Protocol for the performance evaluation of nucleic acid tests for the quantitative detection of hepatitis B virus DNA for WHO prequalification assessment

Protocol for the performance evaluation of nucleic acid tests for the quantitative detection of hepatitis B virus DNA for WHO prequalification assessment

ISBN 978-92-4-007963-2 (electronic version)
ISBN 978-92-4-007964-9 (print version)

Suggested citation. Protocol for the performance evaluation of nucleic acid tests for the quantitative detection of hepatitis B virus DNA for WHO prequalification assessment. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO.
Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.
Sales, rights and licensing. To purchase WHO publications, see https://www.who.int/publications/book-orders. To submit requests for commercial use and queries on rights and licensing, see https://www.who.int/copyright.

Contents

1. Introduction

1.1 Prequalification of In Vitro Diagnostics

World Health Organization (WHO) Prequalification of in vitro diagnostics (IVD) is coordinated through the department of Regulation of Medicines and Other Health Technologies. Focus is placed on IVDs for priority diseases and their suitability for use in resource-limited settings.

WHO prequalification of IVDs is a comprehensive quality assessment of individual IVDs through a standardized procedure aimed at determining whether the product meets WHO prequalification requirements (*1*). Two types of prequalification assessment can take place, depending on the regulatory version submitted and evidence from a previous stringent review by a Recognized Stringent Regulatory Authority.

The full prequalification assessment process includes the following components:

- review of a product dossier
- performance evaluation including operational characteristics
- inspection of the manufacturing site(s)
- labelling review.

The abridged prequalification assessment includes the following components:

- performance evaluation including operational characteristics
- inspection of the manufacturing site(s)
- labelling review.

The performance evaluation will be conducted by a Performance Evaluation Laboratories site following a choice between two different mechanisms described here [1]. Performance evaluations conducted by a laboratory in List 1 will be coordinated and cost covered by WHO. Performance evaluations conducted by a laboratory in List 2 will be coordinated and the costs incurred covered by the manufacturer.

This protocol describes the procedures required to perform evaluations of nucleic acid tests for the quantitative detection of hepatitis B virus (HBV) DNA submitted for WHO prequalification assessment. **This protocol is not intended to replace validation and verification studies that need to be conducted by the manufacturer in order to fulfil WHO prequalification product dossier requirements.**

Given the variety of nucleic acid tests available, this protocol remains generic in nature and some sections may be open to interpretation. Manufacturers are encouraged to contact WHO before the start of the evaluation in order to verify that their preferred approach is in line with WHO expectations.

2. Intended audience

This document is intended to provide Performance Evaluation Laboratories and IVD manufacturers with the performance evaluation procedure for WHO prequalification assessment.

[1] https://extranet.who.int/prequal/vitro-diagnostics/performance-evaluation

3. Study objectives

3.1 Overall objective

The overall objective of the performance evaluation is to evaluate the performance and operational characteristics of commercially available Nucleic Acid Tests (NAT) intended for the quantification of HBV DNA in HBV-infected individuals and verify that they meet WHO prequalification requirements.

3.2 Specific objectives:

The specific objectives of the evaluation of the IVD are:

- to assess the test's agreement with a comparator test on clinical specimens, including:
    - quantitative assessment of bias, limits of agreement and correlation
    - misclassification above or below the clinical thresholds.
- to assess the analytical performance of the test under evaluation:
    - precision (repeatability and within-laboratory reproducibility)
    - linear range on the main genotypes
    - limit of detection
    - cross-contamination or carry-over.
- to describe the operational characteristics and ease of use of the tests and their suitability for use in countries with limited infrastructure (no/limited electricity, no/limited clean water, inadequate means of waste disposal).

4 Study implementation

4.1 WHO Performance Evaluation Laboratories

The performance evaluation will be exclusively conducted by a Performance Evaluation Laboratory (PEL). These laboratories have successfully undergone assessment Mechanism which includes submission of an expression of interest (EoI), Stage 1 audit (assessment of EoI and specific quality management system (QMS) documentation), Stage 2 on-site audit to assess compliance with WHO requirements. The list of PEL can be found here:

https://extranet.who.int/prequal/vitro-diagnostics/prequalified/performance-evaluation-laboratories

The laboratory shall hold one of the following certifications for quality management within the laboratory: ISO15189 (Medical laboratories: Particular requirements for quality and competence), ISO17025 (General requirements for the competence of testing and calibration laboratories), or equivalent.

The person(s) listed in the EoI letter to WHO will act as the Principal Investigator (PI) for the work performed by the PEL.

The evaluation may be conducted in two PELs if required, with one PEL performing the analytical performance evaluation and the other performing the clinical performance evaluation. One PEL will be designated as the leading PEL and will be in charge of doing or verifying the analysis and drafting the report.

4.2 Training, performance evaluation and supervision

The following issues are key to minimizing error and maximizing the value of the evaluation.

- Only personnel having received specific training for the evaluation and showing successful competency testing will be employed in the evaluation.
- Accurate record keeping is crucial to the success of the evaluation and the PI will be responsible for ensuring that all data required for the evaluation are recorded as agreed.
- Worksheets should be prepared and tubes, test devices or plates labelled prior to commencement of any run / test.
- Because objective, machine-generated, permanent results for some of the technologies available may not be feasible, it is essential that the PI emphasizes the need for accurate recordkeeping.
- To minimize the risk of error, results will be directly exported from the platform wherever possible. If this is not the case, results should be entered by one staff member and verified by another.

4.3 Safety

HIV, hepatitis B virus and hepatitis C virus and other viruses are transmissible by blood and body fluids. Therefore, all types of specimens (including venous and capillary whole blood, serum/plasma, oral fluid, etc.) must be handled as potentially infectious. Appropriate precautions to minimize infectious hazards must be taken at all stages from the collection of specimens to the disposal of used materials from the laboratory. The WHO Laboratory biosafety manual (*2*) and the site's guidelines on laboratory safety should be carefully followed by the laboratory staff.

4.4 Storage of reagents

All reagents must be stored as indicated in the Instructions for Use (IFU) document also known as the package insert. Calibrated thermometers must be placed at each location where reagents and specimens are stored, *i.e*. ambient, refrigerator and freezer, and temperatures recorded daily on temperature logs or automatically in a central temperature recording system. The lot numbers of the test kits received/used and their expiry dates shall be recorded.

Two separate production lots (**with different lot numbers and different expiry dates**) will be requested for evaluation, according to the following definition of a lot[2]: "The amount of material that is uniform in its properties and has been produced in one process or series of processes. The material can be either starting material, intermediate material or finished product." Each lot shall comprise different production (or manufacturing, purification, etc.) runs of critical reagents. Furthermore, lots must be sourced from a representative production run and not produced especially for the purpose of this evaluation. WHO will verify this information before the product assessment has been finalized.

[2] ISO 18113-1 :2022: In vitro diagnostic medical devices – Information supplied by the manufacturer (labelling) – Part 1 : Terms, definitions and general requirements

5 Specimen panels

5.1 Clinical performance panel

5.1.1 General description of the panel

Clinically derived specimens should comprise specimens collected from HBV-infected individuals (i.e. HBsAg positive). The panel will also include specimens from known HBV-negative individuals (i.e. negative for HBsAg and HBV DNA). Table 1 represents the minimum number of specimens required.

This panel will be used to assess the performance of the test under evaluation compared with a standard comparator test, including a quantitative assessment of bias and correlation, and agreement between the test results in specimens with HBV DNA levels above and below clinically relevant decision levels. The sample size (i.e. the number of specimens in the panel) and distribution of HBV DNA levels across the panel were determined to ensure acceptable confidence intervals around the estimate of agreement at HBV DNA levels of 20 000 IU and 2000 IU /mL, which may be used for making treatment decisions in certain situations, as well as descriptive analysis at HBV DNA levels of 200 000 IU/mL, which is used for prophylactic antiviral therapy in prevention of mother to child transmission (*3*).

Based on an expected agreement between the test results of 95% in specimens below and above the decision level and confidence interval total width of 7%, the sample size for specimens ≥20 000 HBV DNA IU/mL and <20 000 HBV DNA IU/mL will be at least 180 each.

In addition, 100 specimens from HBV-negative individuals (confirmed by negative results for HBsAg and HBV DNA) will be included for a separate estimate of specificity.

When different specimen types are claimed by the manufacturer with different performance characteristics (e.g. serum/plasma versus DBS), then a full clinical evaluation will be conducted with one specimen type (called main specimen type below) and a reduced clinical evaluation with the other specimen type(s) (called additional specimen type below).

Table 1. Number and characteristics of specimens for the clinical performance evaluation

Type of specimens	**Minimum number of specimens**	
	Main specimen type	**Additional specimen type**
Specimens from HBV-infected individuals		
≤ 20 000 HBV DNA IU/mL, including:	180	100
HBV DNA undetectable	*Approximately 50*	
Detectable <2 000 HBV DNA IU/mL	*Approximately 100*	
2000-20 000 HBV DNA IU/mL	*Approximately 30*	
> 20 000 HBV DNA IU/mL, including:	180	100
20 000 - 200 000 HBV DNA IU /mL	*Approximately 80*	
> 200 000 HBV DNA IU /mL	*Approximately 100*	
Specimens from HBV-uninfected individuals (HBsAg and HBV DNA negative)	100	50

5.1.2 Collection and storage of the clinical specimens

Clinical specimens used for the evaluation may include the following specimen types:

- left-over from specimens submitted for routine testing;
- rejected blood units, rejected because of positivity for blood-borne pathogens or for other reasons, including unused within expiry time for blood transfusion.

Collection and storage of specimens will be carried out in accordance with the current version of the IFU. The laboratory will ensure that:

- appropriate specimen type and anticoagulant, if applicable, is used (*see note below*);
- processing of specimens happens in a timely manner respecting manufacturer's claims for stability, e.g. respecting time between venous whole blood collection and serum/plasma preparation or storage of specimens before testing (including number of freeze/thaws);
- when using archived specimens, these have been processed respecting the previous two points.

Note: rejected blood units may include the following anticoagulants: anticoagulant citrate dextrose (ACD), citrate phosphate dextrose or sodium citrate; and may be stored at or below -20°C for extended periods. **If this specimen type is not acceptable to be used with the test under evaluation, the manufacturer is requested to inform WHO as soon as possible and in any case before the evaluation commences.**

Specimens will be assigned a unique identification number. No personal data will be collected. Further ethics considerations are described in Section 10.

5.1.3 Characterization of the panel

Clinically derived HBV-infected and HBV-uninfected specimens for the clinical performance evaluation will be characterized using the cobas HBV on 5800/6800/8800 Systems, which will be considered the comparator test. The result obtained using this test will serve as the reference result; no additional testing will be required. The plan for resolution of discrepant results is described in 8.2. An alternative comparator method may be selected by the PEL with prior agreement from WHO.

Specimens from HBV-uninfected individuals required for the clinical performance evaluation must be tested using, at least, a state-of-the-art prequalified or stringently assessed HBsAg laboratory-based immunoassay (not a rapid test) and the comparator NAT for HBV DNA. HBV-negative specimens included in the panel will be limited to those with a non-reactive HBsAg result followed by "Target Not Detected" result on the comparator HBV DNA test.

For the purposes of quantitative HBV DNA evaluations, comparator testing should be exclusively conducted on serum/plasma.

5.2 Analytical performance panel

5.2.1 General description of the panel

At a minimum, two specimens representing the most commonly occurring genotypes will be used to construct the panel of specimens for the analytical stage of evaluation. Each specimen will be diluted in defibrinated normal human plasma that is negative for HIV, HBV and HCV markers. Details regarding the panel are shown in Table 2.

Table 2. Specimen requirements for the analytical evaluation

Genotype	Concentration (IU/ml)	Number of replicates of each concentration	Total number of tests
Repeatability (within-run variation) and reproducibility (between-run) precision assessment[a]			
A	Approx. 10^3, 10^5	25	50
C or D	Approx. 10^3, 10^5	25	50
Linear range assessment*			
A	10^7, 10^6, 10^5,10^4, 10^3,5x10^2 and 10^2	5	35
C or D	10^7, 10^6, 10^5,10^4, 10^3,5x10^2 and 10^2	5	35
Genotype detection			
A to G, 1st WHO International Reference panel for HBV genotypes (PEI 5086/08)	Approximately 10^3	1	15
Limit of Detection (LoD)			
3rd or 4th WHO International Standard for Hepatitis B Virus for Nucleic Acid Amplification Techniques	0.5 $\log_{10}$ serial dilutions with: • two concentrations above the manufacturer's stated LoD • one concentration at the stated LoD • two concentrations below the LoD	24	120
Cross-contamination / carry-over			
Any genotype	$\geq 10^7$	20	20
Negative specimens	0	20	20

[a] Concentrations will be modified to suit the claimed limit of detection/quantification of the test: the lowest concentration used should be at least 0.5$\log_{10}$ higher than the claimed lower limit of quantification and the highest concentration should be at least 0.5$\log_{10}$ below the higher limit of quantification.

5.2.2 Specimen preparation and characterization

These specimens may be commercially acquired viral stocks or locally prepared high-concentration viral culture supernatants or clinical specimens with high HBV DNA levels, diluted in a matrix suitable for the test under evaluation (e.g. serum, plasma, whole blood) that is confirmed negative for HIV, HBV, and HCV (by serology and NAT tests).

For commercial acquired preparations, the value assigned by the supplier will be used, where applicable. The concentrations may also be verified using the same comparator test as that used for the clinically derived specimens, for verification purposes. For locally prepared specimens, or if a value is not clearly assigned, then the value will be obtained by testing the preparation on the comparator test in five replicates, in the same run as applicable. The mean value will be used, after identification and exclusion of possible outlier.

6 Laboratory testing

6.1 Review of the instructions for use

Each product under evaluation will be used in accordance with the IFU issued by the manufacturer. The evaluating site will send a copy of the IFU to WHO upon delivery of the reagents and prior to the commencement of the laboratory evaluation. The IFU must be reviewed against the IFU submitted to WHO as part of the application or pre-submission form. If the IFU has been updated since this time, it is the responsibility of the manufacturer to submit to WHO a letter detailing changes made prior to the start of the laboratory evaluation. Records of the IFU version used must be kept.

6.2 Clinical performance evaluation

For the clinical performance evaluation, clinical specimens described in section 5.1 will be tested in singular on the test under evaluation using 2 lots (approximately one half of the specimen panel will be tested with one lot and the other half with the other lot).

Specimens with invalid results will be repeated once on the same reagent lot and both results will be reported in the data spreadsheet. Invalid results will be excluded from the analysis.

6.2.1 Discrepant results

Discrepant results are defined as results that vary by more than 0.5 $\log_{10}$ from results obtained with the comparator test. Specimens with test results discrepant from those of the comparator test will be retested once on the same lot by the same operator on the test under evaluation on the same instrument if sufficient specimen is available and stored according to the condition recommended in the IFU.

The initial result will be used for the analysis and the result of the repeat testing will be used for a description of discrepant results. In the event that the result of repeat testing is significantly different from the initial result and an operator error cannot be excluded, the specimen may be excluded from the analysis.

6.3 Analytical performance evaluation

Given that some of the experiments will share the same dilution for specific specimens, whenever possible and depending on the test under evaluation, specimens may be accommodated on the platform to maximise the throughput and avoid the need to run the same dilution of specimen twice for different purposes, e.g., HBV genotype A at dilutions of 10^3 and 10^5 HBV DNA IU/mL, could be tested once in a manner that would allow the calculated value to be used for multiple purposes. All specimens for the analytical performance evaluation will be prepared as single use aliquots with a volume corresponding to the specific requirements of the test under evaluation.

6.3.1 Precision

Within-run variation and within-laboratory/between-run precision assessments will be carried out as part of the same experiment. Variation will be assessed by measuring, at a minimum, five replicates of four specimens (two subtypes, two different concentrations, Table 2) in the same run over five different days, performed by at least two different users, using at least two different lots and, for low-throughput instruments, on at least two different instruments.

A run will be defined depending on the test's throughput: if the platform can accommodate all specimens in a single run, e.g., in the same test plate, the replicates will be run together. If the test can only accommodate a smaller set of specimens, a run will be defined as a testing session carried out on the same day (or morning/afternoon), using the same lot, on the same instrument/module and by the same user.

If there are two or more invalid results for the same specimen in the same run, then the run should be repeated for this specimen. Invalid results should be reported.

6.3.2 Linear range

Linearity will be analysed using a seven-member dilution series of each of the four different specimens of different genotypes as described in Table 2. Five replicates of each member of the dilution series will be tested using the same lot and in the same run, if applicable. The dilutions will cover a clinically relevant range e.g., from 10^2 to 10^7 HBV DNA IU/ml. The range will be adapted to the test under evaluation, taking into account the claimed limits of quantification. The lowest concentration should be at least $0.5\log_{10}$ above the claimed lower limit of quantification and the highest concentration at least $0.5\log_{10}$ below the claimed higher limit of quantification.

Specimens with invalid results should be repeated and both results should be reported.

6.3.3 Genotype detection

Detection of all main HBV genotypes will be assessed using the 1st WHO International Reference panel for HBV genotypes (PEI 5086/08). All members of the panel will be diluted to approximately 10^3 HBV DNA IU/mL and tested in singular (see Annex 1 for the recommended dilution protocol for samples included in panel PEI 5086/08).

When a manufacturer claims that an IVD will detect a particular genotype, a specimen of that genotype will be repeated once if the initial result is invalid or "target not detected".

6.3.4 Limit of detection

To estimate the limit of detection for the test under evaluation, 24 replicates of a five-member dilution series concentrating on the lower end of the manufacturers' claims for the dynamic range of the test will be used. The five dilutions will be spread across the manufacturers' claimed LoD and centred around the claimed LoD with two dilutions above and below (0.5 and 1 $\log_{10}$ higher and lower). The 24 replicates will be performed over at least three days by at least two users and, for low-throughput instruments, on at least two different instruments using only one lot number. The 3rd or 4th WHO International Standard for HBV-1 DNA preparation will be used for this purpose.

If there are more than four invalid results with the same specimen (i.e. concentration) overall, then this concentration should be retested to get at least 20 valid results.

6.3.5 Cross contamination / carry-over

The cross-contamination experiment will allow the determination of the well-to-well cross-contamination rate of high throughput platforms or potential carry over in low throughput instruments. This will be assessed by alternating one high positive ($\geq 10^7$ IU/ml) with one negative specimen and repeating this sequence twenty times. For high throughput tests, this will be done by alternating the high positive and negative specimens in the same run. For low throughput tests, each sequence of high positive followed by a negative specimen should be done on the same instrument. If more than one instrument is used, each run (i.e same instrument and same day) should include at least two sets of alternating high-positive and negative specimens.

6.4 Interpretation and recording of results

The interpretation of results for each test under evaluation is made strictly according to the manufacturers' instructions within the IFU. Invalid test results and errors will also be recorded.

Results will be recorded in a standardized Excel file.

Wherever possible, all test results will be saved and exported directly from the instrument to standardized test result worksheets for further data analysis. Otherwise, results will be recorded manually in the Excel sheet. In order to avoid transcription errors, a digital or print-out record will be kept of the results in the latter case.

7 Quality control

7.1 Competency panels

A competency panel of routine specimens comprising at least 5 well characterized specimens (one low, one medium, one high positive, at least one negative) must be run successfully for each test by each operator before the evaluation commences. This may be the same panel as that used at the time of test demonstration by the manufacturer or for training purposes. They may be left-over from EQA panels, provided these have been appropriately stored in the interim.

7.2 Internal quality control

Internal procedural controls incorporated into the design of the assay by the manufacturer, must be valid as per manufacturer's instructions. Invalid results will be reported.

7.3 Test kit controls

If applicable, manufacturer-supplied positive and negative test kit controls will be run as indicated in the IFU. Where positive and negative test kit controls are not supplied by the manufacturer, external control specimens will act as the control specimens (see section 7.4).

7.4 External control specimens

A well characterised HBV low positive control (i.e. approximately 3-5x LoD) and an HBV negative control should be run regularly in addition to the test kit controls if provided. For high throughput instruments, these external controls should be run with each run or testing session. For low throughput instruments, where several instruments and/or modules are used for the evaluation, one HBV positive and one HBV negative control should be run every day on one instrument/module with a rotation on the instrument/module used. These specimens will be supplied by the PEL or sourced from a commercial supplier. Results from the control will be recorded and provided in the laboratory evaluation report.

7.5 Limits of acceptability

All results on test kit controls (if applicable) and the external controls shall be documented. Trends will be monitored using a Levey-Jennings (LJ) chart. The mean and standard deviation of controls should be determined based on a minimum of 20 measurements over at least two days. Westgard rules shall be used to interpret the trends. Should the external control specimen result fall between ± 2 and 3 standard deviations (SD), the result will be investigated. Should the control specimen result lie outside ± 3SD, the run will be considered invalid, in which case the run will be repeated after resolution of the problem leading to this result. Appropriate corrective action shall be undertaken to resolve any occurrences before testing resumes. Such problems should be recorded. The PI will be responsible for carefully checking all data entry forms for legibility, accuracy and completeness.

8 Analysis of data

8.1 Invalid runs and invalid individual specimen results

The number of invalid test runs will be recorded as the absolute number of invalid runs and as a percentage of the total number of runs performed for the entire evaluation using all specimens. Other types of readings indicating an invalid run (e.g. errors) may be possible depending on the platform under evaluation. These will also be recorded.

The number of individual invalid specimen results will also be recorded. They will be reported as a percentage of the total number of specimens tested for the entire evaluation, as well as separately for the analytical and clinical parts of the evaluation. In addition, when applicable, invalid results will be classified by cause (e.g. operator, specimen or assay-related, error codes, etc.).

8.2 Clinical performance

The main analysis of clinical performance will be done using only results of HBV-infected specimens. Results of HBV-uninfected specimens will be analysed separately.

Invalid results will be excluded from the analysis. For specimens that showed discrepant results with the comparator test, initial results will be used for the analysis.

8.2.1 Trueness of measurement

The trueness will be assessed using Bland-Altman analysis and correlation using Deming regression model.

The level of agreement between results obtained using the test under evaluation and the comparator method will be evaluated using Bland-Altman analysis (*4*). A graphical representation will be made of the difference between the measurements using the two different methods plotted against the mean of the two measurements for each specimen. The bias is defined as the mean difference between the results of the test under evaluation and the comparator method. The limit of agreement is the 95% confidence interval of the difference between the methods (bias ±1.96 SD).

Correlation between the test under evaluation and the reference method will be measured using Deming regression. The slope and intercept will be estimated together with their standard errors and confidence intervals.

Unquantifiable and negative (or undetectable) results will be reported but excluded from this quantitative analysis.

8.2.2 Agreement between test results at clinically relevant decision levels

Positive percent agreement (PPA) and negative percent agreement (NPA) will be assessed by comparing the results of the test under evaluation with the results of the comparator test, after classification of the results below or above the clinically relevant decision level of 20 000 HBV DNA IU/mL (Table 3).

Table 3. Cross-tabulation of results of the test under evaluation and comparator test at 20 000 IU/mL

		Results of comparator test		
		Positive (>20 000 HBV DNA IU/mL)	**Negative (<20 000 HBV DNA IU/mL)**	
Results of test under evaluation	**>20 000 HBV DNA IU/mL**	**a**	b	a+b
	<20 000 HBV DNA IU/mL	c	**d**	c+d
		a+c	b+d	

Secondary analyses of PPA and NPA will be performed after classification of the results below or above the following other relevant decision levels: below and above 2000 HBV DNA IU/mL; and below and above 200 000 HBV DNA IU/mL. These analyses, as well as corresponding sample size, may be revised if the clinically relevant decision levels recommended by WHO are updated.

Positive percent agreement

PPA will be calculated as the number of specimens classified above the clinically relevant decision level by both methods divided by the total number of specimens classified above the clinically relevant decision level by the comparator method and expressed as a percentage.
$PPA = a/(a + c)$ (see tables 3 and 4)

Negative percent agreement

NPA will be calculated as the number of specimens classified below the clinically relevant decision level by both methods divided by the total number of specimens classified below the clinically relevant decision level by the comparator method and expressed as a percentage.
$NPA = d/(b + d)$ (see tables 3 and 4)

Exact 95% confidence intervals for binomial proportions will be calculated.
This analysis is done using the initial results obtained in the test under evaluation.

Discrepant results

Discrepant results will be described separately, including results of the comparator test and results of repeat testing in the test under evaluation, if applicable.

8.2.3 Specificity among HBV-negative specimens

The specificity and 95% confidence interval among the 100 HBV-uninfected specimens will be reported separately.

8.3 Analytical performance

The following methods will be used to calculate the performance characteristics for each test under evaluation. For all these analyses, the $\log_{10}$ values will be used.

8.3.1 Precision of measurement

Estimation of precision will require, at a minimum the testing of the specimens described in section 5.2.

Within-run and within-laboratory imprecision will be calculated for each specimen from the 25 repeats using a one-way ANOVA analysis. The mean (μ), standard deviation (σ) and percentage coefficient of variation (%CV) will be calculated using the $\log_{10}$ values.

Both the %CV calculated by dividing SD by the mean, and with the following formula for calculating %CV for log-transformed data will be presented (*5*).

$\%CV_1 = 100 * \sigma / \mu$

$\%CV_2$ – for log-transformed data= $100 * \sqrt{(10^{1n(10)*\sigma 2} - 1)}$

Detected but unquantifiable results will be given an arbitrary value of half of the LLOQ, which will be used in the analysis.

8.3.2 Linear range

The statistical analysis for linearity will be performed by linear regression of the measured value against expected value. The slope, intercept and correlation coefficient will be reported.

Detected but unquantifiable results below the lower limit of quantification will be given an arbitrary value of half of the LLOQ, which will be used in the analysis. Unquantified results above the upper limit of quantification will be excluded from this linear regression analysis and reported separately.

8.3.3 Genotype detection

The number of members from the International Reference Panel for HBV Genotypes detected by the test under evaluation will be reported. Details will be provided on re-testing (described in section 6.2.1) in case one or several members of the panel were not detected.

8.3.4 Limit of detection

The limit of detection (LoD) is the lowest concentration of analyte that can be consistently detected in ≥95% of specimens with that concentration, tested under routine laboratory conditions and in a given specimen matrix.

The concentrations of specimens in the LOD dilution series will be reported in IU/mL, as calculated directly from the assigned value of the WHO International Standard. The proportion of successes at each concentration will be reported. The LoD and 95% confidence interval will be determined using a probit analysis.

8.3.5 Cross-contamination / carry-over

Cross-contamination/carry-over will be assessed by calculating the percentage of false-positive results among the 20 repeats of the negative specimen.

9 Operational Characteristics

The operational characteristics of the test will be assessed by the laboratory technicians performing the evaluation testing using a standard evaluation sheet (see Annex 2), in order to give an appraisal of the test under evaluation. Special attention should be paid to the IFU to evaluate whether these instructions are sufficient for WHO Member State end-users. Comments on the IFU must be made in the report if it does not meet an acceptable standard for any of the following criteria: clarity, presentation, content, safety instructions.

10 Ethical considerations

10.1 Compliance with International Standards

This protocol was submitted for review to the World Health Organization (WHO) Ethical Review Committee for ethical clearance and considered exempted from review (ERC.000383). When required by local regulations, this protocol will be submitted to the national and institutional ethical review boards of the PEL. Site specific ethics approval should be acquired as per local Institutional Review Board requirements. Any substantial change to the protocol must be approved by all the bodies that have approved the initial protocol, prior to being implemented, unless it is due to participant's safety concerns. The evaluation will be carried out according to the principles stated in the Declaration of Helsinki as amended in 2013 and any further updates, all applicable national and international regulations and according to the most recent *International Conference on Harmonization of technical requirements for registration of pharmaceuticals for human use – Good Clinical Practice* (ICH-GCP) guidelines.

10.2 Specimen collection and informed consent

As stated in section 5, specimens used in this evaluation may belong to the following categories:

10.2.1 Use of left-over specimens

Residual specimens from routine patient care may be used in the evaluation. These specimens may be part of an already existing panel of frozen specimens, or freshly collected.

If required by local regulation, written informed consent for secondary use of left-over specimens will be acquired from participants prior to collecting the specimen. The participant will be provided with general information about the secondary use of their specimen, both on paper and by staff. In case of illiteracy, the information will be read to the participant and a fingerprint will be used as a signature. PELs are allowed to use their own version of informed consent forms (ICFs). As per good ethical practice, it is advisable that ICFs contain at least the following information in clear understandable language respective to the target audience: *the possibility of specimen storage and secondary use for specified purposes; that participation is voluntary and the participant is free to withdraw without consequences; what (if any) compensation the participant will receive and any other benefits related to the participant; any foreseeable risks involved in participating; provisions made to respect and preserve the participants privacy and confidentiality; that the protocol/ICF was reviewed by relevant ethical bodies.*

In some cases, left-over specimens from research studies can be used, provided the original study protocol and informed consent clearly outlined the secondary use of specimens for this purpose.

Alternatively, informed consent for the use of left-over specimens for non-research purposes can be waived by local ethics committees by local ethical review boards, provided one of the following conditions are met:

Presumed consent

Prior to collection, specimen donors are made aware of the potential secondary use of their specimen for non-research purposes. This may be done through several channels, including: clearly visible pamphlets and posters at the collection site, available information on the institute's website and personal communication through the

treating physician or nurse. The participant will be made aware of the right to refuse and opt out without any consequence for the quality of care.

Anonymization

PEL's may choose to anonymize data wherever reasonably possible, thus removing any chance of reidentification and maximizing confidentiality. It is possible in certain countries to obtain a waiver of informed consent by local ethics review committee when using fully anonymized left-over samples from routine practice for non-research purposes. It is the responsibility of the PEL to check the requirements for ethical and regulatory approval with their own institution.

10.2.2 Externally supplied specimens

Blood-derived specimens obtained from blood banks (infected reject-specimens or uninfected, expired for transfusion) may be used. The collection site is responsible to ensure that the blood bank services providing the blood units obtained a written informed consent for secondary use of the blood units consistent with the current proposal, using the internal procedure and consent forms at the blood bank.

Alternatively, if no written informed consent is used at the blood bank, collecting sites may obtain a consent waiver from their local/national ethics review boards allowing them to use rejected or expired blood units anonymously.

10.2.3 Commercially acquired specimens

Commercially available panels or pooled human plasma intended for dilutions can be requested. These panels are presumed to be collected in an ethical manner and no additional informed consent will be sought.

10.2.4 Non-clinical specimens

The evaluation may use stock specimens of pathogens (i.e. viral supernatants). No additional informed consent will be sought for validation of tests using non-human specimens. These specimens can be derived from routine clinical isolates, provided they cannot be traced back to the original host and were acquired in an ethical manner. Stock specimens may need to be diluted in pathogen-negative human plasma/whole blood. Informed consent may still apply for these blood products, depending on their origins.

10.3 Risk-Benefit assessment

There will be no direct benefits to the specimen donors. The results obtained during this study will be used as part of WHO prequalification assessment, to ensure that the tests meet WHO requirements. There will thus be broader benefits to communities affected by HBV by contributing to the selection of well-performing HBV viral load tests.

Participants may experience minor discomforts during specimen collection, such as bruising at the site of venepuncture. This risk will be minimized by employing trained professionals for the collection of specimens. Risks of breach of confidentiality will be minimized by collecting as little personal information as is necessary and ensuring proper data protection according to general data protection regulations, as well as anonymization of specimens and data as soon as a link to identifiable information is no longer required.

10.4 Storage of data and specimens

10.4.1 Confidentiality

No personal data will be recorded for the evaluation. All non-essential personal identifiers (including direct identifiers such as name, address, etc.) will be removed and specimens will receive a unique identification number at the collection site. If collected by WHO partner sites, specimens are then sent in large volumes to the PEL where they are assigned a specimen identification number and stored for later use. In short, all collected data will be pseudonymized and transformed in order to preserve participant's privacy. The key of the pseudonymization will be kept at the collection site in a locked cabinet or as a password-protected document only accessible to the site principal investigator or a delegated staff authorized by the site principal investigator. When a link with the original identifiable information is no longer required, and at maximum at the end of the evaluation (i.e. when the final report is sent to the manufacturer), this link will be destroyed, and data will be anonymized.

10.4.2 Data storage & biobanking

Personal data will be handled and stored according to the European general data protection regulations (GDPR) or local alternatives where applicable. Any documents containing the names and/or signatures of participants (e.g. consent forms) will be kept in a locked cabinet separately from all other evaluation documents containing participant data. All evaluation documents will be stored in lockable rooms or cabinets, or, for digital documents, in access-restricted folders or as password-protected documents, with access limited to evaluation staff. The final datasets, which will not include any personal information, will also be kept at WHO in folders with access restricted to the PQ In Vitro Diagnostics Assessment team.

Names of the participants will not appear on any reports or publications resulting from this evaluation.

After the evaluation, all source data, data analysis records and all correspondence will be retained at the testing laboratory for five years.

10.5 Results and incidental findings policy

The evaluation will not interfere with the clinical care the participant would normally receive. None of the results generated in this evaluation from evaluated test will be used as a replacement for the gold-standard tests currently in use. As the purpose is to evaluate new tests, incidental findings are unlikely to occur. Should such a finding of significant clinical importance occur, the subsequent actions (e.g. feedback to the patient) shall be considered on a case-by-case basis by the evaluating laboratory and its medical staff according to their incidental findings policy.

11 Report preparation

The preliminary data analysis and drafting of the report will be carried out by the PEL according to pre-defined report templates (see Section 14, Other documents required).

For evaluations coordinated by WHO (option 1), the draft report will be shared with WHO. WHO will verify the data and draft report and send the final draft report to the authorized contact designated by the manufacturer for comment.

For evaluations commissioned by the manufacturer (option 2), the data and draft report will be shared simultaneously with WHO and the manufacturer in copy. WHO will verify data analysis and draft report and produce a final draft, which will be shared with the evaluating laboratory and the manufacturer.

In both cases, manufacturers will have one month to make comments on the draft report after it has been shared by WHO. Comments provided by the manufacturer will be given due consideration, however WHO reserves the right to refuse the changes requested by the manufacturer, if applicable. If no comments were submitted by the manufacturer within one month, the report will be accepted as final. The final report will be prepared and disseminated by WHO. A copy of the final report will be sent to the authorized contact designated by the manufacturer and to the laboratory.

If the test under evaluation successfully meets all WHO prequalification requirements, a summary of these data will be published in the WHO Public Report for the prequalification assessment of the test.

WHO reserves the right to publish the results of the evaluation, irrespective of the outcome. In this case, WHO will share the manuscript with the manufacturer for comments at least 30 days before submission, but WHO will have ultimate authority over the version submitted. Authors will include contributors from WHO and the PEL.

Any publication by WHO of the results of these evaluations and the WHO recommendations derived therefrom will, however, be accompanied by the following disclaimer:

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

WHO and the Performance Evaluation Laboratory, do not warrant or represent that the evaluations conducted with the HBV test kits referred to in this document are accurate, complete and/or error-free. WHO and the Performance Evaluation Laboratory disclaim all responsibility for any use made of the data contained herein, and shall not be liable for any damages incurred as a result of its use. This document must not be used in conjunction with commercial or promotional purposes.

12 Materials and supplies

Manufacturers will provide the products and any equipment necessary for the evaluation free of charge. The number of tests to be performed for this evaluation is presented in Table 4.

Table 4. Number of tests required for the performance evaluation

	Number of tests
Precision	100
LoD	120
Linearity	140
Genotype detection	15
Cross-contamination / carry- over	40
Clinical performance	460
Total	**875**
Total + ~20% (for controls and possible repeats)	**1050**

13 Roles and responsibilities

13.1 Responsibilities of the Performance Evaluation Laboratory

i. If required by national authorities, obtain ethical clearance for the evaluation;
ii. Ensure availability of all specimens and panels necessary for the evaluation;
iii. Conduct the performance evaluation in agreement with the protocol and in accordance with internationally recognized best practice;
iv. Prepare draft report on laboratory evaluation using appropriate templates;
v. Advise WHO on operational characteristics of tests evaluated;
vi. Archive all source data, data analysis records and all correspondence for a period of at least five years.

13.2 Responsibilities of WHO - Prequalification of In Vitro Diagnostics Team

i. Provide technical advice to the PI;
ii. Technical and administrative management of the laboratory evaluation (option 1);
iii. Verify of analysis and draft report;
iv. Communicate the final draft report to manufacturer and seek comments from manufacturer;
v. Prepare and disseminate the final report;
vi. Formal contacts with authorized contacts of the manufacturers.

13.3 Responsibilities of the manufacturer

i. Provide the appropriate number of test kits and instruments free-of-charge for the evaluation;
ii. Ensure that kits and instrument are shipped and installed under appropriate conditions and in time for the commencement of the evaluation;
iii. Provide training on the use of the instrument(s) and test, in agreement with WHO;

iv. Provide comments to WHO on the draft performance evaluation report within one month;

v. For option 2 evaluations, select the PQ evaluation laboratory, agree on terms and conditions of the evaluations in line with the conditions set forth in the Letter of Agreement with WHO, and fund the evaluation.

14 Other documents and tools required

Master Templates

IVD/TP/4/P12a Template report for the performance evaluation of quantitative HBV DNA tests

IVD/TP/4/P12b Template data entry spreadsheet for quantitative HBV DNA tests

15 References

1. World Health Organization. Overview of the WHO Prequalification of In Vitro Diagnostics Assessment. Version 9. Geneva: World Health Organization; 2021.
2. World Health Organization. Laboratory biosafety manual, 4th edition [Internet]. Geneva: World Health Organization; 2020. Available from: https://www.who.int/publications/i/item/9789240011311
3. World Health Organisation. WHO Guidelines on Hepatitis B and C Testing [Internet]. World Health Organization, editor. Vol. 66, Who. Geneva; 2017. 1–170 p. Available from: https://iris.who.int/bitstream/handle/10665/254621/9789241549981-eng.pdf?sequence=1 or http://www.ncbi.nlm.nih.gov/pubmed/28742301
4. Martin Bland J, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;327(8476):307–10.
5. Canchola JA. Correct Use of Percent Coefficient of Variation (%CV) Formula for Log-Transformed Data. MOJ Proteomics Bioinforma. 2017;6(3):316–7.

16 Revision history

Version	Date	Summary of changes	Prepared by/ reviewed by
V1.0	02 February 2023	For submission to WHO ERC	AL Page Reviewed by: Susie Braniff, Sue Best, Keith Perry, Isaac Ssewanyana, Tongai Maponga External reviewers: Emi Okamoto, Saleem Kamili
V1.1	9 November 2023	Added ISBN and Barcode; Minor editing. Addition of new IVD protocol number IVD/PR/4/P12 to replace PQDx_337	AL Page

17 Annexes

17.1 Annex 1. Dilutions of the 1st WHO International Reference panel for HBV genotypes (PEI 5086/08)

The specimens included in the 1st WHO International Reference panel for HBV genotypes include several specimens at high concentration. These will be diluted to approx. 500 -1000 IU/mL to assess genotype detection. The table below shows the dilution factors to be used for each specimen.

Specimen n°	Genotype	Overall mean (log_{10} IU/mL)	Dilution factor	Concentration (log10 IU/mL)
1	A	6.10	1000	3.10
2	A	5.86	1000	2.86
3	A	5.80	1000	2.80
4	B	5.92	1000	2.92
5	B	5.78	1000	2.78
6	B	3.91	10	2.91
7	C	5.96	1000	2.96
8	C	6.09	1000	3.09
9	C	5.94	1000	2.94
10	D	5.99	1000	2.99
11	D	6.01	1000	3.01
12	D	6.03	1000	3.03
13	E	5.86	1000	2.86
14	F	4.76	100	2.76
15	G	3.78	10	2.78

Note: for assays with LoD >100 IU/mL, the dilution factor should be divided by 10.

17.2 Annex 2. Assessment of operation characteristics and ease of use

Indicate name of instruments assessed (e.g. extraction and amplification units)

Instrument 1:		Number of units used for the evaluation	
Instrument 2 (if applicable):		Number of units used for the evaluation	

If only 1 instrument, then indicate N/A for instrument 2 and all questions on instrument 2 below
If several units of the same instruments were used for the evaluation, these should be considered as only one instrument.

Table A2.1. Operational characteristics

1	**Assay characteristics**		
1.1	Need to reconstitute reagents		Yes/No
1.2	Total number of steps for one specimen*		*Each action required to obtain a result for one specimen (excluding specimen collection, instrument management, maintenance/calibration)*
1.2.1	Number of steps requiring timing		
1.2.2	Number of steps requiring precision pipetting		
1.3	Number of steps for instrument management		
1.3.1	Number of daily steps for instrument management (excluding maintenance) **		*Each action required daily or per run to set up and shut down the instrument*
1.3.2	Number of steps for maintenance		
1.4	Number of tests per run		*N/A if single-test instrument*
1.5	Minimum number of tests (if partial runs possible)		
1.6	Is continuous loading of specimens possible?		Yes/No
1.6	Is it possible to run different assays simultaneously?		Yes/No
1.8	Time from start to completion for one test		minutes

1.9	Time from start to completion for one run		*N/A if single-test instrument*
1.10	Operator hands-on time for one run (/one test for single-test instruments)		minutes
1.11	For molecular assays: is extraction automated?		Yes/No
1.12	For molecular assays: are extraction and amplification integrated?		*Integrated = no manual step between extraction and amplification*
1.13	Kit controls		
1.13.1	Controls provided by manufacturer (in the kit / separately)?		In the kit / Purchased separately / Not provided
1.13.2	Frequency of controls recommended in the IFU		eg. Each run / Each day / Not specified
1.13.3	Number of controls		
1.13.4	Type of controls		eg. High positive, low positive, negative
2	**Specimen**		
2.1	Type of specimen collection device provided in the kit		*N/A if none provided*
2.2	Validated specimen types		
2.3	Specimen volume(s) for the assay		
2.3.1	If applicable, minimum volume needed in tube		
2.4	Maximum time between specimen collection and testing (for each specimen type) recommended in the IFU		*Specify also conditions (e.g. temperature) if applicable*
3	**Results and data management**		
3.1	Specimen information entered manually or by scanning a bar code		Manually / Bar code
3.2	Instrument connectivity		USB / Wireless data transfer / Other

3.3	Result display		On device only / Printed / Exported
3.3.1	If printed, is printer provided?		
3.3.2	If printer provided, printer cartridges format		Standard format / Specific for printer
3.4	Compatibility / interfacing with Laboratory Information Management Systems		
3.5	Language options (software)		*Include all languages available*
4	**Kit storage**		
4.1	Number of tests per kit		
4.2	Kit dimensions		*width / depth /height (cm)*
4.3	Recommended storage temperatures for the kit		
4.4	Recommended storage conditions of kit components after opening		
4.5	Shelf-life of kit components after opening		
5	**Instrument(s) and infrastructure**		
5.1	For instrument 1		
5.1.1	Dimensions		
5.1.2	Weight		
5.1.3	Type of instrument		Benchtop / Freestanding / Portable
5.2	If applicable for instrument 2		
5.2.1	Dimensions		
5.2.2	Weight		
5.2.3	Type of instrument		Benchtop / Freestanding / Portable

5.3	Power sources		Main power / Battery / Solar power
5.4	Need for stable electricity (/UPS)		
5.5	Additional requirements, if applicable (e.g. weight to surface area ratio)		
5.6	Operating temperature		
5.7	Is equipment sensitive to dust?		
5.8	Are there any altitude or humidity specifications? If yes, specify		
5.9	What are the requirements for separation of workspace (e.g. specimen processing, extraction, amplification)		
5.10	Is distilled/deionised water required?		
6	**Biosafety and waste disposal**		
6.1	Are there any safety concerns for the user (*outside of infectious specimen handling)*		Yes / No
6.1.1	Biological hazards		*List hazards identified*
6.1.2	Chemical hazards		*List hazards identified*
6.2	Waste volume produced per run (solid / liquid)		
6.3	Does waste require additional treatment before disposal and/or specific disposal procedures (*in addition to usual laboratory waste disposal procedures*)		
6.4	Does disposal of the consumables and waste pose a substantial risk for people?		
6.5	Does disposal of the consumables and waste pose a substantial risk for the environment		
7	**Installation, calibration, maintenance and troubleshooting**		
7.1	Does installation require vendor or engineer?		

7.2	Minimal frequency of calibration		Daily / weekly / monthly/ yearly / No need
7.3	Minimal frequency of maintenance		Daily / weekly / monthly/ yearly / No need
7.4	Number of breakdown/blockage during study period		
7.5	Was an intervention by the manufacturer or representant needed?		

** Steps for one specimen: each action required to obtain a result for one specimen (excluding specimen collection, instrument management, maintenance/calibration), e.g. add specimen to the cartridge, close the cartridge, scan/type specimen ID, load the cartridge on the instrument, press start (5 steps) OR scan/type specimen ID, load the specimen collection tube into the instrument, press start (3 step)*

*** Daily steps for instrument management: each action required daily or per run to set up and shut down the instrument, e. g. switch on instrument, log in, maintain supplies, maintain reagents, discard liquid waste, discard solid waste, archive results, switch off instrument (8 steps)*

Table A2.2. Ease of use

		Not applicable	Strongly disagree	Disagree	Agree	Strongly agree
1	**Instruction for use**					
1.1	The IFU for the assay is clear					
1.2	If applicable, pictures/diagrams are clear					
1.3	The IFU contains all important information*					
1.4	Safety instructions are clear					
1.5	The manual for the instrument 1 is clear					
1.6	If applicable, the manual for the instrument 2 is clear					
2	**Kit packaging and labelling**					
2.1	Kit labelling is clear					
2.2	Safety labelling is sufficient					
2.3	Kit packaging is of good quality					
2.4	All reagents were provided in sufficient quantities					
3	**Assay set-up**					
3.1	If applicable, specimen preparation is simple					
3.2	Loading specimen into the plate/cartridge/sample processing unit is simple					
4	**Use of instruments**					
4.1	The use of instrument 1 is simple					
4.2	If applicable, the use of instrument 2 is simple					
4.3	The steps for trouble shooting, error codes and steps to resolve are clearly documented					
4.4	The software is user-friendly					
4.5	Maintenance of the instrument(s) is simple and rapid					
4.6	The instrument(s) is(/are) robust (not susceptible to breakdowns)					
4.7	If needed, technical support was provided in a timely manner					
5	**Results and interpretation**					
5.1	The test results are presented clearly					

5.2	Interpretation of the test result is simple and does not require any additional calculation/transformation by the user					
5.3	Output (print-out, export format) is user-friendly					
6	**Overall appraisal**					
6.1	Overall, the test is easy to use					
6.2	The test can be used in a laboratory with limited facilities					
6.3	The test can be used in non-laboratory settings					
6.4	The test can be used by trained non-laboratory personnel					

** Refer to TGS-5 Designing instructions for use for in vitro diagnostic medical devices https://apps.who.int/iris/bitstream/handle/10665/259737/WHO-EMP-RHT-PQT-TGS5-2017.05-eng.pdf;sequence=1*

If disagree or strongly disagree, report item number and describe

n°	Item	Comment

General appraisal – advantages and disadvantages of the platform/assay	
Advantages	
Disadvantages	

In Vitro Diagnostics Assessment Team
Prequalification Unit
World Health Organization
Avenue Appia 20
1211 Geneva 27
Switzerland
www.who.int

9789240079632