifying incorrect feeding tube placements. This performance improvement provided the data that supported clinical validation of high confidence cases, as described in the ACP. The new version of this SaMD has a modified healthcare situation or condition in which nursing staff would be notified in parallel with radiologists, for high confidence cases with feeding tube misplacements. This could allow for improved and rapid response and corrective action for that subset of impacted patients. Labeling of the device was changed in accordance with the SPS and ACP. The modified algorithm can be marketed without additional FDA review.

Modification Scenario 3B: Change related to intended use (type iii modification), inconsistent with SPS and ACP

The manufacturer used a new database of images with expert radiologists' annotations to train and evaluate a new AI/ML algorithm to identify pneumonia on chest x-rays. The algorithm development and validation testing are similar to what was originally presented in the SPS and ACP protocols; however, adaptions were necessary given the new and different clinical tasks, requiring, for example, new reference standards. The changes reflect a change in the healthcare situation and condition as well as the significance of information, and result in a new intended use for the product. FDA may require a new premarket submission or application and updated SPS and ACP for this algorithm modification.

VI. Appendix B: Proposed Content for an Algorithm Change Protocol (ACP)

An ACP is a description of the set of specific methods that a manufacturer has in place to achieve and appropriately control the risks of the anticipated types of modifications delineated in the SPS. The ACP provides a step-by-step delineation of the data and procedures to be followed so that the modifications achieve their goals and the device remains safe and effective after the modification. The description below is intended to highlight some components of an ACP, but is not intended to be an exhaustive list of ACP components:

Data management plan addressing how data will be collected, added to existing data sets, and used: This data management plan may include a quality assurance (QA) plan for determining which new data are appropriate for inclusion as part of an expanded training data set; an approach to the reference standard determination; a data augmentation strategy that allows for additional training and independent test data to be added; and an auditing and sequestration strategy to monitor, document test dataset independence, and control access to both the training and test datasets as additional data are being included and any revised algorithm is being retrained and tested.

Protocols for re-training / optimizing the medical device algorithm: These protocols may include a re-training strategy that describes the objective of the retraining; the algorithm components that may be modified as a result of the learning process; and any criteria that must be met during the re-training process to trigger a more comprehensive performance evaluation using the test dataset.

Performance evaluation protocols: These protocols may include a description of the intervals of when a new algorithm may be trained and evaluated to consider updating the medical device algorithm; the delineation of appropriate metrics and analysis procedures; statistical analysis plans; appropriate measures to minimize information leakage about the test data set if part of it is re-used in multiple evaluations; performance targets that the revised algorithm must achieve; and protocols for testing, which may be applicable for that device and type of change, for example, for testing "with clinicians in the loop," as appropriate.

Update procedures that describe how updated medical device algorithms will be tested, distributed, and communicated when released: These update procedures may include a description of the update plan including expected frequency of updates and whether the updates will be global (all devices use the same version of the algorithm) or local (multiple versions of the algorithm targeted for specific sub-populations are distributed); version tracking and control; obsolescence planning; requirements for host software/hardware requirements; any plans for 'beta' release of the updated medical device algorithm concurrent with the previous version; and communication procedures that describe how users will be notified of updates and any information that will be conveyed to users about the update.

Questions / Feedback on the ACP:

- *Are there additional components for inclusion in the ACP that should be specified?*
- *What additional level of detail would you add for the described components of an ACP?*

VII. Questions / Feedback

1. Do these categories of AI/ML-SaMD modifications align with the modifications that would typically be encountered in software development that could require premarket submission?
2. What additional categories, if any, of AI/ML-SaMD modifications should be considered in this proposed approach?
3. Would the proposed framework for addressing modifications and modification types assist the development AI/ML software?
4. What additional considerations exist for GMLP?
5. How can FDA support development of GMLP?
6. How do manufacturers and software developers incorporate GMLP in their organization?
7. What are the appropriate elements for the SPS?
8. What are the appropriate elements for the ACP to support the SPS?
9. What potential formats do you suggest for appropriately describing a SPS and an ACP in the premarket review submission or application?
10. How should FDA handle changes outside of the "agreed upon SPS and ACP"?
11. What additional mechanisms could achieve a "focused review" of an SPS and ACP?
12. What content should be included in a "focused review"?
13. In what ways can a manufacturer demonstrate transparency about AI/ML-SaMD algorithm updates, performance improvements, or labeling changes, to name a few?
14. What role can real-world evidence play in supporting transparency for AI/ML-SaMD?
15. What additional mechanisms exist for real-world performance monitoring of AI/ML-SaMD?
16. What additional mechanisms might be needed for real-world performance monitoring of AI/ML-SaMD?
17. Are there additional components for inclusion in the ACP that should be specified?
18. What additional level of detail would you add for the described components of an ACP?