Annex 6

WHO manual for the preparation of secondary reference materials for in vitro diagnostic assays designed for infectious disease nucleic acid or antigen detection: calibration to WHO International Standards

Guidance documents published by the World Health Organization (WHO) are intended to be scientific and advisory in nature. Each of the following sections constitutes guidance for national regulatory authorities (NRAs) and for manufacturers of biological products.

Abbreviations

Ag	antigen
% CV	percentage coefficient of variation
Ct	cycle threshold
DNA	deoxyribonucleic acid
EQA	external quality assurance
IS	International Standard(s)
IU	International Unit(s)
IVD	in vitro diagnostic
MU	measurement uncertainty
NAT	nucleic acid amplification technique
NRA	national regulatory authority
PT	proficiency testing
RNA	ribonucleic acid
S/CO	sample-to-cut-off (ratio); also signal-to-cut-off (ratio)
SI	*Système international d'unités* (measurement system using metric units)
SoGAT	Standardisation of Genome Amplification Techniques (group)

1. Background

Through its Expert Committee on Biological Standardization, WHO developed its Guidelines for the preparation and establishment of reference materials and reference reagents for biological substances in 1978 (*1*). This document was last revised in 2004 (*2*). In the revised WHO Recommendations document, secondary standards are defined as reference preparations established by regional or national authorities, or by other laboratories, that are calibrated against and traceable to WHO reference materials. Part B of the 2004 Recommendations deals with general considerations for the preparation, characterization and calibration of regional or national biological reference standards.

Feedback from manufacturers and providers of secondary (for example, regional) standards used for in vitro diagnostic (IVD) devices, and from regulatory authorities, international trade organizations, IVD manufacturers, providers of external quality assurance (EQA) or proficiency testing (PT) programmes and laboratories using diagnostic assays, indicated a need for more specific guidance on the preparation of secondary standards; it was therefore concluded that a practical manual focusing on IVD needs would be helpful. This topic was discussed at the 2012 meeting of the WHO Expert Committee on Biological Standardization, at which the proposal to generate a WHO document on secondary standards for use in the IVD field was endorsed.

2. Purpose and scope

This WHO document provides practical guidance on the preparation of secondary biological reference materials and on their calibration to WHO International Standards (IS) where available. The document focuses on the in vitro measurement procedures used for diagnosis, detection and management of infectious diseases where the typical analytes (measurands) are nucleic acid or antigen (Ag). These IVD tests cover nucleic acid amplification technique (NAT)-based assays for detecting the DNA or RNA of infectious agents and immunological tests for the detection of Ag(s) of infectious agents. Currently, there are only a small number of IS with an assigned unitage available where the analyte is an antibody directed to an infectious agent. Due to their complexity (that is, the epitope spectrum represented by polyclonal antibodies in the serum of a patient) this document does not cover antibody-based secondary standards. However, several principles outlined in this manual may also apply to antibody assays. Where applicable, the document integrates existing guidance, referenced accordingly.

The document is intended for use by manufacturers of secondary reference materials, IVD manufacturers, providers of EQA or PT programmes

and other laboratories using reference materials for NAT-based and serological infectious disease assays. Analogous guidance has already been issued by WHO on secondary standards for vaccines (*3*) and chemical reference substances (*4*).

3. Terminology

The definitions given below apply to the terms as used in this WHO guidance document. These terms may have different meanings in other contexts.

Accuracy: (measurement) closeness of agreement between a measured quantity value and the true quantity value of a measurand (*5*).

Biological matrix: a discrete material of biological origin that can be sampled and processed in a reproducible manner. Examples include blood, serum, plasma, urine, faeces, saliva, sputum and various discrete tissues (*6*).

Calibration: a process that, under specified conditions, establishes as a first step a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication (*5*).

Calibration hierarchy: a sequence of calibrations from a reference to the final measuring system, where the outcome of each calibration depends on the outcome of the previous calibration (*5*).

Commutability (of a reference material): a property of a reference material, demonstrated by the equivalence of the mathematical relationships among the results of different measurement procedures for a reference material and for representative samples of the type intended to be measured (*7*).

Control material: a substance, material or article used to verify the performance characteristics of an in vitro diagnostic (IVD) medical device (*8*).

Diagnostic specificity: the probability that the device gives a negative result in the absence of the target marker (*9*).

End-point titre: the reciprocal of the highest analyte dilution that gives a reading above the assay cut-off (*10*).

International measurement standard: a measurement standard recognized by signatories to an international agreement and intended to serve worldwide, for example a WHO International Standard (IS) (*5*).

International conventional calibrator: a calibrator whose value of a quantity is not metrologically traceable to SI units but is assigned by international agreement (*11*).

International Unit(s) (IU): the unitage assigned by WHO to an International Biological Standard (*2*).

Linearity (of a measuring system): the ability to provide measured quantity values that are directly proportional to the value of the measurand in the sample (*12*).

Potency: the specific ability or capacity of a product, as indicated by appropriate laboratory tests or by adequately controlled clinical data obtained through the administration of the product in the manner intended, to effect a given result (*13*).

Precision: (measurement) closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions (*5*).

Relative potency: a measure obtained from the comparison of a test to a standard on the basis of capacity to produce the expected potency (*13*).

Reference material: a material, sufficiently homogeneous and stable with regard to specified properties, which has been established to be fit for its intended use in measurement or in the examination of nominal properties (*5*).

Reference standard: a measurement standard designated for the calibration of other measurement standards for quantities of a given kind in a given organization or at a given location (*5*).

Sample-to-cut-off (S/CO) ratio (also signal-to-cut-off ratio): S/CO ratios are calculated by dividing the signal value (for example, optical density or relative light unit) of the sample being tested by the signal value of the enzyme immunoassay or chemiluminescence assay cut-off for that run. If the signal produced by a given test sample is equal to or greater than the calculated cut-off value then the specimen is considered to be reactive in the test. In competitive assays the relationship between the signal value of the sample and the signal value of the cut-off is reversed (CO/S ratio) (*14*).

Secondary (reference) standards: reference standards established by regional or national authorities, or by other laboratories, that are calibrated against and traceable to WHO reference materials (*2*).

SI, International System of Units: a system of units – based on the International System of Quantities – their names and symbols (including a series of prefixes and their names and symbols), together with rules for their use, adopted by the General Conference on Weights and Measures (*5*).

Traceability: (metrological) property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty (*5*).

Threshold cycle: the polymerase chain reaction (PCR) cycle at which the gain in fluorescence generated by the accumulating amplicon exceeds a threshold over baseline – for example, defined as 10 standard deviations of the mean baseline fluorescence using data taken from cycles 3 to 15 (*15*).

Working standard: a measurement standard used routinely to calibrate or verify measuring instruments or measuring systems (*5*).

Uncertainty: a parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand (*16*).

4. Principles of biological standardization

The aim of metrological traceability is to enable the results obtained by the calibrated routine measurement procedure to be expressed in terms of the values obtained at the highest available level of the calibration hierarchy (*11*). This is usually achieved in the clinical chemistry field by physicochemical reference methods, obtaining values in SI units.

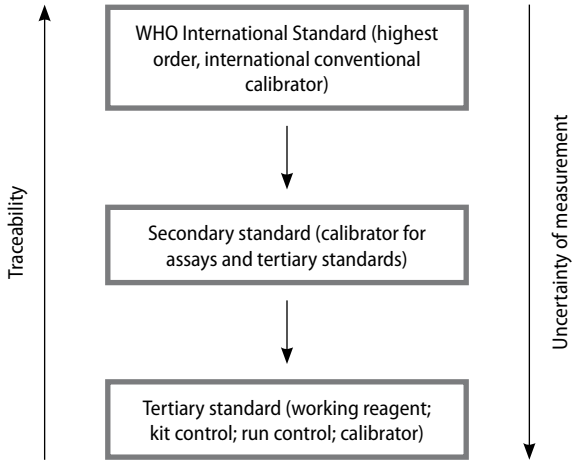
The majority of biological samples, containing for example nucleic acids, antigens or antibodies, are substances that cannot be fully characterized by a physicochemical reference method. Instead, biological assays are used for measurement of the potency or content of the analyte of interest. These methods are heterogeneous and the lack of a reference method does not permit the results to be expressed in absolute values according to the SI system.

The approach taken by the WHO Expert Committee on Biological Standardization to quantify biological materials is to first establish a highest order reference reagent – the IS. The procedure for the preparation, characterization and establishment of WHO IS preparations is described in detail elsewhere (*2*). Such material plays a crucial role in the standardization, harmonization and quality control of IVD assays, as was demonstrated in the 1990s when WHO IS were introduced for human immunodeficiency virus, hepatitis C virus and hepatitis B virus. These reference materials were fundamental in the regulation of IVD assays used for blood safety and for improving patient management in the clinical setting (*17*).

5. Calibration hierarchy of biological standards

"Reference material" is a generic term which refers to a material or substance whose property values are sufficiently homogeneous and stable, and whose fitness for purpose is well established, for its intended use in a measurement process (for example, the assessment of a measurement method or the assigning of values to materials) (*18*, *19*). Biological reference materials for a given analyte can be related through a sequence of comparisons to create a calibration hierarchy traceable to the highest order material – the WHO IS (Fig. A6.1).

Fig. A6.1
Calibration hierarchy and metrological traceability to the WHO IS



Furthermore, all biological standards have a range of key properties as summarized in Table A6.1.

Table A6.1
Key properties of WHO IS, secondary standards and tertiary standards

Property	WHO IS	Secondary standard	Tertiary standard
Alternative names	Highest order, international conventional calibrator	Regional or national reference materials, laboratory or manufacturer's working calibrator	Working reagents or standards, manufacturer's product calibrator, control material
Calibration	Evaluated in an international collaborative study, involving laboratories worldwide, different assays and different types of test laboratories (usually 15–30 participants)	Calibrated against the WHO IS	Calibrated against the secondary standard

Table A6.1 *continued*

Property	WHO IS	Secondary standard	Tertiary standard
Unitage	IU/mL	IU/mL	IU/mL
Traceability	N/A	Yes	Yes
Uncertainty of measurement	No	Yes (assay-specific)	Yes (assay-specific)
Commutability	Must be determined experimentally relative to clinical specimens	Should be determined experimentally relative to clinical specimen	Consideration should be given to experimentally determining relative to clinical specimen
Material	Should resemble, as closely and as feasibly as possible, the analyte being measured – for example, for assays for viral nucleic acids the standard will be the wild-type patient-derived virus in plasma (the normal sample type analysed)	Should resemble, as closely as possible, the analyte to be measured. However, for assay-specific secondary standards, synthetic materials such as armored RNA, plasmids and recombinant proteins, may be used and laboratories are encouraged to address commutability	Should resemble, as closely as possible, the analyte to be measured. Biological material similar to the tested sample, or non-biological materials, such as armored RNA, plasmids and recombinant proteins may be used, and laboratories are encouraged to address commutability
Typical final format of standard	Lyophilized	Lyophilized or liquid	Liquid
Usage	Calibration of secondary standards; initial validation of new assay/platform	Calibration of tertiary standards; working standards; run control; calibrator	Working standards; run control; calibrator

Table A6.1 *continued*

Property	WHO IS	Secondary standard	Tertiary standard
Establishment of standard	International agreement through a WHO international collaborative study, proposal for adoption and subsequent establishment by the WHO Expert Committee on Biological Standardization	May be calibrated in several ways: 1. In parallel with a study to establish the IS Example: **Appendix 1** 2. Regional or national collaborative study similar to the WHO collaborative study but with fewer participants Example: **Appendix 2** 3. Small study by one or a limited number of laboratories with a single assay or a limited number of different assays/ platforms Example: **Appendix 3**	1. Assay-specific study, normally by a single laboratory for use with a specific test/platform. 2. Small study by a limited number of laboratories with a single assay or a limited number of different assays/ platforms

5.1 WHO International Standards

WHO IS are defined by ISO17511 as International Conventional Calibrators and are the highest order of standard for biological references. They are solely established by the WHO Expert Committee on Biological Standardization following specific guidance (*2*).

Establishment of a WHO IS follows a collaborative study involving various users of the material (including national control laboratories, IVD manufacturers and other certified laboratories) and as many different, well-established assays as feasible. The laboratories should be chosen to reflect the global use of the standard and consideration should be given to the expertise of laboratories with a proven track record (perhaps through EQA schemes).

As with all biological references, the material used should resemble as closely as possible the natural analyte of the clinical sample to be measured. An assessment of commutability should be performed as part of the collaborative study where appropriate and feasible (*2*).

By definition, an IS has a specified value expressed in International Units (IU). This value is arbitrarily assigned based on the results of the collaborative study. The assigned IU value of each IS (new and replacement) does not carry an uncertainty associated with calibration (see section 8.1 below). The uncertainty is considered to be the variance of the vial weight determined during the filling process (*20*, *21*). The collaborative study design attempts to ensure continuity of the IU as far as possible. As explained further in section 8.1 there is no uncertainty value associated with replacements.

As the highest-order standard for biological material, the use of a WHO IS should be limited to the calibration of secondary biological reference materials in order to minimize the need to replace the IS on a regular basis. Unfortunately, the limited availability of secondary calibrator standards and the lack of specific guidance on the establishment and calibration of more readily available standards have resulted in the overuse of WHO IS for more routine procedures such as validation of assays and as run controls.

5.2 Secondary standards

A secondary standard is a material[1] that has been directly calibrated against the IS. These preparations usually include regional or national reference preparations. The titre, composition and method of production of secondary standards will vary but should be suitable for obtaining sufficient measurements, when dilution is needed, to achieve an accurate calibration. Regardless of the method of production, each calibration will have a stated measurement uncertainty (see section 8.1 below).

Secondary standards should be used for the calibration of tertiary standards. They should also be used for the calibration and validation of assay systems.

[1] Biological stock materials assigned a potency based on calibration against the IS by using exclusively one specific test, and used, for example, for the preparation of calibrators or run controls for this test, are also considered as in-house secondary standards. Non-biological preparations such as synthetic preparations (for example, plasmid preparations, transcripts, armored RNA and antigens produced by recombinant DNA technology) are often used for test calibration. If this calibration is done against an IS, these materials will also be considered as secondary standards under the scope of this document. Nevertheless, these materials have a number of limitations compared to the biological preparations, for example, in terms of commutability.

5.3 Tertiary standards

Tertiary standards are calibrated against secondary standards using the same calibration procedure. These standards usually include working standards or calibrators established for one specific assay used by a laboratory or other institution.

The standard may be formulated from either biological (for example, patient-derived) or non-biological material. However, regardless of the material used, all references in the traceability chain should also demonstrate commutability to the clinical sample of the tested analyte. Many of the principles discussed in this document also apply to the development of tertiary standards.

Tertiary standards are typically formulated as a liquid preparation and may comprise a concentration of the analyte that is detected without dilution in the linear range of the assay it is intended for. They will often be used as an external control material in addition to that normally supplied by the assay manufacturer. Regular monitoring of such material may allow for the early detection of problems with assay performance.

5.4 Other control material

Control material that does not follow the path of traceability back to the highest-order reference material (that is, to the IS) may be produced by commercial and in-house laboratories where no higher-order standard is available for the analyte.

The material may be used as a run control, whereby the unit of measurement (for example, signal-to-cut-off ratio (S/CO) values, threshold cycle values for real-time nucleic acid amplification technique (NAT)-based assays, copies/mL or genome equivalents/mL) can be used for intra-laboratory monitoring and may provide valuable trending data in a similar way to tertiary standards. However, such material has not been designed to allow for the comparison of results between different assays.

6. Commutability of biological standards

Commutability is a property of a reference material such that values measured for that reference material and for representative clinical samples have the same relationship between two or more measurement procedures for the same measurand (*10*) – that is, reference materials should behave in the same way as the native analyte itself. Producing biological reference materials that are commutable can be challenging because the matrix of the analyte may vary in different clinical conditions, or the analyte may be modified during preparation or processing of the reference material. Commutability can therefore only be demonstrated for particular combinations of assays, with particular clinical samples. It is not a generic property of the reference materials. Thus,

demonstrating commutability for two assay methods does not guarantee that there will be commutability with other methods. Similarly, if a set of samples demonstrate commutability with each other for particular assay methods then this does not guarantee that this will apply to all samples (*22*, *23*).

However, ISO17511:2003 states that calibrators are to be commutable at each step in the traceability chain (*11*). There are established recommendations for the assessment of commutability of reference materials used in laboratory medicine (*7*).

7. Selection and characterization of materials for the preparation of secondary standards – and application to tertiary standards

The basic parameters to be considered when selecting and characterizing a material for the preparation of a secondary standard are described in this section. In general, these considerations also apply to the process of developing tertiary reference materials to be used daily as control materials. In both scenarios, the selection of materials is dependent upon the type of technology to which they are to be applied, for example, NAT-based or immunological test systems. If necessary, expert scientific advice should be sought to support the development of any secondary standard.

The following should be considered:

- analyte – type, source and specificity
- immunological and genetic diversity
- type of matrix
- target concentration
- volume of final aliquots and storage temperature
- diagnostic specificity
- infectivity/inactivation
- physical appearance
- homogeneity
- commutability of the material
- stability – real-time and accelerated degradation.

7.1 Analyte – biological versus synthetic material

Consideration must be given to the most suitable form of the analyte to be established as a secondary standard – for example, whole organism, purified nucleic acid, recombinant protein, or laboratory-derived or clinical isolate.

Ideally, the material selected should resemble the analyte in the IS and in usual clinical specimens as closely as possible. The decision may be based on the availability of a sufficient volume of the material to enable preparation of a single batch of secondary standard that, when frozen, will last several years. In most cases, laboratory strains of microorganisms are better characterized and available in larger volumes than clinical strains. However, the latter may better represent the samples that are routinely tested. Where a whole organism is unavailable in sufficient quantities, a laboratory-derived material (such as a purified nucleic acid preparation) may be the only option. Demonstrating the commutability of such a material to different clinical samples may be challenging. This will need to be addressed on a case-by-case basis and should be done in cases where, through experimental assessment, it is proven that the use of laboratory-derived material improves agreement between assays. Analytes derived directly from human origin (such as a clinical sample) or the matrix of a biological sample (such as plasma or whole blood) should be tested and confirmed to be negative for the presence of pathogens other than the analyte of interest. This should be done to exclude potential cross-reactivity with the specific target analyte. If it is necessary to prepare a bulk material by pooling from more than one source, each component of the bulk material must be characterized and where possible all components should be identical – for example, in molecular detection the sequence of the target regions should be the same. All samples pooled must be mixed thoroughly and the pool should be homogeneous. It should be noted that pooling may not be appropriate in all circumstances. A biological bulk material with a high analyte concentration could if needed be diluted in a suitable matrix.

7.2 Immunological and genetic diversity

The detectability of an analyte by a particular assay may vary due to the immunological and genetic variability (serotypes, strains, variants, genotypes or subtypes and so on) of the organism being tested, resulting in suboptimal detection of particular variants. Therefore, the candidate material that best reflects the samples being tested should be chosen for the preparation of the secondary standard. Consideration should be given to the local geographical patterns of genetic diversity.

Where a standard is being prepared for nucleic acid detection, a well-characterized strain should be used for which the full genome sequence (or at least the sequence covering the most frequently amplified regions) is available. In principle, the same holds true for antigen standards where well-characterized antigen variants should be chosen. In both cases consideration should be given to the diagnostic implications of variant detection.

7.3 Matrix

The matrix in which the standard is formulated is crucial to creating a material that is fit for purpose. For NAT-based assays, the matrix needs to be appropriate for the assay or range of assays for which it is intended. Multiple sample types may be considered. It follows that for commutability purposes, a biological matrix would be preferred over a synthetic matrix. It should be taken into consideration that some sample matrices include inhibitory factors that interfere with the performance of specific types of assays (*7*). Furthermore, the chosen matrix of the reference preparation should be compatible with further matrices into which it may be spiked. For example, where a pathogen may be screened for in plasma, whole blood or urine, it may not be appropriate to formulate a material in a matrix that cannot be further diluted in clinical samples. In the case of plasma, consideration should also be given to any anticoagulant treatment and to any additional treatments such as cryoprecipitation or recalcification.

7.4 Concentration

Secondary standards will often be used in a quantitative capacity. Therefore, the concentration used should be high enough to permit the preparation of dilutions across the dynamic range of the assay and to allow for dilution into further matrices. It must be noted that dilutions performed on a high-titre secondary standard will contribute to the overall uncertainty arising from the dilution process. This will be the case for most secondary standards.

The target concentration of a standard will be dependent upon its final intended use and whether any clinical decision points exist when testing for the analyte. The detectable/quantitative range of all well-established assays for that pathogen must be taken into consideration. Reference materials that perform within the dynamic range of an assay (where changes in signal correspond to changes in analyte concentration) will typically be the norm.

In the case of tertiary standards where the material may be used as an external run control in qualitative tests, the concentration should ideally be at the lower end of the range of detection, at a concentration which will appropriately challenge the assays (for example, three times the 95% limit of detection of a NAT-based assay, or within the dynamic range of serological assays).

7.5 Volume

The aliquot volume may vary depending on the typical assay input volume for that analyte and the final storage temperature. The suitability of the container used for the filling of the aliquots should be validated in terms of the integrity and stability of the analyte. Where the standard is intended for single use, as defined in the Instructions for Use provided with the material, sufficient

volume should be provided for use in the assays for which it is intended and any remaining material must be discarded by the laboratory as no in-use stability testing will have been performed. If the standard is for multiple uses, the volume will depend on the required number of tests, and on stability at the recommended storage temperature. Where the manufacturer intends a material to be used on multiple occasions, suitable stability testing should be carried out to indicate the number of times the material can be removed from its storage temperature, allowed to warm to ambient temperature and then be re-cooled before the stability of the material is compromised.

7.6 Diagnostic specificity

Samples derived directly from human origin (for example, clinical samples or where the matrix is a biological component such as plasma or whole blood) should be tested and confirmed to be negative for the presence of pathogens other than the analyte of interest in order to minimize risk and the potential for cross-reactivity with the target analyte.

Where human material is used as the diluent, the diluent should be tested and found to be negative for both the analyte and common high-risk pathogens.

7.7 Infectivity

Use of infectious materials will impact on processing, handling and shipping of the standard. Where samples are prepared from infectious material, it is important to provide clear information to the end users about the exact nature of the material and origin of the pathogens. Import regulations differ from country to country and can vary depending on the origin of the pathogen. For example, tissue-culture-derived viral specimens may be subjected to different shipping regulations by some countries than a patient sample infected with this virus.

Where standards are prepared from inactivated materials it is important to confirm the success of the inactivation procedure and to determine potential effects of the inactivation on the performance of the final standard. The use of established or proven inactivation methods (such as published methods) is preferred. In addition, in cases where standards are prepared from inactivated material but are diluted in a biological matrix (for example, human plasma) the matrix should be screened by NAT-based and serological assays for the most common bloodborne viruses.

7.8 Physical appearance

A number of factors will determine the most appropriate physical appearance for the standard. Typically, IS are lyophilized preparations which have a better stability and longer shelf-life than, for example, liquid preparations. They are

also required to be shipped around the world, preferably at ambient temperature. This may also be the case for some secondary standards, especially national or regional secondary standards, such as the European Pharmacopoeia Biological Reference Preparations for NAT-based assays. A feasibility study should be conducted to demonstrate that the freeze-drying procedure does not have an adverse effect on the integrity of the target analyte. However, lyophilization is costly and specialized, and, for some pathogens, maintaining a stable product is problematic. In these cases, a liquid preparation stored and shipped at a suitable temperature may be more appropriate. Lyophilization also requires additional validation work to determine the potential impact of such a technique on the biological activity of the standard and on the commutability of the standard with clinical samples. In the case of nucleic acid extraction, lyophilization can lead to the formation of aggregates which reduces extraction efficiency. Lyophilized preparations should be evaluated against the liquid bulk preparation in different assays as part of the commutability assessment. Where standards are used frequently, such as tertiary control materials, a liquid or single-use frozen preparation is probably more suitable so that the standard is ready for use without the need for reconstitution.

7.9 Homogeneity

It is important to confirm the consistency of the filling procedure and to confirm that the bulk was dispersed (for example, stirred) sufficiently throughout. Homogeneity is assessed in two ways – by determining both the biological and the physical content (weight or volume) of multiple vials across the batch. The latter is particularly important prior to lyophilization, and can be addressed by weighing a proportion of vials before and after filling and then calculating the filled weight and associated coefficient of variation as a percentage (% CV). The % CV will be higher for a more viscous matrix. It is also important to determine the biological homogeneity by assessing the concentration of the analyte in multiple vials across the batch. It is known that homogeneity may be impaired by genetic quasi-species heterogeneity or antibody complexation. The number of vials used for testing will depend on the batch size. As a minimum, typically 1–2% of the vials should be tested (*17*).

7.10 Stability

A stability testing programme should be implemented to monitor the potency of the secondary standard over time. Stability monitoring can be based on real-time data. However, additional data from accelerated thermal degradation studies are helpful in characterizing the robustness of lyophilized reference materials. Such data are also important for assessing the suitability of the material after extreme shipping conditions.

Factors affecting stability will be dependent upon the physical appearance and final storage temperature of the material. For example, for final long-term storage at 2–8 °C, preservatives might need to be included in the final formulation to prevent fungal or bacterial growth. Likewise, lyophilized preparations may require different additives, such as inert sugars, which might aid preservation and protect pathogen viability upon reconstitution. It should be noted that the effect of any additives on downstream performance or commutability should be evaluated. Factors affecting the stability of lyophilized preparations also include residual moisture and oxygen, both of which can compromise the integrity of the lyophilized product. It is advisable to assess the levels of residual moisture and oxygen in such preparations following lyophilization and if possible during storage of the material. This is particularly important for stoppered vials which may permit the ingress of air during the lifetime of the product, causing displacement of the vacuum or inert gas which the material is held under. Loss of potency may occur as a result.

7.11 Stability assessment during product lifetime

7.11.1 Real-time stability

Materials should be periodically removed from their designated storage temperatures for testing in the routine laboratory assay of choice. The required frequency of testing will be dependent upon the physical appearance and final storage temperature of the material. For example, a frozen liquid product stored at −20 °C may require more frequent testing than a lyophilized product stored at −20 °C. Real-time monitoring may be more frequent following production of a new material, but by monitoring over time the frequency of testing could be reduced. For example, a liquid-filled preparation stored at −20 °C could be tested every 3 months for the first year following production. If the data suggest good stability then the testing interval could be increased to 6 months. Likewise, a lyophilized product may be tested every 6 months for the first year following production, but further testing could be reduced to annually. Any assessment of stability and associated outcome should be referenced on the Instructions for Use distributed with the materials, for example where an acceptable number of freeze–thaw cycles has been determined, this should be referenced.

7.11.2 Accelerated thermal degradation studies

Real-time stability studies may not demonstrate loss in analyte concentration over the testing period. For lyophilized material, accelerated thermal degradation studies can be used to predict the long-term stability of a product from its performance at elevated temperatures. For example, data demonstrating a loss in titre after 3 months at 37 °C can be used to predict the time it would take

for the same amount of degradation to occur at −20 °C or at another chosen baseline storage temperature of the product. In addition, accelerated thermal degradation studies cover the validation of the use of the reference material after extreme shipment conditions.

A chosen number of samples should be stored in temperature-controlled environments at for example, −20 °C, 4 °C, 20 °C and 37 °C. Studies at 45 °C and 56 °C may also be suitable for some pathogens and matrices but not all. For example, a lyophilized plasma matrix will generally not reconstitute following long-term storage at the higher temperatures indicated; however, this may not fully reflect the degradation of the analyte. Vials should be periodically removed and tested alongside vials that have been continuously stored at the recommended storage temperature (reference point).

The Arrhenius equation can be applied to the resulting data in order to predict the rate of degradation of the material at the recommended storage temperature (*24*, *25*).

While accelerated thermal degradation studies have considerable usefulness in the production of secondary reference materials, they may be less useful for tertiary standards, for which real-time monitoring will provide the most valuable data set. Note that for tertiary standards used as external quality control IVDs, accelerated stability testing is usually performed prior to release. In all cases, stability testing protocols should be designed in conformity with ISO 23640:2013.

7.12 In-use stability

In-use stability testing of the standard measures the stability of the standard once it has been thawed, opened or reconstituted (depending on storage conditions and physical appearance). It is important to establish in-use stability if the material is intended to be used on multiple occasions and if it is stored under different conditions during this period. Where multiple freeze–thaw events are likely, the effect of these should be evaluated.

7.12.1 Lyophilized preparations

Materials should be reconstituted as defined in the Instructions for Use, aliquoted and stored at an appropriate temperature (for example, −80 °C for materials containing free RNA as the analyte).

Reconstituted materials should be subjected to freeze–thawing cycles and tested at predetermined intervals – for example, weekly, monthly, annually or biannually, and monitored by suitable assays (preferably quantitative assays if available). Tests should be performed on at least three replicates per time point to determine any loss in concentration.

7.12.2 Frozen preparations

If the total volume in each vial allows for multiple uses, at least three vials should be freeze–thawed on multiple occasions. At each time point the samples should be tested at least in duplicate in three replicate tests. If no degradation is observed then further time points should be added up to the limit of volume remaining in the vial.

7.12.3 Liquid preparations

Where materials are routinely held at 2–8 °C, exposure to short periods at ambient temperature may occur when the product is in use. Where this is the case multiple exposures to ambient temperatures should be assessed. Three vials should be removed from storage, left at room temperature for up to 1 hour and tested at least in duplicate in triplicate tests. These vials should then be returned to 2–8 °C and the process repeated with the same vials at frequent intervals. Interval frequency can be determined by the accumulation of data points and may be reduced following an observation of good stability, as discussed in section 7.11.1 above.

8. Calibration: testing and statistical analysis

8.1 Principles of calibration

Calibration is the process by which a concentration is assigned to a reference by the direct comparison of measurements with a higher-order reference, and represents one of the crucial stages of the establishment of a secondary standard. Each calibration of a candidate secondary standard has to be performed in parallel with the higher-order reference, in this case the WHO IS. The following sections describe the minimum requirements for the calibration of secondary standards intended for either one specific method in one laboratory (single assay calibration) or for multiple methods (collaborative study calibration). In both cases, several independent runs with the candidate standard and the IS in parallel have to be performed (same assay using the same test conditions). For each run, a new vial of each standard should be used and diluted in the matrix validated for the respective assay (for example, negative human plasma).

This WHO guidance document reflects the common statistical methods used for the calibration of reference materials. Any other statistical method that has been demonstrated to be a reliable approach to the calibration of such materials may also be applied. Appropriate software for the statistical analysis should be available for evaluation of the data, and the statistical analysis should be performed by staff with expertise in this field.

The calibration study data should be analysed using the relevant statistical models for bioassays, for example, using the methods recommended

by the European Pharmacopoeia (*26*). The design of the study should take into consideration that:

- each analyte, IS and the candidate secondary standard should be tested with the same number of dilutions and replicates per dilution;
- the adjacent dilution steps should be equally spaced.

The statistical validity of the fitted model should be assessed for each individual assay. For the parallel-line and Probit models – as the most appropriate and proven statistical methods for this analysis – the linearity and parallelism of the logarithmic dose–response relationships between the IS and secondary standard should be evaluated. If the assay response is linear and the response lines are parallel then the relative potency of the candidate secondary standard against the IS can be calculated. Using the parallel-line model, validity criteria for linearity could include the coefficient of determination (r^2) or a test for nonlinearity (*26*). Parallelism could be demonstrated, for example, by means of a test for non-parallelism or an equivalence approach for the difference or ratio of slopes (that is, the 95% confidence interval for the ratio of slopes must lie entirely between predefined equivalence margins). In addition, the precision with which the potency has been estimated should be provided.

Each calibration will have a stated measurement uncertainty. This estimate can be determined by identifying all sources of variation, calculating the extent of variation and using established methods to combine the uncertainty. The measurement uncertainty associated with assigning a value to the standard is test-system specific (*7*, *8*). It should be noted that an IS, by definition, has a specified value which has been assigned and expressed in IU per millilitre (IU/mL). As a consequence of defining the IU as a fraction of the contents of the container of the current IS, and because the units defined by any previous IS formally cease to exist, an uncertainty value is not given to the assigned IU (*2*). The variability of the vial weight during filling for each IS is quoted in the study report and in the Instructions for Use accompanying the standard.

8.2 Single assay calibration using qualitative tests

8.2.1 NAT-based assays

For qualitative NAT-based assays, four independent runs should be performed. The first run involves testing serial ten-fold dilutions (until negative) of the IS and the candidate secondary standard in duplicate, and is intended to determine the end-point dilution of both standards. In the subsequent three assays, two half-$\log_{10}$ dilutions either side of the predetermined end-point (5 dilutions in total) of the standards should be tested. Each dilution (runs 2–4) should be tested at least in duplicate, giving in total 6 replicates per dilution across all runs. For each assay, data from all runs at each dilution step will be pooled to

give a number of positive results out of the total number of tests performed. The Probit analysis will estimate the concentration at which 63% of the samples tested were positive (that is, the dilution at which on average one single copy per sample tested could be expected under the assumption of an underlying Poisson distribution). The calculated end-point is used to give estimates expressed in NAT-detectable units/mL after correcting for an equivalent volume of the test sample. It should be noted that these estimates are not necessarily directly equivalent to a genuine genome equivalent number or copy number per mL. The software for the Poisson distribution will calculate the proportion of potency of the test sample (candidate secondary standard) relative to the potency of the standard sample (that is, the IS) so long as the dose–response curves fit within the statistical model.

Using a real-time NAT-based assay, the calibration can be determined from the cycle threshold (Ct) values by applying a parallel-line analysis, conditional on the assumption that the slope fulfils the requirement of −1/log2. The number of dilutions and the number of replicates per dilution should follow the instructions given below in section 8.3.1.

8.2.2 Antigen assays

The IS and the candidate secondary standard should be tested in three independent assays. Both standards should be diluted, using serial half-$\log_{10}$ or two-fold dilutions, in the diluent appropriate for the assay. The dilution ranges should be within the detection range of the assays used for the study and should span the end-point titre (intercept with the cut-off of the assays). The analytical sensitivity of each assay can be calculated by linear interpolation using the two dilutions of the dilution series having values below and above the assay cut-off. The (log-transformed) data should be evaluated against the results obtained for the secondary standard using a parallel-line assay analysis to estimate the potency (IU/mL) of the secondary standard relative to the IS. A logarithmic transformation of the assay response may be necessary if the dilution range was chosen around the sample cut-off rather than the dynamic range of the assay.

8.3 Single assay calibration using quantitative tests

8.3.1 NAT-based assays

For the calibration of secondary standards tested in a quantitative NAT-based assay, the candidate material should be tested neat (where possible) and at two or three further (for example, ten-fold) dilutions within the linear range of the assay to obtain at least three concentrations giving quantitative values. The same methodology applies to the IS, with the exception that this material should be diluted starting from a concentration as close as possible to the estimated potency of the secondary standard (as indicated by preliminary tests). All standards

should be tested in duplicate and at least three independent runs should be performed. Where possible, multiple assay lots and reagents should be included in this testing. The calculation of the potency of the secondary standard may be performed in one of several ways:

- The assay output (for example, copies, genome equivalents[2] or IU/mL) should be analysed by the parallel-line method using, if necessary, log-transformed data to obtain a "relative potency" in IU/mL of the secondary standard against the IS (where the slope fulfils the requirement of −1/log2 corresponding to a value of S = −3.322). The parallel-line method should be the preferred option for the data analysis.
- The difference in estimated potency (using the test software) between the candidate secondary standard and the IS (log-transformed data) can be used to determine the potency of the secondary standard. The difference is then subtracted from the log-transformed nominal IU/mL of the IS to obtain the potency of the secondary standard.
- A standard curve generated by the instrument software using the IS as the standard may be used to determine the potency of the secondary standard.

8.3.2 Antigen assays

As with the quantitative NAT-based assays, the candidate material and the IS should be tested neat and at two or three further (for example, ten-fold) dilutions within the linear range of the assay, using the dilution matrix appropriate for the assay. All samples should be tested in duplicate in at least three independent runs. The calculation of the potency of the secondary standard can be performed in one of several ways:

- The results obtained with the parallel-line analysis (if necessary on log-transformed data) should be used to give a "relative potency" of the secondary standard against the IS in IU/mL. The parallel-line analysis should be the preferred option for the data analysis.
- A standard curve generated by the instrument software using the IS as the standard may be used to determine the unitage of the secondary standard.

[2] In the case of assays not yet calibrated against the IS.

8.4 Collaborative study calibration using multiple assays

Secondary standards that are intended to be used in different assays by multiple end users should be calibrated using a collaborative study approach. The amount of work and resources required to perform such a study should not be underestimated. The collaborative study should be organized by/or with advice from a body with experience in this field, such as a WHO collaborating centre. If necessary, a scientific adviser from the field should be identified to support the collaborative study, including the selection of study participants. Owing to the complexity of the reported data, which typically include data from many different types of assays, the statistical analysis should be performed by a statistician. The general principles of planning and executing these types of collaborative study are described in section 6 of the WHO *Recommendations for the preparation, characterization and establishment of international and other biological reference standards* (*2*).

Results from all participants should be analysed by statistical methods described and considered appropriate by the responsible statistician. This analysis typically requires access to suitable computing facilities and statistical software. The testing requirements and protocol of each laboratory/test should follow the protocol described for the single assay calibration depending on the assay type (qualitative or quantitative). The results of each assay method should be analysed separately and should provide an estimate of the relative potency and precision of the candidate secondary standard against the IS.

The variation in results between test methods, and between laboratories, should be described and assessed as part of the statistical analysis (precision and consistency of the results). An assessment should be made of any factors causing significant heterogeneity of the estimated potency, nonlinearity and differences in slopes. Although there is no generic outlier detection rule from the statistical point of view, the exclusion of data should be taken into account in subsequent analysis where striking differences in results within assays, between assays, or between participants or test methods are observed. All valid potency estimates for the candidate secondary standard should be combined to produce a final geometric mean potency/content with 95% confidence limits. The results of the tests should be displayed graphically, for example, as histograms or scatter plots. Tables of the quantitative and qualitative raw data should be included in the report annexes.

8.5 Calculation of uncertainty of measurement

The assignment of an uncertainty value must be applied to secondary reference materials. The uncertainty of an observed value is a property of the test system and is not the effect of mistakes introduced through human error. The

measurement of uncertainty is complex and, where possible, advice should be sought from a statistician. This section aims to give an overview of the area and to highlight that consideration must be given to the assignment of an uncertainty value to secondary references.

The uncertainty – often referred to as "measurement uncertainty" (MU) – expresses the 95% confidence limits either side of the observed value assigned to a product. By estimating the MU of a product the confidence in the final value assigned is shown. Uncertainty should be calculated using log-transformed data. For example, a value of 3 $\log_{10}$ IU/mL may be assigned to a standard. Following estimation of uncertainty this value may be displayed as 3 $\log_{10}$ IU/mL ± 0.2 – that is, the value of the material could range from 2.8–3.2 $\log_{10}$ IU/mL.

There are many aspects to uncertainty and well-documented examples of how to estimate uncertainty (*20*, *27–29*). However, one simple approach to estimating MU for a secondary standard is to test the material multiple times on different occasions (but always using the same test system) in parallel with the WHO IS – that is, test the two standards under exactly the same conditions.

As a guide, a minimum of 30 test results for both the secondary standard and the IS should be generated from a combination of at least three independent tests. The more times the sample is tested the better. The test system used should be of the highest order possible, that is, a commercial assay, or, in the absence of such, a well-validated laboratory-developed test.

After determining the mean and standard deviation of the data points, dividing the standard deviation by the square root of the number of samples tested gives the standard uncertainty (or standard error) of measurement. This approach demonstrates the imprecision involved and does not account for MU derived from inherent bias.

9. Establishment of reference materials in the absence of a WHO International Standard

In the absence of a WHO IS, an established regional or national reference standard may become the standard of comparison for the candidate assay. Current examples include: West Nile virus NAT reference reagent (Health Canada), and Chikungunya virus standard (Center for Biologics Evaluation and Research, United States Food and Drug Administration). These standards should be characterized through extensive analysis. The methods for their characterization, preparation and storage should ideally be published in peer-reviewed journals.

The calibration of reference standards where no IS and secondary standards are available should follow the WHO-recommended principles for

the preparation of international biological standards (*2*). Where no IS exists but a regional standard is available the calibration of a reference standard of lower order should follow the guidance provided in this document.

In the absence of an IS or a national reference standard, a candidate reference material may be assigned a value using either a commercial assay or a laboratory-developed test. Alternatively, the material may be calibrated against other commercially available preparations or samples from EQA programmes with assigned values, or a manufacturer's working calibrator. For such reference materials, reserving a separate proportion of vials (baseline samples) to be used in the calibration of subsequent batches is strongly advised. This batch should be stored at the lowest possible temperature validated for the vials, preferably at a lower temperature than that at which the bulk of the reference material is stored (for example, if a reference material is routinely stored at −70 °C, the baseline samples should be stored under nitrogen vapour or liquid). Reference materials may also be assigned a value based on a range of assays (collaborative study) as described above rather than a single assay. In such cases, the assigned value will be the mean of the results reported by all the assays. Whichever approach is used, the method and assigned value need to be documented (*20*).

10. Post-production considerations

10.1 Storage of the material

Following the development and production of a batch of secondary reference materials, the material should be stored at an appropriate temperature which ensures stability throughout the lifetime of the product. The temperature of the cold-storage unit should be monitored and recorded. A protocol should be developed for the real-time monitoring of each product and should include details of testing frequency, number of replicates, methodology used and statistical analysis (see section 7.11 above).

10.2 Distribution of the material

Consideration should also be given to the method by which the material is to be distributed. Where material is potentially infectious, specialist couriers may be required. Some countries have import-permit requirements for infectious materials. These requirements are country-specific and should be discussed with the recipient in advance of the shipment.

Stability of the product during transportation can be addressed by distributing the material at appropriate temperatures, employing where necessary the use of dry ice or cold packs. The use of dry ice may be considered a hazard by some couriers and may require the use of specialized companies.

10.3 Instructions for Use

Detailed Instructions for Use should be supplied with every shipment. These should include the following information:

- characterization details of the analyte
- storage conditions
- procedures prior to use (for example, reconstitution)
- appropriate use
- stability information
- safety information
- references to any publications relating to the material (for example, study report).

10.4 Replacement batches

The need for replacement batches should be addressed in the development stages of the initial product. Consideration should be given to the acquisition of material for future replacement of the standard. For example, if the analyte comprises tissue-culture-derived material, large batches of stock material could be cultured in the first instance. Detailed records of the production of the first batch should also be documented to allow replication of the production method at some point after the initial material was produced.

11. End user advice

Manufacturers are encouraged to include details of the production and calibration process in the Instructions for Use provided with each material, or to provide a reference to where this information can be found. Any additional information not supplied could be requested from the manufacturer, and could include:

- number of replicates and methods used to assess repeatability and reproducibility;
- the metrological traceability of an assigned unit;
- whether a collaborative study was performed, and if so the number of participants and range of assays evaluated;
- assessment of performance in a different matrix;
- stability assessments including of shelf-life and in-use stability;
- assessment of specificity;
- validation of limit of detection or cut-off.

Authors and acknowledgements

This WHO guidance document was jointly drafted by Dr M. Chudy, Paul-Ehrlich-Institut, Germany; Dr C. Morris, National Institute for Biological Standards and Control, England; Dr J. Fryer National Institute for Biological Standards and Control, England; Dr W. Dimech, National Reference Laboratory, Australia; and Dr J. Saldanha, Immucor, Inc., the USA.

The first draft was then discussed at a Standardisation of Genome Amplification Techniques (SoGAT) workshop in 2015 and presented to the WHO Expert Committee on Biological Standardization in the same year. Further comments on the draft were received from IVD manufacturer associations, individual IVD manufacturers, regulatory bodies and experts in the field. A revised draft was then discussed at a 2016 SoGAT workshop. Acknowledgement is made to all delegates of the SoGAT 2015 and 2016 workshops for their critical reviewing of the draft versions and other inputs.

The second draft version was published on the WHO Biologicals website for a round of public consultation between 16 June and 16 September 2016, and the comments received were incorporated to produce the document WHO/BS/2016.2284.

Further changes were subsequently made to document WHO/BS/2016.2284 by the WHO Expert Committee on Biological Standardization.

References

1. Guidelines for the preparation and establishment of reference materials and reference reagents for biological substances. In: WHO Expert Committee on Biological Standardization: twenty-ninth report. Geneva: World Health Organization; 1978: Annex 4 (WHO Technical Report Series, No. 626; http://apps.who.int/iris/bitstream/10665/41323/1/WHO_TRS_626.pdf, accessed 5 February 2017).
2. Recommendations for the preparation, characterization and establishment of international and other biological reference standards (revised 2004). In: WHO Expert Committee on Biological Standardization: fifty-fifth report. Geneva: World Health Organization; 2006: Annex 2 (WHO Technical Report Series, No. 932; http://www.who.int/immunization_standards/vaccine_reference_preparations/TRS932Annex%202_Inter%20_biol%20ef%20standards%20rev2004.pdf?ua=1, accessed 5 February 2017).
3. WHO manual for the establishment of national and other secondary standards for vaccines. Geneva: World Health Organization; 2011 (WHO/IVB/11.03; http://apps.who.int/iris/bitstream/10665/70669/1/WHO_IVB_11.03_eng.pdf, accessed 5 February 2017).
4. General guidelines for the establishment, maintenance and distribution of chemical reference substances. In: WHO Expert Committee on Specifications for Pharmaceutical Preparations: forty-first report. Geneva: World Health Organization; 2007: Annex 3 (WHO Technical Report Series, No. 943; http://apps.who.int/iris/bitstream/10665/43632/1/WHO_TRS_943_eng.pdf, accessed 5 February 2017).

5. International vocabulary of metrology – Basic and general concepts and associated terms (VIM) third edition. JCGM 200:2012 (http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf, accessed 5 February 2017).

6. Guidance for industry. Bioanalytical method validation. Rockville (MD): Center for Veterinary Medicine; 2013 (http://www.fda.gov/downloads/Drugs/Guidance/ucm070107.pdf, accessed 7 February 2017).

7. CLSI. Characterization and qualification of commutable reference materials for laboratory medicine; approved guideline. CLSI document EP30-A. Wayne (PA): Clinical and Laboratory Standards Institute; 2010 (sample: http://shop.clsi.org/site/Sample_pdf/EP30A_sample.pdf, accessed 7 February 2017).

8. ISO 15198:2004. Clinical laboratory medicine – in vitro diagnostic medical devices – validation of user quality control procedures by the manufacturer. Geneva: ISO; 2004 (abstract: http://www.iso.org/iso/catalogue_detail.htm?csnumber=39751, accessed 7 February 2017).

9. 2009/886/EC: Commission Decision of 27 November 2009 amending Decision 2002/364/EC on common technical specifications for in vitro diagnostic medical devices (notified under document C (2009) 9464) (Text with EEA relevance). European Commission. Official Journal of the European Union. 2009;L318/25–L318/40 (http://eur-lex.europa.eu/eli/dec/2009/886/oj, accessed 7 February 2017).

10. Frey A, Di Canzio J, Zurakowski D. A statistically defined endpoint titer determination method for immunoassays. J Immunol Methods. 1998;221(1–2):35–41 (https://www.researchgate.net/publication/13394025_A_statistically_defined_endpoint_titer_determination_method_for_Immunoassays, accessed 7 February 2017).

11. 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: ISO; 2003 (abstract: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=30716, accessed 7 February 2017).

12. ISO 18113-1:2009. In vitro diagnostic medical devices – information supplied by the manufacturer (labelling) – Part 1: Terms, definitions and general requirements. Geneva: ISO; 2009 (https://www.iso.org/obp/ui/#iso:std:iso:18113:-1:ed-1:v1:en, accessed 7 February 2017).

13. United States Pharmacopeia 38 National Formulary 33 (USP 38-NF 33). United States Pharmacopeial Convention; 2015.

14. Linden JV, Bianco C, editors. Blood safety and surveillance. New York/Basel: Marcel Dekker, Inc.; 2001.

15. Jung R, Soondrum K, Neumaier M. Quantitative PCR. Clin Chem Lab Med. 2000;38(9):833–6 (abstract: https://www.ncbi.nlm.nih.gov/pubmed/11097336, accessed 7 February 2017).

16. ISO/IEC Guide 98-3:2008. Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995). Geneva: ISO; 2009 (abstract: http://www.iso.org/iso/catalogue_detail.htm?csnumber=50461, accessed 7 February 2017).

17. Resolution WHA63.12. Availability, safety and quality of blood products. Sixty-third World Health Assembly, Geneva, 17–21 May 2010. Geneva: World Health Organization; 2010 (http://apps.who.int/gb/ebwha/pdf_files/WHA63-REC1/WHA63_REC1-en.pdf, accessed 7 February 2017).

18. Analytical methods and reference materials program. AMRM Glossary of Terms [website]. NIH, US Department of Health & Human Services (https://ods.od.nih.gov/Research/AMRMGlossary.aspx, accessed 7 February 2017).

19. ISO 15194:2009. In vitro diagnostic medical devices – measurement of quantities in samples of biological origin – requirements for certified reference materials and the content of supporting documentation. Geneva: ISO; 2009 (https://www.iso.org/obp/ui/#iso:std:iso:15194:ed-2:v1:en, accessed 7 February 2017).

20. Dimech W, Francis B, Kox J, Roberts G. Calculating uncertainty of measurement for serology assays by use of precision and bias. Clin Chem. 2006;52(3):526–9 (https://www.researchgate.net/publication/7362486_Calculating_Uncertainty_of_Measurement_for_Serology_Assays_by_Use_of_Precision_and_Bias, accessed 7 February 2017).

21. Principles and methods of validation of diagnostic assays for infectious diseases. In: Manual of diagnostic tests and vaccines for terrestrial animals [web version]. Paris: World Organisation for Animal Health; 2016. Chapter 1.1.6, adopted in May 2013 (http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.06_VALIDATION.pdf), accessed 7 February 2017).

22. Miller WG, Myers GL. Commutability still matters. Clin Chem. 2013;59(9):1291–3 (http://clinchem.aaccjnls.org/content/59/9/1291, accessed 7 February 2017).

23. WHO Consultation on commutability of WHO Biological Reference Preparations for in vitro detection of infectious markers. Meeting Report. Geneva: World Health Organization; 2013 (WHO/BS/2013.2230 Addendum 1; http://www.who.int/entity/bloodproducts/norms/BS_2230_Addendum1_Commutability.pdf?ua=1, accessed 27 March 2017).

24. Kirkwood TB. Predicting the stability of biological standards and products. Biometrics. 1977; 33(4):736–42 (abstract: https://www.ncbi.nlm.nih.gov/pubmed/588659, accessed 7 February 2017).

25. Egan W, Schofield T. Basic principles of stability. Biologicals. 2009;37(6):379–86; and discussion 421–3 (abstract: http://www.sciencedirect.com/science/article/pii/S1045105609001158, accessed 7 February 2017).

26. Statistical analysis of results of biological assays and tests. 01/2008:50300. European Pharmacopoeia 6.0, section 5.3 (http://www.uspbpep.com/ep60/5.3.%20%20statistical%20analysis%20of%20results%20of%20biological%20assays%20and%20tests%2050300e.pdf, accessed 7 February 2017).

27. Konnert A, Berding C, Arends S. Uncertainty calculation for calibrators and controls of laboratory diagnostic assays. Clin Chem Lab Med. 2006;44(10):1175–82 (abstract: https://www.researchgate.net/publication/6763360_Uncertainty_calculation_for_calibrators_and_controls_of_laboratory_diagnostic_assays, accessed 7 February 2017).

28. Bell S. A beginner's guide to uncertainty of measurement. Measurement good practice guide No. 11 (Issue 2 with amendments 2001). Teddington: National Physics Laboratory; 1999 (abstract: http://www.npl.co.uk/publications/a-beginners-guide-to-uncertainty-in-measurement, accessed 7 February 2017).

29. Tang L, Sun Y, Buelow D, Gu Z, Caliendo AM, Pounds S et al. Quantitative assessment of commutability for clinical viral load testing using a digital PCR-based reference standard. J Clin Microbiol. 2016;54(6):1616–23 (http://jcm.asm.org/content/54/6/1616.long, accessed 7 February 2017).

Appendix 1

Example of the parallel calibration of a secondary standard in a study to establish the International Standard

WHO/BS/2016.2291
Page 1

World Health Organization

WHO/BS/2016.xxxx
ENGLISH ONLY

EXPERT COMMITTEE ON BIOLOGICAL STANDARDIZATION
Geneva, 17 to 21 October 2016

Collaborative Study to Evaluate the Proposed 4th WHO International Standard for Hepatitis B Virus (HBV) DNA for Nucleic Acid Amplification Techniques (NAT)

Jacqueline F. Fryer [1,3], Rehan Minhas [1], Thomas Dougall [2], Peter Rigsby [2], Clare L. Morris [1] and the Collaborative Study Group *

*[1] Division of Virology and [2] Biostatistics, National Institute for Biological Standards and Control, South Mimms, Potters Bar, Hertfordshire, EN6 3QG, UK; [3] Study coordinator, E-mail Jacqueline.Fryer@nibsc.org; * Details provided in Appendix 1*

NOTE:
This document has been prepared for the purpose of inviting comments and suggestions on the proposals contained therein, which will then be considered by the Expert Committee on Biological Standardization (ECBS). Comments MUST be received by **16 September 2016** and should be addressed to the World Health Organization, 1211 Geneva 27, Switzerland, attention: Technologies, Standards and Norms (TSN). Comments may also be submitted electronically to the Responsible Officer: **Dr C M Nübling** at email: nueblingc@who.int.

Summary

This report describes the collaborative study evaluation of the replacement 4th WHO International Standard for hepatitis B virus (HBV) DNA for the calibration of secondary reference preparations for HBV nucleic acid amplification techniques (NAT). The candidate 4th HBV International Standard was prepared and evaluated as part of the collaborative study to establish the 3rd HBV WHO International Standard in 2011. The lyophilized preparation comprises a dilution of the same HBV DNA-positive plasma sample as used for the previous HBV International Standards, in pooled human plasma. In this collaborative study, thirteen laboratories from seven countries evaluated the suitability of the candidate using their routine HBV NAT-based assay. The candidate NIBSC code 10/266 (sample 2) was evaluated alongside the 3rd HBV WHO International Standard, NIBSC code 10/264 (sample 1), three HBV reference preparations (samples 3-5) and three HBV-positive plasma samples comprising different HBV genotypes (samples 6-8). A range of HBV NAT assays were used in the evaluation, the majority of which were commercial quantitative assays based on real-time PCR technology.

The overall mean potency estimates for samples 1 and 2, were 5.94 and 5.97 $\log_{10}$ IU/mL respectively. These values are very similar to the values obtained for these samples relative to the pre-existing 2nd HBV WHO International Standard in the 2011 collaborative study (5.93 and 5.98 $\log_{10}$ IU/mL respectively). The standard deviation in individual laboratory mean estimates for samples 1 and 2 was 0.13 $\log_{10}$ IU/mL. The overall mean potency estimate for sample 2, relative to sample 1 was 5.96 $\log_{10}$ IU/mL. The results obtained from ongoing accelerated thermal degradation studies indicate that the candidate sample 2 has remained stable over the 5 years post-manufacture.

The results of the study indicate the suitability of sample 2 as the replacement 4th HBV WHO International Standard. Since the overall mean potency obtained for the candidate in this collaborative study is very similar to the overall mean potency obtained in the 2011 collaborative study, relative to the pre-existing 2nd HBV WHO International Standard, it is proposed that the value assigned to the candidate sample 2 is that obtained in the 2011 collaborative study. This approach would minimize any potential drift in the value of the IU during the replacement. It is therefore proposed that candidate sample 2 (NIBSC code 10/266) is established as the 4th WHO International Standard for HBV DNA for NAT with an assigned potency of 955,000 IU/mL (~5.98 $\log_{10}$ IU/mL) when reconstituted in 0.5 mL of nuclease-free water.

Introduction

Hepatitis B virus (HBV) remains a major public health problem worldwide, despite the availability of an effective vaccine and antiviral therapies. More than 240 million people worldwide are chronically infected, with 0.5-1 million dying annually as a result of serious liver disease [1]. The virus is transmitted in blood and body fluids, perinatally and through close person-to-person contact in early childhood (in regions with high HBV prevalence), and through infected needles and sexual contact (in regions with low HBV prevalence) [1]. Nucleic acid amplification techniques (NAT) for HBV were first introduced for blood screening in 1997, and are now implemented in at least 30 countries worldwide [2,3]. However, there remains a residual risk of transfusion-transmitted infection, through occult HBV infection and vaccine breakthrough infections [4]. NAT is routinely used in the diagnosis and management of HBV infections, particularly, to guide the initiation of and monitor the response to antiviral therapy in chronically-infected patients [5]. A range of both commercial and laboratory-developed NAT-based assays are currently in use. The WHO International Standard for HBV DNA was established in 1999 [6,7], and is used by manufacturers of *in vitro* diagnostic devices (IVDs), blood transfusion centres, control authorities, and clinical laboratories, to calibrate secondary reference materials for NAT in terms of the International Unit (IU).

The 1st and 2nd WHO International Standards for HBV were prepared by dilution of a Eurohep R1 sample [8] (Genotype A2, HBsAg subtype *adw*2, derived from a single donor) in HBV-negative pooled human plasma. Both materials were prepared from the same bulk (filled and freeze-dried on two separate occasions) and evaluated in parallel in a worldwide collaborative study using a range of NAT-based assays for HBV [6,7]. The first candidate (NIBSC code 97/746) was established as the 1st WHO International Standard for HBV DNA in 1999, with an assigned potency of 1,000,000 IU/mL when reconstituted in 0.5 mL nuclease-free water. In 2006, the WHO Expert Committee on Biological Standardization (ECBS) established the second candidate (NIBSC code 97/750) as the replacement 2nd WHO International Standard for HBV DNA following a smaller collaborative study [9,10]. In 2011, two replacement batches (NIBSC codes 10/264 and 10/266) were prepared from the same original HBV Eurohep R1 stock as the 1st and 2nd WHO International Standards, diluted in pooled human plasma, and were evaluated in a worldwide collaborative study in parallel with the 2nd WHO International Standard for HBV (NIBSC code 97/750) [11]. Mean relative potencies for 10/264 and 10/266 were 5.93 and 5.98 $\log_{10}$ IU/mL respectively, when compared against 97/750. The first candidate (10/264) was established as the 3rd WHO International Standard for HBV DNA in October 2011 with a unitage of 850,000 IU/mL. It was noted that the second candidate (10/266) would be a suitable replacement HBV International Standard in due course, depending on ongoing stability assessment.
The established use of the HBV IU as the unit of measurement for HBV DNA highlights the importance of maintaining the availability of this International Standard. This report describes the collaborative study evaluation of the second candidate 10/266 as the replacement 4th WHO International Standard for HBV for NAT. The proposal to replace the 3rd WHO International Standard for HBV DNA was endorsed by the WHO ECBS in October 2015. The collaborative study results were presented to the Scientific Working Group on the Standardization of Genome Amplification Techniques (SoGAT) in London in June 2016. The proposed standard is intended to be used in the *in vitro* diagnostics field and relates to ISO 17511:2003 Section 5.5 [12].

Aims of study

The aim of this collaborative study was to evaluate the suitability of the candidate lyophilized preparation in parallel with the 3rd HBV WHO International Standard (NIBSC code 10/264) using a range of NAT-based assays.

Materials

Candidate standard

The candidate preparation (NIBSC code 10/266) comprises lyophilized human plasma and HBV. The HBV was sourced from a stock of the Eurohep R1 reference material stored at NIBSC and is a genotype A2, HBsAg subtype *adw*2 virus from a single donor [8]. The pooled human plasma diluent was sourced from UK blood donations and had been tested and found negative for HIV antibody, HCV antibody, HBsAg and syphilis. It was also tested at NIBSC and found negative for HCV RNA by NAT. The preparation was lyophilized to ensure long-term stability.

The filling and lyophilization of the bulk material was performed under contract at eQAD, UK NEQAS (Colindale, UK), in March 2011 and has been described previously [11]. The bulk was dispensed in 0.5 mL volumes into 3 mL screw-cap glass vials (Adelphi Tubes, Haywards Heath, UK). The homogeneity of the fill was determined by performing check-weighing of approximately every fiftieth vial, with vials outside the defined specification being discarded. Filled vials were partially stoppered, lyophilized and then fully stoppered in the freeze dryer. A total of 2700 vials were prepared for 10/266. The percentage coefficient of variation (%CV) of

the fill weight was 0.36%. The sealed vials were returned to NIBSC for storage at -20 °C under continuous temperature monitoring for the lifetime of the product. Evaluation of multiple aliquots of 10/266 (n=30) at NIBSC indicated that the HBV content was homogeneous (2SD of 0.12 $\log_{10}$ IU/mL). Comparison of the liquid bulk versus the lyophilized product indicated that there was a minimal loss in potency of approximately 0.04 $\log_{10}$ IU/mL upon freeze-drying. Assessments of residual moisture and oxygen content, as an indicator of vial integrity after sealing, were determined for 20 vials of 10/266 as previously described [11], and were 0.29% (Karl Fischer, 0.45% NIR units, CV=14.66%) and 0.17% (CV=63.81%) respectively.

Stability of the lyophilized candidate

Ongoing accelerated thermal degradation studies have been underway at NIBSC since 2011 in order to predict the stability of 10/266 when stored at the recommended temperature of -20 °C. Vials of lyophilized product have been held at -70 °C, -20 °C, +4 °C, +20 °C, +37 °C and +45 °C. At 11 weeks, 12, 34 and 60 months post-manufacture, three vials have been removed from storage at each temperature and HBV DNA quantified by NAT using the COBAS® AmpliPrep/COBAS® TaqMan® HBV Test, version 2.0 (Roche Diagnostics GmbH, Mannheim, Germany).

Study samples

The lyophilized candidate 10/266 was evaluated alongside the 3rd HBV WHO International Standard (NIBSC code 10/264), a HBV Secondary Reference Reagent (SRR), a HBV National Standard (NS) and Working Reagent (WR), and three individual HBV-positive plasma donations comprising different genotypes. These plasma donations were sourced from HBV-positive plasma packs rejected by the UK NHS Blood and Tranplant authority (Colindale, UK). The HBV genotype was determined using a multiplex PCR assay with genotype-specific primers [13]. Lyophilized and liquid frozen study samples were stored at -20 °C and -70 °C, respectively, prior to shipping to participants by courier on dry ice.

Study samples were coded as samples 1-8 and were as follows:
– Sample 1- Lyophilized 10/264 3rd HBV WHO International Standard in a 3 mL crimp-cap glass vial.
– Sample 2 - Lyophilized candidate 10/266 in a 3 mL screw-cap glass vial.
– Sample 3 - Liquid frozen HBV Secondary Rreference Reagent (genotype A) in a 2 mL Sarstedt tube
– Sample 4 - Liquid frozen HBV National Standard (genotype C) in a 2 mL Sarstedt tube.
– Sample 5 - Liquid frozen HBV Working Reagent (genotype C) in a 2 mL Sarstedt tube.
– Sample 6 - Liquid frozen HBV plasma (genotype D) in a 2 mL Sarstedt tube.
– Sample 7 - Liquid frozen HBV plasma (genotype E) in a 2 mL Sarstedt tube.
– Sample 8 - Liquid frozen HBV plasma (genotype A) in a 2 mL Sarstedt tube.

Study design

The aim of this collaborative study was to evaluate the suitability of the candidate 4th HBV WHO International Standard in parallel with the 3rd WHO International Standard for HBV using a range of NAT-based assays. Three HBV reference reagents were included in the study with the intention of calibrating these reagents in IU. Three HBV plasma samples were included in the study in order to provide a limited assessment of commutability [12,14]. Three vials of each study sample were sent to participating laboratories, with specific instructions for storage, reconstitution and testing. Samples 6-8 were only sent to laboratories performing quantitative HBV NAT.

Study protocol

Participants were requested to test dilutions of each sample using their routine HBV NAT assay on three separate occasions, using a fresh vial of each sample in each independent assay. In accordance with the study protocol (Appendix 2), the lyophilized samples were to be reconstituted with 0.5 mL of deionized, nuclease-free molecular-grade water and left for a minimum of 20 minutes with occasional agitation before use. Samples 6-8 were to be tested neat and were therefore only evaluated by laboratories performing quantitative assays.

For quantitative assays, participants were requested to test samples 1-8 neat and to test samples 1-5 at a minimum of two serial ten-fold dilutions (10^{-1} and 10^{-2}). For qualitative assays, participants were requested to test ten-fold serial dilutions of samples 1-5, around the assay end-point (in order to determine the actual assay end-point). For subsequent assays, participants were asked to test the dilution at the predetermined end-point, and a minimum of two half-$\log_{10}$ serial dilutions either side of the end-point (i.e., at least five dilutions in total). Participants were requested to perform dilutions using the sample matrix specific to their individual assay (e.g. HBV DNA-negative human plasma), and to extract samples prior to HBV DNA measurement.

Participants were requested to report the concentration of each sample in IU/mL (positive/negative for qualitative assays) for each dilution of each sample and to return results, including details of the methodology used, to NIBSC for analysis.

Participants

Study samples were sent to 13 participants representing 7 countries (Appendix 1). Participants were selected for their experience in HBV NAT, geographic distribution and participation in previous evaluation studies. They represented IVD manufacturers, Official Medicines Control Laboratories (OMCLs) and WHO collaborating centres. All participating laboratories are referred to by a code number, allocated at random, and not representing the order of listing in Appendix 1. Where a laboratory returned data using different assay methods, the results were analyzed separately, as if from different laboratories, and are referred to as, for example, laboratory 01A, 01B etc.

Statistical methods

Qualitative and quantitative assay results were evaluated separately. In the case of qualitative assays (from laboratory 12), data from all assays were pooled to give a number positive out of number tested at each dilution step. A single 'end-point' for each dilution series was calculated, to give an estimate of 'NAT detectable units/mL', as described previously [15]. It should be noted that these estimates are not necessarily directly equivalent to a genuine genome copy number/mL. In the case of quantitative assays, analysis was based on the results supplied by the participants. Results were reported as IU/mL. For each assay run, a single estimate of $\log_{10}$ IU/mL was obtained for each sample, by taking the mean of the $\log_{10}$ estimates of IU/mL across replicates, after correcting for any dilution factor. A single estimate for the laboratory and assay method was then calculated as the mean of the $\log_{10}$ estimates of IU/mL across assay runs.

All analysis was based on the $\log_{10}$ estimates of IU/mL or 'NAT detectable units/mL'. Overall mean estimates were calculated as the means of all individual laboratories. Variation between laboratories (inter-laboratory) was expressed as standard deviation (SD) of the $\log_{10}$ estimates and % geometric coefficient of variation (%GCV) [16] of the actual estimates. Potencies relative to sample 1, the current HBV WHO International Standard (10/264), were calculated as the difference in estimated $\log_{10}$ 'units per mL' (test sample – standard) plus the value in $\log_{10}$ IU/mL for the International Standard. Therefore for example, if in an individual assay, the test

sample is 0.5 $\log_{10}$ higher than the International Standard, assigned 5.93 $\log_{10}$ IU/mL, the relative potency of the test sample is 6.43 $\log_{10}$ IU/mL.

For the quantitative assays, variation within laboratories, and between assays, (intra-laboratory) was expressed as SDs and %GCVs of the individual assay mean $\log_{10}$ estimates. These estimates were pooled across all samples. The significance of the inter-laboratory variation relative to the intra-laboratory variation was assessed by an analysis of variance.

Results and data analysis

Validation of study samples and stability assessment

Production data for the candidate 10/266 from 2011 showed that the CV of the fill weight and mean residual moisture were within acceptable limits for a WHO International Standard [17]. The residual oxygen content was within the NIBSC working limit of 1.1%.

Samples of 10/266 were stored at elevated temperatures, and assayed at NIBSC in parallel with samples stored at -20 °C and -70 °C by HBV NAT (as described for the stability of the lyophilized candidate). Three vials of each sample were evaluated after storage at each temperature for 11 weeks, 12, 30 and 60 months. It was not possible to reconstitute the vials stored at +37 °C and +45 °C for 12, 30 and 60 months. The mean estimated $\log_{10}$ IU/mL and differences ($\log_{10}$ IU/mL) from the -20 °C baseline samples are shown in Table 1. A negative value indicates a drop in potency relative to the -20 °C baseline. The 95% confidence intervals for the differences are ±0.08 $\log_{10}$ based on a pooled estimate of the standard deviation between individual vial test results. As there is no observed drop in potency it is not possible to fit the usual Arrhenius model for accelerated degradation studies, or obtain any predictions for the expected loss per year with long-term storage at -20 °C. All available data indicates adequate stability. Stability testing of 10/266 will be ongoing.

The stability of 10/266 when reconstituted has not been specifically determined. Therefore, it is recommended that the reconstituted material is for single use only.

Data received

Data were received from all 13 participating laboratories. Participants performed a variety of different assay methods, with some laboratories performing more than one assay method. In total, 15 datasets were received from 14 quantitative assays and 1 qualitative assay. Apart from the cases noted below, there were no exclusions of data.

Quantitative Assays:
Data were returned with dilutions ranging from neat to 10^{-8} from different laboratories, although only dilutions between neat and 10^{-2} were used in the analysis. For Sample 5, one or two dilutions were removed from the following laboratories' data, 01, 02B, 03, 05, 06, 07, 08, 10 and 13 because the results were below the limit of detection. Samples not demonstrating dilutional linearity (i.e. a linear relationship between reported HBV content against $\log_{10}$ dilution with fitted slope between 0.80 and 1.25) were excluded. Non-linearity was assessed visually and determined in the following cases; laboratory 11A, Sample 5 on day 2, Samples 3 and 5 on day 3; laboratory 11B, Sample 5 on day 3; laboratory 13, Sample 3 on day 2. Laboratories 04, 05, 07, 09, 11B and 13 had one to three samples with slopes outside the range 0.80-1.25, with the majority of these cases (7 out of 10) for Sample 5.

Qualitative Assays:

Laboratory 12 tested multiple dilutions, from 10^{-1} down to 10^{-8} for different samples. Only dilutions that were tested in at least two of the three assays were used in the analysis.

Summary of assay methodologies

The majority of participants prepared dilutions of study samples 1-5 using negative human plasma, however, Basematrix 53 (SeraCare Life Sciences, Inc., Milford MA, USA) was also used. Assay methodologies for qualitative and quantitative assays are summarized in Table 2. Ten different commercial HBV NAT assays were represented. Five new assays were included that were not represented in the 2011 collaborative study (assay codes; c68, APT, VER, SKB and SaB). Only one of the commercial assays was qualitative (assay code cTSM). No participant used laboratory-developed tests.

Estimated IU/mL or 'NAT detectable units/mL'

The laboratory mean estimates of IU/mL (log_{10}) from the quantitative assays and 'NAT detectable units/mL' (log_{10}) from the qualitative assay (shaded in grey) are shown in Table 3. The individual laboratory mean estimates are also shown in histogram form in Figure 1. Each box represents the mean estimate from one laboratory, and the boxes are labeled with the laboratory and assay code. They are also colour coded by assay. For samples 1, 4 and 5, there is good agreement between qualitative and quantitative assays, however, for samples 2 and 3 there is not. This may reflect variability in the dilution of study samples required by the study protocol for qualitative assays rather than actual variability in the detection of HBV in the different study samples. There was good agreement between the estimates from the quantitative assays, particularly for samples 1-3 which all comprise the same genotype A virus. Laboratory 6 appears to underquantify samples 4-8 compared to other quantitative assays. There also appears to be variability in the quantification of samples 4-8 by the different assay methods, with some assays showing either under or over-quantification compared to the overall mean estimate for each of the samples. This may represent variability in the quantification of different HBV genotypes as has been reported previously [18]. However, for sample 8, the pattern of individual laboratory mean estimates is different to that for samples 1-3 despite all comprising genotype A viruses.

Table 4 shows the overall mean estimates of log_{10} IU/mL from the quantitative assays, along with the SD (of log_{10} estimates) and the %GCV (of actual estimates). The overall mean estimates for samples 1 and 2 were 5.94 and 5.97 Log_{10} IU/mL respectively. These values are very similar to the values obtained for these samples in the 2011 collaborative study [11] (5.93 and 5.98 Log_{10} IU/mL respectively), despite some differences in the participants and assays involved in each study. For samples 1-3, the SDs and %GCVs are 0.11-0.13 log_{10} and 29-35% respectively. The overall SDs for samples 1 and 2 are slightly higher than those reported in the 2011 collaborative study. However, this still represents good agreement between laboratories and assay methods. For samples 4 and 5, the SDs and %GCVs are 0.22-0.27 log_{10} and 65-86% respectively. For samples 6-8, the SDs and %GCVs are 0.29-0.42 log_{10} and 95-160% respectively. The increased SDs and %GCVs for samples 4-8 are principally due to the outlying results of laboratory 6 (see Figure 1). The overall mean estimates, SDs and %GCVs for samples 1-8 excluding the dataset for laboratory 6 are shown in Table 5. The SDs and %GCVs for samples 4-8 are all reduced (by approximately 2-fold) when the results for laboratory 6 are excluded. Five laboratories reported results using the cobas® AmpliPrep/cobas® TaqMan® HBV Test, v2.0 (assay code cTM). This assay is over-represented in comparison to the other assays. However, removing datasets from 3 laboratories using this assay did not greatly alter the overall laboratory results (Table 6)

Potencies relative to the 3rd WHO International Standard for HBV (Sample 1)

The estimated concentrations of samples 2-8 were expressed in IU, by direct comparison (relative potencies) to the current International Standard (10/264, sample 1), which has an assigned unitage of 850,00 IU/mL (5.93 $\log_{10}$), as described in the statistical methods section. The laboratory mean estimates are shown in Table 7 for the quantitative and qualitative assays. Units are $\log_{10}$ IU/mL in both cases. The results are also shown in histogram form in Figure 2. The overall mean relative potencies, along with SDs and %GCVs, are shown in Table 8. The overall mean relative potency for the candidate sample 2 is 5.96 $\log_{10}$ IU/mL, based on the quantitative assays. This value compares well to the direct estimate of 5.97 $\log_{10}$ IU/mL from the quantitative assays which are all calibrated in IU/mL.

Figure 2 and Table 8, show an improvement in the agreement between laboratories for samples 2-5 for the quantitative assays. The SD and %GCV between laboratories has reduced for these samples when compared to the values in Table 4. The reduction in these values for samples 4 and 5 (both genotype C) is less marked than for samples 1-3 (all genotype A), possibly because of increased variability in the quantification of genotype C viruses between the assays. For samples 6-8 there is no improvement in the agreement between laboratories when compared to the values in Table 4. This may be due to variability in the quantification of different HBV genotypes, although for sample 8 there is no improvement in the agreement between laboratories, despite samples 1 and 8 comprising genotype A viruses. Again, the over-representation of the cTM assay in the study did not greatly alter the overall laboratory results (Table 9).

Potencies relative to the candidate 4th WHO International Standard for HBV (Sample 2)

The estimated concentrations of samples 3-8 were expressed in IU, by direct comparison (relative potencies) to the candidate International Standard (10/266, sample 2), based on a candidate unitage of 955,000 IU/mL (5.98 $\log_{10}$), as described in the statistical methods section. This candidate unitage was based on the overall mean potency obtained for 10/266, relative to the 2nd HBV WHO International Standard (97/750), in the 2011 collaborative study.

Overall mean relative potencies, along with SDs and %GCVs, are shown in Table 10. Table 10 shows an improvement in the agreement between laboratories for samples 3-5 for the quantitative assays. The SD and %GCV between laboratories has reduced for these samples when compared to the values in Table 4. For samples 6-8 there is no improvement in the agreement between laboratories when compared to the values in Table 4. Again, this may be due to variability in the quantification of different HBV genotypes, although for sample 8 there is no improvement in the agreement between laboratories, despite samples 2 and 8 comprising genotype A viruses. The over-representation of the cTM assay in the study did not greatly alter the overall laboratory results (Table 11).

Inter and intra-laboratory variation

For all samples, the inter-laboratory variation was greater than the intra-laboratory variation (p<0.01). Table 12 shows the intra-laboratory SDs and %GCVs for each laboratory, calculated by pooling the estimates for samples 1-8. There are differences between the repeatability of laboratory estimates across assays. In general, the repeatability is good for assays of this type and the average SD is 0.07 $\log_{10}$ or a %GCV of 18%. These values are slightly improved compared to the equivalent values obtained in the 2011 collaborative study (average SD of 0.08 $\log_{10}$, %GCV of 20%) [11]. These figures represent the variability between individual assay mean estimates of IU/mL. Since each assay tested multiple replicates of samples at different dilutions,

the resulting between-assay variability is lower than would be expected if only a single replicate was tested in each assay. The 'NAT detectable units' from the qualitative assays are obtained by pooling all assay data to give a single series of number positive out of number tested at each dilution. As a result, there is no comparable analysis of intra-assay variation for the qualitative assay.

Conclusions

In this study, a range of NAT-based assays for HBV have been used to evaluate the suitability of the candidate standard (NIBSC code 10/266) as the replacement 4th WHO International Standard for HBV DNA for NAT-based assays. The candidate was prepared from the same virus stock used for previous HBV WHO International Standards and was diluted in a similar pooled human plasma material [6,7,9,10]. Production data suggests that the batch is homogeneous and contains residual moisture and oxygen levels that are within WHO and NIBSC limits for lyophilized standards [11,17]. Comparison of the liquid bulk versus the lyophilized product indicates that there was minimal loss in potency upon freeze-drying (0.04 $\log_{10}$ IU/mL). The results of ongoing accelerated thermal degradation studies at 60 months indicate that the candidate is stable and suitable for long-term use.

In the collaborative study, the lyophilized candidate (sample 2) was evaluated alongside the 3rd HBV WHO International Standard (sample 1). The overall mean estimates for samples 1 and 2 were 5.94 and 5.97 Log_{10} IU/mL, respectively, based on the calibration of quantitative assay kits in IU/mL. These values are very similar to the values obtained for these samples relative to the 2nd HBV WHO International Standard in the 2011 collaborative study [11], despite laboratories reporting results using different HBV NAT-based assays. In the present study, the agreement between laboratories for sample 2 was improved when the potency was expressed relative to the 3rd HBV WHO International Standard (sample 1). There was some evidence for variability in the quantification of different HBV genotypes present in the different study samples. This has been reported previously [18]. Inter-laboratory variability was higher than intra-laboratory variability for the quantitative assays. This highlights the continued need for standardization of HBV NAT-based assays, and the importance of accurate calibration to the WHO International Standard.

A full assessment of commutability of the candidate standard for HBV-positive samples has not been possible in this study, due to the limited number of clinical samples that could be included. Three HBV-positive plasma samples comprising HBV genotypes A, D and E, from rejected blood donations were included in the study. There was no overall improvement in the agreement between laboratories when the estimated concentrations of the three HBV plasma samples were expressed relative to the candidate 4th HBV WHO International Standard (same formulation as previous HBV WHO International Standards), compared to the uncorrected values. This is principally due to variability in the quantification of different HBV genotypes between different assays, and outlying results from one or two laboratories for samples 6-8. There is no evidence of non-commutability with the three plasma samples that were included in the study.

In summary, the results of the study indicate the suitability of the candidate sample 2 as the replacement 4th WHO International Standard for HBV DNA for NAT. Since the overall mean potency obtained for the candidate in this collaborative study is very similar to the overall mean potency obtained in the 2011 collaborative study, relative to the pre-existing 2nd HBV WHO International Standard, it is proposed that the value assigned to the candidate sample 2 is that obtained in the 2011 collaborative study. This approach would minimize any potential drift in the value of the IU during the replacement.

WHO/BS/2016.2291
Page 10

Proposal

It is proposed that the candidate standard, **NIBSC code 10/266**, is established as the **4th WHO International Standard for HBV DNA for use in NAT-based assays**, with an assigned potency of 955,000 IU/mL (~5.98 $\log_{10}$ IU/mL) when reconstituted in 0.5 mL of nuclease-free water. This potency is based on the value that was assigned in the collaborative study in 2011 where the candidate was assessed alongside the 2nd HBV WHO International Standard. The uncertainty can be derived from the variance of the fill weight and is 0.36%. The proposed standard is intended to be used by IVD manufacturers, blood transfusion centres, control authorities, and clinical laboratories, to calibrate secondary reference materials used in HBV NAT assays. Proposed Instructions for Use (IFU) for the product are included in Appendix 3.

Comments from participants

8 of 13 participants responded to the report. There were no disagreements with the suitability of the candidate standard (NIBSC code 10/266) to serve as the 4th WHO International Standard for HBV DNA for NAT-based assays. Some comments suggested minor editorial changes and these have been implemented.

Acknowledgements

We gratefully acknowledge the important contributions of the collaborative study participants.

References

1. World Health Organisation, Hepatitis B, Fact sheet No. 204, updated July 2014. http://www.who.int/mediacentre/factsheets/fs204/en/
2. Roth WK1, Busch MP, Schuller A, Ismay S, Cheng A, Seed CR, Jungbauer C, Minsk PM, Sondag-Thull D, Wendel S, Levi JE, Fearon M, Delage G, Xie Y, Jukic I, Turek P, Ullum H, Tefanova V, Tilk M, Reimal R, Castren J, Naukkarinen M, Assal A, Jork C, Hourfar MK, Michel P, Offergeld R, Pichl L, Schmidt M, Schottstedt V, Seifried E, Wagner F, Weber-Schehl M, Politis C, Lin CK, Tsoi WC, O'Riordan J, Gottreich A, Shinar E, Yahalom V, Velati C, Satake M, Sanad N, Sisene I, Bon AH, Koppelmann M, Flanagan P, Flesland O, Brojer E, Lętowska M, Nascimento F, Zhiburt E, Chua SS, Teo D, Stezinar SL, Vermeulen M, Reddy R, Park Q, Castro E, Eiras A, Gonzales Fraile I, Torres P, Ekermo B, Niederhauser C, Chen H, Oota S, Brant LJ, Eglin R, Jarvis L, Mohabir L, Brodsky J, Foster G, Jennings C, Notari E, Stramer S, Kessler D, Hillyer C, Kamel H, Katz L, Taylor C, Panzer S, Reesink HW. International survey on NAT testing of blood donations: expanding implementation and yield from 1999 to 2009. Vox Sang. 2012;102(1):82-90. doi: 10.1111/j.1423-0410.2011.01506.x.
3. Seo DH, Whang DH, Song EY, Han KS. Occult hepatitis B virus infection and blood transfusion. World J Hepatol. 2015;7(3):600-6. doi: 10.4254/wjh.v7.i3.600.
4. Candotti D, Allain JP. Transfusion-transmitted hepatitis B virus infection. J Hepatol. 2009;51:798-809.
5. Sorrell MF, Belongia EA, Costa J, Gareen IF, Grem JL, Inadomi JM, Kern ER, McHugh JA, Petersen GM, Rein MF, Sorrell MF, Strader DB, Trotter HT. National Institutes of Health Consensus Development Conference Statement: management of hepatitis B. Ann Intern Med. 2009;150:104-10.
6. Saldanha J, Gerlich W, Lelie N, Heermann K, Heath A. An international collaborative study to establish a WHO International Standard for HBV DNA nucleic acid amplification technology assay. WHO ECBS Report 1999; WHO/BS/99.1917. http://www.who.int/bloodproducts/cs/en/991917.pdf

7. Saldanha J, Gerlich W, Lelie N, Dawson P, Heermann K, Heath A; WHO Collaborative Study Group. An international collaborative study to establish a World Health Organization international standard for hepatitis B virus DNA nucleic acid amplification techniques. Vox Sang. 2001;80:63-71.
8. Heermann KH, Gerlich WH, Chudy M, Schaefer S, Thomssen R. Quantitative detection of hepatitis B virus DNA in two international reference plasma preparations. Eurohep Pathobiology Group. J Clin Microbiol. 1999;37:68-73.
9. Baylis SA, Gerlich W, Chudy M, Pisani G, Klotz A, Kerby S, Heath A. WHO collaborative study to establish a replacement WHO International Standard for hepatitis B virus DNA nucleic acid amplification technology (NAT) assays. WHO ECBS Report 2006; WHO/BS/06.2034. https://extranet.who.int/iris/restricted/bitstream/10665/69957/1/WHO_BS_06.2034_eng.pdf
10. Baylis SA, Heath AB, Chudy M, Pisani G, Klotz A, Kerby S, Gerlich W. An international collaborative study to establish the 2nd World Health Organization International Standard for hepatitis B virus DNA nucleic acid amplification technology-based assays. Vox Sang. 2008;94:358-62.
11. Fryer JF, Heath AB, Wilkinson DE, Minor PD and the Collaborative Study Group. Collaborative Study to Evaluate the Proposed 3rd WHO International Standard for Hepatitis B Virus (HBV) for Nucleic Acid Amplification Technology (NAT)-Based Assays. WHO ECBS Report 2011;WHO/BS/2011.2170. http://apps.who.int/iris/bitstream/10665/70779/1/WHO_BS_2011.2170_eng.pdf
12. 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 (ISO); 2003.
13. Kirschberg O, Schüttler C, Repp R, Schaefer S. A multiplex-PCR to identify hepatitis B virus-genotypes A-F. J Clin Virol. 2004;29:39-43.
14. WHO Consultation on Global Measurement Standards and their use in the in vitro Biological Diagnostic Field, Switzerland, Geneva, Switzerland, 7-8 June 2004: http://www.who.int/bloodproducts/publications/en/Minutes-220804.pdf
15. Saldanha J; Lelie N and Heath A B. Establishment of the First International Standard for Nucleic Acid Amplification Technology (NAT) assays for HCV RNA. 1999. Vox Sanquinis. 76:149-58.
16. Kirkwood TBL. Geometric means and measures of dispersion. Biometrics 1979;35:908-9.
17. Recommendations for the preparation, characterization and establishment of international and other biological reference standards (revised 2004). WHO Technical Report Series 2007. Geneva, Switzerland:WHO 2007; 932,73-131. http://www.who.int/bloodproducts/publications/TRS932Annex2_Inter_biolefstandardsrev2004.pdf
18. Chudy M, Hanschmann KM, Kress J, Gerlich W and Nübling CM. Collaborative study to establish a World Health Organization International Genotype Panel for hepatitis B virus nucleic acid amplification technique (NAT)-based assays. WHO ECBS Report 2009;WHO/BS/09.2121. http://whqlibdoc.who.int/hq/2009/WHO_BS_09.2121_eng.pdf

WHO/BS/2016.2291
Page 12

Table 1. Thermal stability of 10/266 at different storage temperatures. * Mean results from single extractions from 3 vials at each time point and temperature.

Temperature (°C)	Mean log_{10} IU/mL * (difference in log_{10} IU/mL from -20 °C baseline sample)			
	11 weeks	12 months	34 months	60 months
-70	6.04	6.05	5.96	5.77
-20	6.03	6.02	5.92	5.77
4	6.07 (0.04)	6.01 (-0.01)	5.91 (-0.01)	5.80 (0.03)
20	6.06 (0.03)	6.01 (-0.01)	5.91 (-0.01)	5.89 (0.12)
37	6.09 (0.06)	-	-	
45	6.13 (0.10)	-	-	

Table 2. Collaborative study assay methods and codes.

Quantitative Assays		
Assay Code	Assay	No. of datasets
kPCR	VERSANT HBV DNA 1.0 Assay (kPCR) (Siemens Healthcare Diagnostics)	1
cTM	cobas® AmpliPrep/cobas® TaqMan® HBV Test, v2.0 (Roche Molecular Systems, Inc.)	5
c68	cobas® HBV test for use on the cobas® 6800/8800 Systems (Roche Molecular Systems, Inc.)	2
AbRT	Abbott RealTi*me* HBV (Abbott Molecular, Inc.)	1
ArQS	*artus*® HBV QS-RGQ Kit, Version 1 (QIAGEN)	1
APT	Aptima HBV Quant assay on the Panther system (Hologic, Inc.)	1
VER	VERIS HBV Assay (Beckman Coulter, Inc.)	1
SKB	HBV DNA real-time PCR detection kit (Shanghai Kehua Bio-Engineering Co., Ltd.)	1
SaB	Hepatitis B Viral DNA Quantitative Fluorescence Diagnostic Kit (PCR Fluorescence Probing) Mag (Sansure Biotech, Inc.)	1
Qualitative Assays		
Assay Code	Assay	No. of datasets
cTSM	cobas® TaqScreen MPX Test, v2.0 (Roche Molecular Systems, Inc.)	1

Table 3. Laboratory mean estimates from quantitative assays (log_{10} IU/mL) and qualitative assays (log_{10} 'NAT detectable units/mL'). Qualitative results are shaded in grey. nd, not determined (either not tested or data excluded).

Lab	Assay	Sample							
		S1	S2	S3	S4	S5	S6	S7	S8
1	kPCR	6.01	6.08	5.74	5.98	3.00	3.41	3.45	3.11
2A	c68	6.00	6.01	5.68	6.05	3.06	3.28	3.24	3.54
2B	cTM	5.95	5.98	5.67	6.16	3.19	3.54	3.51	3.78
3	AbRT	5.86	5.87	5.58	5.91	2.92	3.59	3.27	3.55
4	ArQS	6.15	6.19	5.88	6.14	nd	3.16	2.86	3.94
5	cTM	5.91	5.99	5.68	6.17	3.18	3.52	3.50	3.71
6	VER	5.79	5.81	5.61	5.45	2.33	1.98	2.50	2.39
7	APT	5.86	5.88	5.62	5.93	3.05	3.09	3.11	3.40
8	c68	5.92	5.96	5.63	6.06	3.12	3.40	3.30	3.61
9	cTM	5.95	5.96	5.71	6.15	3.25	3.45	3.47	3.67
10	cTM	5.94	5.95	5.69	6.21	3.20	3.47	3.54	3.87
11A	SKB	6.23	6.26	5.91	6.30	3.10	3.43	3.40	3.44
11B	SaB	5.75	5.78	5.48	5.78	nd	3.70	3.33	3.76
12	cTSM	6.06	6.57	6.3	6.16	2.78	nd	nd	nd
13	cTM	5.88	5.93	5.68	6.16	3.16	3.42	3.42	3.62

Table 4. Overall mean estimates and inter-laboratory variation for quantitative assays (log_{10} IU/mL).

Sample	No. of datasets	Mean	Min	Max	SD	%GCV
S1: 10/264 (gt.A)	14	5.94	5.75	6.23	0.13	35
S2: 10/266 (gt.A)	14	5.97	5.78	6.26	0.13	35
S3: SRR (gt.A)	14	5.68	5.48	5.91	0.11	29
S4: NS (gt.C)	14	6.03	5.45	6.30	0.22	65
S5: WR (gt.C)	12	3.05	2.33	3.25	0.24	75
S6: HBVpl (gt.D)	14	3.32	1.98	3.70	0.42	160
S7: HBVpl (gt.E)	14	3.28	2.50	3.54	0.29	95
S8: HBVpl (gt.A)	14	3.53	2.39	3.94	0.39	145

WHO/BS/2016.2291
Page 14

Table 5. Overall mean estimates and inter-laboratory variation for quantitative assays (log_{10} IU/mL), excluding the dataset from laboratory 6.

Sample	No. of datasets	Mean	Min	Max	SD	%GCV
S1: 10/264 (gt.A)	13	5.96	5.75	6.23	0.13	34
S2: 10/266 (gt.A)	13	5.99	5.78	6.26	0.13	34
S3: SRR (gt.A)	13	5.69	5.48	5.91	0.11	30
S4: NS (gt.C)	13	6.07	5.78	6.30	0.14	39
S5: WR (gt.C)	11	3.06	2.55	3.26	0.19	54
S6: HBVpl (gt.D)	13	3.42	3.09	3.70	0.17	47
S7: HBVpl (gt.E)	13	3.34	2.86	3.54	0.19	55
S8: HBVpl (gt.A)	13	3.61	3.11	3.94	0.22	66

Table 6. Overall mean estimates and inter-laboratory variation for quantitative assays (log_{10} IU/mL), excluding datasets from laboratories 9, 10 and 13 using the cTM assay.

Sample	No. of datasets	Mean	Min	Max	SD	%GCV
S1: 10/264 (gt.A)	11	5.95	5.75	6.23	0.15	40
S2: 10/266 (gt.A)	11	5.98	5.78	6.26	0.15	41
S3: SRR (gt.A)	11	5.68	5.48	5.91	0.13	34
S4: NS (gt.C)	11	5.99	5.45	6.30	0.23	71
S5: WR (gt.C)	9	2.95	2.33	3.19	0.28	91
S6: HBVpl (gt.D)	11	3.28	1.98	3.70	0.47	193
S7: HBVpl (gt.E)	11	3.22	2.50	3.51	0.31	102
S8: HBVpl (gt.A)	11	3.47	2.39	3.94	0.42	165

Table 7. Laboratory estimates of potency relative to the 3rd WHO International Standard for HBV (sample 1) from quantitative assays and qualitative assays (log_{10} IU/mL) - based on an assigned unitage of the International Standard of 850,000 (5.93 log_{10}) IU/mL. nd, not determined (either not tested or data excluded).

Lab	Assay	Sample						
		S2	S3	S4	S5	S6	S7	S8
1	kPCR	6.00	5.67	5.90	2.92	3.33	3.38	3.03
2A	c68	5.93	5.61	5.98	2.99	3.20	3.17	3.47
2B	cTM	5.96	5.65	6.14	3.17	3.52	3.49	3.76
3	AbRT	5.93	5.64	5.98	2.99	3.65	3.34	3.62
4	ArQS	5.97	5.65	5.91	nd	2.94	2.63	3.72
5	cTM	6.01	5.70	6.19	3.19	3.54	3.52	3.73
6	VER	5.95	5.76	5.59	2.48	2.13	2.64	2.54
7	APT	5.95	5.69	6.00	3.12	3.16	3.19	3.47
8	c68	5.97	5.63	6.06	3.13	3.40	3.31	3.62
9	cTM	5.94	5.69	6.13	3.23	3.43	3.45	3.65
10	cTM	5.94	5.68	6.20	3.19	3.46	3.52	3.86
11A	SKB	5.96	5.61	6.00	2.80	3.13	3.10	3.14
11B	SaB	5.97	5.66	5.97	nd	3.89	3.51	3.94
12	cTSM	6.44	6.17	5.86	2.65	nd	nd	nd
13	cTM	5.98	5.73	6.21	3.22	3.47	3.47	3.67

Table 8. Overall mean estimates and inter-laboratory variation for potency relative to the 3rd HBV WHO International Standard (sample 1) log_{10} IU/mL for quantitative assays - based on an assigned unitage of the International Standard of 850,000 (5.93 log_{10}) IU/mL.

Sample	No. of datasets	Mean	Min	Max	SD	%GCV
S2: 10/266 (gt.A)	14	5.96	5.94	6.01	0.02	5
S3: SRR (gt.A)	14	5.67	5.61	5.76	0.04	10
S4: NS (gt.C)	14	6.02	5.59	6.21	0.16	46
S5: WR (gt.C)	12	3.03	2.48	3.23	0.22	67
S6: HBVpl (gt.D)	14	3.30	2.13	3.89	0.41	159
S7: HBVpl (gt.E)	14	3.27	2.63	3.52	0.30	100
S8: HBVpl (gt.A)	14	3.51	2.54	3.94	0.38	138

WHO/BS/2016.2291
Page 16

Table 9. Overall mean estimates and inter-laboratory variation for potency relative to the 3rd HBV WHO International Standard (sample 1) $\log_{10}$ IU/mL for quantitative assays - based on an assigned unitage of the International Standard of 850,000 (5.93 $\log_{10}$) IU/mL, excluding datasets from laboratories 9, 10 and 13 using the cTM assay.

Sample	No. of datasets	Mean	Min	Max	SD	%GCV
S2: 10/266 (gt.A)	11	5.96	5.93	6.01	0.02	5
S3: SRR (gt.A)	11	5.66	5.61	5.76	0.04	10
S4: NS (gt.C)	11	5.98	5.59	6.19	0.16	43
S5: WR (gt.C)	9	2.97	2.48	3.19	0.23	69
S6: HBVpl (gt.D)	11	3.26	2.13	3.89	0.46	189
S7: HBVpl (gt.E)	11	3.21	2.63	3.52	0.31	106
S8: HBVpl (gt.A)	11	3.46	2.54	3.94	0.41	155

Table 10. Overall mean estimates and inter-laboratory variation for potency relative to the candidate 4th HBV WHO International Standard (sample 2) $\log_{10}$ IU/mL for quantitative assays - based on a candidate unitage of 955,000 (5.98 $\log_{10}$) IU/mL.

Sample	No. of datasets	Mean	Min	Max	SD	%GCV
S3: SRR (gt.A)	14	5.69	5.64	5.79	0.04	10
S4: NS (gt.C)	14	6.04	5.62	6.24	0.16	45
S5: WR (gt.C)	12	3.05	2.51	3.27	0.22	66
S6: HBVpl (gt.D)	14	3.32	2.16	3.90	0.41	157
S7: HBVpl (gt.E)	14	3.28	2.64	3.56	0.30	98
S8: HBVpl (gt.A)	14	3.53	2.57	3.95	0.38	138

Table 11. Overall mean estimates and inter-laboratory variation for potency relative to the candidate 4th HBV WHO International Standard (sample 2) $\log_{10}$ IU/mL for quantitative assays - based on a candidate unitage of 955,000 (5.98 $\log_{10}$) IU/mL, excluding datasets from laboratories 9, 10 and 13 using the cTM assay.

Sample	No. of datasets	Mean	Min	Max	SD	%GCV
S3: SRR (gt.A)	11	5.68	5.64	5.79	0.04	10
S4: NS (gt.C)	11	5.99	5.62	6.16	0.15	41
S5: WR (gt.C)	9	2.99	2.51	3.18	0.22	66
S6: HBVpl (gt.D)	11	3.28	2.16	3.90	0.46	186
S7: HBVpl (gt.E)	11	3.22	2.64	3.52	0.31	103
S8: HBVpl (gt.A)	11	3.47	2.57	3.95	0.40	153

Table 12. Intra-laboratory SD of $\log_{10}$ IU/mL and %GCV for quantitative assays.

Lab	Assay	SD	%GCV
1	kPCR	0.05	13
2A	c68	0.03	7
2B	cTM	0.05	12
3	AbRT	0.03	8
4	ArQS	0.17	47
5	cTM	0.04	11
6	VER	0.06	15
7	APT	0.08	21
8	c68	0.03	6
9	cTM	0.05	11
10	cTM	0.07	16
11A	SKB	0.09	24
11B	SaB	0.05	13
13	cTM	0.05	13
Overall		0.07	18

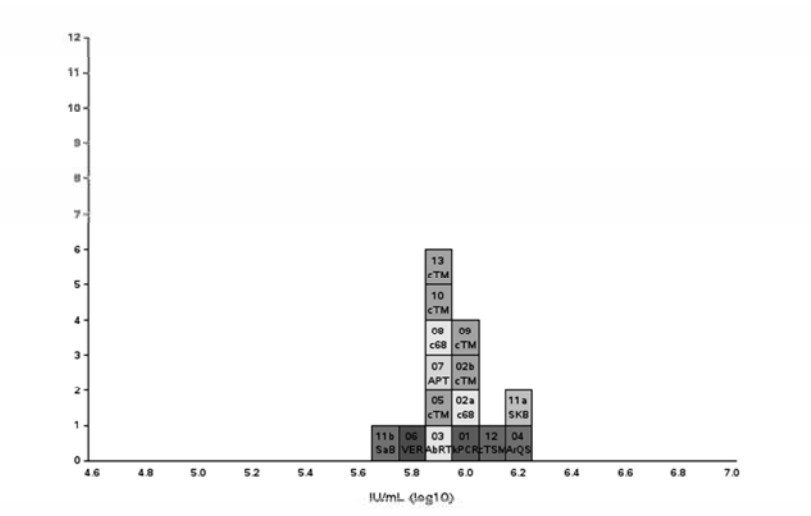
Figure legends

Figure 1. Individual laboratory mean estimates for study samples 1-8 (a-h respectively) obtained using qualitative or quantitative NAT assays. Each box represents the mean estimate from each laboratory assay and is labeled with the laboratory and assay code. Boxes are also colour coded by assay.

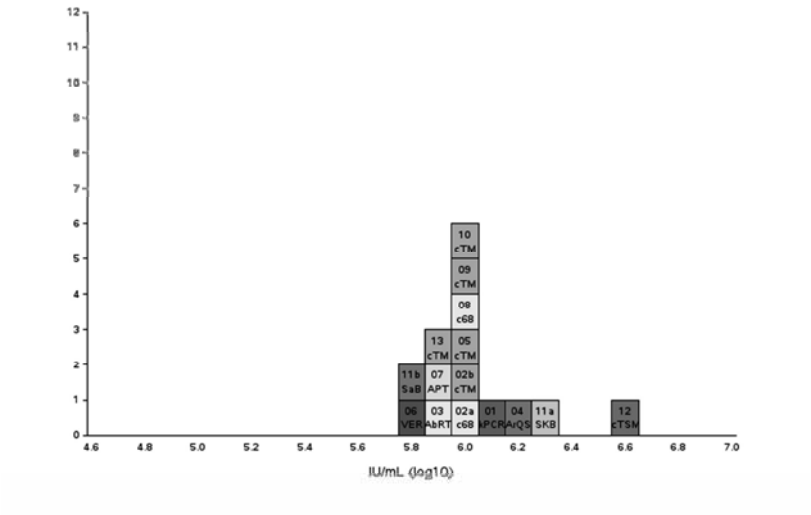
Figure 2. Relative potencies of samples 2-8 against sample 1 (a-g respectively), for each qualitative or quantitative assay. Units are expressed as candidate $\log_{10}$ IU/mL. Each box represents the relative potency for each laboratory assay and is labeled with the laboratory and assay code. Boxes are also colour coded by assay.

Figure 1

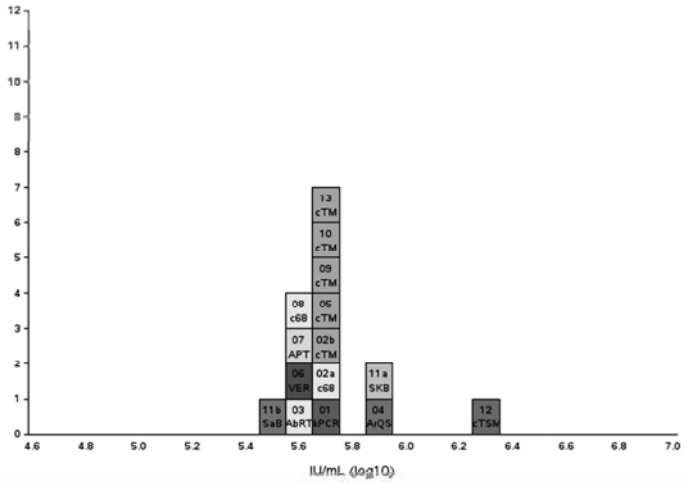
a Sample 1



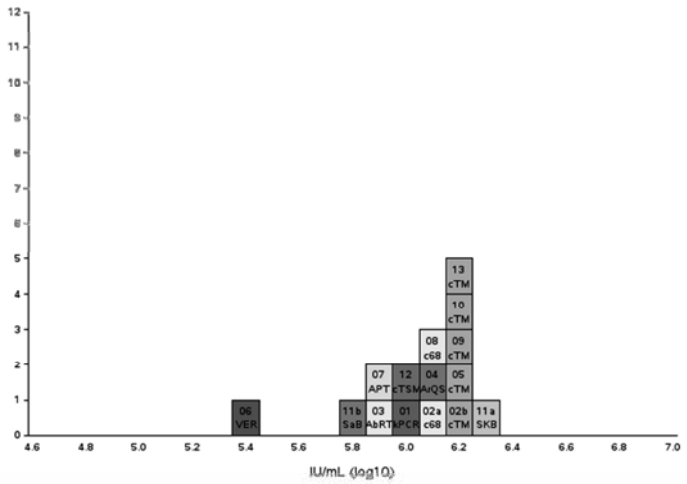
b Sample 2



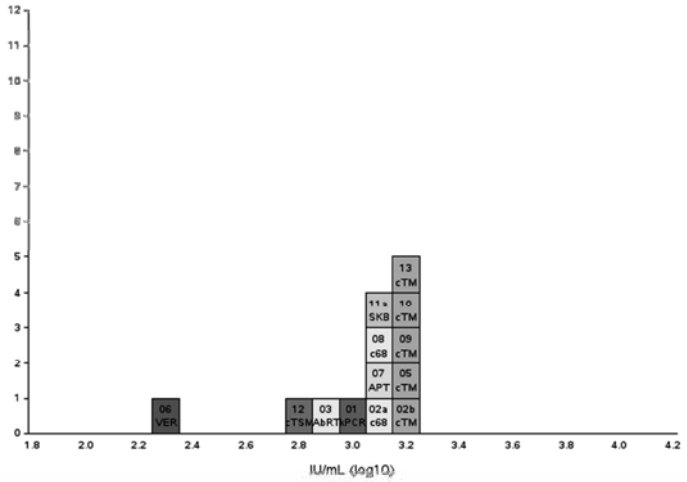
c Sample 3



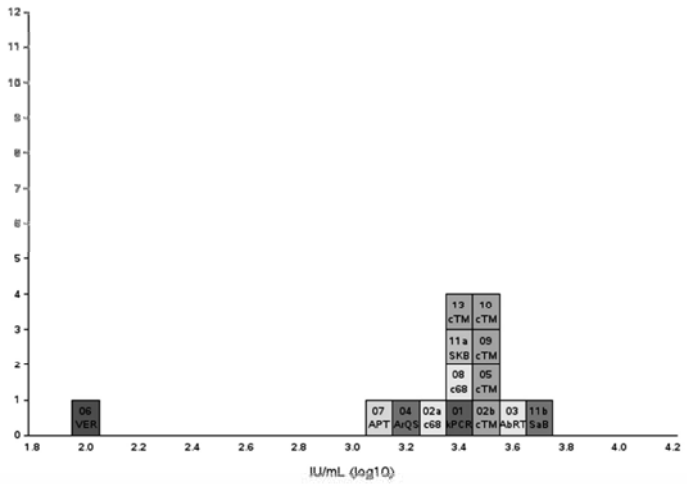
d Sample 4



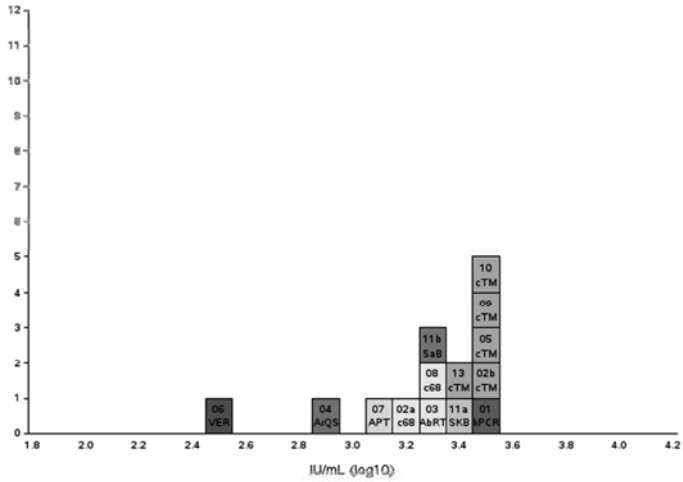
e Sample 5



f Sample 6



g Sample 7



h Sample 8

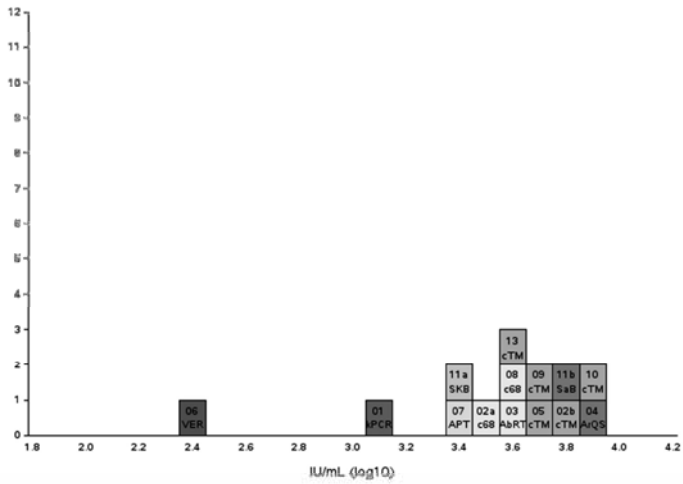
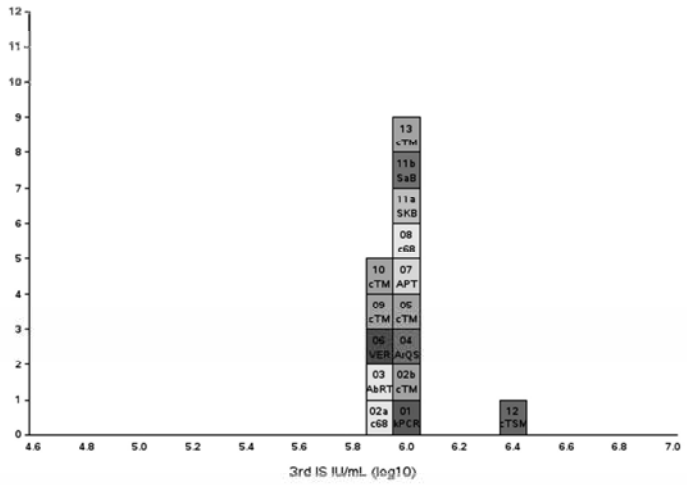
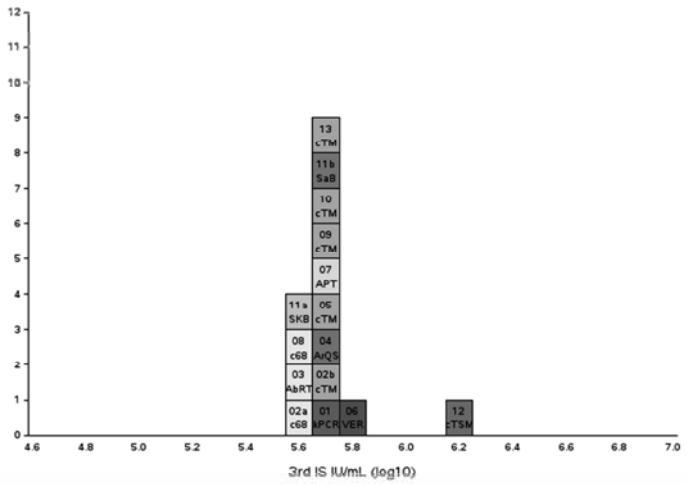


Figure 2

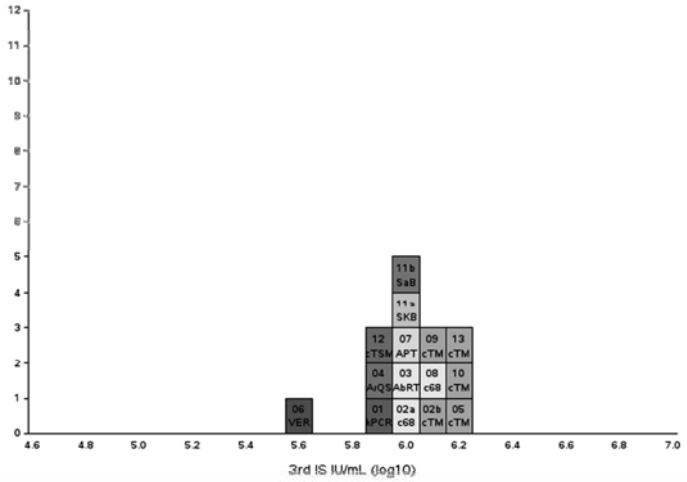
a Sample 2 relative to Sample 1



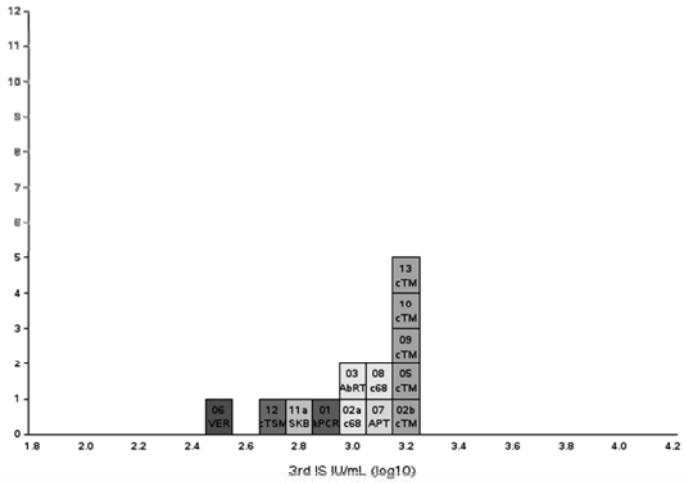
b Sample 3 relative to Sample 1



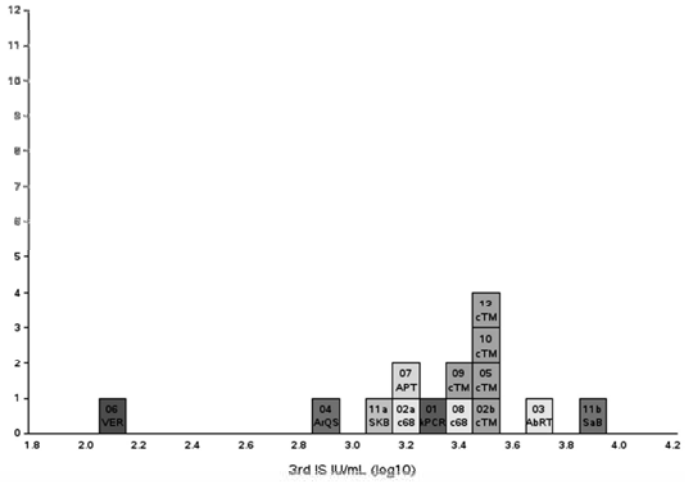
c Sample 4 relative to Sample 1



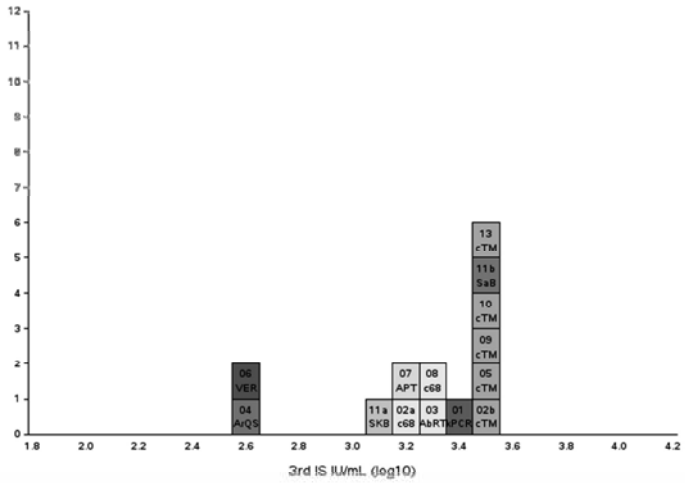
d Sample 5 relative to Sample 1



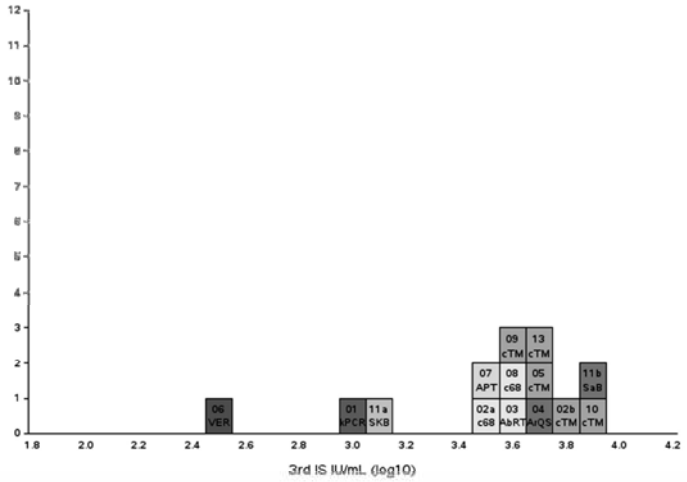
e Sample 6 relative to Sample 1



f Sample 7 relative to Sample 1



g Sample 8 relative to Sample 1



Appendix 2

Example of the calibration of a national standard (collaborative study calibration using multiple assays)

Journal of Virological Methods 191 (2013) 122–127

Contents lists available at SciVerse ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

A collaborative study to establish the first National Standard for HIV-1 RNA nucleic acid amplification techniques (NAT) in Taiwan

Yi-Chen Yang[a,b,*], Daniel Yang-Chih Shih[a], Mong-Hsun Tsai[b], Chia-Hung Cheng[a], Hwei-Fang Cheng[a], Chi-Fang Lo[a], Der-Yuan Wang[a]

[a] Food and Drug Administration, Department of Health, Executive Yuan, Taipei 11561, Taiwan
[b] Institute of Biotechnology, National Taiwan University, Taipei 10672, Taiwan

Article history:
Received 5 January 2013
Received in revised form 26 March 2013
Accepted 4 April 2013
Available online 19 April 2013

Keywords:
Standard
Nucleic acid amplification techniques (NAT)
Human immunodeficiency virus type 1 (HIV-1)

A B S T R A C T

The World Health Organization (WHO) International Standard (IS) for human immunodeficiency virus type 1 (HIV-1) RNA is only available in limited amounts. It is critical to use the most common HIV-1 genotype as source for secondary standards, e.g. a National Standard for Taiwan. The objective of this study was to establish the first National Standard for HIV-1 RNA NAT assays in Taiwan. A collaborative study, including eleven laboratories from five different countries, was carried out to establish the HIV-1 RNA National Standard by calibration, in International Units (IU), against the WHO HIV-1 RNA IS. The HIV-1 RNA content for the candidate was quantitated by each laboratory in three independent assays and the results were collected and analyzed statistically. Overall, a high level of agreement among results was achieved from different laboratories. In addition, the stability study indicated that the candidate was stable for 24 months at $-80 \pm 5\,^{\circ}C$. In conclusion, the candidate standard was established as the first National Standard for HIV-1 RNA for use in NAT assays in Taiwan. The standard is intended to be used for the quality control of HIV-1 NAT assays and as a quantitative reference material for HIV-1 NAT assays.

1. Introduction

Human immunodeficiency virus type 1 (HIV-1) is the causative agent of acquired immunodeficiency syndrome, commonly known as AIDS (Weiss, 1993). According to the UNAIDS report on the global AIDS epidemic in 2012, an estimated 34 million people are infected HIV (UNAIDS report, 2012) after 30 years of a very complex epidemic. Almost 30 million people have died from HIV-related diseases so far. In Taiwan, the number of reported cases of HIV/AIDS was approximately twenty four thousand from 1984 to 2012 December (CDC of Taiwan, 2012). Increasing sexual activity and needle sharing activity among drug abusers has resulted in HIV/AIDS becoming a severe public health problem.

HIV-1 is a member of the Retroviridae family and belongs to the *Lentivirus* genus. The RNA genome of HIV-1 is approximately 9.7 kb, containing three structural genes (gag, pol, and env) and six regulating genes (tat, rev, nef, vif, vpu, and vpr) (Pluta and Kacprzak, 2009; Bolinger and Boris-Lawrie, 2009; Karlsson Hedestam et al., 2008). HIV-1 strains are categorized as major group (M group), outlier group (O group), new group (N group), or P group on the basis of differences in the envelope region. The M group is further divided into 9 genetic subtypes (A–D, F–H, J, and K) and circulating recombinant forms (CRFs). In Taiwan, subtype B was found to be the predominant genotype in homosexual males and in the intravenous drug abuser population. In recent years, subtype CRF 07_BC has been the major group in intravenous drug abuser population (Spira et al., 2003; Robertson et al., 2000; Simon et al., 1998; Lin et al., 2007; Plantier et al., 2009).

Screening of blood and plasma products for blood–borne viruses has usually been performed using sensitive antibody-detection assays. In recent years, the nucleic-acid amplification techniques (NAT) have been widely applied in blood safety screens to enhance the sensitivity of detection of HIV-1 when present in low concentrations and at earlier stages of infection (Piatak et al., 1993; Murthy et al., 1999; Busch and Dodd, 2000).

To improve the safety of plasma products, a requirement that the plasma pools used to manufacture plasma products should be screened for HIV RNA by NAT was announced by the Department of Health in Taiwan on December 19, 2002. The development of a calibrated national reference standard that could be used routinely in assays would give assurance as to the validity of the test results and therefore fulfillment with such regulations. In addition to this national requirement for plasma screening, the HIV viral load assay is very critical in the management of antiretroviral therapy. Recently, numerous in vitro diagnostic devices (IVDs)

* Corresponding author at: 161-2, Kunyang St., Nangang, Taipei 11561, Taiwan. Tel.: +886 2 27877751.
E-mail address: ycyang@fda.gov.tw (Y.-C. Yang).

0166-0934/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jviromet.2013.04.002

based on NAT technology have been developed to detect HIV qualitatively or quantitatively for blood screening or viral load measurement. Such NAT-based IVDs are highly related to blood safety, the quality and performance of the IVDs are of great importance. In order to ensure the continued fitness for purpose of the IVD, both the pre-market approval testing and the performance evaluation are crucial in post-marketing surveillances.

The first and second World Health Organization (WHO) International Standard (IS) for HIV-1 RNA was established by the WHO Expert Committee on Biological Standardization (ECBS) in 1999 and 2008, and the National Institute for Biological Standards and Control (NIBSC) code numbers are 97/656 and 97/650, respectively (Holmes et al., 2001; Davis et al., 2008). One of the main purposes of an International Standard is to facilitate the calibration of secondary working reagents developed at a local level, i.e. by individual laboratories. The calibration of secondary working reagents from a higher order standard would help reduce test result variability from different laboratories, aid in comparing the different commercial and 'in-house' assays, and make it easier to compare the proficiency of different laboratories (Revets et al., 1996; Schuurman et al., 1996). In order to ensure the correct use of the International Standard by the end user, for example for secondary working reagent calibration and not for use as a routine run control, the WHO IS for HIV-1 RNA is available only in limited amounts, several control laboratories (such as National Institute for Biological Standards and Control (NIBSC), Food and Drug Administration (FDA) and Istituto Superiore di Sanità (ISS)) have already prepared an in house or national secondary HIV-1 NAT working reagent themselves (Davis et al., 2003; Lee et al., 2006; Pisani et al., 2007). It is known that the distribution of HIV subtypes may differ by geographic region. It is therefore critical to use the major genotype of the HIV-1 as a source material for a National Standard. Since subtype B was found to be the predominant genotype in Taiwan, the genotype of the National Standard would select to be subtype B, the same as WHO IS. Therefore, the objective of this study was to establish the first National Standard for HIV-1 RNA NAT assays and to calibrate the HIV-1 RNA content of the candidate standard against the WHO IS for HIV-1 RNA NAT assays (97/650). The procedure for the development of a National Standard was based on the previous experience of the development of the National Standard for human parvovirus B19 DNA nucleic acid amplification techniques (NAT) in 2008 (Yang et al., 2008).

2. Materials and methods

2.1. Preparation of the candidate standard

The candidate standard for HIV-1 RNA NAT assays was liquid preparation and stored at or below −70 °C. It was prepared by diluting HIV-1 RNA positive plasma in pooled human cryosupernatant. The proposed titer was approximately 10^4 IU/mL. The original HIV-1 RNA positive plasma had a titer of HIV-1 RNA of approximately 4.7×10^4 IU/mL and was negative for HBsAg, HBV DNA, anti-HCV, HCV RNA, HAV RNA as well as B19V DNA. The genotype of the HIV-1 RNA positive plasma was confirmed as subtype B by sequencing. The cryosupernatant was negative for HBsAg, HBV DNA, B19V DNA, anti-HCV, HCV RNA, anti-HIV 1/2, HIV-1 RNA, and HAV RNA.

2.2. Design of the international collaborative study

The aim of the international collaborative study was to calibrate the titers of the HIV-1 RNA National Standard that was prepared by the Taiwan Food and Drug Administration (TFDA). Including our laboratory, a total of eleven laboratories from five different countries have participated in this study. Participants received the proposed national candidate standard and WHO IS for HIV-1 RNA (97/650) and were requested to perform three independent assays for HIV-1 RNA using the candidate standard and the WHO IS. For each assay, serial dilutions of the WHO IS were prepared using the appropriate diluent. The recommended diluent for the study was HIV-1 RNA negative human plasma. The serially diluted IS were used to calibrate the candidate standard (Sample A) by creating a standard curve. If a commercial kit was used, the IS could be treated as a second unknown sample (sample B) and quantitated in parallel with sample A. A single estimate was obtained for each sample in each laboratory and for each assay method by calculating the geometric mean of repeat data within a single assay. The overall mean and SD were then calculated from the results of all participants.

2.3. Stability study on the candidate standard

Vials of the proposed national candidate standard were incubated at +4 °C, +24 °C, −20 °C, and −80 °C, three vials were removed at regular intervals for three independent tests. Two different commercial assays were used for quantitative analysis: the Abbott RealTime HIV-1 (Abbott Molecular Inc., Des Plaines, IL, USA) and the COBAS Ampliprep/COBAS TaqMan HIV-1 Test, v1.0 (Roche Molecular Systems, Inc., Branchburg, NJ, USA). Here, 500 uL of the HIV-1 RNA candidate standard was used in Abbott RealTime HIV-1 Kit and 1 mL 5× pre-diluted candidate standard was used in COBAS Ampliprep/COBAS TaqMan HIV-1 Test Kit. Both real-time PCR systems, the Abbott m2000 RealTime system and the COBAS Ampliprep/TaqMan 48 system, were used according to the manufacturer's instructions.

3. Results

3.1. Assay methods

Ten of the participants performed quantitative assays: four laboratories used the Abbott RT HIV-1 (Abbott Molecular Inc., Des Plaines, IL, USA); two laboratories used the COBAS AmpliPrep/COBAS TaqMan HIV-1 Test, v1.0 (Roche Molecular Systems, Inc., Branchburg, NJ, USA); two laboratories used the COBAS AmpliPrep/COBAS TaqMan HIV-1 Test, v2.0 (Roche Molecular Systems, Inc., Branchburg, NJ, USA); two laboratories used the COBAS TaqMan HIV-1 Test, for use with High Pure System Viral Nucleic Acid Kit, v1.0 (Roche Molecular Systems, Inc., Branchburg, NJ, USA); two laboratories used the VERSANT HIV-1 RNA 3.0 (Siemens Healthcare Diagnostics Inc., Tarrytown, NY, USA); and one laboratory used the COBAS Amplicor HIV-1 Monitor Test, v1.5 (Roche Molecular Systems, Inc., Branchburg, NJ, USA). Between them, lab code 1 used two different methods to detect HIV-1RNA and analyzed the data separately (1A, 1B), and lab code 2 used three different methods to detect HIV-1RNA and reported the results separately (2A, 2B, 2C). The overall results were therefore based on a maximum of 13 data sets. All these data sets were obtained by commercial assays. The quantitative methods used are summarized in Table 1. The other one participant used the COBAS HIV-1 AmpliScreen Test, v1.5 (Roche Molecular Systems, Inc., Branchburg, NJ, USA), which is a qualitative assay that was only give "positive" results (Detection limit: 78.4 IU/mL) and could not be calculated.

3.2. Estimated value of the HIV-1 RNA for the candidate standard

The estimated values of HIV-1 RNA, relative to the International Standard, for the candidate standards from each laboratory are listed in Table 2 and shown in Fig. 1. All the values have shown a good agreement with each other, except one laboratory has submitted an outlying result. The value of HIV-1 RNA estimate from each laboratory is shown in Fig. 2. Each box represents the estimate from

Table 1
An overview of the quantitative assays used in the collaborative study.

Lab code[a]	Assay method	Region for primer design
1B 2A 3 8	Abbott RT HIV-1	A highly conserved region in HIV-1 *pol-int* genes[b]
1A 6	COBAS AmpliPrep/COBAS TaqMan HIV-1 Test, v1.0	A highly conserved region in HIV-1 *gag* p41 gene
2C 5	COBAS AmpliPrep/COBAS TaqMan HIV-1 Test, v2.0	Two highly conserved regions of the HIV-1 genome–*gag* and long terminal repeat (LTR)
2B 10	COBAS TaqMan HIV-1 Test, for use with High Pure System Viral Nucleic Acid Kit, v1.0	A highly conserved region in HIV-1 *gag* p41 gene
9	COBAS Amplicor HIV-1 Monitor Test, v1.5	A highly conserved regions in HIV-1 gag gene
4 7	VERSANT HIV-1 RNA 3.0 (b DNA)	A highly conserved region in HIV-1 *pol* gene

[a] Two laboratories (lab code 1and 2) returned data from two and three different assay methods, respectively. The results are reported separately.
[b] *pol-int* gene is the integrase region of the polymerase gene.

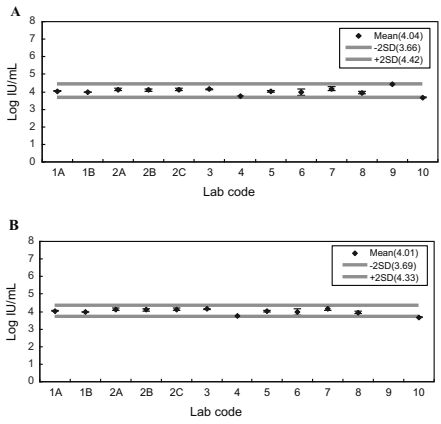


Fig. 1. The estimated values of HIV-1 RNA (Log IU/mL) for candidate standard as determined by 10 laboratories (A) and by 9 laboratories (B). The results showed that all the laboratories were in good agreement with the estimates, except one laboratory submitted an outlying result (A). The data generated from Lab code 9 was excluded from the overall means for the candidate standard (B).

Table 2
The estimated values of HIV-1 RNA (Log IU/mL)[a] for candidate standard from 10 laboratories.

Lab code	Mean	Minimum	Maximum	CV (%)
1A	4.04	4.03	4.06	0.37
1B	3.98	3.96	4.01	0.58
2A	4.13	4.06	4.23	1.37
2B	4.11	4.01	4.18	1.63
2C	4.13	4.06	4.24	1.79
3	4.15	4.12	4.17	0.63
4	3.73	3.67	3.79	1.04
5	4.03	3.99	4.08	1.09
6	3.98	3.84	4.08	4.26
7	4.18	4.04	4.26	2.83
8	3.95	3.90	4.03	1.84
9	–	4.42[b]	4.42[b]	–
10	3.68	3.65	3.70	0.59

[a] The measurements were performed using WHO International Standard for human HIV-1 RNA (WHO IS, 97/650) as the standard.
[b] Only one assay result was available from this laboratory, not three independent assay results.

one laboratory and/or assay method. All data were within a range of 1.0 Log for each sample, indicating that all the laboratories were in good agreement with the estimates. A comparison of the different commercial kit results is shown in Table 3; the results showed that the COBAS Amplicor HIV-1 Monitor Test, v1.5 was significantly different from other commercial assay kits. The data generated from the COBAS Amplicor HIV-1 Monitor Test was therefore excluded from the overall means for the candidate standard. Therefore, the

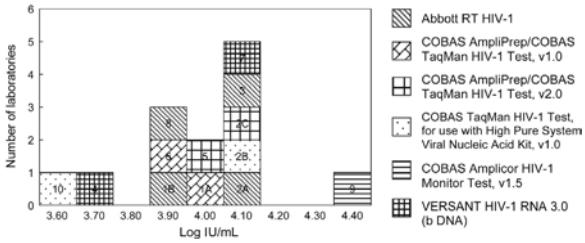


Fig. 2. The histogram of estimated values of HIV-1 RNA (Log_{10} IU/mL) for candidate standard from 10 laboratories. The number labeled in the box represented the laboratory code number.

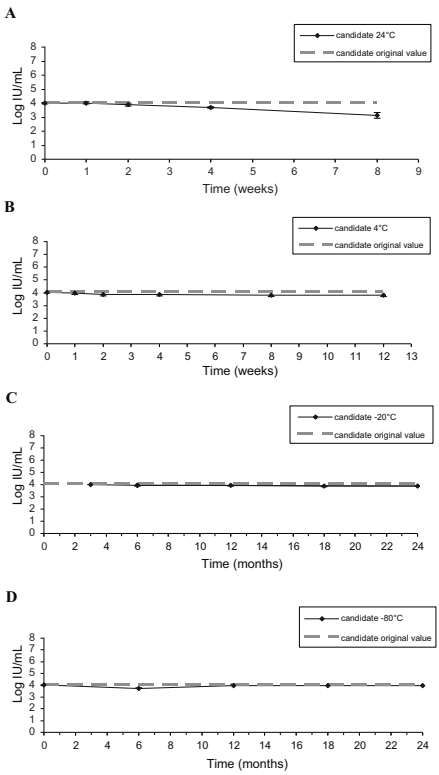


Fig. 3. Stability analysis of HIV-1 RNA present in candidate standard after storage at different temperatures: (A) +24 °C for 8 weeks, (B) +4 °C for 12 weeks, (C) −20 °C for 24 months and (D) −80 °C for 24 months. The solid line and the dotted line represent the tested values and the original values before storage, respectively.

overall mean for the candidate standard is 1.0×10^4 IU/mL, and the 95% confidence intervals is 8.26×10^3 to 1.25×10^4 IU/mL (Table 4).

3.3. Stability analysis of the candidate standard

Several vials of the candidate standard were stored at +24 °C, +4 °C, −20 °C, and −80 °C, three vials were randomly selected for stability tests. Samples were taken after one-, two- or four-week intervals from candidate samples stored at +24 °C and +4 °C and after three- or six-month intervals from candidate samples stored at −20 °C and −80 °C. Triplicate samples were assayed for each time point at different temperatures in three independent tests. The calculated mean concentration (IU/mL) for each time point and temperature is shown in Fig. 3. The results indicate that the candidate samples were stable after storage at +24 °C for 4 weeks, at +4 °C for 8 weeks, at −20 °C for 24 months, and at −80 °C for 24 months. The results suggest good long term stability for the proposed national candidate standard when stored at −20 °C and −80 °C.

4. Discussion

There was a clear variation between the results from earlier HIV-1 viral load assays such as nucleic acid signal branch amplification (NASBA), PCR end point detection and branched DNA (bDNA) signal amplification compared to more recent tests such as the Abbott real-time assay. The limitation of these assays has previously been reported (Church et al., 2011) and it is known that are optimized to target subtype B group M viruses. New-generation real-time PCR assays for HIV-1 RNA quantification include the Abbott RT HIV-1 assay and the Cobas Ampli-Prep/Cobas TaqMan HIV-1 assay (CAP-CTM). These real-time PCR assays have been improved and are now able to detect HIV-1 group M, non-B subtype viruses, group N viruses and O viruses. In addition, the Abbott RealTime HIV-1 assay has been reported to successfully detect HIV-1 group P infection (Plantier et al., 2009). However, it has also been reported that multiple mismatches in gag primers and probe binding regions for the first version of the CTM assay (CTM1) exist, which can result in an underestimation of the CTM1 values for some patients infected with HIV-1 group M, non-B subtypes. To overcome this problem, Roche Diagnostics have already upgraded their test to CTM version 2.0 (CTM2), which uses a dual-target strategy (Damond et al., 2010; Church et al., 2011; Wirden et al., 2011).

In this international collaborative study, most of the participants performed quantitative assays: four laboratories used the Abbott RT HIV-1; six laboratories used the COBAS TaqMan HIV-1 Test (COBAS AmpliPrep/COBAS TaqMan HIV-1 Test, v1.0 & 2.0, and COBAS TaqMan HIV-1 Test, for use with High Pure System Viral Nucleic Acid Kit, v1.0); two laboratories used the VERSANT HIV-1 RNA 3.0; and one laboratory used the COBAS Amplicor HIV-1 Monitor Test, v1.5. The results showed that all the laboratories were in good agreement with the estimates, except one laboratory submitted an outlying result, which is generated from COBAS Amplicor HIV-1 Monitor Test, v1.5. By comparing the results from different commercial kits, indicated that the COBAS Amplicor HIV-1 Monitor Test, v1.5 was significantly different from other commercial assay kits. The data generated from the COBAS Amplicor HIV-1 Monitor Test was therefore excluded from the overall means for the candidate sample. Since there is only one assay result available in this collaborative study, it does not represent the performance of the kit. Interestingly, a similar result was also shown in an earlier collaborative study to establish a replacement International Standard for HIV-1 RNA nucleic acid assays (Davis et al., 2008).

All data points received from laboratories were within a range of 1.0 Log from this collaborative study, furthermore, most of the data were within in a range from 3.9 Log IU/mL to 4.1 Log IU/mL. In conclusion, a high level of agreement among the results obtained from the participating laboratories was observed. The first National Taiwan Standard for HIV-1 RNA NAT assays, with an assigned value of 1.0×10^4 IU/mL, was recognized. In order to reflect the predominant HIV-1 subtype found in Taiwan, this National Standard was formulated from a subtype B plasma. The results of the stability study indicated that the HIV-1 RNA National Standard is stable long-term when stored at −20 °C and -80 °C. Therefore, the first National Standard for HIV-1 RNA NAT assays in Taiwan was established. This standard could be used for quality control of HIV-1 RNA assays and as a quantitative reference material for HIV-1 NAT assays. Moreover, the standard could be used nationally for pre-market approval testing and the performance evaluation in post-marketing surveillances of NAT-based IVDs and facilitating to ensure the continued fitness for purpose of the IVD, either imported or domestic.

In recent years, subtype CRF 07_BC has been the major group of HIV-1 found in the intravenous drug abuser population in Taiwan. As HIV-1 strain diversity and viral recombination events increase, the need for surveillance using commercial assays to

Table 3
Comparison of different commercial kits.

Assay method	Lab code	Result	Mean	SD	CV (%)
Abbott RT HIV-1	1B	3.98	4.05	0.10	2.52
	2A	4.13			
	3	4.15			
	8	3.95			
COBAS AmpliPrep/COBAS TaqMan HIV-1 test, v1.0	1A	4.04	4.01	0.04	1.06
	6	3.98			
COBAS AmpliPrep/COBAS TaqMan HIV-1 test, v2.0	2C	4.13	4.08	0.07	1.73
	5	4.03			
COBAS TaqMan HIV-1 test, for use with High Pure System Viral Nucleic Acid Kit, v1.0	2B	4.11	3.90	0.30	7.81
	10	3.68			
COBAS Amplicor HIV-1 Monitor Test, v1.5	9	4.42[a]	4.42[a]		
VERSANT HIV-1 RNA 3.0 (b DNA)	4	3.73	3.96	0.32	8.05
	7	4.18			

[a] $p < 0.05$. Please note that a single assay result does not represent the performance of the kit.

Table 4
Overall mean estimates of HIV-1 RNA (Log IU/mL) for candidate standard.

Sample	Mean		95% confidence interval (95% CI)	
	Log IU/mL	IU/mL	Log IU/mL	IU/mL
Candidate standard	4.01	1.01E+04	3.92–4.10	8.26E+03–1.25E+04

ensure detection ability and viral load monitoring accuracy in HIV-1-infected patients has increased. To fulfill this goal, an HIV-1 CRF 07_BC National standard will need to be established as the next major step.

Disclaimer

The findings and conclusions in this article have not been formally disseminated by Taiwan FDA and should not be construed to represent any agency determination or policy.

Acknowledgements

We would like to express our sincere thanks to all the participants of the collaborative study group (Appendix A) for their valuable contribution to this collaborative study and to Ai-Zhen Yang and Yu-Chen Pao for their technical assistance for formulating the candidate standard. We would like to thank Clare Morris for providing language help. In addition, we also acknowledge the funding support from Taiwan FDA.

Appendix A.

List of participants:

- Brian Erickson, Abbott Molecular Inc., USA.
- Chien-Feng Sun/Shuan Yang, Chang-Gung Memorial Hospital, Taiwan.
- Clare Morris, National Institute for Biological Standards and Controls (NIBSC), UK.
- Chih-Yuan Yang/Cheng-Feng Kao, Centers for Disease Control (TCDC), Taiwan.
- Der-Yuan Wang/Yi-Chen Yang, Food and Drug Administration (TFDA), Taiwan.
- Eric T. Natoli, Siemens Healthcare Diagnostics, USA.
- Indira Hewlett/Sherwin Lee, Center for Biologics Evaluation and Research/Food and Drug Administration (CBER/FDA), USA.
- John Saldanha/Matthew Lin, Roche Molecular Systems, USA.
- Lena Panagiotopoulos/Stirling Dick, National Serology Reference Laboratory (NRL), Australia.
- Micha Nübling/Michael Chudy, Paul-Ehrlich-Institute, Germany.
- Yi-Li Shih, E-Da Hospital, Taiwan.

References

Bolinger, C., Boris-Lawrie, K., 2009. Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. Retrovirology 6, 8.
Busch, M.P., Dodd, R.Y., 2000. NAT and blood safety: what is the paradigm? Transfusion 40, 1157–1160.
Church, D., Gregson, D., Lloyd, T., Klein, M., Beckthold, B., Laupland, K., Gill, M.J., 2011. Comparison of the RealTime HIV-1, COBAS TaqMan 48 v1.0, Easy Q v1.2, and Versant v3.0 assays for determination of HIV-1 viral loads in a cohort of Canadian patients with diverse HIV subtype infections. J. Clin. Microbiol. 49, 118–124.
Damond, F., Avettand-Fenoel, V., Collin, G., Roquebert, B., Plantier, J.C., Ganon, A., Sizmann, D., Babiel, R., Glaubitz, J., Chaix, M.L., Brun-Vezinet, F., Descamps, D., Rouzioux, C., 2010. Evaluation of an upgraded version of the Roche Cobas AmpliPrep/Cobas TaqMan HIV-1 test for HIV-1 load quantification. J. Clin. Microbiol. 48, 1413–1416.
Davis, C., Heath, A., Best, S., Hewlett, I., Lelie, N., Schuurman, R., Holmes, H., 2003. Calibration of HIV-1 working reagents for nucleic acid amplification techniques against the 1st international standard for HIV-1 RNA. J. Virol. Methods 107, 37–44.
Davis, C., Berry, N., Heath, A., Holmes, H., 2008. An international collaborative study to establish a replacement World Health Organization International Standard for human immunodeficiency virus 1 RNA nucleic acid assays. Vox Sang. 95, 218–225.
Holmes, H., Davis, C., Heath, A., Hewlett, I., Lelie, N., 2001. An international collaborative study to establish the 1st international standard for HIV-1 RNA for use in nucleic acid-based techniques. J. Virol. Methods 92, 141–150.
Karlsson Hedestam, G.B., Fouchier, R.A., Phogat, S., Burton, D.R., Sodroski, J., Wyatt, R.T., 2008. The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. Nat. Rev. Microbiol. 6, 143–155.
Lee, S., Wood, O., Taffs, R.E., Hu, J., Machuca, A., Vallejo, A., Hewlett, I., 2006. Development and evaluation of HIV-1 subtype RNA panels for the standardization of HIV-1 NAT assays. J. Virol. Methods 137, 287–291.
Lin, Y.T., Lan, Y.C., Chen, Y.J., Huang, Y.H., Lee, C.M., Liu, T.T., Wong, W.W., Yang, J.Y., Wang, C.T., Chen, Y.M., 2007. Molecular epidemiology of HIV-1 infection and full-length genomic analysis of circulating recombinant form 07_BC strains from injection drug users in Taiwan. J. Infect. Dis. 195, 1283–1293.
Murthy, K.K., Henrard, D.R., Eichberg, J.W., Cobb, K.E., Busch, M.P., Allain, J.P., Alter, H.J., 1999. Redefining the HIV-infectious window period in the chimpanzee model: evidence to suggest that viral nucleic acid testing can prevent blood–borne transmission. Transfusion 39, 688–693.
Piatak Jr., M., Saag, M.S., Yang, L.C., Clark, S.J., Kappes, J.C., Luk, K.C., Hahn, B.H., Shaw, G.M., Lifson, J.D., 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. Science 259, 1749–1754.

Pisani, G., Marino, F., Cristiano, K., Bisso, G.M., Mele, C., Luciani, F., Wirz, M., Gentili, G., 2007. Collaborative study for the calibration of HCV RNA, HBV DNA and HIV RNA reference preparations against the relative international standards. Ann. Ist. Super. Sanita 43, 69–76.

Plantier, J.C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D.L., Simon, F., 2009. A new human immunodeficiency virus derived from gorillas. Nat. Med. 15, 871–872.

Pluta, K., Kacprzak, M.M., 2009. Use of HIV as a gene transfer vector. Acta Biochim. Pol. 56, 531–595.

Revets, H., Marissens, D., de Wit, S., Lacor, P., Clumeck, N., Lauwers, S., Zissis, G., 1996. Comparative evaluation of NASBA HIV-1 RNA QT, AMPLICOR-HIV monitor, and QUANTIPLEX HIV RNA assay, three methods for quantification of human immunodeficiency virus type 1 RNA in plasma. J. Clin. Microbiol. 34, 1058–1064.

Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S., Korber, B., 2000. HIV-1 nomenclature proposal. Science 288, 55–56.

Schuurman, R., Descamps, D., Weverling, G.J., Kaye, S., Tijnagel, J., Williams, I., van Leeuwen, R., Tedder, R., Boucher, C.A., Brun-Vezinet, F., Loveday, C., 1996. Multicenter comparison of three commercial methods for quantification of human immunodeficiency virus type 1 RNA in plasma. J. Clin. Microbiol. 34, 3016–3022.

Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F., Brun-Vezinet, F., 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat. Med. 4, 1032–1037.

Spira, S., Wainberg, M.A., Loemba, H., Turner, D., Brenner, B.G., 2003. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. J. Antimicrob. Chemother. 51, 229–240.

Statistics of HIV/AIDS 12/31/2012. Center for Disease Control, ROC (Taiwan). 2012, 1.

UNAIDS report on the global AIDS epidemic 2012. Joint United Nations Programme on HIV/AIDS, pp. 8–14.

Weiss, R.A., 1993. How does HIV cause AIDS? Science 260, 1273–1279.

Wirden, M., Tubiana, R., Fourati, S., Thevenin, M., Simon, A., Canestri, A., Ait-Arkoub, Z., Soulie, C., Marcelin, A.G., Katlama, C., Calvez, V., 2011. Upgraded Cobas Ampliprep-Cobas TaqMan version 2.0 HIV-1 RNA quantification assay versus first version: correction of underestimations. J. Clin. Microbiol. 49, 2700–2702.

Yang, Y.C., Chang, C.J., Tseng, C.I., Wang, D.Y., Cheng, H.F., Lin, C.P., 2008. Establishment of a National Taiwan Standard and working reagent for human parvovirus B19 DNA nucleic acid amplification techniques (NAT). J. Food Drug Anal. 16, 9–14.

Appendix 3

Example of the calibration of a reference preparation by a single NAT assay

The reference preparation (RP) is calibrated against the WHO IS by testing dilutions of both samples using a real-time PCR, collecting raw data (threshold cycle values Ct) and performing a valid statistical analysis (e.g. parallel line). Study performed under the supervision of G. Pisani, ISS, Rome, Italy.

Study samples:

- RP for HIV RNA with a presumptive titre of 15 000 IU/mL of HIV RNA
- 3rd WHO IS HIV RNA batch 10/152 with a concentration of 185 000 IU/mL (5.26 log IU/mL).

Test the following dilutions of WHO IS and RP (in triplicate) on three separate days. Collect the raw data.

Day 1

Sample	Final dilution (log)	Concentration	Raw data (Ct value)		
			Replica 1	Replica 2	Replica 3
WHO IS	–1.09	15 000 IU/mL	28.5	28.5	28.3
	–1.59	4 700 IU/mL	29.8	30.0	30.0
	–2.09	1 500 IU/mL	31.6	31.5	31.6
	–2.59	470 IU/mL	32.9	33.4	33.5
RP	Not diluted	–	28.4	28.9	28.7
	–0.50	–	30.9	30.3	30.5
	–1.00	–	32.1	31.9	32.5
	–1.50	–	33.6	34.8	34.1

Day 2

Sample	Final dilution (log)	Concentration	Raw data (Ct value)		
			Replica 1	Replica 2	Replica 3
WHO IS	–1.09	15 000 IU/mL	28.7	28.4	28.4
	–1.59	4 700 IU/mL	29.7	29.9	30.0
	–2.09	1 500 IU/mL	31.5	31.4	31.7
	–2.59	470 IU/mL	33.3	33.2	33.4
RP	Not diluted	–	29.1	28.6	29.1
	–0.50	–	30.1	30.8	31.3
	–1.00	–	32.9	32.6	31.5
	–1.50	–	34.5	33.2	33.3

Day 3

Sample	Final dilution (log)	Concentration	Raw data (Ct value)		
			Replica 1	Replica 2	Replica 3
WHO IS	–1.09	15.000 IU/mL	28.6	28.3	28.2
	–1.59	4.700 IU/mL	29.1	29.3	30.1
	–2.09	1.500 IU/mL	31.1	31.2	31.5
	–2.59	470 IU/mL	33.4	33.3	32.5
RP	Not diluted	–	28.8	28.9	28.7
	–0.50	–	30.1	30.5	30.2
	–1.00	–	32.1	32.2	31.9
	–1.50	–	33.5	33.2	34.3

Perform statistical analysis: parallel-line assay.

Acceptance criteria: linearity and parallelism should be fulfilled.

It is possible that on each day (each experiment) one replicate or one dose may be deleted in order to fulfil the acceptance criteria.

Example calibration: Combistat (EDQM).

Combistat Version 5.0. Thursday, 19 May 2016, 11:25:34 [+01:00]. Page 1 of 1

Substance	HIV RNA
Method	TaqScreen MPX
Assay number	1
Technician	
Date of assay	DAY 1

Remarks: Calibration of Reference Preparation for HIV RNA against the WHO IS.
EXAMPLE DAY 1

Standard			
HCV	WHO IS 10/102		
Ass. pot.	185000IU/mL		
Pre-dil. 1	1 mL/12.33mL		
Doses	(1)	(2)	(3)
1/1	28.5	28.5	28.3
1/3.16	29.8	30.0	30.0
1/10	31.6	31.5	31.6
1/31.6	32.9	33.4	33.5

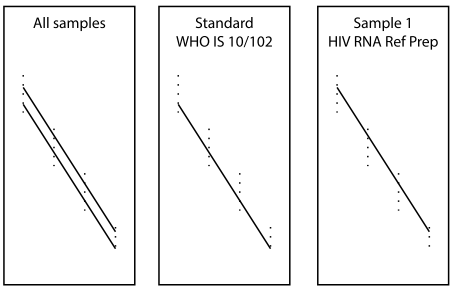
Sample 1			
	HIV RNA Ref Prep		
Ass. pot.	?IU/mL		
Pre-dil. 1			
Doses	(1)	(2)	(3)
1/1	28.4	28.9	28.7
1/3.16	30.9	30.3	30.5
1/10	32.1	31.9	32.5
1/31.6	33.6	34.8	34.1

Model: Parallel lines
Design: Completely randomized
Transformation: $\gamma' = \gamma$
Variance: Observed residuals

Common slope (factor) = -1.48691 (-1.57151 to -1.40232)
Correlation | r |: 0.989630

Source of variation	Degrees of freedom	Sum of squares	Mean square	F-ratio	Probability
Preparations	1	2.10042	2.10042	22.504	0.000 (***)
Regression	1	87.8933	87.8933	941.714	0.000 (***)
Non-parallelism	1	0.290027	0.290027	3.107	0.097
Non-linearity	4	0.112543	0.0281358	0.301	0.873
Standard	2	0.0305883	0.0152942	0.164	0.850
Sample 1	2	0.0819547	0.0409774	0.439	0.652
Treatments	7	90.3963	12.9138	138.362	0.000 (***)
Residual error	16	1.49333	0.0933333		
Total	23	91.8896	3.99520		

Sample 1			
	HIV RNA Ref Prep		
(IU/mL)	Lower limit	Estimate	Upper limit
Potency	8399.12	10078.5	12047.6
Rel. to Ass.	?	?	?
Rel. to Est.	83.3%	100.0%	119.5%



Executed by: Calculated by: Approved by:

ID: ISTISAN/ITA

combistats Version 5.0. Thursday, 19 May 2016, 11:25:38 [+01:00]. Page 1 of 1

Substance	HIV RNA
Method	TaqScreen MPX
Assay number	2
Technician	
Date of assay	DAY 2

Remarks: Calibration of Reference Preparation for HIV RNA against the WHO IS.
EXAMPLE DAY 2

Standard			
HCV	WHO IS 10/102		
Ass. pot.	185000IU/mL		
Pre-dil. 1	1 mL/12.33mL		
Doses	(1)	(2)	(3)
1/1	28.7	28.4	28.4
1/3.16	29.7	29.9	30.0
1/10	31.5	31.4	31.7
1/31.6	33.3	33.2	33.4

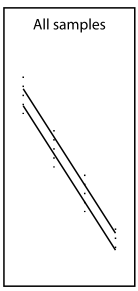
Sample 1			
	HIV RNA Ref Prep		
Ass. pot.	?IU/mL		
Pre-dil. 1			
Doses	(1)	(2)	(3)
1/1	29.1	28.6	29.1
1/3.16	30.1	30.8	31.3
1/10	32.9	32.6	31.5
1/31.6	34.5	33.2	33.3

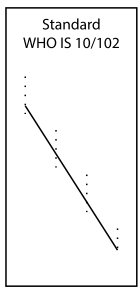
Model: Parallel lines
Design: Completely randomized
Transformation: $\gamma' = \gamma$
Variance: Observed residuals

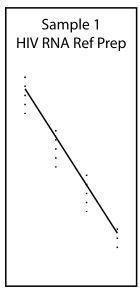
Common slope (factor) = -1.38413 (-1.50796 to -1.26029)
Correlation | r |: 0.978440

Source of variation	Degrees of freedom	Sum of squares	Mean square	F-ratio	Probability
Preparations	1	2.28167	2.28167	11.408	0.004 (***)
Regression	1	76.1616	76.1616	380.808	0.000 (***)
Non-parallelism	1	0.00533467	0.00533467	0.027	0.872
Non-linearity	4	0.289684	0.0724211	0.362	0.832
Standard	2	0.125761	0.0628805	0.314	0.735
Sample 1	2	0.163923	0.0819616	0.410	0.671
Treatments	7	78.7383	11.2483	56.242	0.000 (***)
Residual error	16	3.20000	0.200000		
Total	23	81.9383	3.56254		

Sample 1			
	HIV RNA Ref Prep		
(IU/mL)	Lower limit	Estimate	Upper limit
Potency	7184.66	9609.89	12717.7
Rel. to Ass.	?	?	?
Rel. to Est.	74.8%	100.0%	132.3%







Executed by: Calculated by: Approved by:

ID: ISTISAN/ITA

combistats Version 5.0. Thursday, 19 May 2016, 11:25:30 [+01:00]. Page 1 of 1

Substance	HIV RNA
Method	TaqScreen MPX
Assay number	3
Technician	
Date of assay	DAY 3

Remarks: Calibration of Reference Preparation for HIV RNA against the WHO IS.
EXAMPLE DAY 3

Standard			
HCV	WHO IS 10/102		
Ass. pot.	185000IU/mL		
Pre-dil. 1	1 mL/12.33mL		
Doses	(1)	(2)	(3)
1/1	28.6	28.3	28.2
1/3.16	29.1	29.3	30.1
1/10	31.1	31.2	31.5
1/31.6	33.4	33.3	32.5

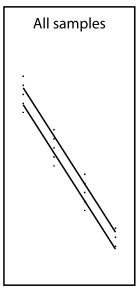
Sample 1			
	HIV RNA Ref Prep		
Ass. pot.	?IU/mL		
Pre-dil. 1			
Doses	(1)	(2)	(3)
1/1	28.8	28.9	28.7
1/3.16	30.1	30.5	30.2
1/10	32.1	32.2	31.9
1/31.6	33.5	33.2	34.3

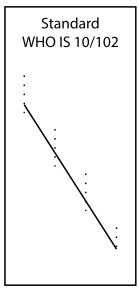
Model: Parallel lines
Design: Completely randomized
Transformation: $\gamma' = \gamma$
Variance: Observed residuals

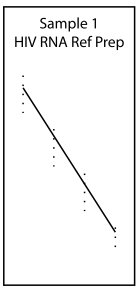
Common slope (factor) = -1.40151 (-1.49990 to -1.30312)
Correlation | r |: 0.984946

Source of variation	Degrees of freedom	Sum of squares	Mean square	F-ratio	Probability
Preparations	1	2.53500	2.53500	20.079	0.000 (***)
Regression	1	78.0867	78.0867	618.509	0.000 (***)
Non-parallelism	1	0.0213320	0.0213320	0.169	0.686
Non-linearity	4	0.441966	0.110492	0.875	0.500
Standard	2	0.386621	0.193310	1.531	0.246
Sample 1	2	0.0553456	0.0276728	0.219	0.806
Treatments	7	81.0850	11.5836	91.751	0.000 (***)
Residual error	16	2.02000	0.126250		
Total	23	83.1050	3.61326		

Sample 1			
	HIV RNA Ref Prep		
(IU/mL)	Lower limit	Estimate	Upper limit
Potency	7518.42	9436.03	11762.6
Rel. to Ass.	?	?	?
Rel. to Est.	79.7%	100.0%	124.7%







Executed by: Calculated by: Approved by:

ID: ISTISAN/ITA

Substance	HIV RNA
Method	TaqScreen MPX
Technician	
Ass. pot.	?IU/mL

Remarks:

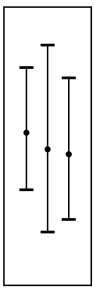
Assay number	Date of assay	Sample	Info	Lower limit	Estimate	Upper limit	df
1	DAY 1	1	HIV RNA Ref Prep	8399.12	10078.5	12047.6	16
2	DAY 2	1	HIV RNA Ref Prep	7184.66	9609.89	12717.7	16
3	DAY 3	1	HIV RNA Ref Prep	7518.42	9436.03	11762.6	16

Geometric combination
Homogeneity: p = 0.879

Standard			
(IU/mL)	Lower limit	Estimate	Upper limit
Potency	8677.41	9779.06	11020.6
Rel. to Ass.	?	?	?
Rel. to Est.	88.7%	100.0%	112.7%

Semi-weighted combination			
(IU/mL)	Lower limit	Estimate	Upper limit
Potency	8659.34	9776.06	11036.8
Rel. to Ass.	?	?	?
Rel. to Est.	88.6%	100.0%	112.9%

Unweighted combination			
(IU/mL)	Lower limit	Estimate	Upper limit
Potency	8918.24	9704.38	10559.8
Rel. to Ass.	?	?	?
Rel. to Est.	91.9%	100.0%	108.8%



Executed by: Calculated by: Approved by:

ID: ISTISAN/ITA