WHO PREQUALIFICATION TEAM:
DIAGNOSTICS

World Health Organization

Technical Guidance Series (TGS) for WHO Prequalification – Diagnostic Assessment

Establishing stability of in vitro diagnostic medical devices

TGS–2

Annex 5

WHO Expert Committee on Biological Standardization - WHO Technical Report Series, 1011 - Sixty-eighth Report

WHO/BS/2017.2304 Rev. 1 (The original version of this document was published for public comment in 2017. This final version has been published as Annex 5 of the sixty-eighth report of the WHO Expert Committee on Biological Standardization (WHO Technical Report Series, No. 1011; http://apps.who.int/iris/handle/10665/272807))

Suggested citation. Technical Guidance Series for WHO Prequalification – Diagnostic Assessment: Establishing stability of in vitro diagnostic medical devices. Geneva: World Health Organization; 2018 Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see http://www.who.int/about/licensing.

Preface

WHO Prequalification – Diagnostic Assessment: Technical Guidance Series

WHO prequalification of IVDs

WHO prequalification is coordinated through the Department of Essential Medicines and Health Products. WHO prequalification of in vitro diagnostic medical devices (IVDs) is intended to promote and facilitate access to safe, appropriate and affordable IVDs of good quality in an equitable manner. The focus is on IVDs for priority diseases and on their suitability for use in resource-limited settings. WHO prequalification is based upon a comprehensive assessment of individual IVDs using a standardized procedure that is aligned with international best regulatory practice. It also involves post-qualification activities for IVDs to ensure their ongoing compliance with prequalification requirements.

Procurement of prequalified IVDs

Products that are prequalified by WHO are eligible for procurement by United Nations agencies. The products are then commonly purchased for use in low- and middle-income countries.

Prequalification requirements

IVDs prequalified by WHO are expected to be accurate, reliable and able to perform as intended for the lifetime of the IVD under conditions likely to be experienced by a typical user in resource-limited settings. Countries in which WHO-prequalified IVDs are procured often have minimal regulatory requirements, and the use of IVDs in these countries presents specific challenges. For example, IVDs are often used by health-care workers who do not have extensive training in laboratory techniques, in harsh environmental conditions, in the absence of extensive pre- and post-test quality assurance capacity, and for patients with a disease profile that differs from the profiles encountered in high-income countries. Therefore, the requirements of WHO prequalification may differ from the requirements of high-income countries, or those of the regulatory authority in the country of manufacture.

About the Technical Guidance Series

The Technical Guidance Series (TGS) was developed following a WHO working group consultation held on 10–13 March 2015 in Geneva, Switzerland. The consultation was attended by experts from national regulatory authorities, national reference laboratories, and WHO prequalification dossier reviewers and inspectors. The guidance series is a result of the efforts of this and other international working groups.

Audience and scope

This guidance is intended for manufacturers interested in WHO prequalification of their IVD. It applies in principle to all IVDs that are eligible for WHO prequalification for use in WHO Member States. This guidance should be read in conjunction with relevant international and national standards and guidance.

The TGS documents are freely available on the WHO web site.

Acknowledgements

The document *Technical Guidance Series for WHO Prequalification – Diagnostic Assessment: Establishing stability of in vitro diagnostic medical devices* was developed with support from the Bill & Melinda Gates Foundation and Unitaid. The document was prepared in collaboration with R. J. S. Duncan, London, the United Kingdom, S. Best, S. Braniff and M. Lanigan, National Serology Reference Laboratory, Victoria, Australia; and D Healy and R. Meurant, WHO, Geneva, Switzerland, with input and expertise from S. Hojvat, Maryland, United States of America (USA); L. Kestens, Institute of Tropical Medicine, Antwerp, Belgium; and D. Lepine, Medical Devices Bureau, Health Canada, Ottawa, Canada. This document was produced under the coordination and supervision of R. Meurant and I. Prat, Prequalification Team, WHO, Geneva, Switzerland.

List of contributors

First-round public comments were received from the following:

J. C. Badciong, Abbott Laboratories, Chicago, the USA; K. De Vore, Bio-Rad Laboratories, France; A. Halim, Celldex Therapeutics, Hampton, New Jersey, the USA; S. Hojvat, Maryland, the USA; L. Kestens, Institute of Tropical Medicine, Antwerp, Belgium; D. Lepine, Medical Devices Bureau, Health Canada, Ottawa, Canada; L. Ochs, Clinical and Laboratory Standards Institute (CLSI), Wayne, Pennsylvania, the USA and members of the CLSI Consensus Committee, ISO T212 WG3 committee; G. Pennello, United States Food and Drug Administration, Silver Spring, Maryland, the USA; J. Pierson-Perry, Siemens Healthcare Diagnostics, Erlangen, Germany; E. Russek-Cohen, United States Food and Drug Administration, Silver Spring, Maryland, the USA; M. Stevens Hardy, Medical Laboratory & Technology Consultants, LLC, Washington, DC, the USA; C. Zang, National Institutes for Food and Drug Control, Beijing, China; and the Japanese Committee for Clinical Laboratory Standards , Tokyo, Japan.

The draft technical guidance document was posted on the WHO Prequalification website for public consultation on 14 December 2015. Various stakeholders, including manufacturers submitting to WHO prequalification of IVDs, IVD manufacturing industry associations, various national and international regulatory bodies, and IVD standards organizations, were informed of the consultation in order to solicit feedback. A two-month response period was provided.

Second-round public comments were received from the following:A. Asahina, Abbott (formerly Alere Medical Co., Ltd.), Chiba, Japan; J. Budd, Beckman Coulter Inc., Chaska, the USA; C. Candia Ibarra, Ministerio de Salud Pública y Bienestar Social, Asunción, Paraguay; N. A. Carrington, Roche Diagnostics, Indianapolis, the USA; M. Dreher, mdc medical device certification GmbH, Stuttgart, Germany; I. Fijalkowska, United States Food and Drug Administration, Silver Spring, Maryland, the USA; J. Goss, Sysmex Partec GmbH, Goerlitz, Germany; C. Hill,

Encinitas, California, the USA; L. Kestens, Institute of Tropical Medicine, Antwerp, Belgium; M. Kondratovich, United States Food and Drug Administration, Silver Spring, Maryland,, the USA; M. Leportier, Beckman Coulter, Marseille, France; K. Máté, European Diagnostic Manufacturers Association, Brussels, Belgium; F Nyberg, Asia Pacific Medical Technology Association, Singapore; S. Ortigoza, Ministerio de Salud Pública y Bienestar Social, Asunciòn, Paraguay; G. P. Payne, BD Diagnostics Point of Care, San Diego, California, the USA; J. Pierson-Perry, Siemens Healthcare Diagnostics, Erlangen, Germany; L. Seixas, ALADDIV, Brasília, Brazil; W-W Tsai and P-W Tu, Asian Harmonisation Working Party TC WG2, China, Hong Kong SAR; Dr NT Wetherall, DAIDS/NIAID, Bethesda, Maryland, the USA; and L. Xu, Theranos, Inc., Palo Alto, California, the USA.

Following incorporation of the second-round public comments, a revised draft was published on the WHO Biologicals website for a final round of public consultation between 18 June and 18 September 2017. The comments received were incorporated to produce the document WHO/BS/2017.2304. The document was adopted by the WHO Expert Committee on Biological Standardization as a WHO written standard on 20 October 2017.

1 Abbreviations

ASTM	ASTM International
CE	Conformité Européenne (European Conformity)
CLSI	Clinical and Laboratory Standards Institute
EIA	enzyme immunoassay
HBsAg	hepatitis B surface antigen
HBV, HCV	hepatitis B virus, hepatitis C virus
IFU	instructions for use
IgG, IgM	immunoglobulin G, immunoglobulin M
ISO	International Organization for Standardization
IVD	in vitro diagnostic medical device
NAT	nucleic acid test
NIBSC	National Institute for Biological Standards and Control
NS3, NS4, NS5	HCV proteins
OD	optical density
PEI	Paul-Ehrlich-Institut
QA	quality assurance
QC	quality control
QMS	quality management system
RDT	rapid diagnostic test
RPM	revolutions per minute
R&D	research and development
SOP	standard operating procedure(s)
TGS	WHO Technical Guidance Series
TP	*Treponema pallidum*

2 Definitions

The definitions given below apply to the terms used in this document. They may have different meaning(s) in other contexts. Common English dictionary definitions apply to non-defined concepts, such as device, constituent, equipment, evaluation, part, product, reaction, signal, substance, etc.

Accelerated stability evaluation: Study designed to increase the rate of chemical and/or physical degradation, or change, of an IVD reagent by using stress environmental conditions to predict shelf-life.

Note: The design of an accelerated stability evaluation can include extreme conditions of temperature, humidity, light or vibration *(1)*.

Acceptance criteria: A defined set of conditions that must be met to establish the performance of a system *(2*, *3*).

Numerical limits, ranges or other suitable measures for acceptance of the results of analytical procedures*(2*, *3*).

Accuracy of measurement: Closeness of the agreement between the result of a measurement and a true value of the measurand.

Note 1: Accuracy of measurement is related to both trueness of measurement and precision of measurement.

Note 2: Accuracy cannot be given a numerical value in terms of the measurand, only descriptions such as “sufficient” or “insufficient” for a stated purpose *(4).*

Arrhenius plot: Mathematical function that describes the approximate relationship between the rate constant of a chemical reaction and the temperature and energy of activation *(2)*

Batch/Lot: Defined amount of material that is uniform in its properties and has been produced in one process or series of processes.

Note: The material can be either starting material, intermediate material or finished product *(5).*

Biocidal products: Active substances and preparations containing one or more active substances, put up in the form in which they are supplied to the user, intended to destroy, deter, render harmless, prevent the action of, or otherwise exert a controlling effect on any harmful organism by chemical or biological means *(6)*.

Characteristic: Distinguishing feature.

Note 1: A characteristic can be inherent or assigned.

Note 2: A characteristic can be qualitative or quantitative.

Note 3: Characterization: a description of the distinctive nature or features of something *(7)*.

Component: Part of a finished, packaged and labelled IVD medical device *(5)*.

Note: Typical kit components include antibody solutions, buffer solutions, calibrators and/or control materials *(5)*.

Constituent: Raw materials used to make a component.

Control material: Substance, material or article intended by its manufacturer to be used to verify the performance characteristics of an IVD medical device *(5, 8)*.

Design input: The physical and performance requirements of an IVD that are used as a basis for IVD design *(9)*.

Drift: Characteristic slow change of a metrological value from a measuring instrument *(910)*.

Environmental factors: Variables that might affect the performance or efficacy of IVD reagents – for example, temperature, airflow, humidity and light *(2).*

WHO note: For WHO purposes, this also includes altitude and microorganisms.

Evidence: Information which can be proved true based on facts obtained through observation, measurement, testing or other means *(modified from (7)*.

Independent lots: lots with different production (or manufacturing, purification, etc.) runs of critical reagents (for example, biological reagents prepared in different syntheses, growths or purifications or other risk-defined critical reagents from different manufactured lots or from different suppliers if applicable).

Instructions for use (IFU): Information supplied by the manufacturer to enable the safe and proper use of an IVD.

Note: Includes the directions supplied by the manufacturer for the use, maintenance, troubleshooting and disposal of an IVD, as well as warnings and precaution *(5)*.

WHO note: In order to avoid confusion, please note that, in the USA, the acronym IFU also stands for "Indications for use", and the acronym IU stands for "Intended use" or "Indications for use" (the acronym PI is often used in the USA to indicate the package insert, which may contain IFU). The International Organization for Standardization (ISO) definition and requirements *(5)* for IFU cover the intended use and the precise method of use and is the definition used by WHO and throughout this and other TGS documents.

In-use stability: Duration of time over which the performance of an IVD reagent within its expiration date remains within specified limits after opening of the container system supplied by the manufacturer and use under standard operation conditions (for example, storage on the instrument).

WHO note: For the purpose of this guidance document, WHO considers that it includes the number of times the reagents can be removed, used and returned to the storage condition without impact on test kit performance. It must reflect the routine conditions of use (for example, on-board stability, reconstitution and open-vial/bottle stability). A single product may have several different types of in-use stability claim, each reflecting different aspects of its usage. For example, an IVD reagent may have one in-use stability claim for unopened storage on board its associated instrument system and another stability claim once it is opened and put into active use. Another type of in-use life is the calibration interval of an IVD reagent *(12)*.

In vitro diagnostic (IVD) medical device: A medical device, whether used alone or in combination, intended by the manufacturer for the in vitro examination of specimens derived from the human body solely or principally to provide information for diagnostic, monitoring or compatibility purposes.

Note 1: IVD medical devices include reagents, calibrators, control materials, specimen receptacles, software, and related instruments or apparatus or other articles, and are used, for example, for the following test purposes: diagnosis, aid to diagnosis, screening, monitoring, predisposition, prognosis, prediction, and determination of physiological status.

Note 2: In some jurisdictions, certain IVDs may be covered by other regulations *(11)*.

IVD reagent: Chemical, biological or immunological components, solutions or preparations intended by the manufacturer to be used as an IVD *(5)*.

WHO note: This document uses the terms IVD and IVD reagent interchangeably.

Life-cycle: All phases in the life of a medical device, from the initial conception to final decommissioning and disposal *(12)*.

Metrological traceability: Property of the result of a measurement or the value of a standard whereby it can be related to stated references (usually national or international standards) through an unbroken chain of comparisons, all having stated uncertainties.

Note: Each comparison is affected by a (reference) measurement procedure defined in a calibration transfer protocol *(4)*.

Performance claim: Specification of a performance characteristic of an IVD as documented in the information supplied by the manufacturer.

Note: This can be based upon prospective performance studies, available performance data or studies published in the scientific literature *(5)*.

WHO note: "Information supplied by the manufacturer" includes but is not limited to: statements in the IFU, in the dossier supplied to WHO and/or regulatory authorities, in advertising and on the internet.

Referred to simply as "claim" or "claimed" in this document.

Precision: The closeness of agreement between independent test results obtained under stipulated conditions *(4)*.

Real-time stability evaluation: Study designed to establish or verify the shelf-life of the IVD reagent when exposed to the conditions specified by the manufacturer.

Note: Conditions that can affect the stability of an IVD reagent include temperature, transport conditions, vibration, light and humidity *(1)*.

Risk management: The systematic application of management policies, procedures and practices to the tasks of analysing, evaluating, controlling and monitoring risk *(12)*.

Risk-management plan: For the particular IVD being considered, the manufacturer shall establish and document a risk-management plan in accordance with the risk-management process *(12)*.

Shelf-life: Period of time until the expiry date, during which an IVD reagent, in its original packaging, maintains its stability under the storage conditions specified by the manufacturer.

Note: Stability and expiry date are related concepts*(5)*.

WHO note: In this document "Labelled life" is considered to be the time up to the expiry date printed on the label of an IVD or IVD component.

Stability: Ability of an IVD reagent to maintain its performance characteristics within the limits specified by the manufacturer.

Note 1: Stability applies to:

- IVD reagents, calibrators and controls, when stored, transported and used under the conditions specified by the manufacturer;
- reconstituted lyophilized materials, working solutions and materials removed from sealed containers, when prepared, used and stored according to the manufacturer's IFU;
- measuring instrument or measuring system after calibration.

Note 2: Stability of an IVD reagent or measuring system is normally quantified with respect to time:

- in terms of the duration of a time interval over which a metrological property changes by a stated amount;
- in terms of the change of a property over a stated time interval.

WHO note: because definition restricts IVD **reagent** only. Refer to *(1)* definition 3.10.

Stability monitoring: Real-time stability testing at certain points in time during shelf-life (or in-use life) to assure that an IVD reagent performs within specified claims *(2).*

Note: A continuing stability monitoring programme (ongoing stability monitoring) is required to verify that the stability claim is maintained over the life-cycle of the product. Data on stability must be obtained at end of shelf-life (see *(1);* section 4.1) and ideally at the halfway point of assigned shelf-life so that any problems that do occur can be dealt with in a timely fashion.

Trueness of measurement: Closeness of agreement between the average values obtained from a large series of results of measurements and a true value *(4)*.

Validation: Confirmation by examination and provision of objective evidence that the requirements for a specific intended use or application have been fulfilled *(7)*.

Verification: Confirmation by examination and provision of objective evidence that specified requirements have been fulfilled *(7, 13)*.

3 Introduction

3.1 Key concepts

Stability is the ability of an IVD reagent to maintain its performance characteristics over a defined time interval *(1*, 2*)*. The purpose of most stability studies is to establish or verify the time interval, and the storage conditions that can maintain stable IVD performance characteristics.

3.2 Rationale of stability studies

The stability of an IVD is fundamental to its reliable performance over a defined period of time. It is a regulatory requirement for the manufacturer to provide objective, scientifically sound evidence to support all claims made regarding the stability of an IVD. In addition, a manufacturer can use stability studies to demonstrate the probability that lots manufactured up to the end of the life-cycle of the IVD will meet predetermined user needs (as identified in design inputs).

3.3 Purpose of this document

The purpose of this document is to provide IVD manufacturers with guidance on possible approaches to determine stability. It also describes the expectations of WHO prequalification in relation to stability studies.

3.4 Standards

WHO recommends the following standards for use in establishing stability claims: International Organization for Standardization (ISO) 23640:2011 *(1)*; Clinical and Laboratory Standards Institute (CLSI) EP25-A *(2)* and ASTM International D4169 - 14 *(14).* It is recommended that manufacturers be familiar with these standards and consider them when designing and planning their stability studies. For other relevant standards see *TGS–1: Standards applicable to the WHO Prequalification of in vitro diagnostic medical devices*[1].

[1] Available at: http://www.who.int/diagnostics_laboratory/guidance/170808_tgs1_standards_2.0.pdf?ua=1

3.5 Limitations of this guidance

This guidance document should not be taken as a prescriptive checklist of the stability testing that must be performed, but as a guide on how to improve processes and generate the evidence needed to ensure a comprehensive and systematic procedure with an appropriate risk-management plan.

Depending on the particular categorization of the product and on the particular jurisdiction, additional regulatory and/or legal requirements, beyond the scope of this document, may apply.

The examples included throughout the document are not exhaustive and apply to the principles outlined in this document only. Manufacturers must still perform their own product-specific risk assessment for each of their IVDs, which may identify other critical characteristics (for example, physical measurements).

4 Considerations when applying for WHO prequalification

WHO requires that reports of studies used in establishing the stability claims for the product be submitted as part of the prequalification application.[1] As part of the WHO prequalification assessment, manufacturers must describe the rationale, the study methods, the stability monitoring programme followed and the testing algorithms used, with references to the relevant standard operating procedures (SOP). The information provided must demonstrate the link to the predetermined user requirements and product development.

The expectations of WHO prequalification may be different from the inputs of the users and from the requirements of the regulatory authority in the country of manufacture. In addition, the expectations set out in this guidance document may be additional to the requirements of ISO 23640 *(141)* and the expectations of CLSI EP25-A *(2).* Wherever possible, this guidance document explains the reasons for these additional expectations. Other approaches to meeting these additional expectations, supported by rigorous risk assessment or other evidence, may also be acceptable in dossiers submitted for WHO prequalification.

4.1 Manufacturer responsibility

It is a manufacturer's responsibility to ensure that the evidence supporting performance claims regarding the end of the IVD shelf-life is objective and scientifically rigorous.

[1] WHO documents PQDx_049 *Product dossier checklist* and PQDx_018 *Instructions for compilation of a product dossier* are available on the WHO Prequalification – Diagnostic Assessment website: http://www.who.int/diagnostics_laboratory/evaluations/en/

4.2 Suitability for use in WHO Member States

The stability studies submitted to WHO prequalification shall accurately reflect the expected environmental conditions and the normal usage conditions/methods encountered by users in WHO Member States, such as:

- extremes of temperature under in-use conditions and during transportation;
- extremes of humidity encountered under in-use conditions and during transportation and storage;
- any affects that light may have on IVD functionality, especially on the length of time for which a result is claimed to be stable; and
- presence of microorganisms.

4.3 Meeting customer requirements

By undertaking well-designed stability studies (including periodic verification activities) the manufacturer can demonstrate that the product meets input requirements (that is, customer requirements), as required by ISO 13485 (see *(15)* section 7.2: Customer-related processes). Meeting predetermined user expectations, not merely evaluating the capability of an IVD, is a fundamental aspect of IVD development (see *(9)* definition (f); and *(15)* section 7.3.4). It is a proactive means for the manufacturer to prevent quality problems at lot release and in the post-production and marketing phase.

5 Basic principles for stability testing

5.1 Critical characteristics or metrics of the IVD

A well-designed stability study must generate evidence of the stability of each of the critical constituents in the IVD (**risk-evaluated critical constituents**), evidence of stability for each of the claimed analytes, and evidence for any particular level of performance, including the precision, sensitivity and specificity of the kit. A documented risk-based approach should be taken to determine which claims and constituents must be evaluated over the stated shelf-life.

Examples:

1. *A hepatitis C virus (HCV) assay containing the critical constituents related to detection of NS3 or core proteins must have the stability of all such constituents proven for the shelf-life of the IVD.*
2. *For an assay designed to detect both immunoglobulin G (IgG) and immunoglobulin M (IgM) by use of protein A and protein L, the stability of both protein A and protein L must be proven in the IVD.*
3. *For an IVD to quantitate CD4, all the constituent antibodies used (for example, anti-CD3 and anti-CD4) must be shown to be stable in the IVD.*

4. *For an IVD claimed to detect particular seroconversion specimens or genotypes, or to have specified precision at particular analyte concentrations, or a particular specificity, each of these claims at risk or that change over time must be proven over the stated shelf-life (see TGS–4: Guidance on test method validation for in vitro diagnostic medical devices (16).*

Other critical characteristics (also called critical metrics) identified in the risk assessments may include physical measurement (for example, volume, pH, flow rate, legibility and adhesion). These characteristics must be shown to meet their specifications for the shelf-life of the IVD but are outside the scope of this document.

5.2 Finalized product presentation

During stability testing, all IVD components (including the IVD, calibrator and/or control material, etc.) must be made and tested to the finalized manufacturing specifications and in the finalized packaging, including intended labels and containers (see section **10.4**). In most circumstances, all presentations (for example, different buffer volumes used for different kit sizes) must be used during stability testing. Where some presentations are not tested, the manufacturer should document the rationale, justifying why all presentations have not been tested.

5.3 Environmental conditions

The stability study must subject the IVD to a combination of conditions that define, with predetermined confidence limits, the stability for lots marketed during the life-cycle of the IVD. The combination of conditions and durations of exposure and number of lots to be used will be driven by a manufacturer's risk assessment for the IVD and by research and development (R&D) data. The risk assessment should, at a minimum, take into account the following:

- the variability of the constituent materials (identifying the most important sources of variation);
- an understanding of the nature of user environments; and
- the extremes of conditions (temperature, humidity, ambient pressure and vibration) potentially occurring during transportation to those users (see also section **4.2**).

Boundary conditions for stability studies must reflect realistic extreme conditions that are consistent with the design input requirements for the IVD. The subsequent stability studies will prove the IVD capable of meeting performance requirements up to the end of its stated shelf-life, after transportation to the users.

5.4 Minimum number of lots

The design of stability studies must take into consideration lot-to-lot variability, with a risk assessment conducted to identify the most important sources of variability. The degree of variation of individual lots affects the confidence that a future production lot will remain within specification throughout its shelf-life. Lot variability is most often caused by minor differences in the biological reagents rather than by lack of reproducibility of the manufacturing process. Although existing standards *(1, 2)* recommend the use of a single lot for certain stability studies, the impact of lot-to-lot variability must be taken into consideration and the use of additional lots may be necessary. Three lots, at a **minimum**, must be used to establish or verify shelf-life; in-use claims require testing on a **minimum** of one lot. To ensure that the potential for lot-to-lot variability is addressed, independent lots must be used – that is, lots containing different batches of critical constituents such as nitrocellulose membranes, recombinant antigens, peptides, nucleic acids and the enzymes used in nucleic acid test-based (NAT-based) testing technologies.

Example:

For NAT-based testing technologies, it is crucial to use independent lots of enzyme for stability studies, as the manufacturing process can affect them. Other components (including primer, probe and buffer) can also be affected by the manufacturing process (for example, in terms of purity, pH, and DNase and RNase contamination). Thus for these other components, the use of independent lots that represent both material and process variability are also recommended.

5.5 Assessment of liquid components

The orientation of the product during storage (that is, upright versus inverted or horizontal) may need to be included in a protocol where contact of the product with the different parts of the container (such as the closure system or the body of the container) may be expected to affect the stability of the products contained (for example, liquid component). This is sometimes referred to as “inverted container stability”. The product orientation may need to be moved occasionally during the stability study to ensure that there is direct contact between the liquid contents and all parts of the container. This aspect requires particular attention during in-use stability studies of components that are diluted or reconstituted from a freeze-dried state before use.

5.6 Specimens for the stability testing panel[1]

The specimens used in the stability testing panel(s) must reflect the performance claims related to the IVD. The specimen types most likely to be used in those WHO Member States in which the IVD is intended to be used must be considered and, as appropriate, included in the specimen panels used throughout the stability studies (see Appendix 2). If a variety of specimen types (for example, serum, plasma, whole blood and saliva) are claimed as being suitable for use in the IFU, the stability study plan must be designed to provide evidence that the IVD will meet its claims (for example, for sensitivity, specificity, proportion of valid runs and precision) for each of the specimen types for the whole of the claimed shelf-life, including during transport to the final users, unless an alternative approach can be justified using a documented rationale. Evidence must be statistically valid (see section 11.5). Regulatory requirements may also dictate the addition of specified panel members.

5.7 Validation of stability testing panel

The stability testing panel(s) must be validated, and rejection and replacement criteria must be established. The validation of the panel members used is crucial. Panel members themselves must be stable and they must monitor parameters that are useful in controlling the characteristic being tested.

Storage of a validated panel for testing stability is not always feasible. For example, this is often the case for assays requiring fresh and/or whole blood specimens (for example, assays for counting CD4 cells). When replacing panel members, particularly for CD4 monitoring, the accuracy of results generated using the replacement material must be confirmed using an appropriate reference method (for example an instrument validated for use in an ISO 15189 *(17)* accredited laboratory). Replacement criteria for unstable panel members must include the duration for which a critical member will give valid results.

5.8 Panel member selection and value assignment criteria

Panel members are chosen specifically to ensure that each member has an attribute relevant to the intended use. The goal of stability testing is to ensure that the test method appropriately monitors functionality at the end of the assigned life (shelf-life or in-use life) of the antigens, epitopes and antibodies, along with any physical specifications relevant to the intended use.

[1] A panel is a collection of well-characterized specimens and other materials that are used to monitor aspects of IVD and component function during stability studies, for in-process control, for some aspects of design validation and at release to sale. The same materials might be used for each of these purposes but be assigned different acceptance criteria for the different functions.

For example, an intended use claim may be that early seroconversion specimens are detected. To show that this claim is true at the end of the product's shelf-life, a stability panel member representative of a very early seroconversion specimen could be included. This might be a weakly reactive IgM specimen, or some other specimen that has been shown to closely mimic the behaviour of the IVD with the critical specimens. Rare and valuable specimens would not be expected to be tested at all time points of stability studies. However, evidence must be provided that key performance claims made in the IFU, published material (including advertising) and dossiers submitted to WHO prequalification are met at the end of the assigned shelf-life and in-use life.

Each panel member is assigned an expected value and this is used to assign the acceptance criteria for that panel member. The expected value for each panel member is assigned in a measurable manner that is relevant to the outputs of the particular methodology. For example, the acceptance criteria for each panel member may be assigned in terms of sample-to-cut-off ratio, cycle time (CT) values or band intensity measured quantitatively/semi-quantitatively.

In the example of a weakly reactive IgM seroconversion specimen, the specimen at the start of shelf-life may have an RDT reading of 1+ out of 4 assigned as its expected value using a semi-quantitative value based on band intensity. The acceptance criteria assigned as a result may be that "all reactive specimens remain reactive, and all non-reactive specimens do not react in the assay".

Panel members must be chosen so that they will not only be relevant in demonstrating the intended use but will also have values that will appropriately detect, and therefore monitor, any deleterious effects of storage. A strong positive specimen that has a 4+ out of 4 semi-quantitative reading may continue to give this reading despite decay in the assay, whereas a specimen with a reading of 1+ out of 4 (with an assigned acceptance criteria of "remaining positive") is more likely to give an indication of the ongoing stability of the assay.

Thus it is essential to know (and document) that whenever a panel member meets the acceptance criteria, this is a true reflection of the stability of the product and not due to the inability of the specimen result output to reflect any change in the IVD.

5.9 Time points

A simple study design requires a minimum of three testing intervals *(2):*

1. an initial baseline test;
2. a test at the time point beyond the claimed stability limit (see section **5.9.1** below);
3. one point in between.

This simple study design is acceptable for submission to WHO prequalification under some circumstances and for some IVDs based on:

- the manufacturer's risk analysis;
- whether the manufacturer has prior-objective documented experience of the stability of the product; and
- whether the statistical confidence in the result is sufficiently great for all lots tested.

The benefits of a simple study design are that a small number of testing intervals and fewer resources are required. However, such a simple design represents a high-risk approach that has the potential to waste time and resources if the IVD does not meet the acceptance criteria with an appropriate margin of statistical confidence at the end of testing. If the acceptance criteria would have been met at another intermediate time point then that might have been acceptable as an assigned shelf-life.

A more effective and well-established approach routinely used is to test at a number of additional predetermined intermediate time point intervals (between 1 and 2 above). Typically, testing is carried out at relatively short intervals (every 10 or 14 days) for the first 3 months, and then at monthly intervals until at least one month beyond the design input-specified shelf-life. This protocol provides information on whether the IVD ages more rapidly in the period just after manufacture than later on in the shelf-life, and usually provides sufficient data to enable the assignment of a confidence interval to the shelf-life.

The manufacturer could identify the most practical intermediate test points from a risk evaluation of a specific IVD and include them in the stability study plan/protocol. Such planning will also help manufacturers to estimate the resources required to implement the testing.

Testing of all panel members is not expected at each of the test/time points. However, testing with all stability testing panel members is expected at the initial, the second to last and the last test/time point for all of the study types. The manufacturer should consider and document the rationale for the selection of intermediate test points, and choose panel members to be tested at these intermediate test points (for example, representative members, specimens that are close to the medical decision points and those at the extremes of the assay range tested).

5.9.1 Duration of testing

Testing conducted in stability studies should extend beyond the shelf-life determined from user needs. At a minimum, testing should extend at least one time point (one testing interval) beyond the predetermined user requirement to provide a margin for uncertainty. The length of the time periods chosen will depend on risk assessment, but should provide a safeguard in the event of unexpected IVD failure during the testing period,

where extrapolation from an earlier time point would not be considered acceptable.

It is recommended that the standard relevant units of measurement are used for the entire study (for example, unopened kit shelf-life is normally measured in months; opened IVD/reagent stability in days or weeks; and allowed reading times for enzyme immunoassay (EIA) and RDT in minutes or hours after performing the assay).

5.10 "Zero time" values and variance

The value of each measured characteristic at the beginning of the stability study and its variability over the course of the study are important pieces of information. They should be measured independently for each lot of material in the stability study. Analysis of the data will indicate if a statistically significant change has occurred to any measured parameter from any lot during the course of the study. A statistically significant change may not be of practical significance. Relevant practical limits will have been predetermined in IVD or process development. However, all statistically significant changes must be thoroughly evaluated to decide whether they represent some important change that would otherwise be undetected.

Zero time values could be obtained by evaluating each measured characteristic for each lot on five or more occasions to establish the value and its variance with freshly made materials. A definition of "occasion", following appropriate consideration, could be specified, for example, as involving a different day, a different operator and a different set of equipment in order to investigate potential sources of analytical variation. Later in the study, apparent differences in the values of the characteristics can be detected reliably, relative to the "zero time" value.

6 Shelf-life studies

6.1 Requirements for determination of shelf-life

The stated shelf-life of an IVD must normally be based on real-time experimental results. Accelerated stability studies are usually not sufficient to support a claimed shelf-life, although they may be used in situations where experience already exists with similar products (see *(1)* section 4.1) or when the stability of very similar products is already known (see *(2)* section 7.3.1).

Note: If at the time of dossier submission for WHO prequalification the real-time study outcome is not available, accelerated studies might be considered. The manufacturer must justify why the accelerated study is acceptable as supportive evidence until real-time experimental results become available. In these cases, the results of real-time stability studies

will be requested as a condition of WHO prequalification. The shelf-life of the IVD could be extended upon WHO review of real-time data.

6.1.1 Real-time stability studies

Real-time stability is determined using storage temperatures derived from user requirements, over a period longer than the required life of the IVD.

Where a broad range of storage temperature is claimed (for example, "Store at 4–40 °C") WHO expects the studies will provide evidence for stability over the whole of the temperature range for at least the length of the claimed shelf-life. However, where claimed stability is restricted to a limited range (for example, "Store at 2–8 °C") it is acceptable for stability studies to be conducted at a single temperature within this range.

It is recommended that a sequential approach be used *(2)* in which IVDs are first submitted to stresses simulating transport before they are placed into a shelf-life or in-use study. This approach best simulates the real-life situation, where products will first be transported to the end user and then stored under the recommended conditions before use, possibly until almost the end of their labelled shelf-life.

It may be routine practice to store IVDs for an extended period after manufacture before shipping. In this case, the IVDs would be kept first for a defined period of time under recommended storage conditions, then taken through the transport stress condition sequences, and finally put back into the recommended storage conditions for the duration of the study *(2).*

6.1.2 Accelerated stability studies

Accelerated stability studies are designed to predict the shelf-life of an IVD using increased rates of chemical and/or physical degradation caused by extreme environmental conditions (for example, elevated temperature at higher humidity).

Accelerated stability studies provide results in a relatively short time. However, the results of these studies are reached using assumptions about the degradation of reagents and other IVD components that may not reflect their observed performance under actual conditions of storage and use.

If the Arrhenius equation is used to calculate the expected life at temperatures other than those actually used then the parameters of the equation must be derived from the experimental data and not assumed *(2).* Manufacturers must ensure that there are sufficient data (for example, for different temperatures and test intervals) to allow for reliable extrapolation.

7 Component stability studies

7.1 General principles

7.1.1 Testing on final specifications

Component stability studies, including antimicrobial and desiccant studies, must be performed using components made according to finalized and approved manufacturing specifications (ideally to validated manufacturing scale) on qualified manufacturing equipment and meeting finalized and approved in-process quality control (QC) specifications.

7.1.2 Considering component stability

IVD components are sometimes prepared in bulk and stored before being used in several different lots of a completed IVD. The design-input documentation should define how long components are likely to be stored before use. With that information, component stability studies should be planned to provide evidence that component shelf-lives will not restrict IVD shelf-life, since **an IVD cannot have a shelf-life beyond that of any of its dependent components**.

The shelf-lives of components manufactured in bulk and used in several different lots of an IVD can be verified using three lots of the component as a minimum for shelf-life studies and, depending on documented risk assessment related to variability, one or more lots subsequent to changes made to the component. It is possible there will be two shelf-lives to evaluate: that of the bulk material stored prior to transferring to the final packaging and that of the component in its final packaging. The final contents of the evaluated lots of the component must differ with regard to the batches of critical constituents used (independent lots) but, subject to documented risk assessment, may all be tested in their final presentation with a single set of the other components that will be used together to constitute the IVD.

Examples of stored components:

Wash solutions and substrates for EIA, amplification reagents for NAT and calibrators for quantitative tests; all manufactured and stored in their final labelled vials ready to be put into a kit.

Component stability can be assessed from the functionality of the lot and also by factors related to the component that might change over time, such as turbidity, colour, microbial contamination and the pH of liquid components. Depending on the IVD and the conditions it is subjected to, it may be necessary to distinguish between turbidity that arises from heat/cold denaturation and turbidity that arises from microbial contamination.

7.1.3 Considering constituent stability

The stability study plan should consider whether components made from freshly made constituents (for example, antigens, recombinant antigens, enzymes, antibodies and membranes) will have the same shelf-lives as components made from stored raw materials. Evidence should be provided to support the use of stored constituents and detailing the lot-to-lot variability of critical constituents.

The stability study plan should also consider the choice of reagents or methods to ensure that the most appropriate are used to measure the performance of the component being studied (whether made from freshly made constituents or from constituents with an already proven shelf-life).

Examples of stored constituents:

Purified recombinant antigens and monoclonal antibodies stored in aliquots ready for use.

7.2 Stability of control materials

Assay-specific control materials provided by the manufacturer are used to show that an IVD has performed as intended during use. These are often referred to as "run controls" and are provided with some IVDs, along with an IFU statement that if the control meets a certain criterion then the IVD will have functioned as expected. "Control materials" does not refer to controls such as international calibrators or those used in external quality assurance (QA) programmes.

The manufacturer must be able to demonstrate that the loss of signal from control materials does not occur at a different rate from the loss of signal from a validated panel member or from genuine, critical specimens; otherwise a failing IVD might be regarded as still functional. Thus, the stability of control materials must accurately reflect the stability of the IVD. The use of a control material that is apparently more stable than the IVD and other components, or the use of incorrectly assigned values for the control material, must be avoided *(18).*

Example:

It is frequently seen in dossiers submitted for WHO prequalification that a positive run control will produce a signal of > 2.0 optical density (OD) in a freshly manufactured lot, and the IFU will state that an OD > 0.8 for the same control qualifies a run. Thus the IVD may have lost more than half its activity and still appear functional, even though some critical specimens are shown in the dossier to have very weak signals on freshly made IVDs. This is not considered appropriate unless data can be provided that demonstrate that the critical specimens will still be detected at the end of shelf-life and with a control material signal of 0.8 OD.

7.3 Biocidal stability and efficacy

7.3.1 Rationale

Bacterial and fungal organisms relevant to the environment of use must be identified in the design input risk assessment, and antimicrobial preservatives should be chosen, based on risk assessment, to prevent contamination of the product in storage and in use. Antimicrobial preservative effectiveness must be demonstrated throughout the shelf-life of the IVD.

If a new or modified preservative (for example, a different concentration) is used as a result of further information on the conditions of intended use, the manufacturer must obtain evidence that the new antimicrobial preservative or concentration chosen does not negatively affect the stability of the IVD.

7.3.2 Study conditions

The studies should reflect expected in-use conditions for opened containers – the stability of the IVD in the user environment, as intended by the manufacturer, must be proven. On-board stability must be tested for an IVD used with an instrument.

See *(18)* sections 51, 61 and 62; and *(19)* Appendix XI for suggested study methods. Examples of bacterial groups to consider are spore-forming bacteria, fungi, indigenous bacteria, bacteria found in the environment of the country of manufacture and those found in the countries of intended use. Specific examples outlined in references *(18)* and *(19)* include *Aspergillus niger*, *Bacillus subtilis*, *Candida albicans*, *Escherichia coli*, Salmonella species, *Pseudomonas aeruginosa*, *Clostridium sporogenes* and *Staphylococcus aureus*.

7.4 Desiccant functionality

Desiccants affect the stability of the entire IVD. Stability studies must show that the desiccant will support the product over the whole claimed shelf-life within the predetermined extremes of transport, storage and in-use conditions.

Note: For WHO prequalification purposes:

1. It is recommended that a self-indicator (a humidity indicator that changes colour upon saturation) be part of the desiccant design. However, WHO strongly recommends against the use of cobalt dichloride, the most commonly used humidity indicator, as it is a carcinogenic substance.

2. Sachets are preferred to tablets, since the labelling instruction “Do not eat” is more visible. There have been reports of desiccants in a tablet formulation being mistaken for antimalarial medicine.

8 Stability during transport

8.1 Rationale

Transport stability studies evaluate the tolerance of an IVD to the types of environmental conditions (for example, temperature and humidity) and physical conditions (for example, inversion, vibration, physical handling and stacking) to which it is likely to be subjected during and after shipping from the manufacturer to the end user. These studies should provide evidence that there will be no impact on IVD performance over the whole of its stated shelf-life as a result of the transportation of the IVD by the recommended methods.

The manufacturer should assess the potential impact of multiple factors and justify and document whether or not to include them in the evaluation. Final transport conditions recommended by the manufacturer should reflect (and the stability study plan document) the assessment of the conditions expected to be encountered in the areas of use. The manufacturer should address any issues that arise as a result of the transportation studies (for example, failing the stressed conditions), and address these limitations in the manufacturer documentation (for example, shipping documents and IFU if applicable).

WHO expects that a transportation challenge would precede the real-time determination of shelf-life, and in-use studies. This will serve to determine that transportation conditions do not reduce the shelf-life of the IVD (see section **6.1.1**).

In some cases it may be acceptable for the product to undergo transportation-stability studies without a subsequent long-term real-time stability study. In this case, shelf-life must be established under specified storage conditions along with a stringent and evidence-based risk assessment of the probabilities of extreme transport stress affecting IVD performance at the end of the claimed life (see *(2)* section 4.2.3).

8.2 Challenge conditions

Determination of the stability of an IVD during transportation should take into consideration the local routes and means of transport used to supply the IVD, which are usually defined in the design input risk assessment. It is not necessary to test the IVD to the point where it is no longer usable, but merely to validate the window of transport conditions within which the IVD will retain its claimed performance to the end of its stated shelf-life. However, knowledge of the possible limitations of an IVD and at what point the IVD becomes unusable is useful to a manufacturer when trouble-shooting post-marketing problems. WHO expects the manufacturer to take into consideration the possibility that the product might continue to be subjected to suboptimal storage conditions by the end user.

Example:

A static challenge of 45 °C for 3 days may represent conditions seen during the actual transportation of an IVD – however, a more stringent challenge of cyclical high and low temperatures (including freezing) for a longer period of time, and followed or preceded by exposure to vibration might better cover a "worst-case scenario" of shipment, storage and subsequent transportation to the end user.

8.3 Number of lots

Where transport stability studies are incorporated into studies to establish shelf-life, as recommended in this guidance document, a minimum of three lots of the IVD must be used. For transport studies alone, a minimum of one lot of the IVD may be used, however, as with shelf-life studies, more lots may be required depending on lot-to-lot variability (see section **10.1**).

8.4 Simulated versus actual challenge

An actual shipping challenge can be used to verify the conditions found in the simulated transportation challenges. However, it may only replace a simulated shipping challenge where there is an appropriate risk evaluation and where experience and data have been actively collected for similar products and documented in detail (for example, it is not sufficient to note "no complaints").

In the R&D phase, actual data from shipping can be used to define the conditions needed for an appropriate simulation of extremes. However, in the post-production phase, actual shipping challenges often do not explore the full range of shipping conditions that could be encountered, including extreme values.

8.5 Multiple stress test sequences (simulated transport challenges)

Proof of IVD performance after actual shipment is generally not sufficient evidence of stability under all conditions and delay hazards. Multiple stress test sequences are typically needed to address the range of transport conditions used for global product delivery. Relevant guidance *(14)* recommends the evaluation of several extreme conditions.

Appropriate stress test sequences may be developed on the basis of data from actual product transport studies. Testing multiple stress sequences allows a manufacturer to identify the most cost- and/or resource-effective transport conditions from a set of alternatives, while ensuring adequate product stability protection (see *(2)* section 4.2.3).

Note: For WHO prequalification, the environmental conditions investigated as part of a stability study must reflect those likely to be encountered in resource-limited WHO Member States. For example, temperatures at some airport tarmacs in sub-Saharan Africa can exceed 40 °C, while temperatures encountered during air transport fall below 0 °C. Significant delays can be

encountered at any time and especially during wet season transport to remote health centres.

See Appendix 1 for an example of a protocol for simulated transport challenges.

8.6 Physical conditions

Physical handling can be both manual and mechanical. The relevant user and commercial factors should be identified as part of the design input risk assessment and the packaging and shipping methods developed accordingly. Reference *(14)* describes a number of factors to be considered, and their evaluation: drop, impact, compression, vibration, repetitive shock, longitudinal shock, cyclic exposure, vacuum, impact and inversion; along with the size, weight and composition of the packaging. This should be regarded as part of stability testing.

9 In-use stability studies

9.1 Rationale

In-use stability of an IVD is the period of time over which components retain adequate performance, after transport to the users, once they are opened, reconstituted and/or diluted and exposed to the environmental conditions in which they will be used.

As far as possible, the study should be designed to simulate the use of the product in practice. If a range of conditions for use is stated in the IFU (for example, “use at 15–40 °C”) evidence must be provided to prove the stability over that range with all the specimen types claimed (for example, serum, whole blood and oral fluid), unless a documented rationale is provided. It is considered best practice for the manufacturer to claim a stability range that includes an appropriate safety margin (for example, test range 2–35 °C, claimed 4–30 °C) to ensure that that the claimed stability range is acceptable. However, where claimed in-use stability is restricted to a limited range (for example, “use at 35–37 °C”) it is acceptable for in-use stability studies to be conducted at a single temperature within this range, subject to evidence from documented robustness studies or risk assessments.

It is good practice to perform the in-use stability testing at both the start and end of the shelf-life of the IVD (or with components at the start and end of their shelf-lives if any of the components have a longer shelf-life than the complete IVD) and after simulated transport challenge (see section **8**). This will confirm that the IVD will have the claimed in-use life throughout its whole shelf-life.

All studies should support precisely defined periods of in-use stability claims.

Example:

An RDT test cassette may be labelled "Use immediately on opening". However, it is still necessary to determine the interval (one hour, one day, etc.) over which IVD performance remains stable after the component is opened.

9.2 Conditions of use

Determination of the in-use stability of an IVD and/or its components must reflect the routine conditions of use of the IVD. Freeze-thaw stability should be considered to address situations in which reagents may be exposed to multiple freeze-thaw cycles during use.

Note: For WHO prequalification, in-use stability studies should take into account the environmental conditions and usage conditions encountered in WHO Member States and by users, such as exposure to extreme temperature, humidity and light and to microorganisms.

9.3 Multiple in-use stability claims

Depending on the way in which the IVD is used it may be necessary to have several in-use stability claims. In situations where multiple stability claims are made, a manufacturer must provide evidence (from testing that investigates routine use) supporting each of the claims.

Examples:

1. *A reagent may have a stated period of stability once it has been placed on board an instrument and another period of stability once it is in active use (that is, during actual use/testing).*
2. *Multiple-use reagents (for example, buffers) may repeatedly be exposed to high temperatures during the day while in use and exposed to lower temperatures when not in use and stored in the refrigerator. The actual use of the multiple-use reagent – squeezing of bottles, exposure of the lid and tip to working surfaces and hands, and exposure to dust and light – may also affect stability. Stability studies and associated risk assessments should take all of these factors into account.*

10 Production lots used in stability studies

10.1 Considering variability

As noted in section **12.3** below, planning for stability studies must take into consideration all possible sources of variation within and between manufactured lots. For most IVDs it is likely that differences between batches of the biological reagents will cause the most variation. Factors to consider include apparently minor and technically uncontrollable differences in the culture and purification of recombinant antigens and antibodies; synthesis and purification of primers, probes and peptides;

undocumented production changes of an outsourced buffer component; and lot variability of nitrocellulose membrane used in lateral-flow IVDs.

At a minimum, lots chosen for stability studies must be independent lots – that is, they must differ in the source lot of their critical constituents, for example, different purification and/or culture batches for all recombinant antigens and monoclonal antibodies. If pilot or small-scale lots are chosen, special attention must be paid to the potential for variability (see also section **12.3**). However, the sources of variation will depend on the particular process, product and component, and should be identified during product development risk analyses.

Use of different batches of critical components ensures that the stability evidence obtained is more likely to be representative of long-term manufacture. Any variability found can be taken into consideration when assessing the outcome of the studies against the design input requirements and when making claims. This minimizes user problems and hence complaints.

10.2 Testing the final configuration

Shelf-life, in-use and transport stability must be determined for the finalized approved product in terms of:

- manufacturing specifications
- release-to-market QA criteria
- packaging and labelling (see section **10.4**)
- validated manufacturing scale on qualified manufacturing equipment.

Note 1: For WHO prequalification, it is important that the stability studies have been conducted using the IVD intended to be prequalified, and not surrogates and/or closely related products. Changes perceived as small (for example, change in production scale, bulk container materials, supplier of a critical biological or vial stopper) can have unexpected effects on stability and other performance characteristics. After such changes, a new documented risk assessment and, if necessary, a stability plan and study, is needed. Manufacturers should have change-control procedures in place compliant with ISO 13485 *(15).*

Note 2: Stability studies undertaken in the R&D phase of the product life-cycle provide an important understanding of how to design the product so that it will meet the final stability requirements identified in the input documentation. However, these studies are usually not sufficient for submission to WHO prequalification assessment since they may not reflect the final design and manufacture of the IVD.

10.2.1 Exceptions

If any of the above criteria are not met (for example if “pilot lots” or small-scale lots are used, or if the method of use described in the IFU is not

finalized), strong evidence must be provided that the materials that were evaluated will perform exactly the same as the final commercial product.

Note: In some exceptional circumstances, where it is not possible to sample from actual production lots, samples from pre-production or development lots might be used. If this is the case, manufacturers should justify why production lots were not used, and provide robust evidence that the lots chosen are expected to behave identically to the production lots. Data concerning lot-to-lot variability must still be submitted. Although WHO will consider the available evidence on its merits, this preliminary information must be followed by stability claims conducted on fully qualified production lots.

10.3 Number of lots required for testing

Current guidance*(1, 2)* recommends that three product lots at a **minimum** must be used to establish or verify shelf-life; in-use claims require testing on a **minimum** of one lot. The actual minimum number of lots to be used must be determined by a stringent risk assessment based on evidence of variability obtained during R&D (see section **10.1**). However, the minimum will never be less than three lots for shelf-life verification.

WHO note: It is **not** acceptable to sample IVDs from a single production lot but to label them so that they appear to have been taken from three separately manufactured production lots. This is true for all performance evaluation and regulatory submission purposes. WHO prequalification investigates batch records during on-site inspections. Non-compliance with this requirement may result in a critical non-conformity grading.

10.4 Components of lots required for testing

Current guidance *(1, 2)* recommends that stability work be performed using materials in their final packaging. Labelling is a significant factor of packaging and is known to present stability issues in some cases. For example, some label adhesives diffuse through some plastics, enter vials and affect the function of the reagents over time. Other label types lose adhesion over time; while some printing inks fade. The physical stability of packaging requires the same degree of risk evaluation and subsequent experimental verification as its chemical stability, with attention given to the countries of intended use. This is most important for primary packaging but must also be considered for secondary packaging, particularly for transport stability studies.

If there is more than one configuration or version of the IVD (for example, pack size differences, or Conformité Européenne (CE) marked and non-CE marked) then any potential effects on performance, including stability, must be assessed. In particular, if different reagent-container sizes are used in packs with different volumes of reagent (for example, different volumes for single use and multiple use), stability evidence should be obtained on all variants, even if the contents of the containers are identical,

unless stringent risk evaluation supported by physical or chemical evidence indicates otherwise.

Once component shelf-lives are assigned, it is expected that both relatively fresh components and components which have progressed into their assigned shelf-life will be used when selecting the different production lots for use in studies to establish the product shelf-life *(1, 2)*

11 Stability study plan

Stability studies should be well designed, scientifically sound, well implemented, well recorded and able to deliver meaningful conclusions concerning IVD performance. This will minimize the time and resources required by the manufacturer to generate appropriate evidence and by the regulatory authority to assess it.

It is good practice to prepare, within the mechanisms of a quality management system (QMS), a plan for the investigation of each characteristic of IVD stability. A well-developed study plan, with clearly defined objectives, responsibilities and pass/fail criteria, should be developed, reviewed and internally approved in advance of testing. The plan should be based on the design input requirements.

It is essential that the stability study plan takes into account the intended use of the product to ensure that the relevant critical characteristics are all captured by the plan. The results of the stability studies should support the claims made in the IFU.

Careful forward planning will help to ensure that sufficient resources are made available, effective experiments are performed, and both experimental results and associated documentation are recorded in an appropriate manner.

11.1 Responsibilities

The study plan should outline the responsibilities and applicable training for all staff involved in the study. The responsibilities for implementing the study plan must be assigned to appropriately qualified and trained staff. Responsibilities to be allocated include study set up, testing, monitoring, validation of equipment and/or processes, sample selection, risk assessment and corresponding documentation.

In addition, the manufacturer must nominate a person responsible for investigating failures and a person responsible for conducting risk assessments if the IVD fails to meet the requirements of the design inputs.

11.2 Preparing the testing plan

A complete, detailed description should be prepared that documents all of the required testing and procedures to be undertaken and the expected

outcomes. Authorization of the plan should be obtained internally in advance of commencing work. The plan should include the following details:

- the qualification and training of technical staff performing the work;
- any biohazard issues identified with reagents;
- aspects of instrumentation, including storage facilities or rooms, validation, calibration, monitoring and servicing;
- the lot/batch numbers of kits to be used, with justification for any manufacturing anomalies or deviations from documented procedures;
- the expected life of the kit from the input documentation;
- any proposal, with justification, to launch a kit with a shelf-life based on accelerated data, or to launch with a shorter shelf-life than in the input documentation while awaiting the conclusion of real-time testing;
- documentation of the nature and extent of in-use testing;
- the justification for the selection of lots and components, taking into account lot-to-lot variability and the critical characteristics;
- the number of units (test cassettes, bottles, tablets, etc.) of each component to be collected and stored under each condition;
- the nature of the panel to be used, justifying each panel member's inclusion and defining the volume and characterization of the bulk specimen to be used, and the aliquot size and number to be stored for the testing;
- the expected criteria for each panel member at the beginning and end of the product's proposed shelf-life;
- the statistical methods to be used for data analysis, including those used to identify outlying values and to establish criteria (see section **11.5**); and
- the methods for approval and justification of any deviations from the plan.

11.3 Product storage

A sufficient number of product components from the identified lots should be reserved and stored separately to ensure that the study will be completed with identified products. Sufficient numbers of the testing IVDs should be retained to allow for additional testing, calculated from estimated invalid result rates.

11.4 Documentation

The plan should make reference to the preparation of a study report that will be used to summarize the interim, and ultimately final, study findings and conclusions. The study plan, the testing protocol, the study report and all associated documentation (worksheets, etc.) should be controlled within the manufacturer's QMS. At the end of the study, the manufacturer should be able to confirm whether or not the design input requirements have been met.

Any changes from the methods identified in the plan must be recorded and undergo risk assessment. The plan should refer to the development of a detailed and valid testing protocol that includes all information and material relevant to testing.

11.5 Statistical methods

Statistical methods are used to support stability claims by providing estimates of the probability of results being as stated. For example, prior to the stability studies on an EIA, it has been documented that if a panel member has at least a particular OD then the IVD will meet a particular claim. Given the results of the stability study using that panel member and showing the variability within and between lots of the IVD, the probability of future similar production of the IVD meeting claims at the assigned life can be estimated. The derivation of valid criteria and the probability of maintenance of all claims can be estimated by appropriate statistical methods.

There is a wealth of information available on the statistical methods used in the R&D of IVDs, from both ISO *(20–22)* and CLSI *(*2, *23–26)*. Although most of these methods apply to quantitative assays, information on statistical methods for qualitative assays is also available *(27)*.

The fundamental considerations for stability testing are the number of replicates required at each time point and the number of different production lots required which together will produce an "acceptable overall probability estimate" of the likelihood of future production lots meeting claims (and hence user input requirements) at the end of the shelf-life.

However, consideration must also be given to what represents "an acceptable overall probability limit". "Acceptability" is a decision critical to quality and must be decided upon in advance based on the input requirements (for example, 80% confidence that 95% of lots will meet the claims). This is a tolerance interval as described in ISO 16269-6:2014 *(23).*The consideration can then be phrased as: "How many replicates and how many different production lots can then be derived from the tolerance interval required?"

It is strongly recommended that manufacturers seek advice from a professional statistician once the quality-critical requirements have been defined and before beginning any experimental work.

The statistical methods to be used must be documented in the plans and protocols of any stability study and consideration given to the treatment of unexpected and atypical results. In general, all results must be used unless there is a documented physical reason that the result can be ignored – for example, known operator error, too little volume, incorrect timing or use of an unqualified instrument (one lacking maintenance or calibration). Any

ignored results must nevertheless be recorded and included in the report of the stability study.

11.6 Stability testing protocol

As part of an approved study plan for the determination of IVD stability, a detailed testing protocol should be prepared as appropriate (examples of stability protocols are provided in Appendix 1 Examples of stability protocols). The protocol should include the following as a minimum:

- QMS identifiers (for example, experiment name, document references, etc.) that allow traceability to both the overarching study plan and to the records/documents generated, such as result worksheets.
- The training requirements for operator(s).
- The expected dates and times when the data will be collected.
- The objectives of the study (that is, determination of shelf-life, determination of in-use stability of a component, etc.).
- The name and lot number of the IVD and/or components to be investigated.
- Specification of how the components will be sampled from the production department.
- The panel members to be used and their characterization, including valid test methods which reflect the IFU claims.
- The experimental method that will be used for testing. This must follow the finalized testing method from the IFU where appropriate. It must describe clearly how the experiment is to be performed in terms of:
  - required storage and/or challenge conditions
  - duration of storage/challenge
  - schedule of testing intervals (see *(2)* section 4.3)
  - stability testing panel
  - numbers of replicate tests performed for each panel member.
- How and where results are to be recorded.
- The acceptance criteria.
- How aberrant, discordant or invalid results will be dealt with.
- How storage/challenge conditions are to be applied:
  Example: *For determination of stability during transportation it should be made clear that each IVD will be subjected to a sequence of stated temperatures*.
- How actual storage/challenge conditions are recorded:
  Example: *Recording of temperature not as "room temperature" but as an actual numerical value obtained from calibrated instrumentation*.

  Note: Statements of a general nature can be unclear to a regulatory or WHO reviewer. For example: "Sample buffer was stored at the required temperature and tested each month". This statement raises questions such as: (a) were the bottles of sample buffer stored open at the required temperature for the entire testing period; or (b) were the bottles stored capped and refrigerated, and only reopened briefly at the

required temperature at each schedule test point? To avoid confusion, the details of actual storage and use procedures are required in the testing report.

11.7 Reading and recording results

11.7.1 Avoiding reader bias

It is good practice to use approaches that make the reading of results as objective as possible, such as using a documented scoring system. For IVDs for which a subjective element forms part of the result (for example, reading the intensity of an RDT band within a specified time frame) the results should always be reviewed by both a first and a second reader to avoid operator bias. Both readers must be blinded to the expected results and the second reader must also be blinded to the first reader's results. If a validated band intensity scoring tool is to be included in the final RDT kit, this should be used to record results.

11.7.2 Recording actual individual results

The results of a test (not only the test interpretation) should be recorded. An interpretation on its own provides insufficient detail to detect the degradation of a signal over time. Photographic records of qualitative tests are recommended, as appropriate.

Some IVDs (for example, line-blots) may require the presence of particular band patterns to allow an interpretation to be reached, and several different patterns may yield the same final result. Recording only the final interpretation of a test specimen may cause the failure of particular bands to go unnoticed, while allowing the IVD to pass stability assessment.

Quantitative EIAs and NATs should be tested with sample panels containing concentrations of analyte across the quantitative range of the assay. Numerical results should be reported and statistical methods should be applied to ensure that the assay is measuring the analyte appropriately across the quantitative range.

Qualitative EIAs and NATs should also be tested with samples at several different analyte concentrations, including samples at low concentration near the cut-off level of the assay. Results should be recorded as positive or negative according to the predetermined cut-off level of the assay.

Example:

Some RDTs may stipulate that the strength of test band is not correlated with the strength of antibody titre. Nevertheless, the following should be recorded: (a) the intensity of observed patterns according to a predetermined and validated intensity scoring system with as fine a gradation as possible; and (b) the final result interpretation.

11.7.3 Retention of records

WHO recommends the retention of photographic records, machine printouts and electronic data, or physical retention of membranes from opened test cassettes, as appropriate. Records should be retained for the period of time equivalent to the commercial lifetime of the IVD but not less than two years (modified from *(15)* section 4.2.4).

11.8 Instability versus imprecision

Testing at more than two time points can be important for avoiding confusion between imprecision and instability. For example, if a 10% decrease (compared to the zero time value) is recorded from testing at the end of the shelf-life, it may not be possible to judge if the difference was due to imprecision or instability. The inclusion of additional test points (for example one or more between the zero time and the end of the shelf-life) allows for fluctuation caused by imprecision to be distinguished from drift due to instability.

Increased clarity between instability and imprecision can be gained by increasing the number of replicates and runs, primarily with reference to the zero time values (see sections **5.9** and **5.10**).

11.9 Testing schedule

Testing intervals should be selected to detect any trending of results over the testing period. Different testing intervals may be required for different components. For example, it may be appropriate to test an IVD test cassette against a panel on a monthly or quarterly basis, but to test for open vial stability on a weekly basis.

11.9.1 Acceptance criteria for results

The acceptance criteria to establish what is acceptable or not acceptable should be defined according to the panel criteria for both qualitative and quantitative test methods. Results from failed (invalid) test runs must not be used in the determination of the stability claim. However, the invalid results should be recorded and included in the report of the stability testing.

12 Stability study report

12.1 General

After testing has been completed, the findings should be summarized in a stability study report. The report should clearly identify the IVD that was tested, the objectives of the study, the conditions under which the IVD was tested and the conclusions that were drawn from the findings. The report should be traceable to the study plan, testing protocol and input

requirements. It should make clear references to other supporting documentation (for example, result worksheets).

12.2 Link to claims

The results and conclusions of stability studies presented in the report must support the claims of IVD stability reported in the IFU and elsewhere in the WHO prequalification dossier.

12.3 Consider variability

An overall stability claim (whether for shelf-life, in-use stability or stability during transportation) must be based on the expected stability when taking into account inter-lot variability.

Example:

*The manufacturer should evaluate the variability between the different lots studied (see section **10.1**) and assume that any differences in shelf-life are inherent to the manufacturing process. The claimed life should be calculated so that a known and stated proportion of all lots (usually > 95%) will meet the claimed shelf-life. Frequently, more than three lots are needed to obtain a realistic idea of the variability of the results.*

12.4 IVD stability versus component stability

A claim of stability for an IVD as a whole must not exceed any individual component stability.

Example:

For an IVD claimed to detect HIV-1 and HIV-2 antibodies – if detection of HIV-1 antibodies is stable to 24 months but that of HIV-2 to only 18 months then the shelf-life must be based on the shorter time of 18 months.

13 Changes to a WHO prequalified IVD

13.1 Dealing with change

Any critical or major modification to a WHO prequalified IVD or to its process of manufacturing will require the provision of new direct evidence of stability.

An appropriate risk assessment and an accelerated stability study comparing the original product and the modified product for usability, performance and lot-to-lot variability may serve to assess the impact of the changes to product formulation or manufacture.

It would be necessary to validate the stability of the modified IVD on a **minimum** of one lot of the IVD (subject to risk assessment) in order to

demonstrate equivalence between the original and modified IVDs. Testing of further lots may be appropriate depending on the product nature, variability of components and failure risk (see *(2)* section 7.1.2). WHO expects the results of accelerated testing to be confirmed by real-time studies.

If there are different presentations, evidence of the stability of each one must be provided (see also section **10.4**).

The following examples illustrate the scope for considering the performance evidence from one IVD as support for the performance of another. It should be noted that the observations discussed here refer specifically to IVD stability. Other aspects of IVD performance should still be validated as appropriate.

Examples:

1. *An HIV RDT uses an identical test cassette and physical components as a manufacturer's existing, fully validated, HCV RDT, but the reagent formulations are different (antigen/antibodies, buffers, conjugates, etc.) – evidence of stability of the HCV RDT would not suffice for the HIV RDT. Even if the manufacturer claims that both IVDs have been sold in a number of countries for several years and no adverse feedback has been reported, this would not constitute evidence in support of the stability of either IVD.*

2. *From an HIV RDT that has been fully validated for detection of HIV-1 antibodies, a new product is developed that includes detection of HIV-2 antibodies. The stability of any sample buffers that are identical between the two IVDs would, most likely, not need to be validated. However, other components (conjugates, antigens or antibodies) that are different between the two IVDs would need to be tested; it would not be sufficient to assume that HIV-1 reagents will have the same stability in the new IVD. An IVD modification of this nature is likely to require substantial new validation of stability.*

3. *An HIV RDT previously intended for testing serum/plasma has a claim added for detection of HIV-1 in whole blood. The only substantive design change associated with the new claim is the addition of a small filter pad near the sample port which acts as a filter for whole blood specimens. Depending on the nature of the material, it may be reasonable to argue that the pad material would not be expected to age; that it is not, in any practical sense, chemically labile. Consequently, shelf-life and in-use stability may not necessarily need to be retested in full. However, stability during transportation may need to be determined to provide confidence that the modification is able to withstand likely shipping conditions (for example, that the extra square of filter pad material does not dislodge when packages are jostled and bumped in transit).*

4. *Based on an HIV RDT that has been fully validated for the detection of HIV-1 antibodies, a new IVD is developed which includes detection of antibodies to Treponema pallidum (TP). Detection of TP-specific antibodies occurs on a completely separate membrane (and associated architecture) to that of HIV-antibody detection. Additional handling steps may have an impact on the stability of the HIV-1 antibodies and retesting may be required. It may be necessary to review evidence of stability during transportation to ensure that new components are not affected by transit (for example, where a new packaging concept is used).*
   - *If a new machine is used for striping of the HIV-1/TP IVD, validation of the new machine (installation qualification, operational qualification and performance qualification) would be required to show that the stability studies are still valid.*
   - *If the IVD is designed in a way that HIV and TP detection occurs either on the same membrane and/or using most of the same architecture (and assuming that sample buffers are identical between IVDs) it is likely that this new IVD would need to be fully validated.*

14 References

1. ISO 23640:2011. In vitro diagnostic medical devices – evaluation of stability of in vitro diagnostic reagents. Geneva: International Organization for Standardization; 2011.
2. Evaluation of stability of in vitro diagnostic reagents; approved Guideline EP25-A. Wayne (PA): Clinical and Laboratory Standards Institute; 2009.
3. Specifications: test procedures and acceptance criteria for biotechnological/biological products. Q6B (Step 4). ICH Harmonised Tripartite Guideline. Geneva: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 1999.
4. ISO 17511:2003. In vitro diagnostic medical devices – measurement of quantities in biological samples – metrological traceability of values assigned to calibrators and control materials. Geneva: International Organization for Standardization; 2003.
5. ISO 18113-1:2009. In vitro diagnostic medical devices – information supplied by the manufacturer (labelling) – Part 1: Terms, definitions and general requirements. Geneva: International Organization for Standardization; 2009.
6. Directive 98/8/EC of the European Parliament and of the Council of 16 February 1998 concerning the placing of biocidal products on the market). OJ. 1998;L 123:1–63 of 24.4.98 (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1998:123:0001:0063:en:PDF, accessed 20 December 2017).
7. ISO 9000:2005. Quality management systems – fundamentals and vocabulary. Geneva: International Organization for Standardization; 2005.
8. ISO 15198:2004. Clinical laboratory medicine – in vitro diagnostic medical devices – validation of user quality control procedures by the manufacturer. Geneva: International Organization for Standardization; 2004.
9. Code of Federal Regulations Title 21. Section 820.3 Definitions. Washington (DC): United States Food and Drug Administration; 2010.
10. ISO/IEC Guide 99:2007. International vocabulary of metrology – basic and general concepts and associated terms (VIM). Geneva: International Organization for Standardization; 2007.
11. Glossary and definitions of terms used in GHTF documents. GHTF/SC/N4:2012 (Edition 2). Global Harmonization Task Force (GHTF) Steering Committee; 2012 (http://www.imdrf.org/docs/ghtf/final/steering-committee/procedural-docs/ghtf-sc-n4-2012-definitions-of-terms-121109.pdf, accessed 21 December 2017).
12. ISO 14971:2007. Medical devices – application of risk management to medical devices. Geneva: International Organization for Standardization; 2007.
13. Quality management systems – process validation guidance. GHTF/SG3/N99-10:2004 (Edition 2). Global Harmonization Task Force (GHTF) Steering Committee; 2004 (http://www.imdrf.org/docs/ghtf/final/sg3/technical-docs/ghtf-sg3-n99-10-2004-qms-process-guidance-04010.pdf, accessed 21 December 2017).
14. Standard practice for performance testing of shipping containers and systems. ASTM D4169 - 14. West Conshohocken (PA): ASTM International; 2014.
15. ISO 13485:2003. Medical devices – quality management systems – requirements for regulatory purposes. Geneva: International Organization for Standardization; 2003.

16. Technical Guidance Series (TGS) for WHO Prequalification – Diagnostic Assessment. Guidance on test method validation for in vitro diagnostic medical devices. TGS–4. Geneva: World Health Organization; 2017 (http://apps.who.int/iris/bitstream/10665/258971/1/WHO-EMP-RHT-PQT-TGS4-2017.04-eng.pdf?ua=1, accessed 21 December 2017).
17. ISO 15189:2012. Medical laboratories – requirements for quality and competence. Geneva: International Organization for Standardization; 2012.
18. United States Pharmacopeia 31 - National Formulary 26 (USP 31-NF 26). Rockville (MD): The United States Pharmacopeial Convention; 2008.
19. Pharmacopoeia of the People's Republic of China English edition. Beijing: State Pharmacopoeia Commission of the People's Republic of China; 2000.
20. ISO 5725-1,2,3,4,6:1994 and ISO 5725-5:1998 Accuracy (trueness and precision) of measurement methods and results – Parts 1–6. Geneva: International Organization for Standardization; 1994 and 1998.
21. ISO 3534-1,2:2006 and ISO 3534-3:2013. Statistics – vocabulary and symbols – Parts 1–3. Geneva: International Organization for Standardization; 2006 and 2013.
22. ISO 16269-4:2010, ISO 16269-6:2014, ISO 16269-7:2001 and ISO 16269-8:2004. Statistical interpretation of data – Parts 4 and 6–8. Geneva: International Organization for Standardization; 2010, 2014, 2001 and 2004.
23. Evaluation of the linearity of quantitative measurement procedures: a statistical approach; approved Guideline EP06-A. Wayne (PA): Clinical and Laboratory Standards Institute; 2003.
24. Interference testing in clinical chemistry; approved Guideline EP07-A2. Second edition. Wayne (PA): Clinical and Laboratory Standards Institute; 2005.
25. Evaluation of detection capability for clinical laboratory measurement procedures; approved Guideline EP17-A2. Second edition. Wayne (PA): Clinical and Laboratory Standards Institute; 2012.
26. Evaluation of precision of quantitative measurement procedures; approved Guideline EP05-A3. Third edition. Wayne (PA): Clinical and Laboratory Standards Institute; 2014.
27. Valcárcel M, Cárdenas S, Barceló D, Buydens L, Heydorn K, Karlberg B et al. Metrology of qualitative chemical analysis. Brussels: Directorate-General for Research and Innovation (European Commission); 2002 (http://bookshop.europa.eu/en/metrology-of-qualitative-chemical-analysis-pbKINA20605/, accessed 22 December 2017).

Appendix 1

Examples of stability protocols

This appendix uses the example of a wholly fictitious IVD to illustrate the kinds of experimental design that would be required to adequately determine:

1. the stability of whole kits during transport followed by the stability of whole kits during shelf-life; and
2. the in-use stability of whole kits including reagents.

The information provided in these examples should be used as a guide to possible approaches for generating evidence of a standard sufficient to satisfy the expectations of WHO prequalification. Further examples can be found in the WHO Prequalification: Sample Product Dossiers available on the WHO Prequalification website[1].

WHO expects that a transportation challenge would precede the real-time determination of shelf-life and in-use studies.

Description of the fictitious IVD

The fictitious IVD used in the examples below is an RDT for the detection of antibodies to HIV-1, HIV-2 and *Treponema pallidum* (TP) in serum, plasma and whole blood, and is referred to as the HIV/TP RDT.

The IVD kit components are: a test cassette sealed in a foil pouch (with desiccant) and a bottle of specimen buffer/diluent for use.

It is recommended that the kit be stored at 8–40 °C and brought to 15–30 °C before use.

It is recommended that once the sealed foil pouch of the test cassette is opened that the test cassette be used immediately.

The specimen buffer is expected to have similar stability to the sealed and pouched test cassette. The stability of the opened bottle of specimen buffer is determined below (see **Example 2: In-use stability protocol**).

Stability study plan:

The manufacturer has developed a stability study plan to determine the stability of the HIV/TP RDT. As part of this plan, a preliminary determination of accelerated stability has been made at several extremes of temperature, which suggests that the IVD would be stable to an equivalent of 12 months following manufacture. The plan calls for the

[1] http://www.who.int/diagnostics_laboratory/guidance/sample_product_dossier/en/

development of real-time stability protocols that will form the basis of subsequent testing of the IVD.

Preliminary work has shown that the variability between lots is minimal. As a result, three independent lots (with no critical constituents in common) will suffice to enable a reasonable estimation of shelf-life, taking lot-to-lot variability into account.

Example 1: Evaluation of transport stability followed by real-time stability

Objective

To determine the stability after transportation of the HIV/TP RDT in real-time using simulated shipping conditions, and to generate components that have already undergone stress testing to be used in real-time shelf-life studies as proposed in Stability Study Plan XZY00001.

Preparation

Acquire sufficient numbers of the IVD kits from three independent production lots using a predetermined sampling protocol (for example, random, first X number of kits in first box, every 100th kit, etc.). Allow at least 10% overage for unexpected requirements and re-testing.

> Note 1: To provide security against unforeseen events, duplicate tests should be performed as a minimum. However, testing in triplicate will provide more statistical confidence in the observed test result.

The IVD kits chosen for testing must be in their final packaging including all labelling (see section **10.4**).

The IVD kits are stored so that the reagents are in contact with all elements of the packaging (for example, the bottles in the IVD kits are stored horizontally, lying flat on their sides, allowing liquids to remain in contact with the bottle closures).

Acquire sufficient volume of each panel member for the duration of the testing schedule (see testing schedule below).

The protocol for these studies specifies the number of IVD kits to be picked, the statistical sampling plan to be used and the required panel members and their volumes.

Documentation

In Worksheet XYZ00001 record the following:

- the lot numbers from which the IVD kits were sampled;
- the number of IVD kits sampled from each lot; and

- details (including manufacturing/lot information) for each of the IVD kit components that will be tested as part of this protocol (test cassette and specimen buffer).

Testing schedule: for transport simulation

Testing will be conducted at 0, 3, 6, 9, 12 and 13 months.

> Note 1: Testing beyond 13 months will allow for an understanding of when, in real-time, the IVD is likely to "fail" and may allow for an extension of the proposed shelf-life.
>
> Note 2: For determination of shelf-life, a fresh bottle of specimen buffer must be opened at each testing point – though there may be circumstances in which multiple sampling could be taken from the same bottle after it has been opened.

The IVD kits will be divided into two groups. One group will be stored at 40 ± 5 °C, the other at 8 ± 2 °C. IVD kits from each group will then be subjected to the following conditions:

Condition 1: Temperature and humidity sequence; all IVD kits will be taken through a temperature and humidity sequence consisting of:

i) Ambient humidity (X% RH)
- Put at IFU storage temperature for 24 ± 4 hours followed by
- 30 ± 5 °C for 24 ± 4 hours, followed by
- 45 ± 5 °C for 24 ± 4 hours, followed by
- 8 ± 5 °C for 24 ± 4 hours, followed by
- IFU storage temperature for 24 ± 4 hours

Followed by

ii) Desert humidity (30% RH)
- Put at IFU storage temperature for 24 ± 4 hours followed by
- 30 ± 5 °C for 24 ± 4 hours, followed by
- 45 ± 5 °C for 24 ± 4 hours, followed by
- 8 ± 5 °C for 24 ± 4 hours, followed by
- IFU storage temperature for 24 ± 4 hours

Followed by

iii) Tropical humidity (85% RH)
- Put at IFU storage temperature for 24 ± 4 hours followed by
- 30 ± 5 °C for 24 ± 4 hours, followed by
- 45 ± 5 °C for 24 ± 4 hours, followed by
- 8 ± 5 °C for 24 ± 4 hours, followed by
- IFU storage temperature for 24 ± 4 hours

Followed by

iv) Ambient humidity (X% RH)
- Put at IFU storage temperature for 24 ± 4 hours followed by

- 30 ± 5 °C for 24 ± 4 hours, followed by
- 45 ± 5 °C for 24 ± 4 hours, followed by
- 8 ± 5 °C for 24 ± 4 hours, followed by
- IFU storage temperature for 24 ± 4 hours.

Note 1: It is important to make clear that the above complete sequence of temperatures will be used, as opposed to separate IVD kits being held at individual temperatures. The actual temperatures, durations and the nature of the sequence will depend on the IVD and on the kinds of conditions expected to be encountered during shipping.

Note 2: Freezing temperatures are not considered in this example but should be included if the IVD kits could be exposed to freezing temperatures during transport.

Note 3: If transport by air is anticipated, the effect of reduced pressure should be included in the protocol[1] for a period of time at least 10% longer than the longest anticipated flight, and at a pressure expected in aircraft holds.

Note 4: The protocol should call for testing of at least five individual IVD kits after each stress condition, using the stability panel members giving the most informative results. This approach will enable verification that the IVD kits are sufficiently stable to progress to the next condition – though this should already be known from preliminary experiments and R&D work.

Condition 2: Transport stress conditions – shaking; each IVD kit will be placed on a shaking table at X revolutions per minute (rpm) for X hours/days at 42 ± 5 °C as defined by ASTM D4169 section 12.[1]

After the simulated shipping challenge, each IVD kit will be returned to its corresponding storage temperature (40 ± 5 °C or 8 ± 2 °C).

Testing schedule for real-time stability studies

Testing will be conducted at 0, 3, 6, 9, 12 and 13 months. At each scheduled time point, the allotted number of IVD kits will be brought to 15–30 °C and used to test each member of the panel in triplicate.

Note 1: The test at 0 months will provide evidence that the IVD kit is stable under extreme conditions of shipping (but similar to those likely to be experienced); the testing at later time points will provide evidence to support the claimed shelf-life after transport; and testing beyond the claimed shelf-life will provide evidence that the IVD kit is stable and not close to a failure point.

[1] See: Standard practice for performance testing of shipping containers and systems. ASTM D4169 - 14. West Conshohocken (PA): ASTM International; 2014.

Documentation for transport stress conditions

In Worksheet XYZ00001 record the following:

- the lot numbers of the IVD kits used to conduct the test;
- the operator(s) name(s);
- the dates of testing;
- identifying details for each member of the panel being tested;
- the temperature at which the IVD kits are stored;
- the values of temperature and humidity for each of the challenge conditions;
- instrument settings for the shaking apparatus and duration of operation for challenge conditions;
- the ambient temperature and humidity during testing;
- each test result as an interpretation according to the IFU;
- each test result as a band intensity – band intensity should be scored using the calibrated scale described in Protocol ZXY00001 (for example, 0; faint/trace; +1; +2; +3;…+10) even though the IFU do not give scores to results);
- any aberrations or deviations from the protocol, the reason for the deviation and any remedial action undertaken. Results from invalid assays must be recorded but not included in the calculations of shelf-life. Apparently aberrant results, unless the underlying cause can be positively identified as not related to a problem with the IVD, must be included in the calculations of shelf-life.

Panel for monitoring stability

See the suggestions in **Appendix 2 Suggested specimens for stability testing panel.**

Acceptance criteria

Each panel member should show a band intensity result that matches its expected result at each tested time point. The expected result must be validated so that if the IVD fails to meet the claims (for example, fails to detect critical specimens, has unacceptable performance at medical decision concentrations or has unacceptable specificity) the panel member would also fail to meet its specified result.

The stability after transportation of the IVD kit will be taken as the time point before the last time point to have met the acceptance criteria – for example, if the IVD is stable to 13 months, the stability after transportation will be deemed to be 12 months.

The stability after transportation should be identical to the claimed shelf-life of the IVD kit – that is, the extremes of possible conditions to which the IVD kit is likely to be subjected during transport must not affect the shelf-life of the IVD.

Calculation of results

Detailed statistical instruction must be obtained from a professional statistician with an understanding of the expectations of the stability study

plan and outcome. Professional statistical input is particularly recommended when calculating confidence limits for discrete data such as readings from a graduated scale.

Each of the following applies at each time point:

The variance of the results for all replicates within and between all the lots must be calculated for each panel member. From the overall variance between lots, the confidence with which future lots of the IVD kit will detect the panel member at that time point after manufacture and transport can be calculated. If the confidence that the panel member will meet its specification is less than some pre-defined value (normally 95%), it must be deemed to have failed at that time point and the shelf-life of the IVD kit should be restricted accordingly.

If regression analysis is used to define the time point at which a panel member would not meet its criterion, then lot-to-lot variability must be included when setting the confidence limits around the regression line. However, real-time data must extend beyond the claimed shelf-life so that the intercept of the regression confidence limit and the expected value must be at a time period longer than the claim. It is usually more appropriate to calculate as discussed in the previous paragraph, particularly if the regression cannot be proven to be linear.

Example 2: In-use stability protocol

Objective

To determine the stability of opened bottles of the specimen buffer used in the IVD kit in real-time when stored at 15–30 °C as proposed in Stability Study Plan XYZ00001.

In this example, the manufacturer recommends that the test cassette be used immediately upon opening; this claim should also be validated in a separate experiment, so that it can be confirmed that the IVD will still perform satisfactorily after the test cassette has been removed from its pouch and left at room temperature for 1, 2, 6 and 24 hours, etc., as appropriate.

Acquire sufficient numbers of IVD kits from one production lot using a predetermined sampling protocol (for example, random, first X number of kits in the first box, every 100th kit, etc.).

Acquire sufficient volume of each panel member for the duration of the testing schedule. Establish a method for randomizing the panel for testing.

In Worksheet XYZ00001 record the following:

- the lot numbers from which the IVD kits were sampled;
- the number of IVD kits sampled from each lot; and

- details (including manufacturing/lot information) for each of the IVD kit components that will be tested as part of this protocol (test cassette and specimen buffer).

Preparation

Two lots of specimen buffer are to be tested. One lot of the component must be freshly made, while the other should be towards the end of the assigned shelf-life of the IVD kit.

The component is to be tested in its final packaging.

The IVD kits are stored so that the reagents are in contact with all elements of the packaging (for example, the bottles in the IVD kits are stored horizontally, lying flat on their sides, allowing liquids to remain in contact with the bottle closures).

Half of each lot will be stored at 30 ± 5 °C, the other half at 15 ± 5 °C. At the start of testing, each bottle will be brought to room temperature (20 ± 2 °C), opened, used for testing and then recapped and returned to the stated storage temperature.

> Note: It is important that the components under test are opened and used under circumstances likely to occur in users' laboratories (that is, not in rooms with HEPA-filtered air) thus mimicking, as far as possible, genuine use.

Testing schedule

At each subsequent scheduled time point the allotted number of bottles will be brought to room temperature and used to test each panel member in triplicate. Testing will be conducted at 0, 1, 2, 3, 4 weeks, etc., up to the end of the claimed in-use life.

Documentation

In Worksheet XYZ00001 record the following:

- the lot number of the IVD kit used to conduct the test;
- the operator(s) name(s);
- the dates of testing;
- the temperature at which the IVD kits are stored;
- the ambient temperature during testing;
- identifying details for each member of the panel being tested;
- each test result as a band intensity – band intensity should be scored using the calibrated scale described in Protocol ZXY00001 (for example, 0; faint/trace; +1; +2; +3;…+10);
- each test result as an interpretation according to the IFU;
- any aberrations or deviations from the protocol, the reason for the deviation and any remedial action undertaken.

Panel for testing stability

See the suggestions in **Appendix 2 Suggested specimens for stability testing panel.**

Acceptance criteria

Each panel member should show a band intensity result that matches its expected result at each tested time point. The in-use stability of the sample buffer will be taken as the time point before the last time point to have met the acceptance criteria – for example, if the IVD kit is observed to be stable to 5 weeks, the in-use stability will be deemed to be 4 weeks.

Appendix 2

Suggested specimens for stability testing panels

Examples in this section

Not all of the specimens in the examples that follow will be necessary for all IVDs, and nor is the list exhaustive. Panels must be composed according to strict risk-management principles and all decisions must be documented and traceable.

The minimum set of specimens recommended for inclusion in a testing panel for different types of products are outlined below.

1. Specimens to monitor NAT-based tests

If a proprietary nucleic acid preparation/extraction system is provided, the recovery must be shown to meet claims for each genotype from each of the specimen types claimed (for example, dried blood spots, whole blood and plasma). Successful removal of inhibitory substances, if intended, must be demonstrated for appropriate specimen types. Unless potentially variable biological reagents are involved, this system would be expected to be verified in manufacture and not necessarily tested at release.

Specimens	Remarks
Specimens to demonstrate maintenance of sensitivity and/or limit of detection, and/or accuracy, and precision	Traceability is required to one of the WHO international standards[1] if available – for example, the Third WHO International Standard for HIV1-RNA for NAT-based assays (National Institute for Biological Standards and Control (NIBSC) code 10/152); or the Fourth WHO International Standard for hepatitis C virus RNA for NAT-based assays (NIBSC code 06/102). More than one genotype may be required to validate these claims: see the First WHO International Reference Panel for hepatitis B virus genotypes for NAT-based assays (Paul-Ehrlich-Institut (PEI) code 5086/08). This may be required on each of the claimed specimen types.
Specimens to demonstrate specificity and validity of runs	Sufficient negative specimens should be included to ensure that the claims will be met at end of shelf-life.
Specimens (or reagents) to demonstrate stability of each of the critical components of the IVD	If more than one part of the genome is to be detected, both systems must be shown to be stable. If both DNA and RNA are measured the complete system must be shown to be stable.

[1] The catalogue of WHO International Reference Preparations is available at: http://www.who.int/bloodproducts/catalogue/en/

2. Specimens to monitor tests that measure CD4 cells

Rationale

CD4 measurements are quantitative, and accuracy at the clinical decision points is crucial. The design input should have information on the accuracy and other parameters required, and the panel must be designed to provide evidence that these parameters are maintained over the assigned life of the reagent and measuring IVD.

Parameters

The panel used in stability work must be able to demonstrate the following:

- stability of all the antibodies used in the IVD (frequently anti-CD4 and anti-CD3 antibodies; any other critical components must also be covered);
- accuracy and trueness of measurement maintained at the critical level (at least five specimens required);
- claimed linearity over the required range of CD4 count (at least five specimens required); and
- measure drift.

Specimens

Artificial specimens (such as stabilized blood specimens) can be used if a risk assessment based on R&D work indicates that they are effective. Fresh specimens are usually required. Measurements should be compared to an approved reference system.

Examples of approaches

Aged or in-use lots may be compared with a reference – for example, a new lot. Precision studies can be performed as described elsewhere.[1]

[1] Evaluation of precision of quantitative measurement procedures; approved Guideline EP05-A3. Third edition. Wayne (PA): Clinical and Laboratory Standards Institute; 2014.

3. Specimens to monitor tests for HIV antibodies

Specimens	Remarks
IgM first seroconversion specimens and IgG first seroconversion specimens	Possible approaches to obtain samples: • Study the early data from commercial seroconversion panels where the seroconversion was frequently monitored by IgM **and** IgG blots. • Study the responses to second and third generation assays or protein A and protein L assays (this approach is less useful).
All other parts of the HIV proteome included – for example, reverse transcriptase (RT)	
Late stage specimens – usually a high-dilution set near the sample-to-cut-off ratio	This might serve to monitor any kit run control. Note: HIV serology is not particularly genotype dependent. It is usually not necessary to include controls for genotype detection unless risk assessment or experiment shows that it is required for a particular IVD.
HIV-2, diluted to near the sample-to-cut-off ratio	Seroconversion specimens are very rare.
HIV-1 (O), if claimed	
Difficult specimens to monitor specificity and invalidity rate	100 negatives at release subject to risk analysis and statistical analysis of the allowable (relative to the claimed) false-reactive rate and invalidity rate.

4. Specimens to monitor tests for HIV-1/2 and *Treponema pallidum* (TP) antibodies

Specimens	Remarks
Specimens to detect HIV	See the above section **3. Specimens to monitor tests for HIV antibodies**.
Specimens to detect all the critical epitopes in the IVD – for example, TpN47, TpN17 and TpN15	Note: Each of these epitopes plays a role in detecting syphilis in different stages of the infection. It is necessary to have a panel member to monitor each epitope system present (and possibly each stage of infection) even if poly-fusion proteins are used. This can be avoided if the manufacturer can demonstrate that each epitope system is equally stable.
Specimens able to show that the invalidity and specificity rates do not fall outside the claims, particularly if whole blood is a claimed specimen type	Note: It would not be sufficient for WHO prequalification to extrapolate to the stability of HIV-2/TP detection by testing only HIV-1-positive specimens.

5. Specimens to monitor tests for HCV antibodies

Specimens	Remarks
NS3 first seroconversion specimens and core first seroconversion specimens	
Specimens to monitor any other antibodies claimed (frequently against NS5 and NS4)	Results can be obtained from line immunoassays that differentiate antibody responses to the different proteins.
A late-stage dilution near the sample-to-cut-off ratio	Note: HCV serology is not particularly genotype dependent in terms of anti-core and anti-NS3, but it is possible to make serotyping assays based on NS4 that mimic genotyping reasonably well. It is usually not necessary to include controls for genotype detection unless risk assessment or experiment for a particular IVD show otherwise.
Difficult specimens to monitor specificity and invalidity rate	100 negative specimens subject to risk analysis and statistical analysis of the allowable (relative to the claimed) false-reactive rate and invalidity rate.

6. Specimens to monitor tests for HBsAg

Specimens	Remarks
Specimens to define sensitivity relative to the claim	Traceability is required to one of the WHO international standards[1] – for example, the Third WHO International Standard for hepatitis B virus surface antigen (genotype B4; HBsAg subtypes *ayw1*/*adw2*); NIBSC code 12/226) for one or more specimens and probably also to the *ad* and *ay* standards available from a commercial supplier. Commercially available seroconversion specimens are almost all of the *adw2* subtype – different from the Third WHO International Standard – so claims of critical threshold specimen detection must be proven by specimens in the panel.
Specimens to monitor the maintenance of the claims for a variety of serotypes/genotypes and mutant forms	These will almost certainly be traceable to the First WHO International Reference Panel for hepatitis B virus genotype for HBsAg assays (PEI code 6100/09).
Specimens to control against prozone/high dose hook effect if found or if theoretically an issue	
If detection of HBsAg in the presence of anti-HBsAg is claimed (current best practice) proof of maintenance of the claim is required	
Specimens to monitor the critical components of the IVD	If the monoclonal antibodies used have a particular function or bias, such as against the *ayr* or *adr* subtypes (not controlled by the standards) or are used to detect mutant forms of the antigen, then each must be monitored to ensure viability at end of shelf-life. These may be the same specimens as mentioned in the previous paragraphs. If there are critical dissociation chemicals or red-cell capture or rupture agents used then these must also be monitored.
Difficult specimens to monitor specificity and invalidity rate	100 negative specimens subject to risk analysis and statistical analysis of the allowable (relative to the claimed) false-reactive rate and invalidity rate.

[1] The catalogue of WHO International Reference Preparations is available at: http://www.who.int/bloodproducts/catalogue/en/

Appendix 3

Summary table of standards relevant to stability studies

Recommendation	Comment	Standard
Studies must be compliant with CLSI EP25-A and ISO 23640:2011	The minimum expected standards.	CLSI EP25-A ISO 23640:2011
Studies must be fully documented with risk evaluations, plans and protocols prior to initiation	Risk assessment must be specific to the analyte, type of physical device and assay format, and previous manufacturing experiences, not generic nor by rote.	CLSI EP25-A (many sections) ISO 23640:2011 (section 2) ISO 14971:2007
Studies and risk management must take into consideration the conditions likely to be encountered in the geographical and health-care settings in which the IVD is intended to be used	This is particularly important for transport stress where extreme conditions must be evaluated.	
IVDs must be subjected to simulation of transport stress before being used to establish any form of stability	This is particularly important to WHO-PQ as transport will always be involved before use of an IVD, and transport conditions cannot be guaranteed nor predicted.	CLSI EP25-A (section 4.2.3) ISO 23640:2011 (section 5.2)
Transport simulation must cover the extremes of environmental conditions ascertained during risk evaluations	It is most unlikely that actual transport will involve all extreme conditions that might occur during the marketing life of the IVD, or that the conditions during actual transport can be adequately documented.	CLSI EP25-A (section 4.2.3)
IVDs used in any stability studies must be made to finalized manufacturing specifications, to final scale and in the packaging (including labelling) in which the IVDs will be made available	If IVDs are not made to final validated and documented manufacturing scales, stringent proof must be presented that the scale change will not affect any parameters of the IVD, nor any of the manufacturer's claims. Pre-production lots can only be used for stability work if these conditions are met.	Good manufacturing practice (GMP) CLSI EP25-A
If several presentations of the IVD are to be presented, all aspects of stability must be shown for each	If, for example, two pack sizes are to be provided then each pack size must be evaluated completely, even though the contents are identical except for vial size.	CLSI EP25-A

Recommendation	Comment	Standard
Sufficient numbers of independent lots of the IVD must be evaluated to enable each form of stability to be evaluated in terms of inter-lot variability	"Independent lots" means lots with different production (or manufacturing, purification, etc.) runs of critical reagents (for example, biological reagents prepared in different syntheses, growths or purifications or other risk-defined critical reagents from different manufactured lots or from different suppliers if applicable). CLSI EP25-A and ISO 23640:2011 specify **minimum** numbers of lots to be used but give no guidance to recommended numbers beyond documented risk evaluation.	CLSI EP25-A (section 4.4)
If critical components of the IVD are assigned lives independently of the life of the IVD, the various forms of stability of the IVD must be proven with those reagents at different stages of their lives	It must be documented that stored materials (for example, freeze-thawed biological reagents) operate as expected during the whole of the assigned shelf-life.	CLSI EP25-A (section 4.4)
Each form of stability must be defined statistically with respect to any inter-independent lot variability, not just assigned to the minimum stability found among the lots that happened to be evaluated experimentally	If any lot-to-lot variability is found, the manufacturer must provide evidence that subsequent lots will not have worse stability than that claimed.	
If any control material with a claim to prove the functionality of the IVD is provided to users that claim must be justified in stability studies in addition to any other studies	If the analytic function of the IVD is out of specification from any cause, including stability failure, the control material must be demonstrated to be able to alert the user to that fact.	
Use of accelerated stability, even to provide interim life assignments, must be justified scientifically	Accelerated stability is acceptable in providing interim life if the parameters of the Arrhenius equation, or any other method used, are adequately proven and documented.	CLSI EP25-A (section 7.3 & Appendix B) ISO 23640:2011 (section 5.3.1; notes 1 & 2)

WHO/BS/2017.2304

The Technical Guidance Series (TGS) for submission to WHO Prequalification – Diagnostic Assessment is intended to assist manufacturers in meeting WHO prequalification requirements for their IVD. For further information on this guidance and other TGS documents email:

diagnostics@who.int