Bradley Merrill Thompson
t 202.861.1817
f 202.861.3517
bthompson@ebglaw.com

FDA-2016-D-2483

VIA ELECTRONIC SUBMISSION

December 9, 2016

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

Re: Docket No. FDA-2016-D-2483: Draft Guidance for Software as a Medical Device: Clinical Evaluation

Dear Sir or Madam:

The Clinical Decision Support Coalition ("CDS Coalition" or the "Coalition")[1] commends FDA for its tremendous efforts in working toward greater consistency in international medical device software regulation, and welcomes the opportunity to provide comments on FDA's "Draft Guidance for Software as a Medical Device: Clinical Evaluation" ("Draft Guidance"). As the Coalition has already submitted substantive/technical comments to the International Medical Device Regulators Forum ("IMDRF") regarding its proposed Software as a Medical Device ("SaMD") clinical evaluation guidance document (which is identical to the Draft Guidance),[2] this letter will largely focus on the Coalition's administrative concerns surrounding the adoption of the Draft Guidance in the U.S.

However, we would like to emphasize that overall, the Coalition is concerned that the Draft Guidance proposes much more demanding standards for the clinical evaluation of medical

[1] The CDS Coalition is a diverse group of stakeholders consisting of software providers, IT infrastructure manufacturers, healthcare providers, medical device and pharmaceutical manufacturers, trade groups and members of the clinical community. Focused on clinical decision support software, the Coalition's goal is to ensure a risk-based and clearly defined regulatory system for such software that appropriately balances the need for regulatory oversight with the need for innovation and access to new technology.

[2] These IMDRF comments are attached as Appendix A.

device software (which includes CDS software) than FDA has historically required. In fact, the Draft Guidance seems to suggest that manufacturers should conduct clinical evaluation for what would equate in the FDA classification structure to nearly all Class II and Class III devices, as well as a subset of Class I devices.

As set forth in further detail below, the Coalition believes that adoption of the Draft Guidance in its current form is both unworkable and inappropriate on multiple levels. Accordingly, we request that FDA formally revoke the Draft Guidance. This letter highlights our key administrative concerns driving this request for revocation, including:

(1) the unclear logistics of how FDA plans to adopt the Draft Guidance, and the scope of what the Agency plans to adopt;
(2) the lack of linkage between the Draft Guidance and current U.S. laws, regulations and policies; and
(3) the clear disconnect between the recommendations in the Draft Guidance and our current understanding of FDA's positions in this space, and the confusion that will result if FDA does not formally revoke the document.

We have also included at the end of the letter some additional specific comments.

I. Improper Guidance Development Process

As it is not typical FDA practice to issue a draft international guidance document for direct application in the U.S., the Coalition has significant concerns and questions surrounding how the Draft Guidance will be adopted by FDA. This section describes these logistical concerns and questions, which relate to: (1) the need for the U.S. public to be able to comment on the final IMDRF clinical evaluation document; (2) the scope of IMDRF documents FDA intends to adopt; and (3) the format of the Draft Guidance.

A. Opportunity for U.S. Stakeholder Comment on the Final IMDRF Document

Because the IMDRF clinical evaluation document is only in draft form in the IMDRF guidance development process, we cannot predict what exactly the final IMDRF document will say. Accordingly, FDA should wait for the IMDRF clinical evaluation document to be finalized by IMDRF and then provide U.S. stakeholders the opportunity to comment on that final document. This opportunity to comment on the final IMDRF clinical evaluation document should be provided before FDA considers adopting any form of the IMDRF document in the U.S.

Based on stakeholder comments on the final IMDRF clinical evaluation document, FDA may decide to develop its own guidance document on this topic, using the final IMDRF document as a guide (but possibly omitting or revising certain principles and concepts). This "clean sheet of paper" approach, which avoids direct adoption of an international regulatory document, may be the most workable option based on the expected political stance of the new administration. Alternatively, FDA may decide to publish the final IMDRF clinical evaluation document as is, but with a cover sheet explaining areas where FDA agrees and disagrees with document. However, if FDA chooses this latter approach, FDA should detail its plan for

managing lifecycle changes to the IMDRF document. For example, if IMDRF decides to make fundamental changes to its clinical evaluation framework, FDA should explain how it will determine whether those changes should also be incorporated into FDA's framework.

While we understand that FDA wants to encourage U.S. stakeholders to offer comments at this stage so their perspectives can be considered in finalizing the IMDRF clinical evaluation document, we believe that FDA's adoption of a draft international guidance document on the extremely complex topic of clinical evaluation for SaMD is premature and inappropriate at this juncture.

B. Scope of the Draft Guidance

With respect to scope, FDA has not provided clarity on which specific IMDRF documents it is proposing to adopt as FDA guidance. While the proposed IMDRF clinical evaluation document is the direct subject of the Draft Guidance, this document references and builds off of three previous IMDRF guidance documents and eight previous Global Harmonization Task Force ("GHTF') guidance documents. Therefore, the reader needs to be familiar with these various other documents in order to understand the Draft Guidance. This dependence on outside documents creates confusion, especially as the Draft Guidance appears to introduce certain fundamental concepts not discussed in the previous documents. Although a self-contained document would be most preferable, we ask that, at a minimum, FDA expressly state which IMDRF and GHTF documents it is proposing to be included as FDA guidance. If FDA does not intend to adopt the previous IMDRF and GHTF guidance documents, the Agency should explain how this impacts interpretation of the Draft Guidance.

C. Format of the Draft Guidance

Because the Draft Guidance was developed as an IMDRF guidance document, it does not follow the standard drafting style of FDA guidance. Specifically, the Draft Guidance is written such that readers may perceive its proposed framework and criteria as requirements, rather than suggestions or recommendations (this is also true of the various IMDRF documents that the Draft Guidance builds upon). Use of mandatory language, such as "must" and "manufacturers are expected to," which appears in the IMDRF documents, is inappropriate in FDA guidance. FDA itself reiterated this principle in its recent third party review program draft guidance, noting that FDA guidance documents do not establish legally enforceable responsibilities, but rather represent the Agency's current thinking on a topic, and should only be viewed as recommendations (unless a specific regulatory or statutory requirement is cited).[3] However, we believe that FDA can address this formatting issue fairly easily by including a cover sheet or specific language in the Draft Guidance that explains that any mandatory language included in the IMDRF document(s) will not be construed as mandatory for FDA purposes.

[3] FDA, *510(k) Third Party Review Program: Draft Guidance for Industry, Food and Drug Administration Staff, and Third Party Review Organizations* at 7 (Sept. 12, 2016), http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM339697.

II. Failure to Link to U.S. Law, Regulations and Policy

The Coalition believes that the Draft Guidance has limited value for U.S. stakeholders because it is not specifically based on and does not link to U.S. laws, regulations and policies. Indeed, the Draft Guidance clearly states that the recommendations in the document are not meant to replace or conflict with the premarket or postmarket regulatory requirements related to the regulatory classification of SaMD in different jurisdictions (see lines 126-131).

Of course, this approach was intentional, and appropriate for an international guidance document designed to create an international consensus around clinical evaluation criteria for SaMD and assist regulators in developing their own country-specific regulations and guidance. However, because U.S. industry stakeholders were not envisioned as the primary intended audience for the Draft Guidance, and the primary purpose of the document is not to directly influence industry behavior, we do not see how the document can seamlessly be translated into FDA guidance. It seems impossible that this document could accurately reflect the Agency's specific expectations and recommendations for U.S. industry on this topic.

Importantly, the missing link between the principles set forth in the Draft Guidance and FDA's current regulatory framework and requirements creates tremendous regulatory uncertainty for U.S. stakeholders, and limits the ability of stakeholders to provide meaningful comments on the principles detailed in the Draft Guidance. For example, the Draft Guidance neither links to FDA's device classification system[4] nor the historical regulatory approaches that FDA has relied upon to determine the need for clinical evaluation. Thus, although the document provides potential background for discussions between FDA and stakeholders regarding future FDA rulemaking and guidance development on this topic, without this critical link to the current U.S. regulatory structure, we cannot reconcile how the Draft Guidance fits within our understanding of FDA's current clinical evaluation expectations (as discussed further in the next section). In addition, we cannot predict how exactly this document will be used by FDA to inform its decision-making.

We note, however, that this linkage issue is not necessarily problematic across all of the IMDRF SaMD guidance documents. Consider, for example, the IMDRF guidance document on quality management systems for SaMD ("IMDRF QMS Guidance"), which also does not directly link to U.S. requirements. The Coalition views the missing linkage in this document as less concerning than the missing linkage in the Draft Guidance because unlike clinical evaluation principles, which are deeply impacted by country-specific legal and regulatory requirements, QMS principles are much more harmonized internationally. As such, the Coalition finds the recommendations in the IMDRF QMS Guidance to be useful despite the fact that they are not tethered in U.S. law. In fact, we sent a letter to FDA in April 2016 to specifically voice our support of the IMDRF QMS Guidance, and recommend adoption of that document in the U.S.

[4] We note that this same issue exists with respect to the IMDRF Risk Stratification Guidance (IMDRF, *"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations* (Sept. 18, 2014)).

III. Difficulty in Reconciling the Draft Guidance with Historical FDA Expectations

Based on the current FDA requirements and guidance in place, as well as our experience with FDA through previous submissions, we have some sense of FDA's expectations with respect to clinical evaluation for CDS. However, we are finding it difficult to reconcile our current understanding of FDA's expectations with the proposed clinical evaluation framework described in the Draft Guidance. This section discusses: (1) specific areas where the Draft Guidance seems to diverge from our understanding of FDA principles on this topic; (2) the lack of alignment between the terminology and concepts in the Draft Guidance and other FDA guidance; and (3) the failure of the Draft Guidance to take into account unique FDA regulatory concepts in establishing its clinical evaluation standards.

A. Diverging Expectations

The Draft Guidance includes certain recommendations that overlap with and diverge from what FDA has previously indicated to be its current thinking. This creates significant confusion as stakeholders cannot definitively tell whether FDA is signaling a departure from its historical approach on these issues. For example, we have repeatedly heard from FDA that clinical validation may be required for CDS that is novel, complex or non-transparent. While the Draft Guidance discusses the concept of novelty (see, e.g., lines 253-257) (and mentions the concepts of complexity and transparency in passing (see line 449 and lines 813-820, respectively)), the test for determining clinical evaluation requirements set forth in the Draft Guidance introduces new factors not traditionally considered by FDA, and overall, departs substantially from the test we understood FDA to currently be applying.

As another example, the Draft Guidance expects the collection of real-world evidence on a postmarket basis for certain CDS that would fall within the lowest IMDRF risk category (Category 1) (see lines 890-893). This type of real-world data collection, while being considered by FDA as a potentially beneficial practice for certain software, has never been a specific FDA requirement in this space (especially for low-risk devices).

In addition, the Draft Guidance seems to veer from the well-established principles that FDA has applied with respect to clinical evaluation requirements for computer-assisted detection ("CAD") technology. In 2012, FDA issued two CAD-related guidance documents that cover in detail when clinical evaluation (e.g., reader studies) is required for CAD products as well as the scope of such evaluation. Because FDA has already adopted these guidance documents, issuance of the Draft Guidance, which offers separate insight on the need for clinical evaluation for SaMD, leaves stakeholders guessing as to what is required with respect to CAD products.

Issuing Draft Guidance that contains expectations that vary from what has historically been required by FDA inappropriately shifts the burden onto U.S. stakeholders to try to figure out which portions of the Draft Guidance FDA actually believes are applicable in the U.S. This drives confusion and regulatory uncertainty within industry – the exact opposite of what guidance is intended to accomplish.

B. Varying Terminology and Concepts

We are also concerned that the terminology and concepts from the Draft Guidance do not line up with the content of existing FDA guidance documents related to medical device software (e.g., "General Principles of Software Validation; Final Guidance for Industry and FDA Staff" ("Software Validation Guidance") and "Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices" ("Software Premarket Submissions Guidance")). For example, FDA will need to reconcile the categories of SaMD from the IMDRF document with the "Level of Concern" categories from the Software Premarket Submissions Guidance. Specifically, FDA should clearly articulate how the Level of Concern analysis fits within the benefit-risk assessment framework detailed in the Draft Guidance, and describe how a product's Level of Concern category impacts the scope of clinical evaluation required. With respect to the Software Validation Guidance, FDA will need to explain how the software verification, validation and hazard analysis concepts described in that document mesh with the analytical validity and scientific validity principles discussed in the Draft Guidance, as these concepts/principles appear to overlap and intertwine.

C. Failure to Account for FDA-Specific Regulatory Concepts

Finally, the Draft Guidance does not give adequate consideration to certain important concepts that can impact clinical evaluation requirements for medical device software. For instance, the Draft Guidance does not take into account the fact that FDA is precluded from regulating the practice of medicine. From our perspective, this is a critical omission as that issue is very important when considering software that merely informs clinical decision making. As detailed in our "CDS Coalition Classification Proposal" (which is included as an attachment to our IMDRF comments in Appendix A), the Coalition believes that if a given piece of software is merely informing clinical decision making and the clinician is not substantially dependent on the software – meaning the software is transparent, the clinician has the education and training to independently make the patient care decision, and the clinician has adequate time to reflect – such software should not be subject to FDA regulation (and certainly not clinical evaluation requirements).

In addition, the Draft Guidance does not account for the Agency's use of enforcement discretion. As a result, confusion can arise when software that falls within the enforcement discretion category also falls within one of the IMDRF categories that per the Draft Guidance, carries clinical evaluation requirements (such as the requirement to collect real-world postmarket data). Although engaging in clinical evaluation activities may be a best practice for a manufacturer of software that falls within FDA's enforcement discretion category, it should be made clear that these activities are not expected for this type of software. Another area where this enforcement discretion confusion may arise is in the statement on line 400 that "all manufacturers" are expected to follow adequate QMS. Again, while we believe following QMS is a best practice for software manufacturers, this should not be construed as an expectation for <u>all</u> software manufacturers.

IV. Additional Comments

- To keep pace with innovation in medical device software development, we recommend that FDA's regulatory approach in this area take into account the integration of medical software and knowledge based systems with other diagnostic and therapeutic products.
- We request that FDA provide guidance on clinical evaluation issues related to special regulatory pathways for CDS, such as the de novo classification pathway and the humanitarian device exemption ("HDE") pathway.
- We ask that FDA provide detail on how the Draft Guidance applies to software being developed as a Medical Device Development Tool ("MDDT").
- It is unclear how the Draft Guidance's recommendations surrounding the appraisal of clinical evaluation evidence (see Section 6.6) fit in with the current Center for Devices and Radiological Health ("CDRH") benefit-risk assessment framework (which we understand is in the process of being updated to incorporate the concept of real-world evidence). Further, it not clear to us how the clinical evaluation principles from this section should be reconciled with the requirements set forth in 21 C.F.R. § 860.7.
- Line 786 references as part of benefit-risk assessment the "[o]bjective consideration of patient preference." We ask that FDA provide clarity on whether this should be evaluated pursuant to FDA's recent patient preference guidance.[5]
- Certain principles in this document conflict with the principles recently articulated in FDA's software modifications guidance, where the Agency tries to draw the distinction between changes that trigger the need for a new 510(k) versus those that do not.
- We ask that FDA clarify how the Draft Guidance applies to provider organizations that internally develop CDS software for their own internal use. Are the principles described in the Draft Guidance applicable to such provider organizations that do not sell or market their internally developed CDS software?

* * * *

Based on the significant and fundamental concerns described in this letter, we are asking that FDA formally revoke the Draft Guidance. As the Agency is aware, many of its guidance documents tend to remain in draft form for an extended, and sometimes indefinite, period of time. During this time, both Agency staff and industry stakeholders will often rely on such draft guidance documents as they represent at least preliminary insight into FDA's current thinking in a given area.

---

[5] FDA, *Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders* (eff. Oct. 23, 2016), http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446680.pdf.

We see this as particularly problematic with respect to the Draft Guidance due to our strong belief that its recommendations are both inappropriate and irreconcilable with FDA's current framework. To avoid potential confusion surrounding the meaning and impact of the Draft Guidance, we believe formal revocation of the document is the most appropriate course forward. If the CDS Coalition can assist in developing revised or new draft guidance on this topic, please do not hesitate to contact us.

Yours truly,

Bradley Merrill Thompson
On Behalf of the Clinical Decision Support Coalition

APPENDIX A

Comments from the Clinical Decision Support ("CDS") Coalition

Document number: **SaMD WG (PD1)/N41R3** Title: **Software as a Medical Device (SaMD): Clinical Evaluation**

Submitted by: Bradley Merrill Thompson Affiliated to: Clinical Decision Support Coalition On: Dec. 9, 2016

High level feedback	Yes/No	Comment and rationale and proposed recommendations
Does the document address the intention captured in the introduction/scope or vice-a versa?	No	The first three IMDRF documents on this SaMD topic produced a useful framework. Unfortunately, this document on clinical evaluation seems fundamentally flawed. Below we provide specific comments, but overall, the clinical evaluation document does not build on the prior work, but rather departs from that work.
Does the document appropriately translate and apply current clinical vocabulary for SaMD?	No	We find that the document introduces certain new concepts and clarifications on terminology and framework that are not addressed in earlier IMDRF documents on SaMD. As discussed below, we believe IMDRF should revisit its earlier documents to incorporate and more fully flesh out these new concepts/clarifications.
Are there other types of SaMD beyond those intended for non-diagnostic, diagnostic and therapeutic purposes that should be highlighted/considered in the document?	Yes	It is not so much that the document omits important intended uses, but rather that the document omits important technology. The document seems to completely ignore software that uses machine learning, and the subtle nuances around software that change over time. In addition, the document does not address how to treat software with multiple functions that carry varying levels of risk. Further, on a more general level, the proposed framework seems to isolate regulation of SaMD from other software, medical devices and therapies. We believe that the regulatory approach for SaMD should incorporate an analysis of the convergence of software and knowledge-based systems with traditional medical devices and drugs.
Does the document adequately address the relevant clinical evaluation methods and	No	The document does not adequately address the importance of assessing the environment where the SaMD "lives." This is problematic as we believe the

High level feedback	Yes/No	Comment and rationale and proposed recommendations
processes for SaMD to generate clinical evidence?		performance of SaMD is as much dependent on the processes it is integrated into (e.g., who it is used by, when it is used, amount of training provided, etc.) as it is on the logic/inherent capabilities of the SaMD itself.
Are there other appropriate methods for generating clinical evaluation evidence that are relevant for SaMD beyond those described in the document?		
Are the recommendation identified in section 7.2 related to the" importance of clinical evaluation evidence" appropriate as outlined for the different SaMD categories?	No	As described further in our detailed comments below, IMDRF is proposing far more demanding standards than we believe are appropriate for the production of clinical validation generally, including requiring premarket clinical validation for low risk medical devices and the collection of real-world evidence for essentially all SaMD after marketing. These proposals are inconsistent with the manner in which software and knowledge-based platforms are created, evaluated and improved. For insight into the scope of data collection requirements that we believe should apply to clinical decision support software, please see the "CDS Coalition Classification Proposal" (attached here as Appendix A). This document also provides a proposed framework for risk stratification of clinical decision support software generally, which borrows from IMDRF's approach.
Are the recommendation identified in section 7.3 related to the" importance of independent review" appropriate as outlined for the different SaMD categories?		
Given the uniqueness of SaMD and the proposed framework -- is there any impact on currently regulated devices or any possible adverse consequences?	Yes	As discussed below, requiring the active collection of real-world evidence throughout the entire commercialization for nearly all SaMD is simply not feasible. More importantly, it is essential to minimize the data collection burden imposed on SaMD users, or they will avoid making use of critically important software.

Detailed Feedback

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
1.	Page 18 / Section 6.1 / Lines 524-535	Technical	We are having difficulty understanding why the new terms "diagnostic SaMD" and "non-diagnostic SaMD" have been introduced. First, the definitions of these terms appear to overlap with each other. Non-diagnostic SaMD, for example, is defined to include products that "help aid in diagnosis." Here, IMDRF seems to be relying too much on the clarity around the word "diagnostic" as being somehow only related to a definitive diagnosis. That asks too much of the English language. The word "diagnostic" has many meanings, and the fact that IMDRF is using it in a hypertechnical way here needs to be more explicitly stated. Similarly, non-diagnostic SaMD includes products that aid in treatment. This seems inconsistent with the language that precedes these defined terms, which states that diagnostic and non-diagnostic SaMD are categories of SaMD that are not intended for treating a situation or condition. So we gather that IMDRF is using the word "treating" in a hypertechnical way that does not include anything used as an aid in treatment. Even if IMDRF can get away with such hypertechnical readings on the diagnostic side, it cannot do so on the treatment side. As already mentioned, we guess that in this new definition, IMDRF is trying to draw a distinction	Delete the new concept of diagnostic SaMD and non-diagnostic SaMD (lines 524-535). Consider creating two new subcategories within the previous category of "to treat or diagnose." But that should only be done if IMDRF can come up with an evidence-based justification for that risk stratification to justify the difference in evidentiary requirements.	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			between software that provides a definitive diagnosis versus software that aids in reaching a diagnosis by providing information which is not by itself conclusive, but needs to be interpreted in the context of other information. But if that is the case, these new defined categories intertwine and overlap with the existing IMDRF principles for differentiating SaMD functionality on the basis of whether the software treats or diagnoses on the one hand, or drives or informs clinical management on the other hand. If IMDRF is merely trying to distinguish diagnosis from treatment within the category of "diagnosis or treat," why doesn't IMDRF just say that? In other words, why not just create subdivisions within the "diagnosis or treat" category for diagnosis versus treating? [Note that below, we frankly do not understand why this distinction is important from a risk perspective. IMDRF has not laid the evidentiary groundwork for asserting that there is a substantial difference in risk between diagnostic decisions and treatment decisions.] This whole issue does expose a weakness in the original concept of "diagnose or treat." We find that when many people read that language, they conclude that IMDRF is referring to a closed-loop system where diagnosis or treatment is automatic, without human intervention. But we do not believe that is		

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			what IMDRF intended. We think IMDRF ought to go back to its second document – the document on risk stratification – and amplify and clarify the meaning of the categories for "diagnosis and treatment" and "driving clinical management." In a narrative, IMDRF ought to explain that there is a difference for regulatory and risk purposes between claims that a product merely provides information that is useful to a clinician – together with other information – in reaching a diagnosis, versus information which is definitive in pronouncing a diagnosis. Medical device companies are accustomed to that distinction when the medical device at issue is an instrument that produces test results. When the topic, however, is SaMD, because such software does not produce any new information but merely manipulates existing information, the concept is a bit muddied. More example language may help. 1. To treat or diagnose a. Example language – "Patient Betty Smith has stage IIIC melanoma." 2. To drive clinical management a. Example language – "Patient Betty Smith has three out of the five common indicators of melanoma." 3. To inform clinical management a. Example language – "Patient Betty		

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			Smith has 350 international units per liter (IU/L) of lactate dehydrogenase (LDH)." Although these different approaches to conveying information are well established within the medical device community for in vitro diagnostic tests, when writing a guidance document that hinges on these differences, it is important for this guidance to be very clear about the foundations for the distinctions the guidance document is drawing. It is easy to differentiate 1 from 3, in that 1 is definitive with respect to the diagnosis, whereas 3 indeed does not even mention the diagnostic implication. It is harder to define category 2 because it on the one hand mentions a disease, but in the context to make it clear that no definitive conclusion is reached. In any event, building off this original framework adopted in the second IMDRF document on risk stratification, it seems as though IMDRF now wants to parse the first category into two subcategories – diagnosis versus treatment and anything that isn't diagnostic. If that's the intent, why not just say so? Why create these new definitions disconnected from the existing risk stratification?		
2.	Page 18 / Section 6.2 / Lines 547-552	Technical	This language seems to suggest that real-world evidence always needs to be collected on a postmarket basis for all SaMD. But in the real world, real-world evidence collection is not always equally	Add language that acknowledges the need to consider the costs and incremental risk of real-world	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			valuable. Sometimes it is extremely informative, while other times it is only marginally useful, if at all. Equally important, it is critical to consider the costs of such collection, and the burden the collection of such information imposes on software users. In assessing these costs and burdens, it is also important to bear in mind providers' responsibility to ensure the privacy and security of their patient's health information, as well as limitations on how that data can be used. These factors may make this type of data collection infeasible, or may render the burden and risk of such data collection greater than the added value. Thus, manufacturers should be able to develop a tailored approach based on an assessment of the value of the postmarket data collection weighed against risks and costs. Specific considerations should include the following. **Security risks associated with increased collection and transmission of data:** Devices connected to networks, particularly the internet, have significantly higher cybersecurity risks (*See* FDA's "Content of Premarket Submissions for Management of Cybersecurity in Medical Devices: Guidance for Industry and Food and Drug Administration Staff"), which negatively impacts both the reliability of the device, the trustworthiness of the data, the patient's privacy, and cybersecurity risks to all other provider systems and devices also connected to the same	data collection, as well as the inconvenience to users. Also add language that allows manufacturers flexibility to choose between premarket validation and postmarket data collection with respect to their clinical evaluation activities. Specifically, we suggest the following revisions to lines 547-552: "**SaMD manufacturers and users should consider conducting routine monitoring of** ~~I~~**i**nformation related to clinical evidence ~~should be monitored routinely by the manufacturer and user~~ once the SaMD is available on the market. **Where appropriate based on the associated costs and benefits**, ~~T~~**t**he manufacturer should plan for the continuous discovery of clinical data related to the safety, effectiveness and performance of the SaMD through appropriate post-market programs (e.g., post-market surveillance, adverse event	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			network. We can envision scenarios where a relevant data set would extend beyond the minimum data needed for clinical purposes (for example the unique ID number of the patient's mobile device). This would create additional privacy concerns for the patient and could require additional approval paperwork from the patient to agree to the use of the SaMD and the collection of data not directly relevant to the provision of care to the patient. Collected data may also contain personally identifiable information, which must be de-identified for analysis by and reporting to third parties. Given these concerns, providers, who have responsibility for protecting their patients' information and meaningfully communicating with patients about how their information will be used, may be hesitant to recommend or utilize software if the perceived privacy and security risks are too high. **Utilizing the potential of passive data collection:** It is also important to recognize the various methods by which data might be collected, and the relative costs and benefits of each method. The model laid out in the proposed document is for active collection of real-world evidence. By "active," we mean the data collection requires some action of the user to provide the information. This is in contrast to passive data collection where the involvement of the user is	reports, scientific publications, etc.) as part of their QMS to ensure the SaMD continues to meet the intended safety, effectiveness and performance. **However, recognizing that active postmarket data collection may not be a feasible or desirable strategy based on the costs of collection, privacy concerns and the undue burden placed on users (and also may not be necessary in the case of certain low risk SaMD), manufacturers have the option of alternative approaches, including managing risk through premarket validation strategies and/or passive postmarket data collection activities. The SaMD manufacturer should utilize the least burdensome approach to secure the data necessary to assure the safety, effectiveness and performance of the device**."	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			minimal, if any, and where data can be collected, communicated, and analyzed in real-time or asynchronously. It can be stored and analyzed within the device, or transferred over an enterprise network or the public internet. Postmarket data collection has already proven very problematic for other medical devices, and the language in the document does not seem to be focused on the unique capabilities of software to passively collect information. We endorse the use of the inherent qualities of software to collect information that allows for the improvement of the software over time, so long as it does not become a burden on the user. For purposes of U.S. policymaking, we have been in discussions with physicians regarding data collection in the context of clinical decision support software. The physician groups with whom we have spoken have been very much in favor of collecting and analyzing data, so long as the data collection process does not become a burden on physicians. More specifically, physicians do not generally support any additional forms to be filled out as a part of routine use of SaMD. **The need to balance value and burden:** We are concerned that the data collection from most SaMD will not be automatic, timely, convenient, or inexpensive. And the effort and cost to protect		

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			patient privacy will be higher for both providers/physicians and the SaMD manufacturer. From a policymaking standpoint, it is important for policymakers to recognize that it becomes self-defeating to impose on users significant data collection obligations, because those obligations then operate as a disincentive to use valuable software. Clinical decision support software offers tremendous public health advantages. To discourage the use of such software by requiring users to fill in added information to allow for real-world evidence collection would be to injure – not advance – the public health. Further, it needs to be recognized that sometimes the value of collecting information is marginal at best. And that marginal value needs to be weighed against incremental cost. There is a cost not only to the collection of the data, but also the analysis of it. That cost-benefit evaluation needs to be more explicitly factored into the IMDRF model. Our most important point, however, is that the evidence for SaMD needs to be considered holistically, meaning both premarket validation and postmarket data requirements need to be considered together. All data costs money to produce. All data costs time to produce. Not all data is equally valuable. Not all data is worth the money it costs to produce.		

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			If postmarket real-world evidence is useful for validation, the utility of that ought to be weighed against premarket clinical performance evaluation. In the end, all regulators should require only the least burdensome amount of data necessary to assure the safety and effectiveness of the device. So if software is adequately validated premarket, there may be little or no need to collect real-world evidence after the fact. On the other hand, policymakers may wish to allow products to get to market more quickly by minimizing the premarket clinical performance validation required, and instead using ongoing collection of real-world evidence to assure the safety and effectiveness of the software. But this issue isn't as simple as more is always better. More may be better, but sometimes marginally so and sometimes at great expense. The IMDRF proposed document seems to pay virtually no attention to weighing all of these factors out and coming up with the least burdensome approach to clinical validation.		
3.	Pages 18-19 / Section 6.2 / Lines 561-564	Technical	There seems to be a grammatical error and/or missing content in these lines. Based on other sections of the document, it appears that the authors wanted to state that diagnostic SaMD has a higher risk profile and therefore requires clinical performance evaluation. More fundamental than the grammatical error, it is not clear to us why diagnostic testing and therapeutic interventions would be treated in such a dramatically different way. Choosing the wrong therapy can be	Delete the references to diagnostic SaMD (lines 562-564) unless IMDRF can come up with an evidence-based justification for the distinction.	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			just as risky as making the wrong diagnosis. IMDRF needs to come up with an evidence-based distinction for its significantly different treatment of diagnostic and therapeutic decision-making. Again, to justify this new addition to the risk framework, IMDRF needs to identify factors that were not already adequately addressed in its original risk stratification document. For example, there are temporal elements to both the role of the software and the seriousness of the condition already in the risk model. So an incorrect treatment decision, just like an incorrect diagnosis, when faced with time pressure, can both result in injury. It isn't clear what incremental risk IMDRF is trying to control with the introduction of the diagnostic versus non-diagnostic decision-making.		
4.	Page 20 / Section 6.3.1 / Line 627-629	Technical	We disagree with the concept that data generated through literature searching need to directly relate to the specific SaMD in question. Instead, we believe the literature search should focus on the algorithm that forms the basis of the SaMD, but the data certainly need not be produced by the particular embodiment of the algorithm in the specific SaMD in question. Further, we believe that the literature search could also apply to a collection of algorithms where the output of one is used as the input to another well-known/used algorithm. The interfaces between these algorithms would need to be assessed, but not the	"The data generated through literature searching should relate directly to the **algorithm(s) that forms the basis of the** SaMD in question or earlier versions with justification as to why the data for the earlier versions are applicable (e.g. reports of clinical studies that have been performed by third parties). **Data from the literature can be used to support a collection of algorithms**	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			algorithms themselves.	**where the output of one is used as the input to another algorithm. In such case, the interfaces between these algorithms would need to be assessed, but not the algorithms themselves**."	
5.	Page 22 / Section 6.3.2 / Lines 680-682	Technical	The document does not appear to consider design validation activities to constitute assessment of clinical performance. However, we believe that for some SaMD, studies can be set up to accomplish both goals (i.e., design validation and clinical performance). In fact, the document itself recognizes in lines 265-267 that "[c]linical performance is evaluated and determined by the manufacturer during the development of a SaMD…" (emphasis added). Lines 539-546 also reference clinical evaluation (which would include collection of clinical performance data) being conducted during "the development phase of the SaMD lifecycle."	We recommend revising lines 680-682 to read: "NOTE: testing performed as part of the software development cycle verification and validation activities (customer feedback from focus groups, external analytical validity studies, and research studies) ~~is~~ **will** not **generally be** considered a clinical performance study **as clinical performance goals, requirements, and validation processes are often distinct from purely functional goals, requirements, and processes. However, the entire end-to-end SaMD development process is governed by the QMS. Components of the QMS development process (such as studies and user testing) may be synergistic and contribute to multiple**	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
				goals. As such, for certain SaMD, it may be possible for manufacturers to develop studies that assess both design validation and clinical performance."	
6.	Page 26 / Section 6.6.2 / Lines 840-845	Technical	We believe the document is lacking in its discussion of determining the benefits of SaMD. In particular, it is important to recognize more comprehensively the benefits such software can offer in comparison to the status quo. Those benefits need to be more explicitly factored into the equation so that they can be assessed versus whatever risk cannot be controlled. Likewise, the document fails to properly account for the risk associated with the unavailability of useful technologies to diagnose and treat rising global infectious and chronic disease burdens.	After line 845, we recommend adding the following language: "The broader public health benefits tied to the SaMD should also be considered in the benefit/risk determination. In particular, the consequences of keeping valuable technologies for diagnosing and treating rising infectious and chronic disease burdens out of the hands of clinicians should be taken into account."	
7.	Page 28 / Section 7.2 / Lines 880-921	Technical	This section seems to provide a fairly prescriptive description of the evidence required for the different categories of SaMD, but it does so with very little justification and connectivity with the previous 27 pages. While the previous pages generally describe the clinical evaluation framework and define the relevant terms discussed here, they fail to provide specific justification for the level of evidence required for the particular SaMD categories, which is essential. On a more specific level, we have the following	Using concepts such as novelty and the distinction between diagnostic and non-diagnostic software to determine clinical evaluation requirements for SaMD needs to be either justified by evidence or removed from the document. Further, clinical performance evaluation should not apply to any category II or III product without a very specific reason	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			concerns: • This section uses the concept of novelty as a core driver of risk and a core differentiator for what level of evidence is required. However, the document fails to provide a comprehensive explanation of the concept of novelty, including what constitutes novelty and why it is such an important risk factor as to trigger this added evidence requirement. On a more general level, we note that this is a concept that was not addressed in earlier IMDRF guidance documents on SaMD. If this novelty concept, and any other concepts (such as diagnostic v. therapeutic SaMD) are critical pieces of IMDRF's thinking on the subject of risk stratification, we believe that IMDRF should go back to its earlier document on risk stratification to incorporate these concepts. Introducing new ways of thinking about risk into this latest document without conforming previous documents creates significant confusion. • In addition, beyond novelty, this section is built on the concept of diagnostic SaMD, which as previously explained, we find to be unclear. It seems to overlap with the columns that seek to differentiate treatment and diagnosis from driving clinical management, for example. • Perhaps most fundamentally, we do not understand why after developing a nine	necessitating such evaluation, and only if the cost-benefit considerations demonstrate that such validation would be appropriate. Instead of taking the nine categories developed in the grid, lumping them into three categories, and trying to describe the level of evidence needed in each category, a more effective approach may be to discuss more broadly the types of risks that need to be addressed as we move along the spectrum toward higher significance and more critical health states. Such a framework would allow regulators in each jurisdiction, as well as manufacturers, to develop a cost-benefit analysis and a tailored approach to determine the level of evidence needed for particular SaMD products.	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			square matrix of different categories, four categories (i.e., II.ii, III.i III.ii and IV) are all lumped together and treated equally. It appears based on the rest of the document, as well as on the IMDRF risk categorization document, that category II and category IV represent dramatically different risk profiles. If that is the case, it does not seem to make sense to treat them the same here. • IMDRF might respond that the risk stratifications are for different regulatory approaches other than clinical validation. While we accept that regulators have other controls beyond imposing clinical performance testing, we would note that clinical performance testing should be reserved only for the highest risk software products. Fundamentally, IMDRF has not justified anywhere in this document the imposition of clinical performance testing on any products in category II. Indeed, we believe that only a small subset of category III SaMD products should be subjected to clinical performance testing. The benefits of clinical performance testing for these low risk software products simply does not outweigh the substantial cost associated with performing this testing. • Indeed, this is a fundamental flaw of the whole document. IMDRF simply does not take into account a cost-benefit analysis when		

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			setting the bar on when clinical performance testing is required. IMDRF has selected a rather arbitrary point without any evidence based justification.		
8.	Page 32 / Section 7.4 / Lines 1006-1015	Technical	As noted already, it is not realistic to require the active collection of real-world evidence throughout the entire commercialization of all SaMD. As a Coalition, we have spent a considerable amount of time thinking about how real-world evidence could be collected through software. Through discussions with physician groups, we have learned that physician groups feel it is absolutely essential to minimize the burden on SaMD users, or they will avoid making use of important software. We simply cannot afford to have users avoid helpful software because of burdensome data collection requirements.	The guidance should only contemplate passive collection of data that does not require active involvement of the end user, or at most, include one question posed to the user that could be answered on the basis of multiple-choice, not requiring a narrative response. We propose adding the following text after the sentence that ends on line 1015: "It is not necessary for such collection of real-world observational data by the SaMD manufacturer to rely on the active involvement of the end user. The SaMD manufacturer should aim to impose the least burdensome approach possible in its data collection, with at most, requiring the end user to answer one multiple choice question (without requiring any additional narrative	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
				detail)."	
9.	Page 33 / Section 8.1 / Line 1024	Technical	Section 8.1 discusses the "SaMD Definition Statement," which seems to encompass the intended use and indications for use of the SaMD, but perhaps more. While this phrase may have intentionally been selected to avoid using the terminology of any specific regulatory jurisdiction, it would be helpful if IMDRF included some additional examples to clarify its interpretation of this phrase.	Delete section 8.1 and address that topic only in the earlier definitions document.	
10.	Page 35 / Section 8.2 / Line 1113	Technical	We do not believe that IMDRF should be using the fourth document in the IMDRF SaMD guidance series to clarify the definition of SaMD. If IMDRF believes that the definition requires clarification, it should go back to the first document and provide that clarification. While it may be more expedient to simply add the clarification in the present document, it makes much more sense to have foundational concepts such as this fully described in the first document of the series (where the reader would expect to find it).	Move section 8.2 to the definitions document, and then integrate more fully these concepts into that document.	
11.	Page 37 / Section 8.3 / Lines 1170-1196	Technical	It appears that IMDRF has created nine different categorizations for SaMD, with then binary differentiations between diagnostic and non-diagnostic, and novel and not novel. In terms of combinations, that allows for 36 different possible categorizations. But then after making all of those different categorizations, most of them are grouped together into only a few levels of differentiation. In effect, after coming up with the risk stratification in the second document in the SaMD series, IMDRF is now treating most of category II the same as category	Move this discussion to the risk stratification document to more fully integrate this discussion into that document.	

Comment Number	Page / Section / Line	Editorial or Technical	Comment and Rationale	Proposed Revised Text	IMDRF Decision (& date)
			IV. That does not seem to comport with the U.S. principle of applying the least burdensome regulatory requirements based on risk. It seems to us that IMDRF published its risk stratification first – finalized that document – then set out to create this document, only to depart from some of the distinctions that IMDRF drew in the earlier document. All four documents need to be harmonized and rationalized, and the subjects need to be addressed in their respective documents.		
12.	Page 39 / Section 8.4 / Lines 1208-1242	Technical	While these examples are useful, it is difficult to discern what generalized learning we should take and apply to other specific scenarios. These examples all seem very fact-specific, and it is difficult to draw from them more generalized principles to apply elsewhere.	Identify the more generalized learning to be gleaned from these examples.	

Draft October 4, 2016

CDS Coalition Classification Proposal

The CDS Coalition would propose that this classification model be codified at a high level in a classification regulation published in the Code of Federal Regulations, as well as in more detail in an FDA-issued guidance.

1. Dividing Line between Regulated CDS and Unregulated CDS

Proposal: To be regulated, the software must meet the four following requirements.

A. Is the software a medical device and 'CDS'?

To be regulated by FDA under this framework, a particular piece of software would have to meet both the statutory definition of a medical device (*i.e.*, be intended for use in the diagnosis or treatment of diseases or other conditions in man) and the following proposed definition of CDS:

Clinical decision support software is software that:

- Uses patient-specific information and organized clinical knowledge;
- Performs some analysis (rather than simply display or transmit the information);
- Produces a particular actionable result for the diagnosis, treatment or management of a disease or condition for a particular patient (as opposed to multiple options); and
- Is standalone software, and not an accessory to a medical device.

B. Does the CDS' intended use cause enough risk to merit regulation?

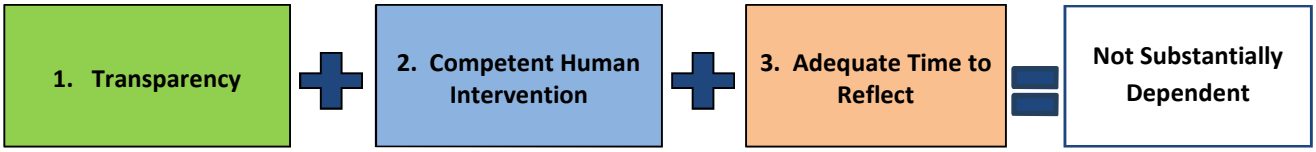
A specific CDS program should present sufficient potential risk to the patient/user to merit FDA regulation. FDA should not regulate low risk CDS. We identify low risk CDS in the table found in section II.B of this paper – the area shaded in green – based on definitions provided in section II. A.

C. For professional use CDS, is the CDS intended to provide information upon which the user would be "substantially dependent"?

By law, FDA regulation does not extend to the practice of medicine. In the case of CDS, this means distinguishing risks that come from the practice of medicine (which are unregulated by the FDA) versus risks that come from a medical device (which FDA has responsibility for regulating.) This is a jurisdictional limitation the agency must face, and it exists in part because state boards of medicine regulate the practice of medicine.

The dividing line between medical device regulation and the practice of medicine involves determining whether the medical professional is intended to be *substantially dependent on the CDS to make a diagnosis or treatment decision.* If there is no substantial dependence, FDA may not regulate. And instead, the state boards of medicine are responsible.

FDA and industry can use three criteria to determine if the CDS is intended to result in the "Substantial Dependence" of the user.



Each of these criteria would need to be met to avoid making the user substantially dependent on the CDS, as explained below.

1. **<u>Transparency</u>**. Does the software provide enough information for the user to understand and be able to evaluate the clinical basis for the software recommendation or other output? This includes disclosure of the following:

    a. What the software does and does not do. It is important in this regard that the labeling for the software communicate to the user the limits of the software's functionalities.

    b. The information inputs used by the software. This includes (i) patient specific information, and (ii) the source of the clinical information or decision rules such as practice or professional guidelines that the software uses to analyze the patient information.

    Furthermore, depending on the design of the software and the risk associated with the intended use, the developer may also need to provide:

    c. An indication of the certainty or reliability of the output, including as appropriate confidence levels and/or ranking of alternatives

    d. The clinical rationale for the recommendations. This goes beyond merely identifying the <u>source</u> of the clinical rules, and includes a reasonable explanation of the clinical logic by which the software arrived at its specific recommendation based on patient specific information.

    Most CDS is intended to aid a trained user in decision-making, but not to be a substitute for the user's expertise and judgment. On the other hand, if the CDS does not enable the intended user to sufficiently understand the recommendation made by the software and equally importantly, the basis for recommendation, it would appear that such CDS is intended to be a substitute for the user's expertise and judgment.

2. **<u>Competent Human Intervention</u>**. Is the intended user competent – through training, experience or otherwise - to make the clinical decision in question without the CDS? The education and experience required to be competent depends on the nature of the

decision. A nurse, a primary care physician, and a specialist are each competent to make different types of decisions, as are pharmacists, home health aides and other health professionals and care-givers.

CDS intended to be used to extend a user's decision-making ability beyond his/her qualifications could create substantial dependence. However, CDS that merely assists the user in applying her existing qualifications does not. CDS that, for example, collects, calculates, sorts, or otherwise gathers and presents information which the user is competent in interpreting (while easing the burden of data gathering or processing) should not, by itself, result in substantial dependence.

It is important to note that it may be prudent for decision-makers to also consider data outside of what the CDS has collected. Competent decision-makers will recognize that need and incorporate it into their decision making process.

3. **Sufficient Time to Reflect**. Based on the intended use, is the user expected to have enough time to reflect on the software output before making a decision? The amount of time available to reflect will depend on the acuity of the condition, and how much time can lapse before the patient receives medical care without risk. The amount of time needed to reflect may also depend on the complexity of the decision.

   If the intended user does not have enough time to independently consider the data inputs, the user may be substantially dependent.

From this point forward, we are going to refer to the regulated portion of CDS as Regulated CDS, or RCDS.

Background: It is important to note that the definitions of CDS and RCDS are narrower than the definition of Software as a Medical Device used by the IMDRF.

IMDRF defines the term "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] As thus defined, SaMD is a broader category than CDS, let alone RCDS. Examples of SaMD taken from the IMDRF risk stratification document[2] that would not qualify as CDS include:

- SaMD that uses the microphone of a smart device to detect interrupted breathing during sleep and sounds a tone to rouse the sleeper.

[1] IMDRF SaMD Working Group, *Software as a Medical Device (SaMD): Key Definitions* (Dec. 9, 2013), http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf.

[2] IMDRF SaMD Working Group, *"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations* (Sept. 18, 2014), http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf.

- SaMD that is intended to provide sound therapy to treat, mitigate or reduce effects of tinnitus for which minor therapeutic intervention is useful.
- SaMD that interpolates data to provide 3D reconstruction of a patient's computer tomography scan image, to aid in the placement of catheters by visualization of the interior of the bronchial tree; in lung tissue; and placement of markers into soft lung tissue to guide radiosurgery and thoracic surgery.
- SaMD that sends ECG rate, walking speed, heart rate, elapsed distance, and location for an exercise-based cardiac rehabilitation patient to a server for monitoring by a qualified professional.
- SaMD that collects output from a ventilator about a patient's carbon dioxide level and transmits the information to a central patient data repository for further consideration.

That difference is important, because it means that CDS – and even RCDS – has a lower risk profile generally than some types of SaMD. CDS only functions to inform a caregiver of recommendations based on evidence. Some SaMD goes beyond that function to operate directly on patients. In the list of examples above, the SaMD examples present an entirely different risk profile than does CDS.

D. For consumer use CDS, is the CDS intended to guide self-diagnosis and/or treatment without the oversight of a health professional?

Consumers, because they lack specialized healthcare education and experience, necessarily need to depend more on the software. For consumers, the critical risk mitigation is to work with a healthcare professional who can oversee the use of the software, and answer questions. If consumers are using the software under the watchful eye of a health professional, as above the healthcare professional is able to independently discern the correct decision. Without healthcare professional oversight, the consumer is substantially dependent on the output of the software. Exactly what constitutes oversight is left to the judgment of the healthcare professional, based on the risk. But it needs to be clear that a particular health care professional is in fact responsible for the consumer's use of the software for the risk to be mitigated.

2. US Classifications for RCDS

The purpose of this section is to come up with an overall approach to classifying RCDS into class I, class II and class III for US FDA purposes, borrowing as much as possible from the IMDRF risk stratification approach.

Proposal:

A. Assess the Significance of the Information and the Seriousness of the Condition

The international community, through the IMDRF, has concluded that the two most important drivers for appropriate risk stratification of software used as a medical device are the significance of the information provided as well as the seriousness of the healthcare condition targeted. Below we have endeavored to borrow the definitions from the IMDRF framework, with as few changes as possible.

1. **Significance of information provided by RCDS to healthcare decision**

a) To treat or to diagnose

Treating and diagnosing infers that the information provided by the RCDS will be used to take an immediate or near term action:

- To treat/prevent or mitigate a disease or condition based on a definitive therapeutic determination the RCDS provides
- To diagnose/screen/detect a disease or condition based on a definitive diagnostic determination the RCDS provides.

b) To drive clinical management

Driving clinical management infers that the information provided by the RCDS will be used to aid in treatment, aid in diagnoses, to triage or identify early signs of a disease or condition that will be used to guide next diagnostics or next treatment interventions:

- To aid in treatment by providing enhanced support to safe and effective use of drugs or a medical device.
- To aid in diagnosis by analyzing relevant information to help predict risk of a disease or condition or as an aid to making a definitive diagnosis.
- To triage or identify early signs of a disease or conditions.
- To provide output specific to the course of action rather than informing on the probabilities of outcomes.

c) To Inform clinical management

Informing clinical management infers that the information provided by the RCDS will not trigger an immediate or near term action:

- To inform of options for treating, diagnosing, preventing, or mitigating a disease or condition.
- To provide clinical information by aggregating relevant information (*e.g.*, disease, condition, drugs, medical devices, population, etc.)
- To inform on the probabilities of outcomes.

2. Healthcare Situation or Condition

a) Critical situation or condition

Situations or conditions where accurate and/or timely diagnosis or treatment action is vital to avoid death, long-term disability or other serious deterioration of health of an individual patient or to mitigate impact to public health. RCDS is considered to be used in a critical situation or condition where:

- The type of disease or condition is:
  - Life-threatening state of health, including incurable states,
  - Requires major therapeutic interventions,
  - Sometimes time critical, depending on the progression of the disease or condition that could affect the user's ability to reflect on the output information;
- Intended target population is particularly fragile with respect to the specific disease or condition (this is more than simply pediatrics, but patients who are particularly frail with regard to the specific disease or condition); and
- Intended for specialized trained users.

b) Serious situation or condition

Situations or conditions where accurate diagnosis or treatment is of vital importance to avoid unnecessary interventions (*e.g.*, biopsy) or timely interventions are important to mitigate long term irreversible consequences on an individual patient's health condition or public health.

RCDS is considered to be used in a serious situation or condition where:

- The type of disease or condition is:
    - Moderate in progression, often curable,
    - Does not require major therapeutic interventions,
    - Intervention is normally not expected to be time critical in order to avoid death, long-term disability or other serious deterioration of health, whereby providing the user an ability to detect erroneous recommendations;
- Intended target population is NOT fragile with respect to the disease or condition; and
- Intended for either specialized trained users or lay users.

Note: RCDS intended to be used by lay users in a "serious situation or condition" as described here, without the support from specialized professionals, should be considered as RCDS used in a "critical situation or condition".

c) Non-Serious situation or condition

Situations or conditions where an accurate diagnosis and treatment is important but not critical for interventions to mitigate long term irreversible consequences on an individual patient's health condition or public health. RCDS is considered to be used in a non-serious situation or condition where:

- The type of disease or condition is:
    - Slow with predictable progression of disease state (may include minor chronic illnesses or states),
    - May not be curable; can be managed effectively,
    - Requires only minor therapeutic interventions, and
    - Interventions are normally noninvasive in nature, providing the user the ability to detect erroneous recommendations;
- Intended target population is individuals who may not always be patients; and
- Intended for use by either specialized trained users or lay users.

B. Deciding the U.S. Classification Based on Those Two Factors

Using the determinations from section A above, we can place the RCDS in the proper FDA classification based on the following table adapted from IMDRF.

Determining US Medical Device Classification for RCDS

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

As a definitional matter, software that appears in the lower right-hand side of that matrix in green is not RCDS as explained more thoroughly in section I.

As a part of our conversion from the IMDRF risk stratification, we adjusted the matrix in one important respect. In the box associated with a critical healthcare condition that is used to inform clinical management, we placed that box in the unregulated category. In the original IMDRF proposal, that box as a matter of risk was the same as the box for a serious condition where the software drives clinical management, and the non-serious healthcare condition, where the software is used to treat or diagnose.

We changed the critical/inform clinical management box in the conversion for a legal reason. In our assessment, any software that merely informs clinical management is regulated by states under the practice of medicine, and not by FDA. Hence we put that entire column where the software merely informs clinical management in the unregulated space.

Background: One of the keys to understanding this approach is to understand the definition of RCDS, and by extension, to understand that even in the first column – treating and diagnosing – the software is not part of a closed-loop system where the treatment or diagnosis is automated. The definition of CDS excludes closed-loop systems. Instead, for RCDS, treatment or diagnosis simply means that the software provides a definitive treatment or diagnostic recommendation that a human still has to implement. Software nonetheless gets placed in a highly regulated category – class III – if the software provides a definitive diagnosis or treatment recommendation, is not transparent (remember, transparent software gets excluded from the definition of RCDS), requires immediate action, and the disease or condition is critical.

While this risk-based framework is based on the IMDRF model for software, and consequently we have not sought to reinvent all of the analysis that supports it, for purposes of understanding the specific regulatory controls that are appropriate, we conducted an informal risk assessment through which we compared medium risk and low risk RCDS with in vitro

diagnostic tests and in vivo diagnostic tests. We did that comparison in a separate document that is 17 pages long. We did that work to help us better articulate the similarities and differences between RCDS software, on the one hand, and, on the other hand, medical devices that come into contact either with a patient specimen or the patient himself.

That comparison leads us to the conclusion that there are, at a high level, two primary differences between RCDS and diagnostic tests (both in vitro and in vivo). Those differences are summarized in the table below.

Differences in Risk Profile

RCDS vs. Diagnostic Tests (In Vitro and In Vivo)

Risk	RCDS Software	Diagnostic tests performed with software driven hardware (in vitro and in vivo)
Safety risk is the risk that the medical device could itself harm the patient without regard to issues of effectiveness	RCDS software does not drive any hardware that comes into contact with the patient. It functions as stand-alone software on a general purpose computing platform.	An in vitro diagnostic, if based on a blood specimen, requires sampling which carries the risk of infection. In vivo diagnostic tests involve touching the body or sending energy into the body.
Hardware risk	Because RCDS only operates on general purpose computing platforms, the hardware risk is well-understood and generally well-recognized by the user.	Diagnostic tests all include hardware used for measurement purposes, and perhaps reagents, that all carry with them risk of malfunction and/or inaccuracy due to a lack of calibration.

The differences described in the in the table above are on top of the primary difference between RCDS and diagnostic tests: *the fact that steps can be taken through heightened transparency to ensure that the user is not substantially dependent on RCDS, whereas no such risk mitigation is possible with a diagnostic test*. We did not include that in this table simply because here we are focused on regulated RCDS, which by definition under our model would be CDS that creates substantial dependency.

3. FDA Regulatory Controls Specific To RCDS Classification

Background:

The task is to translate that general regulatory stratification and classification into specific US regulatory controls that are appropriate and necessary to assure the safe and effective use of RCDS. The informal risk assessment tells us that we can focus the required regulatory controls on the software risk, without needing to employ the types of controls that would be necessary for hardware used to measure clinical signals.

Institutions that have looked at this issue, including the FDASIA Workgroup[3] and earlier NIH committees,[4] have concluded that we lack sufficient data to reliably assess the risk of RCDS. Thus, a priority for regulatory policy in this area is to put systems in place that are designed to improve the collection and reporting of information on risks associated with RCDS.

At the same time, and partly in recognition of that lack of information, policy groups have been reluctant to impose regulatory barriers that might choke off future innovation. The lack of information cuts both ways: there is little information to suggest that the risks are significant and therefore need to be controlled, while at the same time, there is little information to assure us that there are few risks.

On the other side of the coin, with regard to benefits, those data are only now starting to emerge. Intuitively, the clinical value of clinical decision support software seems great, but we do not have hard evidence in most cases as to exactly how much benefit there may be. The one aspect in all of this that seems clear is the need for useful information to support clinical decision-making. The fact that doctors are inundated with new information and are presently, in many ways, unable to cope with the flow of that information seems well-established. Further, the basic human traits of fallacy are equally well-established.

In the face of that assessment, we submit that from a policy standpoint, the best strategy at this point is a set of regulatory controls designed to improve the collection of information, and in particular real-world data. More specifically, we propose regulatory controls focused on collecting information on risk, and at least some measures of effectiveness.

---

[3] FDASIA Workgroup, *FDASIA Committee Report* (2013), https://www.healthit.gov/facas/sites/faca/files/FDASIARecommendations030913_Final.pptx.

[4] See e.g., Health IT and Patient Safety: Building Safer Systems for Better Care, November 8, 2011, Institute of Medicine of the National Academies, http://www.nationalacademies.org/hmd/Reports/2011/Health-IT-and-Patient-Safety-Building-Safer-Systems-for-Better-Care.aspx#sthash.XGy0h8pJ.dpuf; and Enabling Health Care Decisionmaking Through Clinical Decision Support and Knowledge Management, Apr. 24, 2012, Agency For Healthcare Research And Quality, http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=919

Proposal:

We propose regulatory controls in two different buckets – with the common theme of reporting product problems to FDA – to serve as the primary vehicle for the oversight of RCDS.

1. **Adverse event reporting**. As with any other regulated medical device, the adverse event reporting system provides an important information-gathering tool for assessing the product risk. Having a clear and well-understood program in place for collecting adverse events associated with RCDS will give FDA not just the ability to spot individual products that may be causing harm, but also the ability to assess the category as a whole for future policymaking purposes. **To make the MDR program functional, manufacturers of RCDS would need to comply with facility registration and product listing.**
2. **Specific elements of the quality system and the associated part 806 Medical Device Correction and Removals reporting requirements related to data collection**. Here, we want to propose something special and innovative that takes advantage of the fact that the product is software. Software presents certain opportunities for data collection not found in other regulated medical devices. Software can be designed to both actively collect information on a technical level, and to actively solicit information from the user regarding their experience and the effectiveness of the software. We propose that FDA capitalize on these unique traits of software to come up with a tailored approach that sets certain standards for data collection that manufacturers would have to design their products to meet. Of course, in designing these data collection standards, FDA will need to be mindful not to overburden the ultimate user or interfere with patient care The data collected would then be fed into the manufacturer's quality system, on a trimmed down and focused basis, so that the manufacturer constantly assesses that information – together with all other types of information already collected such as complaints – for issues that could impact safety and effectiveness. Finally, to complete the process, if issues arise that do engender risk such that they lead to a class I or class II recall, the manufacturer would be obliged to report those circumstances to FDA. Obviously, the purpose of this approach is more than just feeding information to FDA: the first line purpose of this proposed approach is to feed information into the manufacturer's quality system to ensure continuous improvement and risk management. To help achieve the objective of not unduly inhibiting innovation and market access, we propose focusing the quality system on that data collection and data analysis activity, and exempting RCDS manufacturers from the other quality system requirements.

In totality, those regulatory controls would capture the spontaneously reported adverse events, but also more proactively require manufacturers to collect information on both software performance and effectiveness, that ultimately would ensure that FDA becomes aware of learned risks.

Regulatory Controls over RCDS by Classification

Regulatory Class	Associated Regulatory Controls
Class I	1. General controls, including: • Facility registration with a guidance document that explains how to do facility registration in the context of software that is developed in a collaborative, virtual world • Product listing • Adverse event reporting with a guidance document that explains how the serious injury and malfunction provisions are to be interpreted in the context of RCDS where there is substantial physician decision-making involved • Medical device corrections and removal reporting 2. Exempt from certain portions of the quality system requirement (see section 5 below), but subject to any international standards that are an embodiment of the applicable quality system requirements such as CLSI AUTO 13-A2, IEC 62304, and AAMI TIR57 3. Exempt from inspections, but instead subject to for-cause quality record document requests in connection with a class I or class II recall 4. Exempt from 510(k) premarket notification 5. Add in new low-level data collection requirement that would inform the manufacturer of how often the software fails in some regard
Class II	1. Same general controls 2. Same quality system requirements 3. Add in a heightened postmarket data collection requirement that includes feedback from the user. Include a requirement that this data be periodically summarized and shared with FDA. See section 4 below. 4. Exempt from premarket notification, with certain limits
Class III	1. General controls 2. No exemption from quality system 3. Premarket approval application 4. Same heightened postmarket data collection as required for class II

4. Postmarket Data Collection Requirements for RCDS

Proposal:

Regulatory Class	Postmarket Data Collection Standards
Class I	The goal of these postmarket data collection requirements is to monitor the technical performance of the software. We need to assemble a list of technical information that we would expect a company to design their software to collect. The requirements should be written at an objective level, as opposed to a specific, prescriptive level, to allow for evolution in software technology. For example, the requirements might include collecting data on: • The frequency with which the software fails to operate • The types of error codes • The frequency with which users contact technical support • The frequency with which the software is used repeatedly by individuals (we don't need to know their identities, but it is a positive sign if individuals use it over and over again) • The types of patients for whom the software is used to see if those patients reflect the intended use case • Analytics on the software's self-measurement of its degree of confidence • If machine learning is involved, analytics on the degree to which that learning is progressing
Class II and III	In addition to having to meet the same goals as class I products with regard to collecting data on poor technical performance, class II and class III products would have to meet goals for clinical experience data collection. To collect this data, the software would need to provide an avenue for the user to provide specific feedback on the clinical usefulness of the information provided, as well as incentives designed to encourage the provision of that feedback. As already explained, this approach would need to be tailored to ensure that it does not overburden users or interfere with patient care. Thus, the software could request information on the user's satisfaction with the advice given at the time the advice is given. There is no ground truth or gold standard typically against which the advice could be compared, but rather the only real standard is the user's assessment of the advice. That assessment necessarily needs to cover a range of characteristics such

	as whether the advice is clear, whether the advice adds value by identifying factors that the physician might otherwise have missed, whether the instructions for use were clear, whether the ultimate recommendation is, in fact, what the health care professional believes she should follow. Here we are very mindful of the physician's limited time to invest in providing feedback, as well as our desire to get maximum participation. The most efficient way of getting at that information is a more open-ended question designed to give the physician both an overall method for communicating a range of satisfaction/dissatisfaction, as well as a way to communicate any concerns the physician might have. We would propose that the software ask the physician, after giving the recommendation, "Did the software work as expected?" We picked the word "expected," in contrast, for example, to the word "intended," because the user knows what she expects based on the labeling. The user could then be presented with a range of five possibilities from completely disagree to completely agree. After that, we propose a text box where the physician can explain the rating. We anticipate that physicians will be far more likely to explain low ratings than high ratings. The use of the text box means that the software developer will need to have in place an effective way of evaluating the information that it collects. Obviously there are software programs that could be very helpful in monitoring the information collected via these text boxes to prioritize issues that need to be addressed. The quantitative answer to the question also gives an objective basis for prioritizing responses, focusing on the completely disagree end of the spectrum. Software developers will also need the human resources necessary to complete the task. We envision this question being present over the lifecycle of the software. As everyone knows, software frequently gets updated. This is true not only to improve the functioning of the software itself and the user experience, but clinical intelligence is likely to be added over time. This system provides a sentinel watch for problems that might arise as the result of updates, in addition to driving continuous improvement. The text information is extremely important, because dissatisfaction could range in cause from someone who simply feels like using the software is a waste of time because the answers are obvious, to someone who observed that the software is giving an incorrect answer because, for example, it is based on outdated clinical intelligence. In addition to building in that opportunity for feedback, the manufacturer would have to design incentives for the user to provide the feedback. We would not want to prescribe what those incentives look like; only that they be effective in meeting certain objectives. The actual incentives could range from gimmicks like bonus points that provide some value to the user, to

	more prescriptive designs like not allowing a physician user to enroll more patients until they have answered questions from prior patients. This approach would then need to specify how this data should be best summarized for FDA, and the frequency of providing such summaries to the Agency. This is not unlike existing adverse event reporting regulations and PMA reporting obligations, except that it would not focus on individual reports, but rather exclusively upon trends in the reporting. On the one hand, FDA needs useful data at the earliest possible time. On the other hand, FDA does not have the resources to review large volumes of data. So there may need to be triggers. For example, if it looks like there may be a significant problem, that fact would have to be reported as soon as possible. On the other hand, more routine experience might be reported quarterly or even annually.

5. Applicable Quality System Requirements for RCDS

Proposal:

Applicable Quality System Requirements	Exempt Quality System Requirements
Subpart A--General Provisions § 820.1 - Scope. § 820.3 - Definitions. § 820.5 - Quality system. Subpart B--Quality System Requirements § 820.20 - Management responsibility. § 820.22 - Quality audit. § 820.25 - Personnel. Subpart C--Design Controls § 820.30 - Design controls. Subpart D--Document Controls § 820.40 - Document controls. Subpart F--Identification and Traceability § 820.60 - Identification. § 820.65 - Traceability. Subpart I--Nonconforming Product	Subpart E--Purchasing Controls § 820.50 - Purchasing controls. Subpart G--Production and Process Controls § 820.70 - Production and process controls. § 820.72 - Inspection, measuring, and test equipment. § 820.75 - Process validation. Subpart H--Acceptance Activities § 820.80 - Receiving, in-process, and finished device acceptance. § 820.86 - Acceptance status. Subpart L--Handling, Storage, Distribution, and Installation § 820.140 - Handling. § 820.150 - Storage. § 820.160 - Distribution. § 820.170 - Installation.

§ 820.90 - Nonconforming product. Subpart J--Corrective and Preventive Action § 820.100 - Corrective and preventive action. Subpart K--Labeling and Packaging Control § 820.120 - Device labeling. Subpart M--Records § 820.180 - General requirements. § 820.181 - Device master record. § 820.184 - Device history record. § 820.186 - Quality system record. § 820.198 - Complaint files. Subpart N--Servicing § 820.200 - Servicing.	Subpart O--Statistical Techniques § 820.250 - Statistical techniques.
An important part left in the quality system includes those sections that lay out the requirement of risk management. Risk management is an essential element for all medical devices, including RCDS. In line with that, cybersecurity risk would be addressed both through premarket design controls, as well as postmarket risk management. Specifically, manufacturers would implement comprehensive cybersecurity risk management programs and documentation consistent with the Quality System Regulation, including but not limited to complaint handling (21 CFR § 820.198), quality audit (21 CFR § 820.22), corrective and preventive action (21 CFR § 820.100), software validation and risk analysis (21 CFR § 820.30(g)) and servicing (21 CFR § 820.200).	