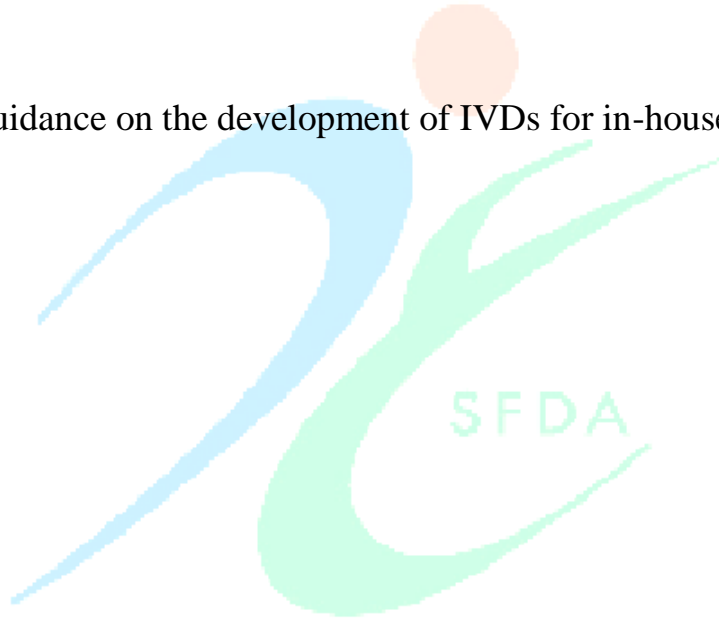MDS – G022

Guidance on the development of IVDs for in-house use

Version Number: 1.0
Version Date: 05/03/2024

Table of Content

Introduction

The aim of this guidance is to support the creation of design and performance data and documents required to support the development of an in-house IVD, the documentations describe in this guidance are required to be submitted to SFDA upon request and should therefore be available.

Scope

Guidance MDS-G009 defines and clarifies the requirements of manufacturers of point of care (POC) medical devices. In-house in vitro diagnostic (IH-IVD) medical devices are a subset of POC medical devices. This guidance outlines the development and post market activities for In-house in vitro diagnostic (IH-IVD) medical devices.

Background

The SFDA has issued this guidance document in accordance to the "Medical Devices Law" issued by the Royal Decree No. (M/54) dated 6/7/1442 AH through the following:

- Article 8 stipulating, "Medical devices cannot be marketed/used unless obtaining a registration and Marketing Authorization, and The SFDA may exempt some medical devices from the requirement to obtain a Marketing Authorization, after ensuring their safety, and not using them for commercial purposes, in accordance with rules approved by the Board".
- Article 26 stipulating, "The SFDA shall monitor the compliance of healthcare providers with technical regulations within healthcare facilities in order to ensure the safety and efficacy of medical devices and supplies in diagnosis and treatment".
- Article 28 stipulating, "The manufacturer, authorized representative, and healthcare provider shall report to the NCMDR any adverse event relating to their medical devices and supplies".

The "Implementing Regulation of Medical Devices Law" issued by Saudi Food and Drug Authority Board of Directors decree No. (3-29-1443) dated 19/2/1443 AH through the following:

- Article (8/3) stipulating that "The SFDA may exempt certain medical devices from the condition of obtaining Marketing Authorization for humanitarian and research purposes

upon verifying its safety in accordance with the following rules", and mentioned "Custom-Made Medical Device".

- Article (28/2) stipulating that "The Manufacturer, Authorized Representative and Healthcare Provider shall adhere to the Requirements of Post-Market Surveillance of Medical Devices, report to the NCMDR about incidents related to the medical devices and provide the NCMDR with all necessary information and documents including supply and distribution data".

General Requirements
Essential Principles

In annex 2 of [MDS-REQ1 Requirements for Medical Devices Marketing Authorization](#) there are a list of requirements to cover the Essential Principles of Safety and Performance for In-Vitro Medical Devices. It has two functions, firstly it ensures that all essential principles have been considered and secondly acts as an index to the objective evidence that supports compliance to the essential principles. It is an important document and should be compiled and approved by the lab before a test is put into service. The Essential Principles checklist ([Annex 3](#)) should be submitted and accompanied with a justification for any principle which is not met by the manufactured IVD medical device.

The checklist includes a table listing all the essential principles. This checklist is designed to be appropriate for all devices and therefore not all the requirements will apply to your test.

In the first column you should identify which essential principles apply to your device. For example, if your test consists of reagents the electrical safety requirements do not apply. It is ideal not only to state not applicable but also say why as this makes it really clear to the reader.

If standards have been applied or key guidance such as ISO standards, CLSI standards or national guidance these can be listed in the standards applied column.

Method of conformity solutions applied refers to verification and validation testing that you will have completed or if applicable certificates you may hold. For example, the essential principle 6 focuses on stability, you would state the titles of the stability studies

that you had conducted. In the final column you should add a reference to the report from that study so that the objective evidence is easily retrieved.

More comprehensive information and description of the Essential Principles can be found in [MDS – REQ 1 Requirements for Medical Devices Marketing Authorization](#)

Quality management System (QMS)

This section describes the general requirements for in-house assays in KSA and creation of performance data and documents required. This documentation may be required to be submitted to SFDA on request and should therefore be available.

- In-house IVDs should be manufactured and used under an appropriate quality management system.
- The quality management system is required to direct design, production, and use of the in-house assay and by using the processes in the system, generate evidence to support that the Essential Principles have been addressed.
- The following are useful sources of information on the creation of a quality management system (QMS)
  - ISO 13485:2016 Medical devices Quality management systems Requirements for regulatory purposes is used as best practice by IVD manufacturers.
  - ISO15189:2022 Medical laboratories. Requirements for quality and competence has been specifically written for the medical laboratory and is another source of best practice.
- Key areas to be fully documented by the quality system should include,
  - design,
  - production,
  - use,
  - storage,
  - packaging and transport of in-house IVDs within the laboratory.
- The QMS should be used to generate documentation for each in-house IVD. There should be a file containing documents defining,

- o Intended purpose,
- o Specifications,
- o Production processes and quality control requirements.
- o Data to support the design verification and validation,
- o Labelling.

Design and Development

These are the core requirements that should be defined and documented in the quality system in order to generate data and records that provide objective evidence of the design process.

- The IVD shall be designed and produced so that when used under the conditions and for the purposes intended, all reasonable measures have been taken to minimise the risk of compromising the health and safety of the patient, the user or any other person.
- The design and construction of the IVD shall conform to the Essential Principles and best practices considered. This includes identifying and eliminating risks associated with use (including disposal) and ensuring that adequate protection measures are in place. The laboratory shall ensure that the safety of the patient, the operator, and other staff is not compromised by the design, production, or the use of the validated IVD.
  - o ISO 14971:2019 Medical devices Application of risk management to medical devices is the standard used by IVD manufacturers on how to perform risk management, it is supported by a Technical Report ISO/TR 24971:2020 Medical devices — Guidance on the application of ISO 14971 which contains a specific IVD Annex H Guidance for in vitro diagnostic medical devices.
  - o ISO 22367:2020 Medical laboratories. Application of risk management to medical laboratories. ISO 22367 discusses risk management for medical laboratories. ISO 22367 outlines a process for a medical laboratory to identify and manage the risks to patients, laboratory workers, and service providers associated with medical laboratory examinations. The process includes identifying, estimating, evaluating, controlling, and monitoring the risks. The requirements in ISO 22367 apply to all

aspects of the examinations and services of a medical laboratory, including the pre-examination and post-examination aspects, examinations, accurate transmission of test results into the electronic medical record, and other technical and management processes described in ISO 15189.

- The device shall be designed and produced in a way that ensures it is safe to use for the entire intended life of the device. It is important to have stability data to support the defined shelf life of the test.

- Where the in-house IVD uses components purchased from an external source, the laboratory shall ensure that the products purchased will meet the requirements as specified in the design protocol. The variability of the materials being purchased should be considered. Communication of changes made by the supplier to the laboratory and how the laboratory will manage such changes shall be described in the design process.

Intended Purpose

Tests need to be designed to meet the needs of patients, clinicians and should be state of the art. This does not mean that they need to be the best in class, but they need to produce clinically relevant information to support medical practice.

To do this it is important to consider the Essential Principles throughout the design of the test according to the QMS and the data generated must support the intended use of the test as described in labelling and on Medical Device Application Form MDS – G009 Annex 1. There are many elements to consider when describing the intended use. Table 1 includes a non-exhaustive list of considerations.

Table 1 What should be included in an intended purpose statement

Intended Purpose Description	What to consider
What is detected and/or measured	The biomarker, gene, organism Are there structurally similar analytes What could interfere with the measurement of the analyte?
its function	Screening

	• Is this to screen blood or tissue donations? • Is this an asymptomatic population? Monitoring, • The sensitivity of a monitoring test may be different to a diagnostic test because you may be operating close to the cut off Diagnosis or aid to diagnosis, • Is the test the only decision-making test or is it systematically routinely used in conjunction with other tests and procedures? Prognosis, Prediction, • These will need data to demonstrate the result versus the outcome of the patient Companion diagnostic, • Evidence will be needed to support the ability of the test to appropriately select patients
the specific information that is intended to be provided in the context of: • a physiological or pathological state; • congenital physical or mental impairments; • the predisposition to a medical condition or a disease; • the determination of the safety and compatibility with potential recipients;	What are the specific information to be provided? What will the clinician do with the information as this impacts the risk? What characteristics will be needed for the device to achieve its intended purpose? In particular the analytical performance, such as, • analytical sensitivity, • analytical specificity, • trueness (bias), • precision (repeatability and reproducibility), • accuracy (resulting from trueness and precision), • limits of detection and quantification, • measuring range, • linearity,

• the prediction of treatment response or reactions; • the definition or monitoring of therapeutic measures	• cut-off, • handling and control of known relevant endogenous and exogenous interference, cross-reactions the clinical performance, such as • diagnostic sensitivity, • diagnostic specificity, • positive predictive value, • negative predictive value, • likelihood ratio, • expected values in normal and affected populations.
whether it is automated or not	If a test is automated or semi-automated data is needed to support the performance on each device combination used in the laboratory Suitable combinations should be recorded
whether it is qualitative, semi-quantitative or quantitative	Whether the qualitative, semi-quantitative or quantitative data is required to support performance across the anticipated range of the assay and around cut off or decision points.
the type of specimen(s) required	Appropriate sample types should be defined and data should be available to support each sample type. Appropriate criteria for specimen collection should be described. There should be data to support the combination of collection devices or swabs described Data to support the stability of the sample as well as the reagents is required. There should be sufficient to support the testing pathway in use. If intersite shipping is needed as part of the testing pathway data is needed to support

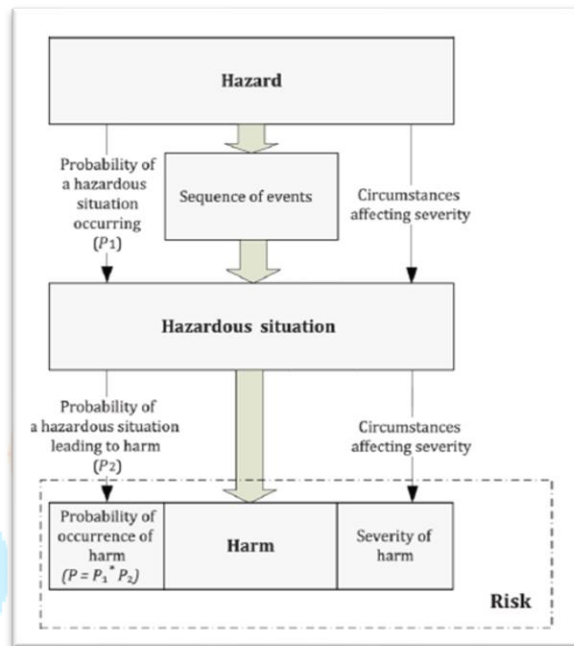|	the stability of the sample under the shipping conditions described.
where applicable, the testing population	Is the test intended for • All patients, adults only and or pediatric? • Are there any special criteria relating to age, sex, ethnicity?
for companion diagnostics, the International Non-proprietary Name (INN) of the associated medicinal product for which it is a companion test.	
the intended user	Define is this is a laboratory test for professional use or whether this could be used by a doctor, nurse of pharmacy. What qualification/ training should they have to use the test? Are the instructions appropriate for the user?

Risk

The first step in risk management starts in the design process with the identification of potential harms to the patient and an analysis of the potential hazards that could lead to that harm this starts from early design although in the early design phase not all the hazards will be controlled, as the design advances risk management will help guide the design and determine what verification and validation data is needed to provide evidence that risks have been controlled. For example, stability of reagents is a potential hazard that can be identified in early design, improved during the design process and evidence the risk is controlled provided by the stability study.

A hazard is a potential source of harm and can include a false positive result, a false negative result or a delay in result. Hazards do not always lead to harm to the patient, user

or third party this is why risk management considers the probability that the hazard causes a harm

It is important to consider what harm could result to the patient, user or third party this is different for every test. For example, a false positive troponin test could lead to a patient receiving invasive treatment unnecessarily whereas a false negative could mean that a potentially lifesaving stent was not fitted. A delay in the result; for example, because the test failed to generate a valid result and had to be repeated would also delay treatment which in this case could be significant.

Hazard

Probability of a hazardous situation occurring (P_1)

Sequence of events

Circumstances affecting severity

Hazardous situation

Probability of a hazardous situation leading to harm (P_2)

Circumstances affecting severity

Probability of occurrence of harm ($P = P_1{}^* P_2$)

Harm

Severity of harm

Risk

Risk is the combination of the severity of harm and the probability the hazard would occur and then lead to the harm. Because failures leading to harm often have multiple causes and may have more than one contributing factor, ISO14971 calls these hazardous situations. Risk management tools have developed to consider the probability of each step. This can be very useful to get a realistic picture for IVD assays that act indirectly on the patient.

Tests are designed to prevent harm occurring, the strongest control measure is to design an assay to prevent a failure occurring. For example, when a test is developed potential interfering substances should be considered. These can be structurally similar compounds or substances that could contaminate the sample e.g. lipemic samples; tests are designed to optimise reagent formulation to minimise these effects. Verification studies are run to challenge with levels of the substance that may be encountered in routine samples to prove that the interference is minimised, this data should form part of the technical documentation.

The Clinical and Laboratory Standards Institute (CLSI) https://clsi.org is a not-for-profit organization that develops laboratory standards worldwide. These standards are recognized by laboratories, accreditors, and government agencies world-wide as the best way to improve medical laboratory testing and provide a wealth of knowledge on how to perform analytical and clinical performance studies.

Performance

IVD performance is composed of analytical and clinical performance. Table 2 includes a list of appropriate terms and definitions. Analytical performance focuses on the ability to accurately generate a result whereas clinical performance focuses on the population and the ability for users to consistently generate an accurate result. Data to support both analytical and clinical performance will be generated using patient samples that should have been acquired under appropriate ethics.

Commercial assays must have data to support that any qualified user should be able to run the test without additional training using the instructions for use provided. In-house tests may be used on one or a limited number of network sites and may be run by a small group of users, the risk associated with the size and diversity of the user group should be considered for each test to ensure that consistent results are generated. Commercial assays typically run clinical studies at three sites to gather data on consistent use, this may not be appropriate for an in-house assay so the approach taken should be justified in the technical documentation.

Table 2 Definition consideration of Analytical and Clinical Performance criteria to be considered based on the define intended purpose of the test.

	Definition	Comments
Analytical performance	Ability of a device to correctly detect or measure a particular analyte;	
Analytical sensitivity	Analytical sensitivity is defined as the limit of detection, i.e., the smallest amount of the target marker that can be detected above the system noise (Lob).	The LoB defines the system noise of an assay. The LoD defines the concentration of the analyte that will consistently be above the

	Definition	Comments
		LoB (95% of the time) based on the variability of the assay. The LoQ defines the concentration where the variability and bias of the assay meet a specified total error that is required for the intended purpose of the assay (quantitative reporting only).
Analytical specificity	Analytical specificity describes the extent to which the testing method measures only what it intends to measure.	If similar substances in the matrix influence the lab measurement, it is called an interfering substance. This may be a cross reactive entity or an inhibitory entity.
Trueness (bias)	Regression model (with noise assumed in both axes) will be used to estimate the bias at relevant levels per MDP or a upper and low bias limit across the range of data.	This is applicable only if there is a certified reference material or a certified reference method.
Precision (repeatability and reproducibility)	Agreement between the different independent measurement results of a sample with the lowest imprecision conditions (e.g. single operator, instrument and reagent lot).	Imprecision is indicated by the standard deviation along with the coefficient of variation for assays without a cutoff. For assays with a cutoff, imprecision is reported as the C95 value (analyte concentration above the assay cutoff) and may also include the C5 value (analyte concentration below the assay cutoff) as well.
Accuracy	Observed results based on bias and imprecision of the assay.	
Limits of detection	The limit of detection (LOD) is the smallest value that can be detected by this method above the system noise.	The LoD and LoQ are assessed for quantitative reporting methods.

	Definition	Comments
		The LoD is assessed for qualitative reporting methods
Limits of quantification	The lower limit of quantification (LLOQ) is the smallest determinable quantitative value that can be determined with an acceptable level of bias and precision.	Bias and precision = accuracy
Measuring range	Is defined by the LOQ and the linearity	
Linearity	Represents the proportional relationship between expected concentration and the observed concentration by an quantitative reporting assay.	Important for quantitative methods, since this, together with the LOQ, indicates a method's measuring range.
Cut off	For qualitative reporting assays, the cut-off value divides the range of measured values into results where the test condition is present (positives) and the test condition is absent (negatives).	For a qualitative reporting assay to work, a cutoff is established to define a binary result. This is typically done by establishing the cutoff with decisions on allowed false positive and negative results as compared to clinical truth.
Clinical performance	Ability of a device to provide results that correlate with a specific clinical condition or physiological or pathological process or state in a specific target population and in specific intended users	
Diagnostic sensitivity	The sensitivity of a diagnostic test is its ability to detect certain characteristics (e.g., disease). Sensitivity is defined by the quotient: Sensitivity = (true positive) / (true positives + false negatives)	Use for assay that are qualitative reporting, that is the reported results indicates that the subject has the test condition that is represented by the analyte detected by the assay.
Diagnostic specificity	The specificity of a diagnostic test is its ability to identify persons lacking certain characteristics (e.g., disease)	Use for assay that are qualitative reporting, that is the reported results indicates that the subject does not have the test condition

	Definition	Comments
	as non-patients. Specificity is defined by the quotient: Specificity = (true negative) / (true negatives + false positives)	that is represented by the analyte detected by the assay.
Positive predictive value	A device's ability to separate true positives from false positives for a given attribute in a given population;	The predictive value is not only influenced by the sensitivity and specificity of a diagnostic test, but crucially also by the prevalence of the disease in the tested population.
Negative predictive value	A device's ability to separate true negatives from false negatives for a given attribute in a given population;	The predictive value is not only influenced by the sensitivity and specificity of a diagnostic test, but crucially also by the prevalence of the disease in the tested population.
Likelihood ratio and expected values in non-affected and affected populations	The likelihood of a particular result occurring in a person with the clinical or physiological target condition relative to the likelihood of the same result occurring in a person without that clinical or physiological condition.	

Assays generally fall into two groups, quantitative and qualitative, Annex 2 provides more detailed information and references to supporting guidance on how to generate objective evidence to support that the design has been met. As every device is different, guidance cannot be prescriptive this Annex provides information on important considerations, sources of detailed guidance and illustrative examples.

Manufacture/ Production,

In-house IVDs may have been designed in-house, could use components that have been brought together and given a medical intended purpose or could be a commercial kit where the lab has changed the intended purpose that includes changing sample type or population.

These present different challenges but rely on similar principles to control manufacturing risks.

Purchasing

Planning, scheduling resources, and purchasing are all part of the manufacture of in-house devices. It is important to ensure the consistency of the materials being purchased to ensure that they do not influence the performance of the assay. It is important to consider the following,

- The laboratory shall establish documented procedures to ensure that products sourced for use in the routine production of in-house IVDs conform to specified requirements. Specifications should be documented, and records kept so that shipments are traceable so that in the event of a performance issue, it is possible to determine which batches of materials were used.

- All sub-contracted services that affect the quality of the in-house IVD should be controlled including the development of any software. To be an in-house device there has to be some elements of manufacturing; however, key raw materials such as antibodies may be critical to quality but are purchased from a supplier. This could also apply to specialist services such as the development of software; for example, software that includes algorithms. These suppliers are critical to the safety of the device and can be controlled in different ways. For example, testing each batch of polyclonal antibody when it is received to determine the titre. Selecting suppliers with ISO 13485, ISO 9001 or equivalent and checking that the certificate covers the service you receive. Acceptance testing of software for example entering know data and ensuring the correct result is generated. The level of control needed will depend on the risk of the material to the safety and quality of the material or service and should be considered in the risk management.

Manufacturing environment

- The laboratory shall plan and carry out the production of in-house IVDs under controlled conditions from design, manufacture and validation of the proposed product to its release for routine use. Procedures should be created to describe these processes and the associated responsibilities and authorities in the lab.

- Where work environment conditions are critical for the handling, production, validation, or use of the in-house IVD, the laboratory shall monitor and control these work environment conditions.

- Controlled conditions include:

    o identification of different stages of assay development, including routine production

    o implementation of review, verification and validation of the above by designated personnel

    o documented requirements (e.g. batch size, acceptable performance limits)

    o identification and use of reference materials and reference measurement procedures

    o identification and use of suitable equipment (i.e. appropriate for use and calibrated)

    o identification and use of monitoring and measuring processes and devices,

    o planned and documented specifications and processes for the release of the product, and for the receipt of the product at the site of routine storage before product use

    o planned and documented specifications and processes for any packaging and labelling of the final product, accessories or components,

    o planned and documented specifications, processes and records if cleanliness of the environment is critical (e.g. production step where sterilisation is required).

- Records relating to production batches and the status of an in-house IVD, or a component thereof, must be traceable and retained for a minimum of 4 years beyond the date of their valid use or 5 years from the date of manufacture

- The following records should be created and maintained,

- the laboratory must maintain records of each batch produced. This must include records of all components of the batch, and any other information relevant to the successful use of the batch in the routine environment.
- each batch, and each component within the batch, must be assigned a unique identifier and must be available for the purposes of traceability.
- at each stage of production, the product status must be identifiable. A batch of an in-house IVD awaiting final release validation must be clearly identified as such. A batch that has been deemed acceptable for routine use (i.e. validation criteria have been passed) should be identified as such.
- if a component within a batch is changed, then that batch must be considered as a new batch and will require re-validation.

- The laboratory must ensure that any validation and monitoring measures required for verification of the processing steps are identified, validated and documented.

- The approved manufacturing site/laboratory is the only site/laboratory allowed to manufacture Kit/device and has the right to disturb it within the laboratory network

- If a batch, or part thereof, is distributed within a laboratory network, then procedures or instructions must also be issued relating to the transport, receipt and use of the test at its destination. Such instructions should refer to the packaging, transport, handling, storage and identification of the in-house IVD.

Competence

- Senior staff must have sufficient diagnostic or research experience with new test development and validation. The depth and complexity of this experience must be commensurate with the range and complexity of IVD development undertaken in the laboratory. Records should be kept to support their experience.

- The presence of experienced supervisors and trainers is essential, given their critical involvement in error detection, error correction and problem solving.

Before putting the test into routine use

- The laboratory must establish procedures to verify the suitability of the in-house IVD for use in the setting in which it will normally be used.

- Medical device manufacturers shall conduct the necessary technical tests to prove their products' compliance with the regulatory requirements for safety, performance and quality including electrical, mechanical, biological, usability and stability tests, in addition to other tests according to the nature of the medical product.
- Before the test is put into routine use this data should be reviewed and approved before the declaration can be prepared.

Labelling

The following should be included in the labelling

- The device is labelled as an IVD
- The lot/serial/batch number is indicated on the device
- Expiry date
- Quantity of content and storage conditions are indicated on the device
- If it is a kit list the contents
- Instructions for use
- Manufacturer
- If the device is sterile then this should be clear in the labelling including the method of sterilization.

Packaging and transport of in-house IVDs within the laboratory

In-house assays are outside the scope of MDS-REQ 12 Requirements on Transporting and Storage for Medical Devices; however, the general principles apply. As part of risk management, the common risks associated with storage and transport should be considered, for example

- Storage is a clean location
- First in first out storage system (FIFO) to ensure reagents are always used within their expiration date
- Design and validate packaging for transport within the network if appropriate. Document how it should be packaged and shipped to ensure that the reagents do not deteriorate and that third parties are not harmed by broken or leaking containers.

- As part of risk management consider any risks that could be posed by humidity (e.g. on membranes) vibration (packing down of solid phase particles)

Post market surveillance

Post market surveillance is proactive and reactive. The aim of proactive PMS is to ensure that the test remains state of the art and suitable to support medical practice. If changes in state of art are reported in journals or if the intended use of a test has changed; for example, because of the development of a new treatment, then the risk should be assessed, and the assay should be validated if changes are made.

Reactive PMS considers what to do when the test goes wrong, or the results are unexpected. Once the test is in routine use measures should be in place to monitor the performance and ensure that any adverse event or failure in performance is investigated.

- The laboratory must establish documented procedures to ensure that conditions preserving the effectiveness of an in-house IVD are not compromised.
- The health institution reviews experience gained from clinical use of the devices and takes all necessary corrective actions.
- MDS-REQ 11 Requirements for Post-Market Surveillance of Medical Devices is a useful reference

Annex (1) Requirements

Requirements for the manufacturer	1	Manufacturers shall obtain an establishment license from SFDA and meet requirements in: - Requirements for Medical Devices Establishments Licensing (see MDS-REQ 9)
General requirements for the MD	2	- For the purpose of obtaining a Medical Devices Marketing Authorization (MDMA) manufacturers shall comply with the requirements mentioned in " requirements for Medical Devices Marketing Authorization" (see MDS-REQ1) - For Manufacturing medical devices at Points of Care (POC) refer to Guidance for Points of Care (POC) Medical Devices Manufacturing (see MDS-G009)

Additional requirements for the MD	3	For biotechnology medical device , the following additional requirements shall be addressed and provide more prescriptive detail to the manufacturer on what is required:
		- review of technical documentation
		- manufacturer shall consider the full medical device lifecycle e.g.: design & development process, safety, design Verification and Validation, Post market activities… etc.) for more details (see MDS-REQ1)
		   • Annex (3) Medical Device Technical Documentation
		   • Annex (4) IVD Technical Documentation
		   • Annex (6) Clinical Evaluation and Post-Market Clinical Follow-Up
		   • Annex (7) Performance Evaluation, Performance Studies and Post-Market Performance Follow-Up
		- Performance evaluation /clinical trials
		   • Sponsors of a medical device clinical investigation or an IVD clinical performance study must apply to the SFDA for approval
		   • must have obtained approval by a local research ethics committee (EC) prior to SFDA application
		for more details (see MDS-REQ2)
		- Monitoring of the inhouse development site
		- Patient specific needs
		- Post market activities
		- Labelling
		- Environmental impact

Annex (2) Analytical and clinical evaluation of in-vitro diagnostic (IVD) assays

<u>General requirements</u>

Most IVD assays fall into one of two types:

Quantitative reporting assays: these report out a quantitative concentration value of the analyte evaluated in the assay. The increase or decrease of this value has meaning in terms of clinical utility.

Examples: sodium, glucose, or cholesterol levels in the blood.

Qualitative reporting assays: these report out a binary result based on whether an observed value of the analyte from an internal continuous response (ICR) is above or below an establish cutoff that determines whether the subject presents with a test condition (positive) or does not present with a test condition (negative) that the analyte represents. An increase or decrease in the value of the ICR does not have clinical utility, it is a matter of being above or below the establish assay cutoff.

Examples: Indication that subject has covid, or that a specific mutation has been found in a tumor.

Both types of assays will evaluate subjects across a required reportable range given the intended purpose of the assay. Analytical evaluation of these assays is slightly different because the evaluation of the accuracy of the assays is different. It is therefore essential to define based on the intended use whether the assay will be quantitative or qualitative reporting when defining the assay product requirements prior to the start of development.

During development and prior to the initiation of the analytical studies it is critical that the assay will be capable of covering the required clinical range, and for quantitative reporting assays this also considers any medical decision points that have been established by the medical community at large.

In the early phase of design, you should also determine and document what the acceptable accuracy will be needed for the assay to fulfill its intended purpose. This will define the requirement of the assay; future analytical studies will then confirm that the test is capable of meeting these criteria in routine use.

One of the goals of any development effort is to ensure that the required accuracy for the assay given it's intended purpose is met. Accuracy is influenced by bias and imprecision. The analytical studies allow developers to evaluate aspects of both of these performance metrics. The clinical study allows developers to put the product in the hands of the "user" to collect data to demonstrate that the design of the product will result in the required accuracy before it is put into routine use.

Quantitative reporting assays do not have a cutoff and report a continuous range of values. As such, performance of these assays is evaluated throughout the intended use range of the assay. In the case where there is one or more established "medical decision points" (MDP) that have been established by the medical community for an assay analyte, performance of the assay should also be evaluated around these MDP as well.

Qualitative reporting assays have one or more established cutoffs. For sake of simplicity only binary reporting assays (those with one cutoff) will be discussed here. The cutoff determines the reported outcome, (i.e. positive or negative) and as such the performance evaluation for these assays is done around the cutoff as this has a direct effect on the achieved accuracy of the product.

Table X identifies the studies which are conducted to support analytical performance. The Clinical Laboratory Standards Institute (CLSI) guidance documents have been developed to provide detail on how to design and run these studies.

Table X The table below indicates the analytical studies that need to be evaluated for each type of assay.

Study	Quantitative	Qualitative	Guidance
Limit of Detection LOB/LOD/LOQ	X	X	CLSI EP17-A2

Precision
  Repeatability
  Reproducibility (within lab and site to site)
	X	X	CLSI EP05-A3
			CLSI EP12-A2

Interfering Substances	X	X	CLSI EP07-A2 CLSI EP37 (Supplemental Tables)
Normal Range	X	NA	CLSI EP28-A3
Linearity	X	NA	CLSI EP06 -2nd Ed

These studies are conducted to verify that the design meets the design input requirements for the product based on its intended purpose, other studies are then required to ensure that the performance of the assay will be consistent. Some of these would be done during the design work, or shortly thereafter, but must be completed before the assay will be put into routine use.

Stability is essential and should consider both the assay including calibrators and controls but also the samples. If reagent kits are transported between locations (post-delivery from the manufacture), then data to show that the assay will still perform at the end of its expiration following the maximum exposure during shipping is required. Sample stability should be considered for the maximum time from taking the sample to running the assay. Transport and storage of samples during development needs to be defined as early as possible to ensure that there are no artifacts caused (ex. Too many freeze/thaw cycles).

Study	Quantitative	Qualitative	Guidance
Stability Sample, reagent, transport	X	X	CLSI EP 25
Guard-banding	X	X	None
Establishment of Cutoff	NA	X	EP24-A2

Guard-banding will help establish any limitations or warnings in the instructions for use for the clinical lab and should be completed prior to the start of the analytical testing.

Since there are no quantitative values reported for a qualitative reporting assay, linearity and normal range studies are not required. Likewise, for quantitative reporting assays these are needed but there is no establishment of a cutoff. Any medical decision point that may be associated with a quantitative reporting assay has usually been established by the medical community at large and not the manufacturer of the assay and should be defined in the design with reference to the literature or medical opinion that supports it.

In the case of studies that require large volumes of samples or levels of samples (e.g. LoD and within lab precision), pooling of samples at the same level can be done. When pooling, care should be taken to minimize any matrix effect (example, plasma should be

pooled with plasma). A clear justification of any pooling strategy should be part of the technical file.

The next section of this document will discuss each of the analytical studies in more detail.

To understand the performance of a test in routine use consider the potential variables including

- The expected use case, how many data points (samples or replicates of a sample) can be done in a run or over runs in a day?
- The combination of an instrument, reagents and software is a configuration and any difference in any of these is not the same assay system or can be expected to have the same performance. (Different models of an instrument, or versions of software for example.)
- Lot to lot variability in reagents (made in-house or purchased)
- Time, studies should not all be run on the same day

Samples/levels from the intended use population, or where samples may be difficult to acquire "mock" or "contrived" samples may be used in some cases. It will be important to evaluate the contrived samples to ensure that they are not more robust than patient samples.

Precision: Repeatability and Reproducibility (within lab)

The objective of this section is to discuss how to establish repeatability and within lab precision within a single site through the evaluation of sources of variability (or factors) within the assay (instruments, operators, reagent lots, runs, days, replicates, or anything else the developer designates).

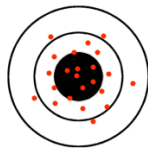Precision and trueness should be understood and appropriate for the use of the test,
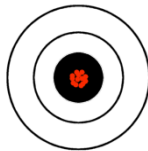
Precise and true
The archer shoots precisely (always around the same point) and true (in the centre of the black target area)

Precise but untrue
This archer shoots precisely (always around the same point) but untrue (always to the lower right of centre)

True but imprecise
This archer shoots true (the hits are located all around the black target area) but imprecisely (the hits are considerably scattered)
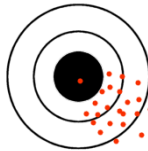
Untrue and imprecise
The archer shoots both untrue (the hits are next to and below the black target area) and imprecisely (the hits are considerably scattered)

taken from: http://www.med4you.at/laborbefunde/allgemeines/lbef_qualitaet.htm#Pr

For an assay that will be performed only at a single site, the within lab study is characterizing the variability within that lab. Therefore, the sources of variability should include the assay instruments (including backups) in the lab, as well as the operators that will be processing samples. If there are different sample types (whole blood vs plasma), then each sample type will require a separate evaluation (it can be the same study protocol or different ones).

Sample pre-treatment steps should also be considered as they often give rise to variability in the final assay.

Study Design Consideration

For Either Quantitative or Qualitative

To design the study adequately we need to understand what can be done in a day (for instance, some systems cannot run more than one lot of reagents in a day). This will define what factors can be cross (done each day) and what will be nested (done on alternate days). This will have an impact on the number of days and number of replicates per day per level (per analyte) that will need to be evaluated.

Samples and Levels

When designing a within lab study the following should be considered.

Quantitative Reporting

- Three levels may suffice for an assay with a limited reportable range and only one medical decision point or with consistent precision across all or most of the reportable range.
- Five or more levels will be needed to accommodate an assay with a wide reportable range or multiple medical decision points or important medical levels near the lower limit of quantitation (LoQ). When considering these levels, the interfaces of the normal range data may be important (high and low).
- For assays intended for monitoring, concentration levels should be chosen carefully to reflect the intended purpose. The consistency of results close to the medical decision point will be important.

Qualitative Reporting

- The study will typically utilize 4-5 levels to evaluate the analyte that are around the cutoff of the assay.

  Level 1: Positive sample (pos result 100% of the time)

  Level 2: Positive sample (pos result 95% of the time)

  Level 3: Positive sample at cutoff (pos result 50% of the time)

  Level 4: Negative sample (neg result 20 to 80% of the time)

  Level 5: Negative sample (neg result 100% of the time)

- Whilst tests may report a positive or negative result it is important to understand the performance around the cut off.

Number of Observations

Sample size for a within lab study is determined by calculating the degrees of freedom (DF) of the study design itself, that is within a single day, how many instruments, reagent lots and replicates can each operator perform on each sample level that needs to be evaluated? To evaluate the variability between the sources, each configuration needs to be done twice (for each sample level) and each level needs to have at least two replicate observations (2) within a configuration. There is no one design but there are typical designs that can be followed (see EP 05 from CLSI) and working with a statistician is always recommended.

For the study to be adequate the "replicate" DF needs to be ≥ 40.

To determine the total number of observations, multiply the sources of variability per level and then by all levels to be evaluated.

For example

2 instruments, 2 operators, 3 reagent lots, repeating each configuration 3 times (days) with replicates at each configuration is 2x2x3x3x2 = 72 per level and 360 across 5 levels in total.

What to include in the technical documentation

Quantitative

A summary table with each of the random factors as well as the total variability is generated at each level of analyte.

Qualitative

The same table is determined for the data from the internal continuous response (ICR). This data is used to determine the C5 and C95 values. ICR data will also be used to determine the call rates at each level of analyte tested and summarized in a roll up table.

Limit of Blank (LoB), Limit of Detection (LoD) and Limit of Quantitation (LoQ)

It is important to understand the LoB and LoD on order to understand the capability of the assay. The aim is to establish the system noise (limit of blank, LoB) and using this, establish the level of the analyte that will consistently be above the LoB 95% of the time. This is the limit of detection or LoD.

The LoQ is the level of the analyte that will meet an allowed level of total error (TE) and is not related to either the LoB or the LoD.

The LoB is established by running blank samples (no analyte present).

The LoD is established by running low levels of the analyte and using this data to assess the variability of these levels to determine the analyte concentration at which repeated measurements would result in 95% of the distribution being above the LoB and 5% being below the LoB.

There are several different approaches to determining LoD and they are not related to whether the assay is qualitative or quantitative, but rather the nature of the quantitative data or the ICR data (in the case of the qualitative reporting assays).

If the data are from a statistical perspective, "well behaved" (normally distributed) and a standard deviation can be determined for all levels then LoB and LoD are determined using the underlying data (classical or precision profile).
If the data is not well behaved, then either a probit or an empirical approach may be used. Probit considers probability and unit. A probit model is a type of regression where the dependent variable can take only two values.

Levels and Sample Size used in LoB studies

Blank samples using the classical approach for LoB: Run 4 different blank samples with a total of 60 observations (15 per sample) for each of 2 reagent lots (total of 120 observations) over several days. The blank data from each lot is evaluated to determine the reported concentration at the 95th percentile of the data distribution (each lot separately), this is the LoB.

Blank sample using the probit approach for the LoB: Run 30 different blank samples with a total of 60 observations (2 per sample) for each of 2 reagent lots (total of 120 observations) over several days. The data from each lot is evaluated to determine the proportion of observations that are greater than zero. If this meets an allow value (typically <5%) then the LoB is determined to be zero.

Low level samples with the classical approach: Run 4 levels of low analyte (two above and two below the presumptive LoD) with 60 observations across all levels (5/level) per reagent lot with two reagents lots used over 3 days (total of 120 observations). This approach assumes that the variability between the levels is the same and the data is pooled to determine the LoD.

Low level samples with the precision profile approach: Run 5 levels of low analyte (above and below the presumptive LoD) with 40 observations per level per reagent lot with two reagent lots used over 3 days (total of 400 observations). This approach assumes that the variability between the levels is different, and a precision profile is used to determine the LoD.

Low level samples with probit approach: Run 2 sets of 5 levels of low analyte with 20 replicates per level per reagent lot with two reagent lots over 3 days (400 total observations). This approach uses the LoB to determine hit rates for each level and this is used in a probit analysis to determine the LoD.

Low level samples to determine the LoQ. This is different from the LoD and is only done for quantitative assays. Here the presumptive LoQ level is generated in 4 separate samples of the same level. Over 3 days each level is run with 3 replicates per reagent lot with 2 lot, resulting in a total of 36 observations across the levels per lot and a total of 72 observations for both lots. The data is analyzed to determine the LoQ.

What to include in the technical documentation

The LoB and LoD are determined for either quantitative or qualitative reporting assays. The LoQ is reported for only qualitative assays.

Interfering Substances

Interfering substances should be considered for every assay and the reasons for including or excluding them from the study should be documented. It is important to evaluate the assay performance in the presence of exogenous and endogenous substances. These substances may interfere with either sample preparation or other system interactions within the sample resulting in a modified test result.

The interference screening study is typically done by adding an interferent to a sample or sample pool (test samples) and evaluating observed results as compared to a control portion of the same sample or sample pool without a spiked interferent (control samples).

If the observed (averaged) difference between the paired test and control sample results is less than a predetermined allowable shift, no further testing is performed. The allowed shift will depend on the assay but should represent a difference that is not biologically relevant to the assay that would change the clinical significance.

If the allowed shift is exceeded, then testing is repeated until the largest level of the interferent that is less than the allowed shift is determined, and the limitation is noted in the assay labeling. Some interfering substances cannot be avoided, the labelling provides warning to the end user:

- Physician: to help prevent samples that cannot be processed from being submitted
- Lab staff to help prevent sample processing of samples that could report an incorrect result
- Lab director when reviewing results

Likewise with qualitative reporting assays the call rate of the control group is compared to the call rate of the test group and an allowed difference in call rate is assessed. It is possible to take the ICR data and look at the result quantitatively, but most regulatory agencies will ask for the call rates for these assays as well.

Levels and Samples

The study is carried out in the same manner for quantitative and qualitative assays.

For quantitative reporting assays several levels across the reportable range should be assess as well as around any medical decision point that may be established.

For qualitative reporting assays if the interferent is expected to reduce the observed level of the analyte, then the level corresponding to the C99 is typically used (this level is expected to be above the assay cutoff 99% of the time so it reduces errant data due to variability of the assay and not the interferent). If the interferent were to increase the observed analyte rather than reduce it, potentially misreporting "negative" samples then the C1 level (this level is expected to be above the cutoff 1% of the time) would be used.

Sample Size to use in interfering substance studies

The number of replicates for the control and test groups depends on several factors:

- The amount of difference the study is attempting to detect
- The confidence we want to have in detecting that difference
- The amount of variability that we have in measuring that difference

The smaller the difference the study is trying to detect and/or the larger the variability is at the level that evaluated the larger the number of replicates that are needed to conduct the assay. A statistician can help determine this.

Another consideration is whether to test one or two levels of interferent. This judgement should consider the confidence that there will be no or minimal interference. This may be based on literature, experience, design work etc. if no issue is anticipated then the then the highest level only can be evaluated. If this fails then more testing will be needed, so this does not affect the sample size of any one interferent but does affect the total estimation of materials that could be needed to carry out the study.

What to include in the technical documentation

If the reported result from the assay is quantitative, then determining the difference between the test and control groups will be done and this must meet the allowed shift. If the reported result from the assay is qualitative (with an ICR) then the percent correct call will be determined and must meet a pre-defined amount. A secondary evaluation can be done on the ICR to determine if an allowed shift has been met.

Linearity

Linearity is the ability (within a given interval) to provide results that are directly proportional to the concentration (or amount) of the analyte in the test sample. These studies are only done for quantitative reporting assays. An assay is linear if the observed results "on average" are proportional to the analytes' true quantity values, meaning that the observed results agree with the true value given a constant multiplicative factor. It defines the "linearity interval" where the results meet the allowable deviation from linearity.

Linearity is not directly related to clinical truth as "accuracy" would be. The evaluation of linearity is in terms of an allowed deviation from linearity or ADL. Linearity can have an impact on accuracy as it does influence the bias in the assay. As such, linearity is evaluated to ensure that any assay bias that could affect accuracy is acceptable.

In a linearity study, a series of levels with relative concentrations of the analyte are prepared. Each level is measured several times, and the average value of these replicate observations is calculated. The response (mean observed values) generated by the assay (y-axis) is plotted against the expected concentration of each level (x-axis), and a linear regression analysis is performed on the data. The slope and intercept of the regression line are used to calculate the concentration of the analyte at each level (predicated values) in the assay. Acceptable linearity is demonstrated when the deviation from the fitted line is within the ADL at each level.

Samples, Levels and Sample Size

When multiple sample types (e.g., serum, plasma, whole blood) are used in an assay, it is recommended that a linearity study be conducted for each sample type, when possible.

The number of concentrations needed for a linearity study can vary depending on the range to be evaluated. In general, a minimum of nine levels are recommended for a linearity study. However, when the linearity interval spans a wide range of concentrations, more than nine levels may be required or when the linearity interval spans a narrow range, fewer than nine levels may be sufficient.

The concentrations should cover the reported range of the assay. At least one concentration should be below the lower limit of quantification (LoQ), and one level above the upper limit of quantification, because the mean results from the high and low samples constitute the upper and lower bound for a claimed interval. It is recommended that the concentrations be evenly spaced across the evaluated range. For assays with a range that is several orders of magnitude (perhaps requiring log transformation of the data), it is highly recommended the levels are equally spaced.

The sample size needed for a linearity study is determined based on the information from the assay's precision profile. The number of replicates for each level is determined as follows:

- The variability of the assay (i.e. within-run precision) at a specific level
- Allowable deviation from linearity (ADL)
- Probability of detecting a value less than or equal to the ADL (95% or greater)

Check CLSI EP-06 for calculations and Appendix D in the document for more information on determining the sample size per level.

Sample levels can be done either by dilution of the highest level with a blank sample (no analyte) or by mixing a high sample with a low sample.

The linearity study only needs to be conducted at one site with a minimum study design that includes one instrument, operator and a single lot of each reagent, controls and calibrators (if required). However, two or more lots of reagents/calibrators are recommended.

The acceptance criteria for linearity is the allowable deviation from linearity (ADL). The ADL is the maximum deviation that is acceptable for an assay to be considered linear within a given range of concentrations.

The following points should be considered when setting up the ADL:

- Define the intended use of the assay and the range of concentrations.

- Consider the clinical or biological significance: the ADL should be based on the assay's clinical use and the risk to the subject.
- Deviation from linearity represents a component of systematic error. Therefore, the ADL should be only a fraction of the allowable bias.

What to include in the technical documentation

A justification on the levels that were determined for the study design should be generated in the study report. The report should also summarize the data from each level in a table with information as outlined below.

Level	Replicates	Relative Concentration (percentage)	Observed Value	Expected Value	Predicted Value	Deviation	Percent Deviation	Within ADL
1		100		120				
2		90		108				
n								

Observed value is what was reported by the assay under evaluation

Expected value is the percentage of the starting concentration for a particular level

Predicted value is the value based on the fit of the model to the data.

Deviation is the difference between the observed value and the predicted value

The report should indicate that the ADL for each level is withing the acceptance criteria for the study.

Normal Range

Normal range data is only evaluated for quantitative reporting assays.

The normal range should be established for the intended population. To do this evaluate healthy donors in the intended use populations in order to determine reference intervals that meet the minimum requirements for reliability and usefulness for the parameters reported by the assay. The objective is to determine if within the population of "healthy" donors there are any differences between individuals that are not related to health issues, that should be communicated to the medical community using the assay that may impact the interpretation of the result.

This study should be representative of the healthy population but should also be representative of the population based on partitions such as, gender, age, ethnic group. Other groups to consider could be pregnant women vs non-pregnant women for example but would depend on the assay. A specific number of individuals need to be evaluated within each group.

Samples and Grouping and Sample Sizes

Each sample represents an observation within a partition that the developer needs to evaluate in order to establish the required reference ranges for the assay. The table below outlines the relationship between samples, partitions and groupings within a partition. Each group within a partition requires 120 observations (or samples). Depending on the intended purpose of the assay there could be other meta data to consider.

Meta Data	Groupings	Partitions	Total observations
Gender	2	Males and Females	240
Age	3	< 18, 18-50, >50 years	360
Ethnicity	4	White, African, Asian, Hispanic	480

What to include in the technical documentation

The range of results for each partition is reported. If there are no differences between the different partitions, then this should be noted as well.

Stability

Stability studies are done to determine whether time or environmental condition (e.g. freeze thaw cycles) or both causes the material under study to degrade, changing the performance of the material either in terms of bias or variability.

Data is collected either over a number of time points or through a set of environmental changes (freeze/thaw cycles) or both. The observed values from the samples run in the study are plotted versus the points the data were generated. The mean of the observed replicates generated at each time point are plotted (y-axis) versus the time/environmental change at which the observations were made (x-axis).

A linear regression is fitted to the data, and it is determined whether the slope of this regression is significantly different from zero ($p < 0.05$). If it's not, then the last time point in the study plus one (Tn+1) is the extent of the stability. If the slope is significantly different from zero, then the stability duration is determined using the allowable drift along with a one sided 95% confidence interval of the regression line. When the drift value on the y-axis intersects the 95%CI of the fitted data (at this point we read down to determine) the corresponding time point for that stability evaluation. CLSI EP-06 has examples of these plots and this analysis.

There are several different types of stability studies that need to be addressed.

- Sample: Determine the time from when the sample was taken to when the sample processing must start so that the level of analyte still reflects the clinical status of the patient.

- Intermediate products: Storage and time for any "intermediate products" (e.g. extracted DNA). The lab needs to define the storage conditions, the time, number of cycles (either frozen to thaw or refrigerated to room temperature) that are allowed. This is for any "product" that is in the middle of sample preparation to allow a break point for long preparation procedures.

- Reagent: Determine the expiration dating for the reagent. Testing is evaluated in terms of time from the lot manufacturing date to a point in time when the difference

between the observed result compared to the original is considered biologically relevant. Typically this is done by evaluation of three reagent lots when establishing this dating.

- In use Reagent: Studies done to determine the amount of time in which a reagent retains its performance characteristics after having been placed in use. These studies can evaluate time and temperature, or changes in environmental conditions, such as freeze/thaw cycles (or other temperature cycling) typically over a 30-day period. Typically this is done with one reagent lot when establishing this dating.

- Transport: These studies evaluate the product packaging during product distribution. Most of the time this is done through simulated transport studies, where the product is "stressed" by exposure to worst case transport conditions. Typically this is done for a single reagent lot when establishing this dating. For a single lab, the manufacturer of the component should have done this, if the material will not be shipped by the clinical lab to any other location, no further work need be done.

For labs where the reagents will be purchased from a vendor, the lab has the responsibility to determine if the established stability claims will be acceptable given the intended purpose of the assay designed in the lab. For any other reagents used by the lab where the manufacturing of that reagent or reagent component has been designed by the lab or an outside vendor has not established dating, the dating needs to be established.

Samples, Levels and Sample Size

For quantitative reporting assays several levels across the reportable range should be assess (considered interfaces of the normal range) as well as around any medical decision point that may be established.

For qualitative reporting assays the level corresponding to the C1 level (this level is expected to be below the cutoff 1% of the time) would be used to represent samples below the assay cutoff and the level corresponding to the C99 (this level is expected to be above the assay cutoff 99% of the time) would be used to represent samples above the assay cutoff.

The sample size or number of replicates per level per time point depends on several factors:

- The amount of difference the study is attempting to detect (needs to be biologically relevant not statistically significant).
- The amount of variability that we have in measuring that difference (what is the SD or CV for the level(s) that will be run in the study).
- The number of time points in the study
- The confidence we want to have in detecting the difference (95% confident or less?)

The smaller the difference the study is trying to detect and/or the larger the variability is at the level that evaluated, the larger the number of replicates that are needed to conduct the study. A statistician can help determine this.

What to include in the technical documentation

Each type of stability study is conducted and reported as a separate study.

The dating/storage conditions for sample and any intermediate product(s) become part of the instructions for use for the assay as does the in-use stability results. The dating from the reagent expiration study becomes part of the labeling for each lot of reagents (or components) that are made (vial as well as box labeling). This may be only 6 months at the time of launch but as the study goes to completion (12, 18 or 25 months or more) the reagent expiration dating can be increased once the reports are written and reviewed by the lab staff and signed off. For a single site lab there may not be the need for a reagent kit transport study unless the lab were to open a second site.

Clinical Evaluation Study

Typically this study is done by the "customer" for the product not by the developer. For a distributed kit this would be considered a "clinical trial" done outside of the developer area. For single lab developed assays this study should be conducted by the individuals that will actually run samples on a daily basis.

There are a few guidances that can be useful in setting up these studies:

Quantitative: CLSI EP9: Measurement Procedure Comparison and Bias Estimation Using Patient Samples

Qualitative: CLSI EP12: Evaluation of Qualitative Binary Output Examination Performance

The clinical evaluation study is done by processing samples from individuals that have either information on the clinical truth status of the subject or other test results (state of the art/reference assay that are designated as clinical truth ("SOA/RA").

For quantitative assays evaluation is about assessment of bias between the test assay and the reference. There is no way to convert a quantitative value into a binary result as there is no cutoff, so the acceptance criteria here would be allowable bias throughout the reportable range. This would involve running another assay that would measure the same analyte in the same intended use population. For quantitative assays where the medical community has established a medical decision point, the data from this study could be evaluated quantitatively (bias estimates throughout the reportable range) as well as qualitatively using the medical decision point(s) to generate a qualitative outcome.

For qualitative reporting assays the reported result is a binary outcome, and this can be compared to clinical truth for the subject or the outcome of a SOA/RA that is representative of clinical truth. In this case the study criteria are typically sensitivity, specificity, positive and negative predictive value. There are other parameters that may be used depending on the design of the assay (e.g. positive and negative likelihood ratios).

Samples are run according to the instructions for use that will be used when the assay is commercialized and with reagents, software, and instruments that are in final configurations for clinical use. Operators running the samples are blinded to any information about the subject (medical or demographic).

In both cases qualitative or quantitative reporting assays care is given in the sample selection to allow for representation of gender and age groups along with any other population attributes relevant to the intended use population.

Unless the test or the SOA/RA (either quantitative or qualitative) requires multiple replicates to determine the result, each sample is tested only once (this does not include repeat testing due to assay QC failures).

Samples and Sample Size

Quantitative

Samples from subjects equally distributed throughout the reportable ranges of the test and SOA/RA from the intended use population are run taking care to consider the respective normal ranges for each assay (with as similar distributions as possible) . For an evaluation of bias, typically ≥100 unique samples should be evaluated, consult with a statistician for larger than 100 samples.

No data may be removed from the data set once it's been determined that the results have passed assay QC for both assays. Any values that exceed allowable bias may be investigated and discussed in the development report. A second analysis may be done without the discordant data point(s).

Qualitative

If the test assay can be compared to clinical truth, the reported outcome of the test assay is compared to the outcome of the clinical truth (e.g. subjects with the test condition are "positive" and without the test condition are "negative"). Samples from subjects with and without the target condition (TC) from the intended use population are run with the test assay. Samples should be equally distributed throughout the underlying internal

continuous response (ICR) range for the test assay. Uing the comparison of the outcomes in both cases (i.e. sensitivity and specificity as well as positive and negative predictive value). Discordant observations (outcomes do not match) are either false negatives or false positives. No data may be removed from the data set if the assay QC has passed for the test assay. If there is a question on how the clinical determination was done this may be footnoted in the data discussion in the study report. A second analysis may be done without the discordant data point(s).

In the case where a SOA/RA is use as the basis of clinical truth then the samples run in the study should also be equally distributed throughout the underlying ICRs for the test and "reference" assays (allowing for samples that overlap between both as well as distributions in the intended use population). Sensitivity and specificity as well as positive and negative predictive value can be determined by comparing the respective reported outcomes. The SOA/RA is "clinical truth" in this case so disagreements with the test assay are either false negatives or false positives. If there are samples with discordant results, then these samples may be additionally tested by a "referee" assay if there is another way to further evaluate the sample (this should be set up in the study protocol). However, the original results remain, and the data tables are footnoted with the refereed results. Any other investigation of discordant results can be presented in the study report, but the primary analysis of all data cannot be altered but footnoted.

The sample size for qualitative reporting assay is linked to the acceptance criteria for the required sensitivity and specificity outcomes and having sufficient data to conduct the analysis. The acceptance criteria are written as the lower bound of a two-sided confidence interval (typically 95%). In addition, the developer needs to consider the actual performance of the test assay versus the SOA/RA since the more discordant observations the larger the sample size. In the case where the new assay is more advanced it may be expected that discordant observation will occur, the study design needs to account for this before the study is done as it can't justify the outcomes after the study has collected the data. Just as an example, the sample size for only positive samples within an assay where no discordant results were expected (very comparable to the SOA/RA) with an acceptance criteria where the lower bound of a 2 sided 95% CI is ≥

95%, would require 73 observations (no discordant samples allowed to pass the study), if 1 discordant expected then the sample size would need to increase to 109 prior to conducting the study.

The sample size for positive and negative sample evaluations do not require the same acceptance criteria. The criteria are dependent on the intended purpose of the assay (is it more important to reduce false positives or to reduce false negatives).

What to include in the technical documentation

The study protocols will include justification on the sample size and overall distribution of the samples required for the study as well as how discordant observations may be further evaluated. The study report will present the outcome of the analysis.

For quantitative reporting assays this would include the modeling of the data along with one or more tables summarizing the observed biases for the data points run in the study. Line listing of all of the data would also be part of the reported results.

For qualitative reporting assays, the outcome of the study would include 2 x 2 tables that summarize the comparison of the outcomes. These tables would also have the criteria results for sensitivity, specificity as well as positive/negative predictive value (along with any other accuracy analysis depending on the design of the assay).

For the either type of reporting assays, if there is more than one analyte evaluated by the assay then each analyte could be presented as a separate analysis, and presentation of a confusion table should be considered as well for assays that are multiplexing. A line listing of all the data would also be part of the reported results.

Annex (3) The Essential Principle checklists

Essential Principal	Applies to the device?	Standards applied	Method of Conformity & solutions applied.	Identity of Specific Documents
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				
16.				
17.				
18.				
19.				
20.				

Relevant documents: Requirements, Guidelines and Standards and (examples)

Requirements for Medical Devices Marketing Authorization (MDS-REQ1)	https://www.sfda.gov.sa/en/regulations/68759
Requirements for Clinical Trials of Medical Devices (MDS-REQ 2)	https://sfda.gov.sa/en/regulations/66129
Requirements on Importation and Shipments Clearance of Medical Devices and Supplies (MDS-REQ5)	https://sfda.gov.sa/sites/default/files/2023-07/MDS-REQ5E.pdf
Requirements for Medical Devices Establishments Licensing (MDS-REQ 9)	https://www.sfda.gov.sa/sites/default/files/2023-03/RequirementsLicensingMDEstablishments_0.pdf
Requirements for Inspections and Quality Management System for Medical Devices (MDS – REQ10)	https://sfda.gov.sa/en/regulations/87120
Requirements for Post-Market Surveillance of Medical Devices (MDS-REQ11)	https://sfda.gov.sa/en/regulations/87494
Requirements for Transporting and Storage of Medical Devices (MDS – REQ 12)	Requirements for Transporting and Storage of Medical Devices (MDS – REQ 12) | Saudi Food and Drug Authority (sfda.gov.sa)
Guidance on recognized standards	https://sfda.gov.sa/en/regulations/93287
Guidance for Points of Care (POC) Medical Devices Manufacturing (MDS-G009)	https://sfda.gov.sa/en/regulations/87669

Contact us

For more information regarding standards, requirements and guidelines, kindly contact

Products registration support section: md.rs@sfda.gov.sa

Important Links

SFDA Standards Web Store	https://mwasfah.sfda.gov.sa/
SFDA Requirements and Guidelines	https://sfda.gov.sa/en/regulations?tags=3
GCC Standardization Organization (GSO)	https://www.gso.org.sa/en/
International Organization for Standardization (ISO)	iso.org/home.html
International Electrotechnical Commission (IEC)	https://www.iec.ch/