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*

Page 5. In Figure 1, I suggest inverting the order of the row and column categories so that cell entries have risk categories ascending from top left to bottom right, not the reverse (we don't count backwards!). Also, you may wish to have just three risk categories to correspond to the three device classes: Class 1 General Controls, Class II General Controls and Special Controls, and Class III General Controls and Premarket Approval.

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?*

I recommend that performance not be listed as a device modification. Performance is not a modification to a device. Change in performance could be due to evolving medical practice or evolving intended use population, even though the device is unchanged.

•*What additional categories, if any, of AI/ML-SaMD modifications should be considered in this proposed approach?*

I recommend you list "change in outputs" as a modification that is distinct from "change in inputs". A change in inputs would certainly change the device. However, a change in the how the inputs are combined (e.g., weighted) to obtain the device output would also change the device. As a simple example, a cut-off may be applied to an underlying continuous signal to obtain binary device output indicating either presence or absence of a clinical condition. If the cut-off is changed, the device also changes.

•*Would the proposed framework for addressing modifications and modification types assist the development AI/ML software?*

The three modifications listed – performance, inputs, and intended use – aren't likely to be part of a learning algorithm that modifies device outputs. More likely, types of modifications made by a learning algorithm would include updating model coefficients on inputs or changing the cut-offs applied to the inputs.

Questions / Feedback on GMLP:

•*What additional considerations exist for GMLP?*

In Figure 3. analytical validation should come before clinical association and validation. Thus, In Figure 3, I would position analytical validation as the leftmost validation.

Clinical validation means that the association between device result and target condition is clinically significant (CDRH Pivotal Studies guidance). Thus I recommend not distinguishing between validation and association. However, clinical utility means that the device output, when used as intended per

instructions of use, improves clinical outcomes. A diagnostic device is validated in a clinical performance study and is demonstrated to have clinical utility in a clinical outcomes study (Pivotal Studies Guidance).

•*How can FDA support development of GMLP?*

For GMLP, I recommend model training, tuning, and validation be conducted in separate datasets. Sometimes, multiple models are developed in the training dataset, one is selected in the tuning dataset, and the one selected is validated in the validation set.

Training and validating a model on the same dataset is not recommended. It is subject to potentially enormous "resubstitution bias". At the very least, perform "cross-validation" (average validation performance across repeated splits of the dataset into training and validation sets).

•*How do manufacturers and software developers incorporate GMLP in their organization?*

Questions / Feedback on SPS and ACP:

•*What are the appropriate elements for the SPS?*

SPS elements should include any device specifications (e.g., model parameters or features) eligible for change by learning algorithm.

•*What are the appropriate elements for the ACP to support the SPS?*

Risk mitigations might include conditions under which modifications are not allowable, e.g., substantial modifications based on new data, new data that are "markedly different" from training data, insufficient quality of new data, indeterminate data, etc.

•*What potential formats do you suggest for appropriately describing a SPS and an ACP in the premarket review submission or application?*

Just like any other device, a software device employing a learning algorithm needs to be validated. The process used by the learning algorithm to update the device output, i.e., the SPS, needs to be validated as safe and effective in a well-conducted pre-market pivotal study. Each potential change needs to be validated. To validate the learning algorithm, demonstrate device progression and performance as data accumulate.

Questions / Feedback on premarket review:

•*How should FDA handle changes outside of the "agreed upon SPS and ACP"?*

Device sponsor could define a new SPS and ACP that includes the changes outside of the existing SPS and ACP. Have sponsor validate the new SPS and ACP for future device updating.

•*What additional mechanisms could achieve a "focused review" of an SPS and ACP?*

•What content should be included in a "focused review"?

Benefit-Risk evaluation