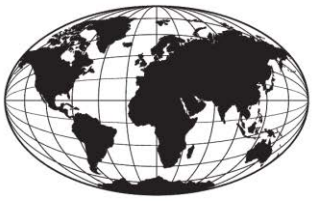


IMDRF International Medical Device Regulators Forum

PROPOSED DOCUMENT

International Medical Device Regulators Forum

Title: Software as a Medical Device (SaMD): Clinical Evaluation

Authoring Group: Software as a Medical Device Working Group

Date: 5 August 2016

Table of Contents

Preface

The document herein was produced by the International Medical Device Regulators Forum (IMDRF), a voluntary group of medical device regulators from around the world. The document has been subject to consultation throughout its development.

1.0 Introduction

The International Medical Device Regulators Forum (IMDRF) seeks to establish a common and converged understanding of clinical evaluation and principles for demonstrating the safety, effectiveness and performance of software intended for medical purposes as defined in the *IMDRF/SaMD WG/N10* document *Software as a Medical Device (SaMD): Key Definitions (SaMD N10)*.

For all medical devices, clinical evaluation, a process activity that is conducted during a product's lifecycle as part of the quality management system, is the assessment and analysis of clinical data pertaining to a medical device to verify its safety, effectiveness and performance.[1] The principles for clinical evaluation are the same for all medical devices and the expected rigor in current clinical guidance is intended to be technology agnostic.

SaMD, a type of medical device, also has significant patient and public health impact and therefore requires reasonable assurance of safety, effectiveness and performance.

This assurance for a SaMD is expected to be provided through a systematically planned clinical evaluation approach that generates adequate scientific evidence to create transparency, and to assure confidence in the SaMD's clinical validity for the intended purpose and indications for use, namely the claims, of the SaMD. This evaluation along with the evidence helps demonstrate that the SaMD is safe, that it performs as intended, and that the risks associated with the use of the SaMD are acceptable when weighed against the benefits to patients.

Global regulators expect that clinical evaluation and the evidence generated for a SaMD have the same scientific level of rigor that is commensurate with the risk and impact of the SaMD, to demonstrate assurance of safety, effectiveness and performance.

SaMD however is unique in that it operates in a complex highly connected-interactive socio-technical environment in which frequent changes and modifications can be implemented more quickly and efficiently. Development of SaMD is also heavily influenced by new entrants unfamiliar with medical device regulations and terminology developing a broad spectrum of applications.

Most SaMD's, except in limited cases, do not directly affect or have contact with a patient, instead only performs computation on data input and provides data output to a user to inform clinical management, drive clinical management, or in the diagnosis or treatment of the patient. Data input received by a SaMD typically relies on other physiological measuring medical device output or an in-vitro diagnostic device. However as healthcare decisions increasingly rely on information provided by the output of SaMD, these decisions can impact clinical outcomes and patient care.

[1] Global Harmonization Task Force (GHTF) Clinical Evaluation, Page 4, May 2007. http://www.imdrf.org/docs/ghtf/final/sg5/technical-docs/ghtf-sg5-n2r8-2007-clinical-evaluation-070501.pdf

Based on the significant impact SaMD has on clinical outcomes and patient care, a SaMD manufacturer is expected to gather, analyze, and evaluate data, and develop evidence to demonstrate the assurance of safety, effectiveness and performance of the SaMD. This evaluation should focus on how well the information provided by the SaMD meets the clinical needs within the intended healthcare situation and condition that includes consideration for the target population, characteristics of the disease or condition, and type of user. This document discusses addressing these clinical needs by demonstrating the analytical validity (the SaMD's output is accurate for a given input), and where appropriate, the scientific validity (the SaMD's output is associated to the intended clinical condition/physiological state), and clinical performance (the SaMD's output yields a clinically meaningful association to the target use of the SaMD) of the SaMD.

In addition to these general clinical evaluation expectations, this guidance considers the uniqueness of indirect contact between patients and SaMD and presents the principles of clinical evaluation with recommendations to address this uniqueness. Additionally, this document highlights the uniqueness of SaMD that can leverage the connected-interactive socio-technical environment to continuously learn from real world use information. SaMD manufacturers can use this real world information to support the assurance of safety, effectiveness and performance, in a continuous and agile clinical evidence gathering paradigm. This paradigm shifts the focus towards observed real world performance as part of post-market monitoring.

	***Clinical evaluation*** *is the assessment and analysis of* ***clinical*** *data pertaining to a medical device in order to verify the safety, effectiveness and performance of the device.* ***Clinical evaluation*** *is an ongoing process conducted during the lifecycle of a medical device.*

This document primarily references previous Global Harmonization Task Force (GHTF[2]) and IMDRF guidance documents to provide a common understanding and application of terminology, concepts and principles for performing a clinical evaluation to demonstrate the performance of a SaMD.

This application of clinical evaluation principles and concepts for a SaMD also relies on the principles and processes described in IMDRF IMDRF/SaMD WG/N23FINAL:2015 Application of Quality Management Systems (QMS) (*SaMD N23*). Specifically *SaMD N23* describes how clinical evaluation is also a process within the lifecycle activities, and the larger quality management systems framework that includes organizational support, lifecycle support processes and realization software development lifecycle processes.

[2] GHTF was a voluntary group of representatives from national medical device regulatory authorities and industry representatives. GHTF was disbanded in 2012 and its mission has been taken over by the IMDRF.

As with other medical devices, the level of documented clinical evidence expected by a regulator will depend on regulatory laws in their individual jurisdictions where the SaMD is intended to be made available. This document does not opine on the individual jurisdiction's requirement; instead this document provides guidance on the relative importance and expectations, based on the impact to health, for conducting clinical evaluation and documented evidence for the different categories of SaMD as described in IMDRF IMDRF/SaMD WG/N12FINAL:2014 (*SaMD N12)*.

This is a companion document to *SaMD N10*, *N12* and *N23* documents, further enabling convergence in vocabulary, approach, and a common thinking for regulators and industry. It should also be noted that this document does not provide guidance on the adequacy of meeting regulatory requirements or "essential principles" that are the basis of GHTF classifications. Rather this guidance provides the relative importance of required clinical performance for the different categories of SaMD as categorized in the *SaMD N12* document.

2.0 Scope

The objective of this document is to provide guidance on clinical evaluation by describing:

- Relevant clinical evaluation methods and processes which can be appropriately used for SaMD to generate clinical evidence;
- The necessary level of clinical evidence for different categories of SaMD; and
- SaMD categories where independent review is important or not important.

The principles discussed are intended to assist SaMD manufacturers and regulators. The principles are based on a common goal to provide confidence to the users of SaMD (patients, providers, consumers, clinical investigators) who rely on the output of SaMD for patient care.

The description of appropriate clinical evaluation methods and processes for SaMD, and recommendations for how much evidence (or degree of certainty of the evidence), and independent oversight is appropriate for SaMD, is not meant to replace or conflict with pre-market or post-market regulatory requirements related to the regulatory classification of SaMD in different jurisdictions. Similarly, the information is not meant to replace, or conflict with, technical or international standards.

In achieving the above objectives, this document relies upon and does not repeat the concepts and principles found in *SaMD N12* (risk categorization of SaMD), and *SaMD N23* (application of quality management for SaMD), but is a continuum to those documents, and this document should be used in conjunction with those.

The categories of SaMD are limited to the definition in *SaMD N10* and the categories of intended use described in *SaMD N12* where the information provided by SaMD is intended to inform clinical management, drive clinical management, or diagnose or treat a disease or condition in non-serious, serious or critical healthcare situations or conditions.

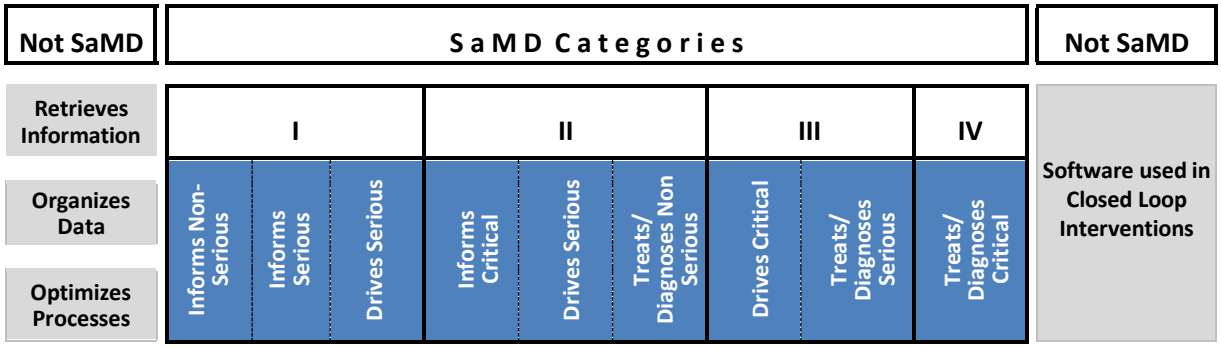


Figure 1- What is / is not a SaMD

Note: Refer to Sections 8.2and 8.3 for more information and examples related to what is a SaMD and what is not a SaMD.

This document specifically does not include in its scope or address other types of software used in health care for retrieving information from devices or systems, organizing the collected data, or optimizing healthcare workflow by automating healthcare provider's care protocols. The scope of SaMD also does not include software that is embedded in a physical medical device or software that is used to provide closed loop intervention (see Section 9.1 Clarifying SaMD Definition for more information and examples).

The guidance provided in this document specifically does not address the regulatory classification of SaMD and does not address whether a premarket clearance is required for a specific SaMD.

This guidance also does not address issues that are generic to all medical devices or specific to a country or jurisdiction such as the following:

- Off-label use or foreseeable misuse;
- Device classification of specific SaMD;
- Whether a pre-market approval or certification is required for specific SaMD.

3.0 References

IMDRF Documents:

SaMD N10	Software as a Medical Device (SaMD): Key Definitions
SaMD N12	Software as a Medical Device (SaMD): Possible Framework for Risk Categorization and Corresponding Considerations
SaMD N23	Software as a Medical Device (SaMD): Application of Quality Management System

GHTF Documents:

GHTF SG5 /N6	Clinical Evidence for IVD medical devices – Key Definitions and Concepts
GHTF SG5 /N7	Clinical Evidence for IVD medical devices - Scientific Validity Determination and Performance Evaluation
GHTF SG5 /N8	Clinical Evidence for IVD Medical Devices - Clinical Performance Studies for In Vitro Diagnostic Medical Devices

GHTF SG5 /N3	Clinical Investigations
GHTF SG5 /N2	Clinical Evaluation
GHTF SG5 /N1	Clinical Evidence – Key Definitions and Concepts
GHTF SG5 /N4	Post-Market Clinical Follow-up Studies
GHTF SG1 /N68	Essential Principles of Safety and Performance of Medical Devices

International Standards:

ISO 14155-1:2011	Clinical investigation of medical devices for human subjects -- Good clinical practice
ISO 14971:2007	Application of risk management to medical devices
IEC 80002-1:2009	Medical device software -- Part 1: Guidance on the application of ISO 14971 to medical device software

4.0 Definitions

This document does not introduce any new definitions but rather relies on the following:

- Definition of SaMD as identified in *SaMD N10*.

  Software as a Medical Device (SaMD)

  The term "Software as a Medical Device" (SaMD) is defined as software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.

  NOTES:

  - *SaMD is a medical device and includes in-vitro diagnostic (IVD) medical device.*
  - *SaMD is capable of running on general purpose (non-medical purpose) computing platforms*
  - *"without being part of" means software not necessary for a hardware medical device to achieve its intended medical purpose;*
  - *Software does not meet the definition of SaMD if its intended purpose is to drive a hardware medical device.*
  - *SaMD may be used in combination (e.g., as a module) with other products including medical devices;*
  - *SaMD may be interfaced with other medical devices, including hardware medical devices and other SaMD software, as well as general purpose software*
  - *Mobile apps that meet the definition above are considered SaMD.*

- Definition of Clinical Evaluation and associated terms and vocabulary as identified by the Global Harmonization Task Force (GHTF) and interpreted for a SaMD not included in Section 4.0 Definitions below can be found in Appendix A of this document.

4.1 Clinical Validity of a SaMD

For purposes of this guidance, the term clinical validity is used to refer to the combination of:

a) The association of the output of a SaMD to a clinical condition/physiological state (scientific validity); together with
b) The ability of a SaMD to yield a clinically meaningful output associated to the target use of SaMD output in the health care situation or condition identified in the SaMD definition statement (clinical performance).

Depending on the type of SaMD, clinical validity can be expressed as follows:

- For SaMD that is intended to treat a disease or condition, clinical validity is the evidence of effectiveness of the SaMD output to the treatment or prevention.
- For non-diagnostic SaMD, clinical validity is the evidence of scientific validity that shows the usefulness of the SaMD output in clinical care.
- For diagnostic SaMD, clinical validity is the evidence of scientific validity in addition to the clinical performance evidence of the SaMD.

4.2 Scientific Validity of a SaMD

Scientific validity is the association of the SaMD output to a clinical condition/physiological state.

Scientific validity is often identified from academic research, and is often supported by studies evaluating the inputs along with the algorithms for an association of the SaMD's output to a clinical condition/physiological state. *Example: Hemoglobin concentration is associated with anemia (clinical condition). Congestive heart failure, Hypertension, Age, Diabetes, prior Stroke (CHADS-2) score is associated with predicting the risk of stroke in patients with non-valvular atrial fibrillation.*

Scientific validity establishes how well the output of the SaMD accurately correlates to the intended clinical health care situation or condition of the intended use of the SaMD. The evidence demonstrates objectively the clinical association of the SaMD's use of inputs, algorithm and outputs as compared to a recognized reference standard (i.e., gold standard), to another SaMD or medical device, to a well-documented method, to the current clinical practice or standard of care, or as compared to a composite reference standard. When comparing to other devices, including other SaMD's, the original reference standard used by the other device to determine the scientific validity of the intended clinical condition is typically used rather than the device itself.

Scientific validity also determines if the association of the SaMD's intended use to a clinical condition/physiological state is well-known (i.e., known clinically acceptable analytical validity standards, and where the analytical validity assessment has determined that the SaMD meets those standards), based on available review of information such as peer reviewed literature, textbooks, historical data and experience based evidence, academic research, or is supported by previous studies.

At the conclusion of scientific validity appraisal, a SaMD can generally be segregated in one of the following categories:

a) **Well-known association:** These SaMD's have output with a well-known association to identified clinical guidelines, clinical studies in peer reviewed journals, consensus for the use of the SaMD, international reference materials or other similar sources. *Example:*

Computation of the Acute Physiology and Chronic Health Evaluation II (APACHE II) score is a well-known association to stroke risk.

b) **Novel association:** These SaMD's involve, new inputs, algorithms or outputs, new intended target population, or a new intended use, and they are not well-known. *Example(s): use of non-standard input such as gait, blood pressure or other physiological and environmental signals using novel algorithms to detect early onset of a deterioration of health or diagnosis of a disease. .*

4.3 Clinical Performance of a SaMD

The clinical performance of a SaMD is the ability of a SaMD to yield a clinically meaningful output associated to the target use of SaMD output in the health care situation or condition identified in the SaMD definition statement (disease type, target user, and intended population). Clinically meaningful means the positive impact of a SaMD on the health of an individual, to be specified as meaningful, measurable, patient-relevant clinical outcome(s), including outcome(s) related to diagnosis or a positive impact on patient management of public health.

Clinical performance is evaluated and determined by the manufacturer during the development of a SaMD before it is distributed for use (pre-market) or after distribution while the SaMD is in use (post-market).

Clinical performance of a SaMD can also be viewed as the relationship between the verification and validation results of the SaMD algorithm and the clinical conditions of patients. This performance can also be determined using real world data, where the data is useful in identifying less common use situations.

The clinical performance of a SaMD may be characterized by demonstrating:

- Sensitivity - ability of the SaMD to correctly identify across a range of available measurements patients with the intended clinical disease or condition;
- Specificity - ability of a SaMD to correctly identify across a range of available measurements patients that do not have the intended disease or condition;
- ROC curve - a graphical plot that shows the tradeoff between sensitivity and specificity as the decision threshold that separates SaMD's negatives and positives is varied;
- Positive predictive value – which indicates the likelihood of the patient having a disease or condition given that the SaMD's output is positive;
- Negative predictive value – which indicates the likelihood of the patient NOT having a disease or condition given that the SaMD's output is negative;
- Likelihood ratio - the likelihood that a given result would be expected in a patient with the target condition compared to the likelihood that the same result would be expected in an individual without that condition; and
- Cut-off thresholds, indices or scales – should be meaningful for the intended use of the SaMD and established prior to validation.

NOTE: The sensitivity and specificity depend on the choice of a cut-off value (e.g., to separate negative from positive values).

NOTE: Predictive value depends on the prevalence of the disease or condition in the population of interest.

4.4 Analytical Validity of a SaMD

The analytical validity of a SaMD is the ability of a SaMD to accurately and reliably generate the intended output, from the input data, i.e., analytical validity measures the SaMD's ability to correctly and reliably process input data and generate output data with accuracy, and repeatability and reproducibility, i.e., precision. Analytical validity may also include measures for analytical sensitivity (e.g., limit of detection), and linearity or behavior of output across the range of input data that is allowed by the SaMD.

Analytical validity is generally evaluated and determined by the manufacturer during the verification and validation phase of the software development lifecycle using a QMS. Analytical validity is always expected for a SaMD.

Analytical validity confirms and provides objective evidence that (a) the software meets its specification, in other words, "is the software being built right?", and (b) software specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled, in other words, "is the right software being built?"

The analytical validity of a SaMD will include measures to demonstrate the following:

- Accuracy - degree of closeness of measurements of a quantity to that quantity's true value. When the output of the SaMD and true value are binary, accuracy is the proportion of true results (both true positives and true negatives) among the total number of output values examined;
- Precision - related to reproducibility and repeatability, is the degree to which repeated measurements under unchanged conditions show the same results;
- Limit of detection - ability of the SaMD to discern between information-bearing patterns of a clinical condition and random patterns that distract from the information;
- Linearity or associated transfer function - the behavior of the output across the range of input data that is allowed by the SaMD; and
- Analytical sensitivity - degree to which the algorithm's output is affected by the input data (e.g., parameters affecting input data may include perturbation, image resolution, illuminations, data spatial distribution, data amount, etc.).

5.0 General Principles and Context of SaMD Clinical Evaluation

At the highest and simplest level of abstraction a SaMD can be described as a software that utilizes an algorithm (logic, set of rules, or a model) that operates on data input (digitized content) to produce an output that is information intended for medical purposes as defined by the SaMD manufacturer as represented in Figure 2 below.

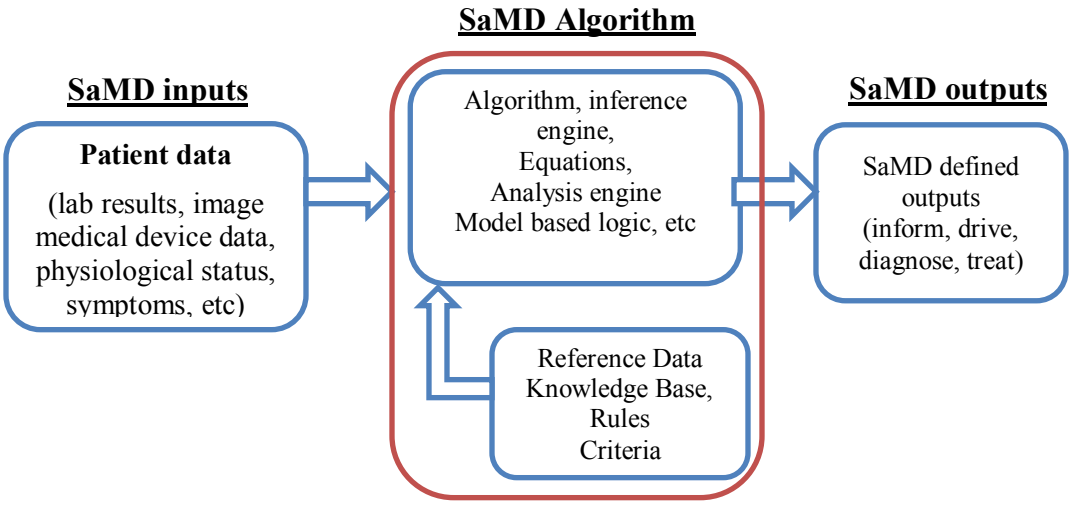


Figure 2: High Level SaMD Components

The risks and benefits posed by a SaMD are largely related to the risk of the output of the SaMD if not accurate (or correct) which in turn impacts the clinical management of a patient; rather than the risk from direct contact between the SaMD and the patient. As covered in SaMD Risk Framework (() many aspects affect the importance of the output information from SaMD. Generally these aspects can be grouped into the following two major factors that provide adequate description of the intended use of SaMD:

A. Significance of the information provided by the SaMD to the healthcare decision, and
B. State of the healthcare situation or condition.

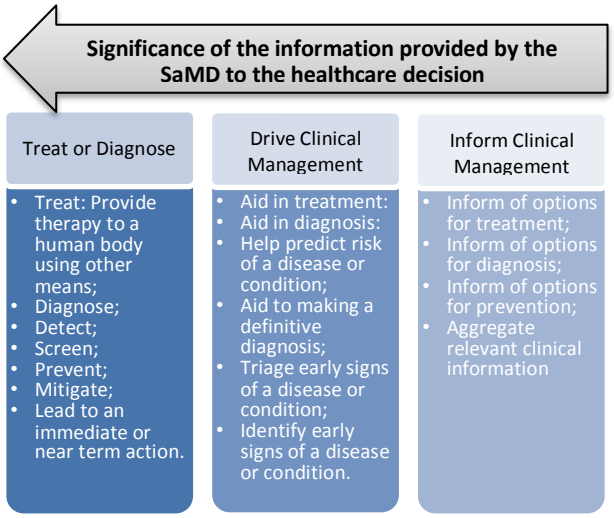


Figure 3 – *SaMD N12* components of "significance" of SaMD output (See Section 8.1 of this document)

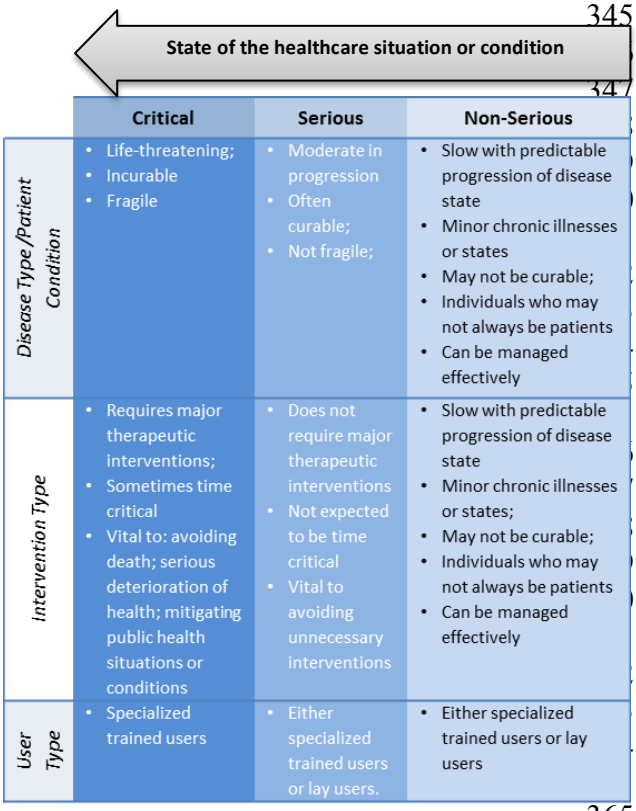


Figure 4- *SaMD N12* Components of healthcare situation or condition (see Section 8.1 of this document)

When these factors are included in the manufacturer's description of intended use, they can be used to categorize SaMD. *SaMD N12 Section 6.0* provides a structured approach for a SaMD definition statement to describe the intended use. *SaMD N12 Section 7.0* provides a method for categorizing SaMD based on the major factors identified in the definition statement. (See section 8.3 for the SaMD categorization)

In limited cases -- where SaMD may have the functionality to accept user inputs or to "treat" using general purpose computer peripherals to impart sound, light, pictures on a display or in some cases low energy vibrations -- such SaMD can be considered to provide therapy to patients (e.g., SaMD used for cognitive behavioral therapy).

These categories include functionality that has an increasing significance of the output to the patient care.

Illustrative examples of SaMD along this spectrum include:

- A SaMD that performs analysis of cerebrospinal fluid spectroscopy data to diagnose tuberculosis meningitis or viral meningitis in children. Such SaMD is used to diagnose a disease in a fragile population with possible broader public health impact that may be life threatening, may require major therapeutic intervention, and may be time sensitive *(SaMD N12 Category IV.i)*.
- SaMD that is intended as a radiation treatment planning system as an aid in treatment in a critical condition that may be life threatening and requires major therapeutic intervention *(SaMD N12 Category III.ii)*.
- SaMD that uses data from individuals for predicting risk score for developing stroke or heart disease for creating prevention or interventional strategies *(SaMD N12 Category II.iii)*.
- SaMD that analyzes images, movement of the eye or other information to guide next diagnostic action of astigmatism. Such SaMD provides aggregation of data to provide clinical information that will not trigger an immediate or near term action for the treatment of a patient condition that even if not curable can be managed effectively and whose interventions are normally noninvasive in nature *(SaMD N12 Category I.i)*.

Other aspects that affect the safety, effectiveness and performance of a SaMD include considerations for:

- Socio-technical environment consideration (*SaMD N12* Section 9.1) when identifying effects/implications and appropriate measures for safety, effectiveness and performance of SaMD throughout the product's design, development and installation including:
  - Usability of the application - How integrating SaMD within real-world clinical workflows.
  - Transparency of the inputs, outputs and methods to the user.
- Technology and system environment consideration (*SaMD N12* Section 9.2).
- Information security with respect to safety consideration (*SaMD N12* Section 9.3).

These other aspects influence the identification of considerations that are unique to a specific approach/method used by the manufacturer of a particular category of SaMD. For example, the type of a platform, that is constantly changing, used in the implementation of SaMD may create considerations that are unique to that implementation. These considerations can also vary by the capabilities of the manufacturer or by the process rigor used to implement the SaMD. This rigor as outlined in N23 expects that all manufacturers of SaMD follow adequate QMS that include risk management processes to manage technological, use environment and clinical risks.

- The governance structure (*SaMD N23* Section 6.0) should provide support for creating and establishing appropriate processes that are important for maintaining the quality objectives and policies[3];
- The elements of SaMD lifecycle support processes (*SaMD N23* Section 7.0) that are common processes and activities that should be considered throughout the SaMD lifecycle regardless of specific software product development approach or method used by the organization. These processes -- product planning; risk management: a patient safety focused approach; document and record control; configuration management and control; measurement, analysis and improvement of processes and product; managing outsourced processes and products – that should be applied throughout the SaMD realization and use processes; and
- Aspects of realization and use processes (*SaMD N23* Section 8.0) commonly found in software engineering lifecycle approaches (process, activities, tasks, etc.) that are important for an effective SaMD QMS include: requirements management, design, development, verification and validation, deployment, maintenance, decommissioning (retirement or end-of-life activity).

QMS rigor when applied correctly is expected to have adequate rigor in generating evidence towards:

- Managing uniqueness of short development cycle for SaMD development and changes (*SaMD N23* Section 8.6).
- Control over distribution channels (*SaMD N23* Section 8.5).

[3] These processes, policies and objectives should be tailored for the needs, type, size and nature of an organization.

- Controlling design/specification changes, versioning, monitoring installed base, managing recalls, remote updates (*SaMD N23* Section 8.5).
- Quality – Usability (including user interface), conformance to specifications, "fitness for use", and reasonably free from the possible serious effects of defects with a plan in place to detect and correct the defects to ensure the SaMD continues to meet the intended safety, effectiveness and performance.
- Objectively verified and validated to show conformance to customer requirements.
- Managed quality while in use through timely maintenance and continuous improvement.

5.1 Clinical Evaluation Principles

Like other high-quality products, a SaMD manufacturer implements on-going lifecycle processes to thoroughly evaluate the product's performance in its intended market. Prior to product launch, the manufacturer continues to collect evidence of the product's accuracy, specificity, reliability, limitations, and scope of use in the intended use environment with the intended user. Once the product is on the market the manufacturer continues to gather evidence to further understand the customer's needs in a real world environment and to ensure the product is meeting those needs. This real world information allows the manufacturer to identify and correct any problems and to enhance the product by expanding functionality to stay competitive or meet user demands.

Lifecycle activities, including clinical evaluation, should follow appropriate planning processes as part of an organization's lifecycle activities and processes. This means clinical evaluation, similar to other SaMD lifecycle activity and process, also needs to be planned prior to conducting the evaluation. Risk assessment done as part of the SaMD's lifecycle activities and processes should also be considered when conducting clinical evaluation. Risk, including the impact of hazards and hazardous situations identified while conducting clinical evaluation should be incorporated into the overall risk management processes of SaMD. The following are examples of considerations for risk management that may impact clinical evaluation:

- Level of clinical evidence available and the confidence of the evidence;
- Complexity of the clinical model used to derive the output information;
- Known specificity of the output information;
- Maturity of clinical basis of the software and confidence in the output;
- Benefit of the output information vs. current standard of care;
- Feasibility (*SaMD N23* Section 7.1);
- User and patient needs intended use (*SaMD N23* Section 8.3); and
- Clinical evidence that product meets clinical end user expectations (*SaMD N23 Section* 8.4).

SaMD clinical evaluation includes the gathering and assessment of scientific validity, analytical validity and clinical (real-world, obtained from patients) performance of a SaMD. A combination of the results of these activities generates clinical evaluation evidence for a SaMD.

The extent of clinical evaluation evidence necessary for a SaMD will depend on parameters including but not necessarily limited to the underlying algorithm, the transparency of the algorithm along with the ability for a user to detect erroneous output, the degree of variability of

the subject population and disease state (intended use target population), and the intended user(s) of the SaMD. Clinical evaluation of SaMD is expected to be iterative and continuous.

While not intended to impose unnecessary burden, clinical evidence should support the intended use of the SaMD as stated by the manufacturer while addressing the relative risks to the patient associated with the use of the SaMD. The intended use for a SaMD defines the medical purpose and determines the type and depth of the clinical evaluation. This statement of intention is the most important starting point for considering the level of evidence necessary and in the choices made to perform appropriate clinical evaluation.

For purposes of this document, performing clinical evaluation and generating data for SaMD assumes the following prerequisites:

- Clinical evaluation scope is dependent on "intended use" as defined by the manufacturer of SaMD.
    - The intended use of the SaMD is dependent on the product claims. The product claims, along with the SaMD definition statement determines the level of clinical evidence needed. Performance, functionality, and features as defined by the manufacturer are expected to be consistent with the claims.
    - While the SaMD is on the market, claims should reflect the actual performance and functionality of the SaMD (real world performance.)

6.0 SaMD Clinical Evaluation Methods, Evidence and Appraisal

Clinical evaluation is a systematic and planned process to continuously generate, collect, analyze, and assess the clinical data pertaining to a SaMD in order to verify the scientific validity, and the analytical validity and clinical performance of the SaMD when used as intended by the manufacturer. The level and extent of clinical evaluation necessary is determined by the role of the SaMD for the target clinical condition. The quality and breadth of the clinical evaluation assures that the output of the SaMD is scientifically valid and can be used reliably and predictably.

While a prospective (e.g., randomized controlled) trial may satisfy the requirements for real-world performance, prospective trials may not be required to generate patient data. The term 'clinical evaluation' should not be understood to be limited to conducting a prospective randomized clinical trial.

This section explains the goal of clinical evaluation in generating evidence, what techniques are available for a SaMD manufacturer to generate that evidence and when such evaluation is conducted in the product lifecycle.

6.1 What are the Evidence Goals of Clinical Evaluation?

The outcome of the clinical evaluation process of a SaMD is essential to the SaMD's value for the user and ultimately patients. The clinical evaluation evidence of a SaMD, as expressed in the intended use by the manufacturer, is generated from and validated by performing clinical evaluation and demonstrating the following:

- Scientific validity – showing with evidence on the association of the SaMD output to a clinical condition/physiological state;
- Analytical validity – showing with evidence the technical performance related to accuracy, reliability, repeatability and reproducibility; and if necessary
- Clinical performance – typically for diagnostic SaMD (see box below), showing evidence of the ability of a SaMD to yield a clinically meaningful output associated to the target use of SaMD output in the health care situation or condition.

Analytical validity addresses how well the device measures what it claims to measure whereas clinical performance addresses how useful that measurement is.

For most SaMD the goal of clinical evaluation is to establish clinical validity and to create evidence with evaluation methods that use of patient data to understand the analytical validity and clinical performance. In most cases since SaMD's output has an influence on a user's decision, clinical evaluations are typically focused towards the user's ability to use the output as intended by the manufacturer. In certain instances when SaMD is intended to treat a healthcare situation or condition, clinical evaluation is conducted using patients or data that is representative or related to the patient's situation or condition to demonstrate effectiveness of the treatment. For example, a SaMD that is intended to provide sound therapy to treat, mitigate or reduce effects of tinnitus for which minor therapeutic intervention is useful would require that the manufacturer provide analytical validity that assures that the treatment output is in accordance with all appropriate performance specifications and limitations. The manufacturer would also

demonstrate that there is a well-known scientific validity that associates specified sounds with an intended treatment.

Generally, SaMD that is not intended for treating a situation or condition can be grouped as follows:

- **Diagnostic SaMD**: These SaMD typically differentiate patients or their physiological conditions and are intended to drive clinical management and / or diagnose. Such SaMD are typically intended to identify early signs, triage, predict risk, screen, detect or diagnose a healthcare situation or condition.
- **Non-diagnostic SaMD**: These SaMD have generic functionality that can be used across various health care situations or conditions. Such SaMD typically provide data to help aid in diagnosis, aid in treatment, inform of options. Examples of such SaMD include calculators (radiation treatment planning SaMD), search and match, filter, user defined rules based matching, processing a signal (e.g., spectral analysis of a sound signal), a memory test that gives a score but no interpretation, etc.

6.2 Determining the Required Level of Clinical Evaluation

Clinical evaluation is an ongoing process throughout the lifecycle of a SaMD. It is based on data collected during the pre- or post-market of the product lifecycle for the SaMD intended use.

During the development phase of the SaMD lifecycle, clinical evaluation allows the manufacturer to objectively assess and demonstrate that the SaMD achieves its intended purpose during normal conditions of use and the known and foreseeable risks associated with the SaMD are minimized. The residual risks are acceptable when weighed against the benefits of the SaMD based on its intended use, and that any safety, effectiveness and performance claims made about the SaMD are supported by suitable evidence. Clinical evaluation also provides opportunities to assess the SaMD design characteristics, algorithm, and technological features to optimize its clinical effectiveness while minimizing any potential risks.

Information related to clinical evidence should be monitored routinely by the manufacturer and user once the SaMD is available on the market. The 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 reports, scientific publications, etc.) as part of their QMS to ensure the SaMD continues to meet the intended safety, effectiveness and performance.

The following provide an overview of steps for generating clinical evidence.

1. Determine if scientific validity of the SaMD is already well-known with clinically accepted analytical validity standards, and where the analytical validity assessment has determined that the SaMD meets those standards:
   - If Yes: document evidence as outlined in Section 7.2;
   - If No: generate scientific validity evidence as outlined in Section 6.3
2. Perform analytical validity: As part of SaMD verification and validation activities generate analytical evidence as highlighted in Section 6.40.
3. Establish the need for clinical performance:

- For Diagnostic SaMD and has a higher risk profile (refer to SaMD expectations in section 7.2 to determine the need) – conduct clinical performance evaluation as outlined in Section7.2.

4. If clinical performance evidence is necessary, and scientific validity is not well-known the following questions should be considered when planning the clinical performance evaluation for a SaMD:
   - Is patient data available to conduct performance evaluation or is new patient data required to support the intended claim?
   - If new patient data is necessary to support the claim, what type of clinical performance evidence is necessary to pursue?
   - Refer to Section 6.3.2 below for approaches and considerations.

6.3 Generating Scientific Validity Evidence for a SaMD

Generating scientific validity evidence for a SaMD is not necessary where association of a SaMD's output to a clinical condition/physiological state is already well-known, based on available information. An example of a well-known association is the Congestive heart failure, Hypertension (CHADS-2) score used for risk stratification of ischemic stroke in patients with non-valvular atrial fibrillation.

Scientific validity evidence should be derived from a critical appraisal of its merits and limitations and appraised to determine each piece of information on its relevance and quality for establishing the association between the output and algorithm of the SaMD and the clinical condition/physiological state.[4] Scientific validity evidence for a SaMD can be generated from following methods:

- Conducting literature search:
  - Review of information found in peer reviewed articles, regulatory guidance documents, conference proceedings, case reports, etc.; Literature sources used to identify data may include: Scientific databases; specialized databases; systematic review databases; clinical trial registers; or reference tests.
  - Review of expert opinions: this information might be found in sources that include textbooks, clinical guidance documents, and position statements from academic and professional organizations;
  - Results from proof of concept studies: these studies are usually smaller scale scientific studies to identify the fundamental association of the algorithm with the clinical condition/physiological state;
  - Results from previously conducted clinical studies that provide association of a signal or output of an algorithm with a healthcare situation or condition or a physiological state.

[4] See GHTF SG5 /N7:2012 Scientific Validity Determination and Performance Evaluation Section 6.0 and 7.2.2, 7.2.3 Scientific Validity Determination for additional details related to potential sources for the identification of scientific validity information and the appraisal and analysis of scientific validity information

- Identifying scientific validity from manufacturers experience data[5]
    - Customer feedback including complaints, adverse events, and other data that can be systematically and scientifically provides an association of the intended SaMD output with a healthcare situation or condition or a physiological state
    - Real world data generated outside of clinical performance studies provides real world experience obtained in larger, heterogeneous and more complex intended use scenarios. The data are most useful in identifying less common but potentially serious SaMD related adverse events. The source of this additional data may include:
        - Manufacturer generated post-market surveillance data (e.g., customer testing results);
        - Complaint handling databases; and
        - Details of clinically relevant software modifications (e.g., recalls, customer notifications, hazard alerts).
- Conducting a scientific validity study
    - These methods of establishing an association is a planned, designed and purposefully conducted when a SaMD manufacturer is establishing an association of the intended SaMD output with a healthcare situation or condition or a physiological state. These studies commonly include prospective studies, observational studies, retrospective and longitudinal studies that establish the clinical association. (See section 6.3.1 for further considerations)

Note: Some low risk SaMD's are developed when the scientific validity of the output and the algorithm is still emerging. An example would be a software application that manages heart failure with medication compliance, diet and activity education, and that is subsequently shown to reduce hospitalization in those that use it fully. As the scientific and medical knowledge further develops, the initially established scientific validity might change and/or expand.

6.3.1 Considerations for Literature search to support scientific validity

Literature searches may be useful in circumstances in which the scientific validity of the SaMD is not initially apparent to the manufacturer.

- The data generated through literature searching should relate directly to 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).
- When considering the relevance of data from literature searches, the SaMD manufacturer needs to consider the quality of the literature source and assess the differences between the published clinical studies and the intended SaMD use (e.g., device inputs, intended

[5] See GHTF SG5 /N7 Scientific Validity Determination and Performance Evaluation Section 7.2.3 Experience Gained by Routine Testing and GHTF SG5 /N2 Clinical Evaluation Section 6.2 Data Generated Through Clinical Experience for additional details related to these sources of data

user, patient population, intended use). Specifically such considerations should take into account the:

- Severity, disease prevalence, and natural history of the healthcare situation or condition being diagnosed or treated;
- Intended target population;
- Intended users; and
- Availability of alternative diagnostic tests and current standard of care.

- The scientific validity evidence cited in literature can provide the manufacturer in establishing acceptable clinical performance for a SaMD.

See GHTF SG5 /N7 Scientific Validity Determination and Performance Evaluation Section 7.2.2 Literature and GHTF SG5 /N2 Clinical Evaluation Section 6.1 Data Generated Through Literature Searching for additional details related to literature searches.

6.3.2 Considerations for Scientific Validity Studies

This section applies to scientific validity studies carried out by or on behalf of a manufacturer specifically for the purposes of conformity assessment in accordance with applicable regulations. Such studies are generally expected to be designed, conducted and reported in accordance and in compliance with local regulations and guidance.

Scientific validity studies are studies carried out by or on behalf of a manufacturer specifically for the purpose of demonstrating the safety, effectiveness and performance of the SaMD. SaMD with little or no relevant literature or clinical experience may require observational studies to validate the SaMD algorithm and demonstrate applicability to the target patient population. Observational studies are studies in which test results obtained during the study are not used for patient management and do not impact treatment decisions. The design of studies needs to be created to minimize bias and confounding and be risk-based. The design types for these studies include:

- Cross-sectional studies where correlation of test results to the clinical condition are established at a single point of time. In some cases, testing is performed at the initial time point, but patients are evaluated at later time points (e.g., the SaMD is used to evaluate the likelihood of future states, or there exists no applicable method to establish the clinical state at the time of testing);
- Longitudinal studies involve multiple patient measurements with the same SaMD over time to validate the clinical performance of the SaMD;
- Retrospective studies where the condition of the patient and the clinical association of the output of the SaMD is known;
- Retrospective multi-clinician multi-case studies where multiple clinicians evaluate each case, which allows clinician variability to be taken into account;
- Prospective studies where the SaMD is tested during the study. In the case the SaMD is used for the determination of a patient's future state, the study will often be based on a prospective design; and
- Prospective-retrospective studies where the clinical status is known but the clinical association of the output of the SaMD is established during the study. As a prospective-

retrospective study will use test data that was previously generated, the manufacturer should ensure that the data is segregated to ensure there is no confounding or bias by other test results.

See GHTF SG5 /N7 Scientific Validity Determination and Performance Evaluation Section 7.2.1 Clinical Performance Studies, GHTF SG5 /N8 Clinical Performance Studies for In Vitro Diagnostic Medical Devices for additional details related to these studies.

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 not considered a clinical performance study.

6.4 Generating Analytical Validity Evidence for a SaMD

Analytical validity evidence of a SaMD is generated during the verification and validation activities in a manufacturer's quality management system process and is always expected for a SaMD.

For more details refer to *SaMD N23*.

Verification and validation activities to determine analytical validity for accuracy of the SaMD should consider one or more of the following:

- Algorithms described in a recognized standard (e.g., any well-known clinical assessment, method, procedure, intervention or measurement of known validity and reliability which is generally taken to be the best available, against which new tests or results and protocols are compared) that exists in literature or current standard of care (e.g., insulin dosing for a given blood glucose level);
- Comparison with a reference standard (e.g., reference standard for the detection of focal lung disease in computer aided diagnosis);
- Comparison with reference material (e.g., Coumadin[6] dosing for a given International Normalized Ratio (INR)); and
- Comparison to another device or SaMD that have similar association of the output to the clinical condition.

The use of reference databases in verification and validation activities to show analytical validity should be qualified. In addition training data sets used during the development of the SaMD algorithm should be kept separate and independent from the data set used to generate analytical validity.

Where the above described methods are not readily available, it may be possible to perform a comparison with an already available SaMD or a comparison to a recognized method.

Where there are no comparative approaches that can be used, then different approaches can be used such as comparison to a well-documented method, or comparison to a composite reference

[6] Coumadin is an anticoagulant normally used in the prevention of thrombosis and thromboembolism.

method. If using a composite reference, then assurances must be provided that the reference remains accurate if the parts of the composite are readjusted.

6.5 Generating Clinical Performance Evidence for a SaMD

In addition to analytical validity evidence, clinical performance evidence should be generated using process and activities that are planned, designed, conducted, analyzed and evaluated so that the best possible representation is achieved with the target population in accordance with the intended use. Optimal design, execution and analysis of such evaluation will ensure the greatest possible generalization of results (e.g., for different demographic or ethnic groups, multiple sites in different health care and geographical settings).

In most circumstances, clinical performance for SaMD can be generated using real or simulated data sets (e.g., automated segmentation of retinal vessels is a generally well understood problem, aided by the public availability of the annotated STARE (Structured Analysis of the Retina) and DRIVE (Digital Retinal Images for Vessel Extraction) datasets with hundreds of papers published[7]) that reflect real patient conditions. The SaMD manufacturer is responsible for identifying relevant data and determining the types and amount of data needed to establish clinical performance, and considering the advantages and limitations of each data type. Data relevant to the clinical performance of a SaMD may be held by the manufacturer (e.g., studies sponsored by the SaMD manufacturer) or in scientific literature (e.g., published articles of clinical performance studies related to the use of SaMD algorithms for intended clinical conditions.)

Before proceeding to validate the clinical performance of the SaMD in question, the manufacturer should consider:

1. Is there published clinical performance data that is not in possession of the manufacturer that may assist the manufacturer in establishing acceptable clinical performance of the SaMD?
2. Are there types of performance data available that are generated in real world use conditions that are outside the conduct of clinical performance studies?
    - The value of such data is that it provides real world experience obtained in larger, heterogeneous and more complex SaMD use scenarios. This type of data is also most useful for identifying less common but potentially serious device-related adverse events. It is also a particularly useful source for low risk SaMD that are based on long standing, well-characterized inputs, algorithms and outputs.
3. Are there existing SaMD or devices that have shown clinical performance for a similar association of the SaMD output to the clinical condition?
    - The manufacturer should determine clinical performance on both the reference device/software and the SaMD against a source of truth (i.e., gold standard) used by the original device. For example, if you were developing a software tool for

[7] Abràmoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal Imaging and Image Analysis. IEEE Transactions on Medical Imaging, 3, 169–208. http://doi.org/10.1109/RBME.2010.2084567

identifying a heart murmur based on an electronic stethoscope input, there may not be a way to evaluate the clinical performance of that tool only against an existing murmur detection software package. Rather, you would test both the old and new software tools against echocardiography as the reference method.

When selecting information for clinical performance it should be evaluated to determine its relevance and quality to address questions about the SaMD, and its contribution to demonstrate the clinical performance of the SaMD (including any specific claims about performance).

- To be relevant the information source should be specific to the SaMD in question and reflect its intended use;
- The information provided should be of sufficient quality to enable a rational and objective assessment of the clinical performance of the SaMD;
- The different data sets should be reviewed for consistency of results across multiple studies and as appropriate, the intended target populations of the SaMD;
- If the different data sets report comparable performance characteristics, certainty about the clinical performance increases. If different results are observed across the data sets, it will be helpful to determine the reason for such differences. Regardless, all data sets relevant to the SaMD should be included;
- Any risks associated with the use of the SaMD are acceptable when weighed against the benefits to the patient.

For novel SaMD that have no known scientific validity it may be important to generate clinical performance evidence by conducting a clinical performance study (see section **Error! Reference source not found.** for details.) Clinical performance studies do not necessarily imply "prospective randomized controlled trials". Rather, depending on the risk profile of the SaMD, data (see Section 7.2) may be collected by conducting an "observational study" which is usually performed in parallel with the use of an existing SaMD, routine diagnostic testing performed for patient management care, passively collecting data while using medical devices, or in general patient care. However, for SaMD intended to diagnose or treat a healthcare situation or condition where there is a high patient risk (see Section 8.1) for inaccurate results, the study should manage risks associated and remove any bias or other confounding assumptions.

The following sections highlight aspects of current GHTF guidance that can be applied by taking into consideration the unique aspects of SaMD. Readers are encouraged to rely on principles and expectations in the GHTF guidance.

6.6 Appraisal of Clinical Evaluation Evidence

The SaMD manufacturer and the user(s) of SaMD should be able to reach the following conclusions through clinical evaluation: the SaMD is appropriate for its intended use; the SaMD achieves the expected performance for its intended use; the safety[8], effectiveness and

[8] For more information on the concept of safety refer to GHTF SG1 /N68 Essential Principles of Safety and Performance of Medical Devices Section 5, and 6 and 7 as appropriate for software. It should be noted that the assessment of the safety of a SaMD may require more than an assessment of the clinical evaluation of the SaMD.

performance of the SaMD are supported by sufficient evidence; and the SaMD risks[9] are acceptable balanced with expected benefits[10]. This appraisal should consider:

- Matching SaMD intended use to the clinical evaluation evidence; and the
- Benefits and risk of the SaMD; which includes:
    - Objective consideration of patient preference in the use of the SaMD; and
    - Benefits as compared to current standard of care for the disease or condition.

The purpose of the appraisal of the evidence is to select information based on its merits and limitations to demonstrate that the clinical evaluation evidence matches the SaMD's intended use and related claims.

Each piece of information should be appraised to determine its relevance and quality. To be relevant, the information should show a clear link between the output of the SaMD to its intended use as stated in the SaMD definition statement, namely its relationship to the healthcare decision and healthcare situation or condition intended by the SaMD. The information provided should be of sufficient quality to enable a rational and objective assessment of the certainty with which the clinical evaluation evidence matches the intended use of the SaMD. It is expected that SaMD manufacturers (and third parties as appropriate) appraise the evidence generated by the clinical evaluation.

Specifically, appraisal of the evidence generated from clinical evaluation should address the relevance and quality of all SaMD aspects including the following:

- SaMD definition statement;
- Risk assessment and associated documentation;
- Labelling including claims, warning, limitations, contraindications, etc.;
- SaMD requirements in the QMS system; and
- Verification and validation.

6.6.1 Matching Clinical Evaluation Evidence to SaMD's Intended Use and Related Claims

When the clinical evaluation evidence isn't adequate for the intended use and claims of the SaMD, it may be necessary to modify the intended use and claims to mitigate or prevent the risk of incorrect results harming patients, and to provide users with confidence in the SaMD. There should be adequate transparency to the users on the clinical validity and any limitations on the SaMD's intended use by providing appropriate contraindications, precautions, and warnings to the users in the SaMD's labeling.

[9] SaMD risks include the risk of an intervention or an unnecessary intervention or the consequences of failing to intervene as a result of inaccurate or incorrect output from a SaMD.

[10] For more information on the concept of safety and benefit/risk refer to GHTF SG1 /N68 Essential Principles of Safety and Performance of Medical Devices Section 6 as appropriate for software.

In cases where data sets used for generating clinical performance evidence are of limited availability or do not cover the desired range of the algorithm, or outputs, limitations of performance should be made transparent to the user and patients as part of the labelling. Alternatively, altering the original intended use statement and claims to match the actual performance is also considered to be adequate. For example – a novel SaMD that intends to diagnose patients with a certain condition, finds out that there is limited evidence on the acceptable analytical validity measures (accuracy, limit of detection, precision, etc). Clearly indicating in a transparent manner such actual performance is needed for the user to have confidence in the output of the SaMD, and to minimize risk to patients from inadequate results of the SaMD.

As stated in Section 7.2, there is flexibility regarding the type of clinical performance evidence required to establish validation of a SaMD claim(s).

6.6.2 Benefit/Risk Determination

Benefit/risk determination should incorporate evidence and knowledge from the assessment of scientific validity, analytical validity, and clinical performance, but also considerations for patient preferences and alternative methods for standard of care associated with the healthcare situation or condition that the SaMD operates in. The risk tolerance varies among patients and affects the individual patients' decisions and willingness to accept such risk with the SaMD in exchange for the benefit. The assessment should focus on relevant facts, uncertainties, and key areas of judgment.

SaMD generally only poses risks associated with decisions made based on the output provided by the SaMD. In cases of a false positive output by the SaMD, an unnecessary test or procedure may occur, resulting in associated procedural risks, the most serious of which may include deterioration of the patient's healthcare situation or condition, the need for surgical intervention, and death. In cases of a false negative, there is risk of failure to diagnose and properly treat a significant situation or condition, which could also be associated with the same adverse events mentioned above.

The probable benefits of the SaMD are also based on the output provided by the SaMD. These include improved sensitivity and specificity for detecting the healthcare situation or condition compared to other available methods of care. The benefit/risk assessment should determine and evaluate the likelihood of false positives and false negatives in the intended use population and where possible compare these to known standards for sensitivity and specificity of the condition being evaluated.

Other methods available for accomplishing the intended use of the SaMD should be identified. The results of the evidence should indicate that the SaMD performs favorably compared to other available technologies. The benefit/risk determination should consider the impact of results that cannot be generalized to a broader population than that studied. Patients may be willing to accept the risks associated with the SaMD because of its noninvasive nature.

In conclusion, the available information should support the quantitative and qualitative analysis of the SaMD results for the intended use, and demonstrate that the probable benefits outweigh the probable risks for the SaMD.

7.0 Level of Evidence According to SaMD Category

7.1 Categories of SaMD

SaMD N12 describes an approach to categorize SaMD based on the factors identified in the SaMD definition statement. The determination of the categories is the combination of the significance of the information provided by the SaMD to the healthcare decision and the impact of the information provided by the SaMD to the healthcare situation or condition as shown in the table below and in Section 8.3.

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

Figure - SaMD Categories

The four categories (I, II, III, and IV) are based on the levels of impact on the patient or public health where accurate information provided by the SaMD to treat or diagnose, drive clinical management or inform clinical management is vital to avoid death, long-term disability or other serious deterioration of health.

7.2 Importance of Clinical Evidence and Expectations by SaMD Category

As described previously, clinical evaluation evidence is generated to show adequate analytical validity along with clinical validity; the level of evidence should be risk based.

The following factors are used to determine the level of clinical evaluation evidence and where needed:

- The category of the SaMD[11] - Category I and Category II SaMD are considered lower risk compared to higher risk SaMD in Categories III and IV as the latter include SaMD that provide a diagnosis or recommendation for treatment for critical and serious situations or conditions; and
- The intended use of the output of the SaMD - As identified in Section 6.1, SaMD can treat a situation or condition, or can be grouped as either non-diagnostic SaMD or diagnostic SaMD.

[11] See Appendix 8.3 – SaMD Categorization.

The following summarizes where clinical evaluation evidence is needed to demonstrate the clinical evaluation of the SaMD based on the clinical evaluation that was performed using the above factors, and based on the impact of the SaMD's output to patients and public health:

SaMD in Category I:

- For all SaMD in this category:
  - Analytical validity evidence (generated through verification and validation QMS activity) based on and in conjunction with scientific validity information is sufficient to demonstrate the clinical evaluation evidence of the SaMD.
- For Novel SaMD in this category:
  - Manufacturers are expected to collect real world performance data to generate scientific validity evidence in addition to analytical validity evidence (generated through verification and validation QMS activity).

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

Figure 5 - Evidence for Category I SaMD

SaMD in Category II

- For all SaMD except for category II.*ii*:,
  - Analytical validity evidence (generated through verification and validation QMS activity) based on and in conjunction with scientific validity information is sufficient to demonstrate the clinical evaluation evidence of the SaMD.
- For Diagnostic SaMD in II.*iii*:
  - Clinical performance evidence is expected in addition to analytical validity and scientific validity evidence.
- For Novel SaMD in this category:
  - Manufacturers are expected to collect real world performance data to generate scientific validity evidence in addition to analytical validity evidence (generated through verification and validation QMS activity).

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

Figure 6 - Evidence for Category II SaMD (except Category II.*ii* SaMD)

SaMD in Categories II.*ii*, III **and** IV**:**

- For all SaMD in these categories (well-known or novel):
  - Analytical validity evidence (generated through verification and validation QMS activity) based on and in conjunction with scientific validity information is sufficient to demonstrate the clinical evaluation evidence of the SaMD.
  - In circumstances where the scientific validity is novel, manufacturers should generate appropriate association of the SaMD output to the clinical condition/physiological state using approaches described in scientific validity as described in Section 6.3.
- For Diagnostic SaMD in these categories:
  - Clinical performance evidence is expected in addition to analytical validity and scientific validity evidence.

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

Figure 7 - Evidence for Category II.*ii*, III, and IV SaMD

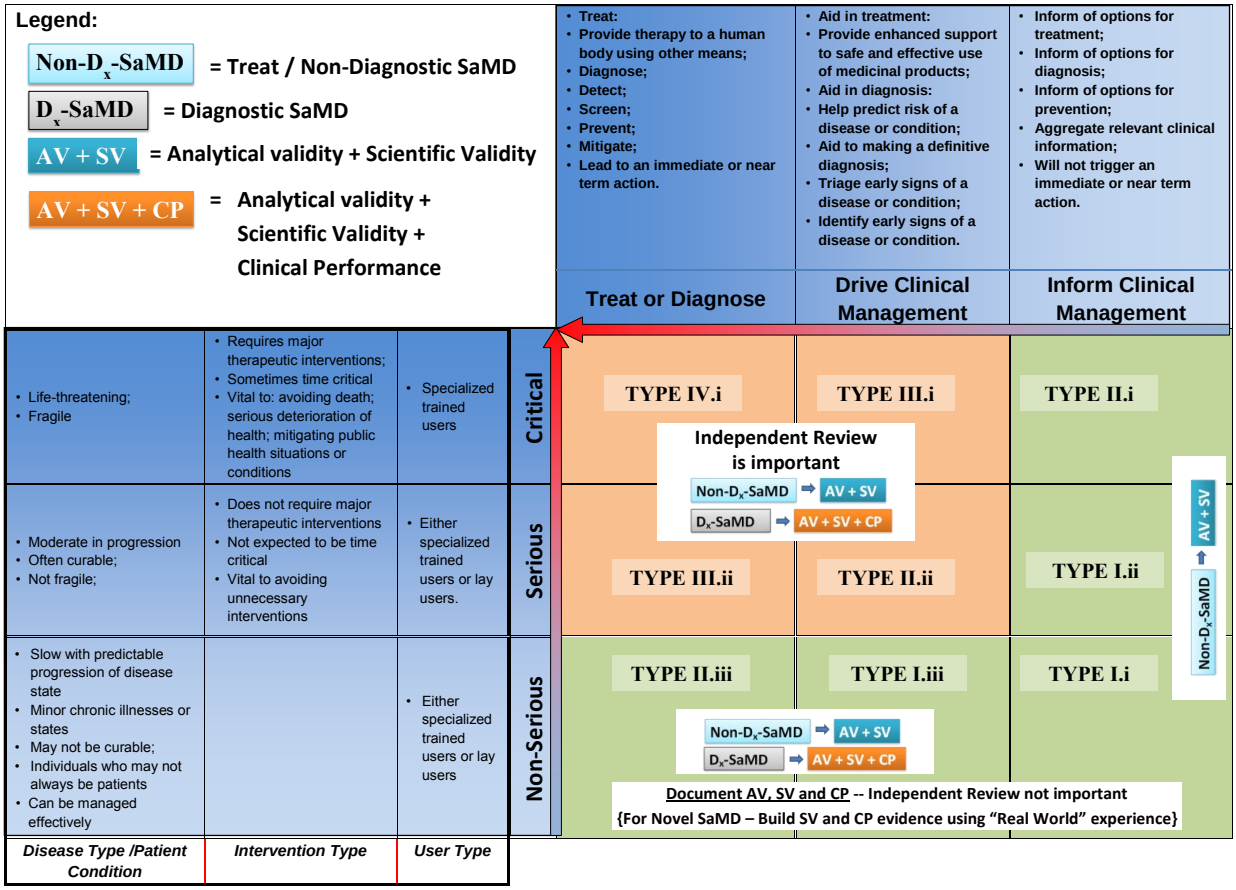


Figure 8 - Summary of Clinical Evidence and Expectations by SaMD Category (See appendix 8.5 for full page image)

7.3 Importance of Independent Review of Evidence by SaMD Category

Similar to the importance of evidence, certain SaMD categories may require independent review of the evidence to provide users the confidence in the SaMD's clinical validity. The concept of independent review is analogous to having peer review of journal articles or the concept of design review performed in the QMS system.

The recommendation for independent review for certain categories of SaMD does not imply the need for premarket review (authorization) by a regulatory authority which is outside the scope of this document. Regardless of the category of SaMD, the level of regulatory oversight (premarket review/market authorization) may depend on an individual jurisdiction's regulatory laws where the SaMD will be made available.

The recommendation for independent review highlights where the evidence generated from the clinical evaluation of the SaMD should be reviewed by someone other than the SaMD manufacturer to objectively appraise the SaMD's intended purpose and the conformity with the overall clinical evaluation evidence.

The following is a possible recommendation where independent review of clinical evaluation evidence is of importance.

SaMD in Category I:

- Independent review of evidence not important
- Manufacturers should document their appraisal of the clinical evaluation evidence with the SaMD definition statement and associated claims.

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

SaMD in Category II (except for category II.*ii*):

- Independent review of evidence not important
- Manufacturers should document their appraisal of the clinical evaluation evidence with the SaMD definition statement and associated claims.

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

SaMD in Categories II.*ii*, III and IV:

- Manufacturers should document their appraisal of the clinical evaluation evidence with the SaMD definition statement and associated claims.
- Independent review of evidence is important

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

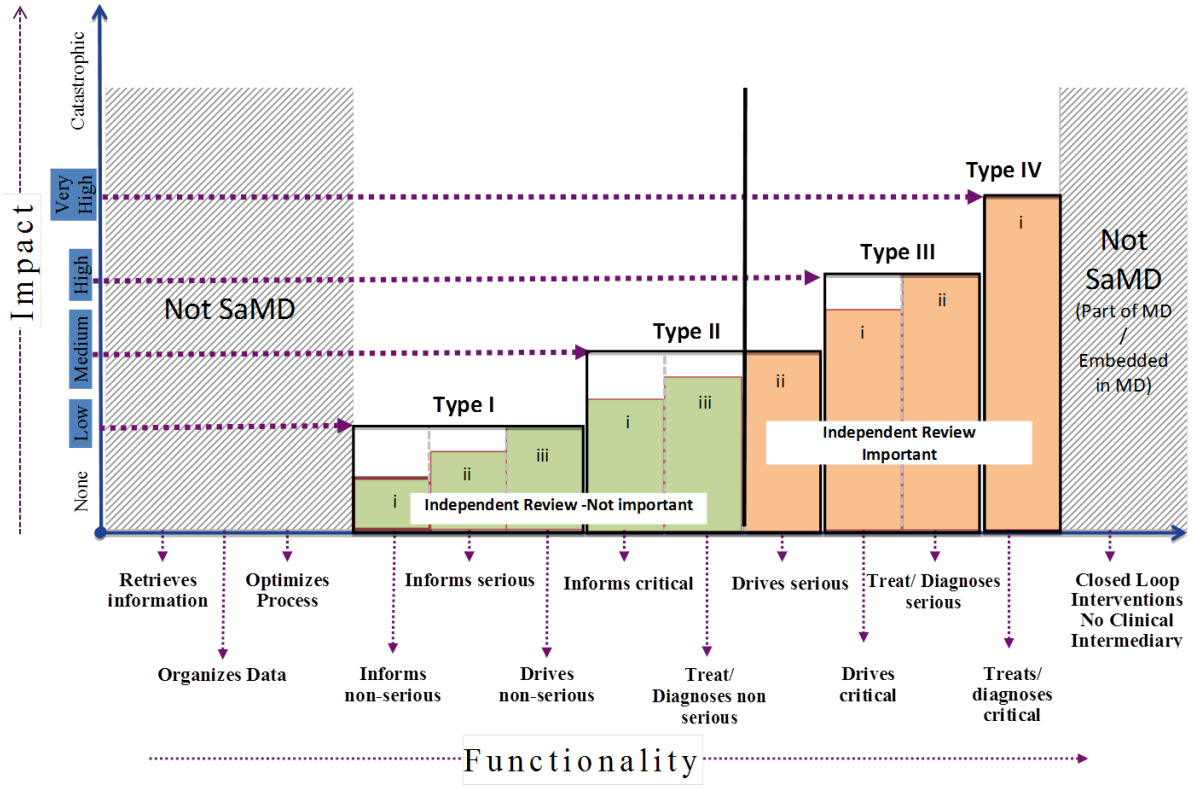


Figure 9: Importance independent review

7.4 Pathway for Continuous Learning Leveraging Real World Clinical Evidence

It is anticipated that one of the unique aspects that differentiate SaMD from other medical devices is the way SaMD may leverage technology and connectivity i.e., the seamless communication between devices, technology and people to continuously monitor the safety, effectiveness and performance of the SaMD. Unlike many other medical devices where real world experience is often difficult to gather as it comes in many forms (e.g., longitudinal follow up data that may be in a registry or insurance claims) and quality (e.g., missing data, variable definitions, etc.), with the connectivity of a SaMD this is easier.

Ideally the SaMD manufacturer has an idea early on regarding the longer term possibilities for the functionality and claims that may be supported by learning about the SaMD over time. As additional clinical data to support the new claims is gathered, the SaMD manufacturer will update the clinical evaluation. In practice, the clinical evaluation is a dynamic summary that changes as knowledge of the SaMD increases.

The "continuous learning" referred to here is not 'machine learning software', i.e., where software device keeps learning automatically after it has been released into the market; rather it refers to collecting post-market information.

Continuously collecting and analyzing post-market information (e.g., safety reports, including adverse event reports, results from performance studies, published literature) can help the SaMD manufacturer understand the real world performance of the SaMD. Manufacturers should appropriately review this information to determine if there are any changes to the safety, effectiveness or performance, or possible impact on benefits and risks of the SaMD that would indicate a need for a design change or a labeling change regarding contraindications, warnings, precautions or instructions for use.

It is also anticipated that if planned correctly, as a SaMD manufacturer learns by monitoring real world experience it can help the SaMD evolve after introduction into the market. This may potentially lead to a substantial change to the SaMD intended use and claims supported by the clinical data gathered, analyzed and appraised from the continuous monitoring.

Learning may impact the original category of a SaMD in the following ways:

- Real world performance provides evidence that analytical or clinical performance is superior than the performance initially evaluated by SaMD manufacturer, or
- Real world evidence indicates that analytical or clinical performance is lower than the performance initially evaluated by SaMD manufacturer.

An example is shown in scenario 1 in Figure 10 below. In this scenario, a SaMD manufacturer can conduct a retrospective clinical evaluation based on real world data and incorporate new information into the SaMD claims to enhance its clinical validity by further clarifying the SaMD's performance.

In the example shown as scenario 2 in Figure 10 below, a SaMD manufacturer can conduct a clinical evaluation based on gathering prospective real world data and incorporating the new information into the SaMD's intended use and definition statement, modifying design features to minimize risk, provide transparency by further clarifying the SaMD's performance and validity,

and minimize risk of incorrect results resulting in patient harm. Such data can potentially result in modification of the impact (risk) category of a SaMD from high to medium.

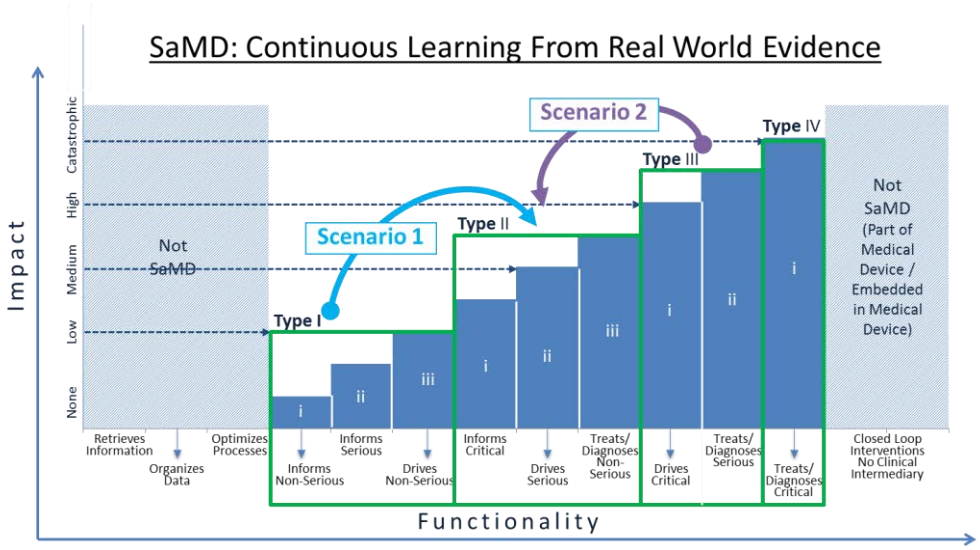


Figure 10: Continuous learning from real world evidence

This document encourages SaMD manufacturers to leverage SaMD's unique capability to capture user's interactions with the SaMD to conduct well planned clinical performance observational studies in addition to ongoing monitoring of technical and clinical performance.

A SaMD manufacturer can conduct an observational study that takes into consideration the healthcare situation or condition, and support a higher level significance of the information. For example, the output of a SaMD that is initially in the market to "inform" a serious healthcare situation or condition can collect evidence and provide the input data set to support claims for the output of the SaMD to either "drive" or "diagnose" a serious healthcare situation or condition. It would be expected that when moving up in significance from "inform" to either "drive" or "diagnose", that the same rigor be applied in evaluating scientific validity, analytical validity and clinical performance where appropriate as recommended in Section 7.3 . The advantage for the SaMD manufacturer is that they would access the data set that can support the evaluation with real world observational data and a retrospective analysis.

To summarize, one can envision a "building block" approach or an agile clinical evidence gathering approach to assimilating clinical evidence for a SaMD based on its risk categorization. Risk categorization of the SaMD is an evolving phenomenon through the lifecycle of the SaMD based on the on-going clinical evaluation process for the SaMD. All modifications that result from real world experience should also follow the framework for evidence requirements as outlined in Section 7.2 and level of independent review as outlined in Section 7.3.

8.0 Appendices

8.1 SaMD Definition Statement

All manufacturers should, as highlighted below and in (Section 6.0, start with a SaMD definition statement that is clear and strong about the intended use of the SaMD. Generally these aspects can be grouped into the following two major factors that provide adequate description of the intended use of SaMD:

A. The "**significance of the information provided by the SaMD to the healthcare decision**" which identifies the intended medical purpose of the SaMD. The statement should explain how the SaMD meets one or more of the purposes described in the definition of a medical device, e.g., supplying information for diagnosis, prevention, monitoring, treatment etc. structured in following sub categories:

a. **To treat or to diagnose** – the information provided by the SaMD will be used to take an immediate or near term action:

i. To treat/prevent or mitigate by connecting to other medical devices, medicinal products, general purpose actuators or other means of providing therapy to a human body

ii. To diagnose/screen/detect a disease or condition (i.e., using sensors, data, or other information from other hardware or software devices, pertaining to a disease or condition)

b. **To drive clinical management** - the information provided by the SaMD will be used to aid in treatment, aid in diagnoses, to triage or identify early signs of a disease or condition will be used to guide next diagnostics or next treatment interventions:

i. To aid in treatment by providing enhanced support to safe and effective use of medicinal products or a medical device.

ii. 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.

iii. To triage or identify early signs of a disease or conditions.

c. **To Inform clinical management** – the information provided by the SaMD will not trigger an immediate or near term action:

i. To inform of options for treating, diagnosing, preventing, or mitigating a disease or condition.

ii. To provide clinical information by aggregating relevant information (e.g., disease, condition, drugs, medical devices, population, etc.)

B. The intended "**state of the healthcare situation or condition**" that identifies the intended use for a disease or condition taking into account the patient's state of health, progression of the disease and associated type and immediacy of interventions, target population and type of users (trained or lay users). This portion of the statement should be expressed in the following structured sub categories:

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 mitigating impact to public health. SaMD is considered to be used in a critical situation or condition where:

    i. The type of disease or condition is:

        1. Life-threatening state of health, including incurable states,
        2. Requires major therapeutic interventions,
        3. 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.

    ii. Intended target population is fragile with respect to the disease or condition (e.g., pediatrics, high risk population, etc.)

    iii. 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. SaMD is considered to be used in a serious situation or condition when:

    i. The type of disease or condition is:
        1. Moderate in progression, often curable,
        2. Does not require major therapeutic interventions,
        3. 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.

    ii. Intended target population is NOT fragile with respect to the disease or condition.

    iii. Intended for either specialized trained users or lay users.

    Note: SaMD 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 SaMD 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. SaMD is considered to be used in a non-serious situation or condition when:

    i. The type of disease or condition is:
        1. Slow with predictable progression of disease state (may include minor chronic illnesses or states),
        2. May not be curable; can be managed effectively,
        3. Requires only minor therapeutic interventions, and

4. Interventions are normally noninvasive in nature, providing the user the ability to detect erroneous recommendations.

ii. Intended target population is individuals who may not always be patients.

iii. Intended for use by either specialized trained users or lay users.

C. **Description of the SaMD's core functionality**[12] which identifies the critical features/functions of the SaMD that are essential to the intended significance of the information provided by the SaMD to the healthcare decision in the intended healthcare situation or condition. This description should include only the critical features. (See applicability of this in (Section 6.0).

For more details and information related to the two major factors and formulating the SaMD Definition Statement refer to (Sections 5.0 and 6.0.

8.2 Clarifying SaMD Definition

This Appendix provides a representative list of features and functionalities that either meet or don't meet the definition of SaMD. This list is not exhaustive; it is only intended to provide clarity and assistance in identifying when a feature or functionality is considered to be SaMD.

Examples of software that are SaMD:

- Software with a medical purpose that operates on a general purpose computing platform, i.e., a computing platform that does not have a medical purpose, is considered SaMD. For example, software that is intended for diagnosis of a condition using the tri-axial accelerometer that operates on the embedded processor on a consumer digital camera is considered a SaMD.

- Software that is connected to a hardware medical device but is not needed by that hardware medical device to achieve its intended medical purpose is SaMD and not an accessory to the hardware medical device. For example, software that allows a commercially available smartphone to view images for diagnostic purposes obtained from a magnetic resonance imaging (MRI) medical device is SaMD and not an accessory to MRI medical device.

- The SaMD definition notes states that "SaMD is capable of running on general purpose (non-medical purpose) computing platforms." SaMD running on these general purpose computing platform could be located in a hardware medical device, For example, software that performs image post-processing for the purpose of aiding in the detection of breast cancer (CAD - computer-aided detection software) running on a general purpose computing platform located in the image-acquisition hardware medical device is SaMD.

- The SaMD definition notes states that "SaMD may be interfaced with other medical devices, including hardware medical devices and other SaMD software, as well as general purpose

[12] These could include specific functionality that is critical to maintain safety, effectiveness and performance profile attributes identified by risk management process undertaken by the manufacturer of SaMD.

software." Software that provides parameters that become the input for a different hardware medical device or other SaMD is SaMD. For example, treatment planning software that supplies information used in a linear accelerator is SaMD.

Examples of software that are not SaMD:

- The SaMD definition states "SaMD is defined as software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device". Examples of software that are considered "part of" include software used to "drive or control" the motors and the pumping of medication in an infusion pump; or software used in closed loop control in an implantable pacemaker or other types of hardware medical devices. These types of software, sometimes referred to as "embedded software", "firmware", or "micro-code" are, not SaMD".

- Software required by a hardware medical device to perform the hardware's medical device intended use is not SaMD even if/when sold separately from the hardware medical device.

- Software that relies on data from a medical device, but does not have a medical purpose, e.g., software that encrypts data for transmission from a medical device is not SaMD.

- Software that enables clinical communication and workflow including patient registration, scheduling visits, voice calling, and video calling is not SaMD.

- Software that monitors performance or proper functioning of a device for the purpose of servicing the device, e.g., software that monitors X-Ray tube performance to anticipate the need for replacement; or software that integrates and analyzes laboratory quality control data to identify increased random errors or trends in calibration on IVDs is not SaMD.

- Software that provides parameters that become the input for SaMD is not SaMD if it does not have a medical purpose. For example, a database including search and query functions by itself or when used by SaMD is not SaMD.

8.3 SaMD Categorization

(describes a method for categorizing SaMD based on two major factors representing aspects that can raise or lower a SaMD's potential to create hazardous situations to patients:

- State of the healthcare situation or condition; and
- Significance of the information provided by the SaMD to the healthcare decision.

With consideration of these two parameters, the table below displays SaMD categories:

State of Healthcare situation or condition	Significance of information provided by SaMD to healthcare decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV.*i*	III.*i*	II.*i*

Serious	III.*ii*	II.*ii*	I.*ii*
Non-serious	II.*iii*	I.*iii*	I.*i*

Criteria for Category IV –

i. SaMD that provides information to treat or diagnose a disease or conditions in a critical situation or condition is a Category IV and is considered to be of very high impact.

Criteria for Category III –

i. SaMD that provides information to treat or diagnose a disease or conditions in a serious situation or condition is a Category III and is considered to be of high impact.
ii. SaMD that provides information to drive clinical management of a disease or conditions in a critical situation or condition is a Category III and is considered to be of high impact.

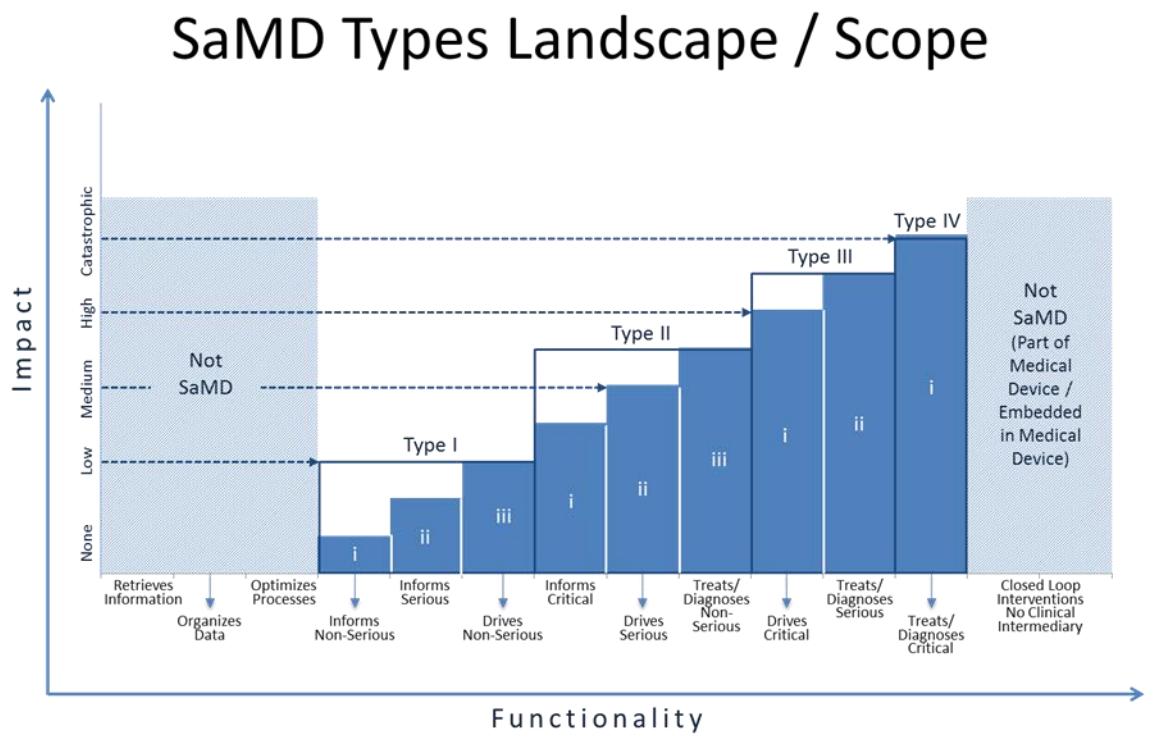
Criteria for Category II –

i. SaMD that provides information to treat or diagnose a disease or conditions in a non-serious situation or condition is a Category II and is considered to be of medium impact.
ii. SaMD that provides information to drive clinical management of a disease or conditions in a serious situation or condition is a Category II and is considered to be of medium impact.
iii. SaMD that provides information to inform clinical management for a disease or conditions in a critical situation or condition is a Category II and is considered to be of medium impact.

Criteria for Category I –

i. SaMD that provides information to drive clinical management of a disease or conditions in a non-serious situation or condition is a Category I and is considered to be of low impact.
ii. SaMD that provides information to inform clinical management for a disease or conditions in a serious situation or condition is a Category I and is considered to be of low impact.
iii. SaMD that provides information to inform clinical management for a disease or conditions in a non-serious situation or condition is a Category I and is considered to be of low impact.

The figure below depicts the categories of SaMD based on the impact and functionality. As displayed in the table above, the impact of the SaMD on patient or the public health is divided into four categories (Categories I, II, III, IV) while functionality (to inform or drive clinical management, to treat or diagnose) includes three categories. This categorization framework builds on the principles underlying the classification rules established in the GHTF classification principles documents, covering individual risks, public health risks, user skills, and importance of the information provided. While the categorization framework itself is not a regulatory classification, it sets a path towards a common vocabulary and approach to such classification aimed at determining appropriate levels of regulatory oversight.


SaMD Types Landscape / Scope
Impact
Catastrophic
High
Medium
Low
None
Not SaMD
Type I
Type II
Type III
Type IV
i
ii
iii
i
ii
iii
i
ii
i
Not SaMD (Part of Medical Device / Embedded in Medical Device)
Retrieves Information
Organizes Data
Optimizes Processes
Informs Non-Serious
Informs Serious
Drives Non-Serious
Informs Critical
Drives Serious
Treats/ Diagnoses Non-Serious
Drives Critical
Treats/ Diagnoses Serious
Treats/ Diagnoses Critical
Closed Loop Interventions No Clinical Intermediary
Functionality

1206
1207

8.4 Illustrative Examples of Clinical Evaluation Concepts for SaMD

The following illustrates a series of questions for different examples that may help to determine the required level of clinical evaluation.

Example: Algorithm to Detect Atrial Fibrillation

The SaMD demonstrates with certainty (success criteria) that the algorithm is able to detect atrial fibrillation with PPV of 65%.

- Is this a clinically adequate criterion for the intended use?
- What are the other clinical performance specifications that are necessary in order to fully assess this criteria (NPV, sample size, etc.)?
- What is the population for which this detection is intended and does this have an impact on the success criteria?
- Does this provide a clinically meaningful outcome/result in the current standard of care?

Example: Algorithm interprets Myocardial Infarction

The SaMD demonstrates with certainty (success criteria) that the algorithm can interpret Myocardial Infarction with 90% accuracy.

- What is the sensitivity and specificity of the result?
- How does this impact clinical workflow?
- How does 90% accuracy fit into current standard of care or when compared to the existing interpretation devices/SaMD?
- What is the comparator/gold standard?
- What is the health care situation (environment) of use and the importance of the SaMD to clinical management?
- What is the severity of the condition and what are the risks associated with an inaccurate result?

Example: EEG Analysis

The SaMD demonstrates with certainty that the SaMD can determine the location of a seizure based on EEG?

- What is the scientific validity for the association of EEG signals to the location of the seizure?
- If no existing gold standard, what is/are the criteria for diagnosis or management and is this clinically meaningful in the context of use for the device?
- Did the testing results demonstrate adequate clinical performance (specificity, selectivity, PPV, NPV, etc)?
- How does the availability of such SaMD output show benefits compared to current standard of care?

8.4.1 Illustrative Example of Clinical Evaluation Concepts – Skin Disorders

Example – Skin Disorder 1

Definition Statement

The SaMD provides generic information on moles, benign and atypical nevus, and malignant skin lesions. The SaMD uses photos with rulers next to them. The user manually identifies the location of the suspect skin lesion on a human body map, and tracks the changes over time in terms of size and appearance. The user is prompted to seek a medical professional's opinion. The SaMD allows the user to send the photos to their family doctor.

Based on the above definition statement the SaMD *informs clinical management*. Because the spectrum of the skin conditions includes information related to malignant skin lesions, the SaMD is used in a *critical healthcare situation or condition*.

This is an example of a Category II.*i* SaMD used for non-diagnostic purposes.

Clinical Evaluation

As a Category II.*i* non-diagnostic SaMD it is recommended that the manufacturer perform a clinical evaluation providing evidence for the scientific validity and analytical validity of the SaMD.

- Evidence of the scientific validity may be found in literature searches and clinical research and may include for example the use of well-known diagnostic rules in dermatology such as the ADCDE (may also be referred to as ABCD) Rule for mapping the mole.
- Evidence of the analytical validity may include thoroughly checking that the results from multiple executions of the SaMD processing the input and output satisfy the expected or desirable properties derived from the software specification or user expectations.

Example – Skin Disorder 2

Definition Statement

The SaMD provides lesion-specific information and flags suspect lesions that have a higher likelihood to progress to an atypical nevus state or are clearly abnormal. The SaMD tracks lesions with the use of color-calibrated photos of a tested minimal image quality and promptly detects any changes to margins, size, color, reflectivity, texture, and numbers. The SaMD automatically maps the skin lesions, highlights new lesions, counts them, and sends photos to a dermatologist or dermatopathologist without user intervention. The SaMD drives the next diagnostic action of a dermatologist, who's primary goal is to decide what lesions need interventions (excision and biopsy), and which lesions are OK to observe and monitor.

Based on the above definition statement the SaMD *drives clinical management*. Because the spectrum of the skin conditions includes information related to malignant skin lesions, the SaMD is used in a *critical condition*.

This is an example of a Category III.*i* SaMD used for diagnostic purposes.

Clinical Evaluation

As a Category III.*i* diagnostic SaMD it is recommended that the manufacturer perform a clinical evaluation to provide evidence of clinical performance in addition to evidence for the scientific validity and analytical validity of the SaMD.

- Evidence of the scientific validity may be found in literature searches and clinical research and may include for example the use of well-known diagnostic rules in dermatology such as the ADCDE (may also be referred to as ABCD) Rule for mapping the mole.
- Evidence of clinical performance demonstrating that the SaMD can stratify lesions into high and low-risk category as efficiently as a dermatologist is necessary to demonstrate the clinical performance. This could be prospective trial or retrospective clinical evaluation of a validated database of skin lesions (assuming the input to the SaMD will be of the same high quality photos as found in the validated database).
- Evidence of the analytical validity may include thoroughly checking that the results from multiple executions of the SaMD processing the input and output satisfy the expected or desirable properties derived from the software specification or user expectations.

For this kind of diagnostic SaMD, the clinical validity evidence that includes scientific validation and clinical performance should be independently reviewed along with the analytical validity evidence that will provide input to assurance of safety, effectiveness and performance of the SaMD.

Example – Skin Disorder 3

Definition Statement

The SaMD replaces the histo-pathology microscopic evaluation of a biopsy/excised sample through the use of a high magnification lens and an external UV light source that detects cytologic atipia (very large cells, poor maturation of cells, growth patterns) or cells typical of malignant melanoma.

Based on the above definition statement the SaMD provides a *diagnosis*. Because the spectrum of the skin conditions includes information related to malignant skin lesions, the SaMD is used in a *critical* condition.

This is an example of a Category IV.*i* SaMD used for diagnostic purposes.

Clinical Evaluation

As a novel Category IV.*i* diagnostic SaMD it is recommended that the manufacturer perform a clinical evaluation providing evidence for the scientific validity along with clinical performance evidence to show clinical validity in addition to analytical validity of the SaMD. such evaluation should include:

- Evidence of the scientific validity may be found in literature searches and clinical research that shows evidence that include using high magnification of images taken under UV light combined with image recognition to detect malignant skin lesions

- Evidence of clinical performance that is generated through a study (e..g. prospective study) comparing specificity and sensitivity of the SaMD based on histo-pathology microscopic or some genetic testing of excised lesions to confirm the diagnosis. Such study should include considerations for removing skin color, ambient light, contrast and other biases that show definitively the detection of malignant lesions. This may also require an adequate follow-up of lesions not excised/biopsied to confirm patient outcomes. There may be a need to consider that some cases may not present with skin lesions, but metastatic disease.
- Further real world experience from user feedback should be gathered post-market on an ongoing basis to continue to evaluate the SaMD's clinical performance.

Alternative claims and additional considerations

The above examples either specifically address melanoma or melanoma is within the spectrum of the claims.

- If the SaMD claims that it intends to detect furuncles, burns, frostbite, psoriasis, neurofibromatosis, chickenpox skin lesions, etc. the SaMD would be intended to be used in a serious situation or condition rather than intended to be used for a critical situation or condition thus lowering the risk profile of the SaMD.
- If the SaMD claims to detect benign skin lesions, such as eczema, acne, cellulitis, keloids, warts, etc. – the SaMD would be used in a non-serious situation or condition lowering the risk profile of the SaMD even further.

An example of scientific validity and acceptable "reference standard" for clinical performance includes an agreement between dermatopathologists reading histology slides under microscope. According to identified studies, there is only 35-58% concordance for grading of dysplasia (Duncan 1993), and dermatopathologists often did not agree with their own assessment of the same slide 6 months later (Piepkorn 1994); there is only 33% agreement on all benign versus all malignant in a sample of 37 "clear-cut" cases (Farmer, 1996).

Example – Coronary Physiological Simulation Software

Definition Statement

The software provides simulated functional assessment of blood flow in the coronary vascular system using data extracted from medical device imaging to solve algorithms and yield simulated metrics of physiological information (e.g., blood flow, coronary flow reserve, fractional flow reserve, myocardial perfusion). The SaMD is intended to generate results for use and review by a qualified clinician. This is a post-processing software for the clinical quantitative and qualitative analysis of previously acquired Computed Tomography (CT) DICOM[13] data for clinically stable symptomatic patients with coronary artery disease. The software displays the coronary anatomy with functional information using graphics and text,

[13] Digital Imaging and Communications in Medicine (standard for the communication and management of medical imaging information and related data).

including a computed and derived quantification of blood flow to aid the clinician in the assessment of coronary artery disease.

Based on the above definition statement the SaMD *drives clinical management* for in a critical *situation or condition*.

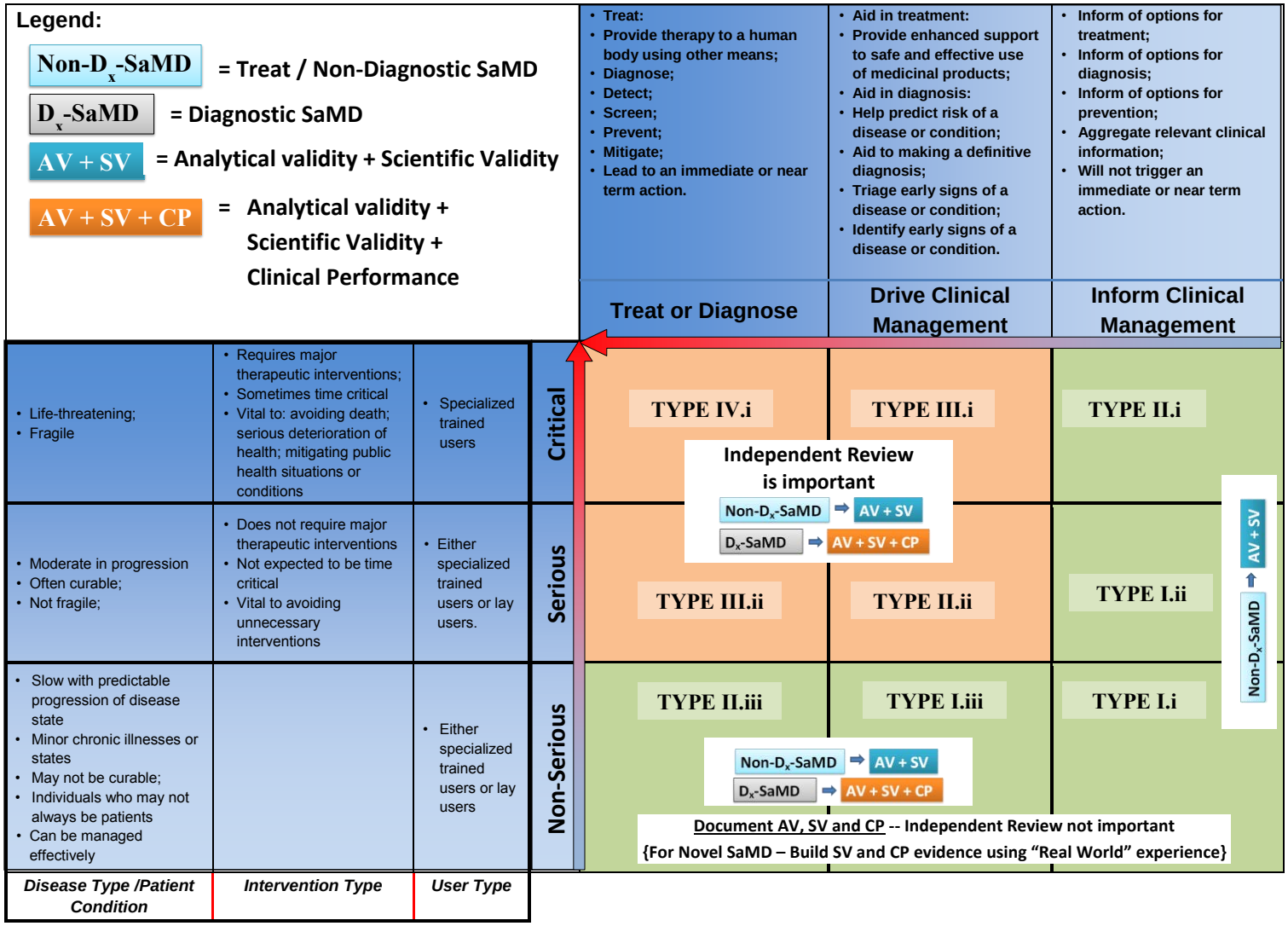
This is an example of a Category III.*i* SaMD used for non-diagnostic purposes.

Clinical Evaluation

As a Category III.*i* SaMD it is recommended that the manufacturer perform a clinical evaluation providing evidence for the scientific validity and analytical validity of the SaMD.

- Evidence of scientific validity may be found in literature searches and clinical research that shows that fractional flow reserve (FFR) has been validated through a number of clinical studies as a safe and effective means for measuring the extent of ischemia in the coronary arteries.
- Evidence of the analytical validity may include thoroughly checking that the results from multiple executions of the SaMD processing the input and output satisfy the expected or desirable properties derived from the software specification or user expectations:
    - Testing demonstrated the appropriate functionality of the SaMD and the basis of the computational methods;
    - Evidence demonstrated the functionality and accuracy of the SaMD output compared to ground truth data sets of specific modules and components such as automatic and semi-automatic image analysis and segmentation tools;
    - Testing demonstrated the reproducibility of the SaMD output using CT scans from various image acquisition systems by the SaMD;
    - Quantitative evidence demonstrated the validity of the computational modeling measurement methods of the SaMD by comparing the computational flow velocity solutions to Laser Doppler Anemometry and phase-contrast Magnetic Resonance Imaging (MRI) flow data in an in vitro model under steady-state and pulsatile flow conditions.
    - Evidence of clinical performance was generated by conducting a prospective, international, multicenter study. Evidence generated from the study demonstrated that the diagnostic accuracy of the lower boundary of the one-sided 95% confidence interval exceeds 70%.

8.5 Summary of SaMD Clinical Evaluation recommendation



8.6 Glossary of Terms Interpreted for SaMD from GHTF Documents

Accuracy	The degree of closeness of measurements of a quantity to that quantity's true value. When the output of the SaMD and true value are binary, accuracy is the proportion of true results (both true positives and true negatives) among the total number of output values examined.
Precision	The degree to which repeated measurements under unchanged conditions show the same results (related to reproducibility and repeatability).
Limit of detection	The ability of the SaMD to discern between information-bearing patterns of a clinical condition and random patterns that distract from the information.
Linearity or associated transfer function	The behavior of the output across the range of input data that is allowed by the SaMD.
Analytical sensitivity	The degree to which the SaMD's output is affected by parameters affecting input data including perturbation, image resolution, illuminations, data spatial distribution, data amount, etc.
Sensitivity	The ability of the SaMD to correctly identify across a range of available measurements patients with the intended clinical disease or condition (also called true positive rate).
Specificity	The ability of a SaMD to correctly identify across a range of available measurements patients that do not have the intended disease or condition (also called true negative rate).
ROC curve	A graphical plot that shows the tradeoff between sensitivity and specificity as the decision threshold that separates SaMD's negatives and positives is varied.
Positive predictive value	The likelihood of the patient having a disease or condition given that the SaMD's output is positive.
Negative predictive value	The likelihood of the patient NOT having a disease or condition given that the SaMD's output is negative.
Likelihood ratio	The likelihood that a given results would be expected in a patient with the target condition compared to the likelihood that the same results would be expected in an individual without that condition.
Cut-off thresholds or indices or scales	Cut-off values in relation to the clinical condition and on PPV, NPV and likelihood ratio. These should be established prior to validation and must be justified as to how they were determined and clinically validated.
True positive	A SaMD output which correctly indicates that a particular condition or attribute is present.
True negative	A SaMD output which correctly indicates that a particular condition or attribute is absent.
False positive	A SaMD output which incorrectly indicates that a particular condition or attribute is present.
False negative	A SaMD output which incorrectly indicates that a particular condition or attribute is absent.