WHO PROTOCOL FOR PERFORMANCE EVALUATION OF TREPONEMAL AND NON-TREPONEMAL SEROLOGY RAPID DIAGNOSTIC TESTS

PQDx_326 Version: 1.0
3. Introduction

3.1. Prequalification of in-vitro diagnostics
The World Health Organization (WHO) Prequalification In Vitro Diagnostics Assessment is coordinated through the Prequalification Team - Diagnostics. The aim of the WHO Prequalification of in vitro Diagnostics Assessment is to promote and facilitate access to safe, appropriate and affordable diagnostics of good quality in an equitable manner. Focus is placed on products for high burden diseases and their suitability for use in resource-limited settings.

The WHO Prequalification of in vitro diagnostics process includes three main components:
- review of an application form and product dossier;
- performance evaluation of the product;
- inspection of the manufacturing site(s).

This document pertains to the objectives and processes of the performance evaluation component of the prequalification assessment. This document intends to provide information for manufacturers on the process for performance evaluation.

3.2. WHO Performance Evaluation of Treponemal and non-Treponemal Rapid Diagnostic Tests (RDTs)
The performance evaluation determines the accuracy of *Treponema pallidum* (TP) and non-treponemal (non-TP) serology Rapid Diagnostic Tests (RDT) used in diagnosis of syphilis in comparison with established performance criteria. These characteristics include: sensitivity, specificity, negative and positive predictive values. In addition, a number of operational characteristics are assessed including the suitability for use in smaller or lesser equipped laboratories and/or testing settings with limited infrastructure including community-based testing and antenatal services.

All RDTs for the detection of antibodies to *Treponema pallidum* (hereafter referred to as TP assays) and non-treponemal antibodies submitted for performance evaluation are assessed at a WHO Prequalification Evaluating Laboratory (PEL).

The WHO syphilis specimen reference panel comprises of approximately 1200 serum/plasma specimens from <Europe, Africa, Asia, Australia, America etc> which will be characterized for both TP and non-TP antibodies. Of these 450 are TP seropositive and 750 are TP seronegative for performance evaluation of TP RDTs. Also efforts will be made to ensure that among the 1200 specimens, 450 will be non-TP seropositive and 750 will non-TP seronegative to cover the performance of non-TP RDTs. Other panels will include well-characterized commercial TP serocconversion panels, mixed titer panel, TP (and non-TP) lot-to-lot variation panels, and the WHO TP international biological reference preparations.

4. Study objectives

4.1. Overall objectives
The overall objective is:
1. To evaluate the performance of currently available RDTs for the detection of syphilis antibodies (TP or non-TP) against established performance criteria.

4.2. Specific objectives
The specific objectives are:
1. To determine the sensitivity and specificity of commercially available TP RDTs for the detection of TP antibodies as compared to a reference result obtained using at least two well validated treponemal assays including the *Treponemal pallidum* Passive Particle Agglutination assay [TP-PA].
2. To determine the sensitivity and specificity of currently available non-TP RDTs for the detection of non-TP antibodies as compared to a reference result obtained using established non-TP flocculation assay.

3. To evaluate the operational characteristics of TP or non-TP assays, e.g. ease of performance, utility of specimen type, inter-reader variability, reaction endpoint stability, rate of invalid devices, and suitability for use in countries with limited infrastructure (no/limited electricity, no/limited clean water, and inadequate means of biosafety disposal).

5. Study Design

5.1. WHO Prequalification Evaluating Laboratory
The WHO Prequalification Evaluating Laboratory shall be one that has undergone assessment using the WHO Alternative Laboratory Evaluation Mechanism which include submission of an Expression of Interest (EoI). The assessment procedure of the laboratories will include Stage 1 audit: (Assessment of EoI submission and specific Quality Management System (QMS) documentation) followed by Stage 2 audit: include on-site audit of the laboratory to assess compliance with WHO requirements and lastly listed as WHO Prequalification Evaluating Laboratories. These processes are described at: http://www.who.int/diagnostics_laboratory/evaluations/alternative/en/

The laboratory shall hold one of the following certification for quality management systems: ISO17025:2017 (General requirements for the competence of testing and calibration laboratories), ISO15189:2012 (Medical laboratories: Particular requirements for quality and competence) 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 WHOPEL.

5.2. Training, performance evaluation and supervision
The following issues are key for minimizing errors and maximizing the value of the performance evaluation:

- The PI will be responsible for training the laboratory technicians on the performance evaluation protocol and in the performance of each assay undergoing performance evaluation;
- Only those personnel who have received specific training on the specific assay will be involved in the evaluation;
- Accurate record keeping is crucial to the success of the performance evaluation and the PI will be responsible for ensuring that all data required for the performance evaluation are recorded on the agreed data collection forms, are accurate and up-to-date;
- It is important to plan the work in advance and follow standard operating procedures prepared and controlled by the WHO PEL;
- To reduce the risk of adding an incorrect specimen to a test device, before starting the test run, the operator will prepare work forms and label all test devices with the specimen’s unique number;
- Because objective, machine-generated, permanent results for simple/rapid diagnostic tests are not feasible, it is essential that the PI emphasizes to the operator performing the tests the need for accurate recording of results and recordkeeping;
- To minimize the risk of error, it is recommended that the results are read and recorded independently by three trained staff members;
- To allow immediate correction of erroneous recording of results (rather than differences in visual interpretation), the PI or designee should assess the results at the end of specified reading time to allow return to the original test device to investigate apparently discordant readings;
• For the investigations performed at the WHO PEL, at least one representative result from TP or non-TP positive and negative specimens will also be recorded by taking electronic images. Unexpected test results will be digitally recorded.

5.3. Safety
HIV, hepatitis B and hepatitis C and other blood borne pathogens are transmissible by blood and other 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 Guidelines on HIV Safety Precautions, and Guidelines for the Safe Transport of Specimens (WHO/EMC/97.3) and the PEL guidelines on laboratory safety should be carefully followed by the laboratory staff.

5.4. Storage of assays
All reagents must be stored as indicated in the instructions for use. In case some of the components of the kits does not need refrigeration and the refrigerated storage space is inadequate, labile reagents can be refrigerated separately from the non-labile supplies. Calibrated thermometers are placed at each location where reagents and specimens are stored, i.e. ambient, refrigerator and freezer. Temperatures are recorded daily on the WHO PEL forms. The lot numbers of the test kits received/used and their expiry dates are recorded on the individual run worksheets.

Two separate production lots will be requested for the performance evaluation, according to the following definition1 of a lot: “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”.1 Furthermore, lots must be sourced from a representative production run and not produced especially for the purpose of the performance evaluation. The manufacturer shall send batch production record for the two lots to WHO for verification before sending the kits to the PEL.

6. Specimens

6.1. WHO syphilis performance evaluation panel

6.1.1. Collection of specimens for the WHO syphilis performance evaluation panel
Specimens are collected as serum or plasma (with appropriate anti-coagulant) from patients suspected or confirmed to have syphilis or from the blood bank (mainly for negative specimens). Where clinical information is available, this should be captured.

Newly collected specimens are assigned a unique identification number at the collection site and then assigned with a WHO specimen identification number upon arrival at the WHO PEL. The specimens are processed and aliquoted into working volumes of 250µl and stored at -20°C or -80°C until testing commences. During the testing period, the specimens are stored at 2 to 8 °C and this time period does not exceed one week. After the completion of testing, they are again stored at -20°C or -80°C. Each aliquot does not undergo more than two freeze/thaw cycles.

6.1.2. Characterization of the WHO syphilis performance evaluation panel
The panel consists of approximately 1200 serum/plasma specimens of <European, African, Latin American, Australian and Asian> origin characterized for presence of TP or non-TP antibodies.

---

1ISO 18113-1:2009 In vitro diagnostic medical devices -Information supplied by the manufacturer (labelling) – Part 1: Terms, definitions and general requirements
The WHO syphilis performance evaluation panel is characterized for TP antibody status using a standardized combination of assays i.e. a testing algorithm. The results are used to determine the TP serological status of each specimen for the purpose of the performance evaluation. (Figure 1).

![Diagram of TP antibody status characterization]

**Figure 1. Characterization of the WHO syphilis performance evaluation panel for TP antibody status**

Initially, each specimen is tested using a **validated Treponema pallidum** Chemiluminescence immuno assay. Specimens that are **non-reactive on the** Chemiluminescence immuno assay are labelled as anti-TP negative and are not tested further.

Specimens that are **reactive on the** Chemiluminescence immuno assay are further tested by a **Treponema Pallidum Passive Particle Agglutination assay (TP-PA)** (SERODIA-TP.PA, Fujirebio Inc). Specimens that are reactive on both assays are labelled as anti-TP positive.

Specimens that are **reactive on validated Treponema pallidum** Chemiluminescence immuno assay but **non-reactive on TPPA** are labelled as anti-TP indeterminate and are excluded from the panel.
6.1.3. Characterization of the WHO non-TP specimen reference panel

The panel consists of approximately 1200 serum/plasma specimens of <European, African, Latin American, Australian and Asian> origin if possible. There are 450 anti-non-TP positive and 750 are anti-non-TP negative.

The WHO syphilis performance evaluation panel is characterized for non-TP antibody status using an established RPR flocculation test (Figure 2).

![Figure 2 - Characterization of the WHO syphilis performance evaluation panel for non-TP antibody status](image)

**Note:** The characterization of the WHO syphilis performance evaluation panel will be conducted with an objective of obtaining specified number specimens with different levels of RPR reactivity as shown in Table 1.

Initially, each specimen is tested using BD Macro-Vue™ RPR Card Tests (Becton Dickinson). Specimens which are reactive on BD Macro-Vue™ RPR Card Tests (Becton Dickinson) are assigned anti-non-TP positive but they are tested further using the quantitative method of the same assay to determine the non-TP titer.

Specimens that are non-reactive on BD Macro-Vue™ RPR Card Tests are labelled as anti-non TP negative. The outcome of the characterization of the WHO syphilis performance evaluation panel is shown in Table 1.

<table>
<thead>
<tr>
<th>TP positive specimens</th>
<th>Non-TP positive specimens</th>
<th>TP &amp; non-TP negative specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RPR reactivity</strong></td>
<td><strong>Number of samples</strong></td>
<td><strong>Number of samples</strong></td>
</tr>
<tr>
<td>Negative</td>
<td>50</td>
<td>NA</td>
</tr>
<tr>
<td>Neat</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>1:2 dilution</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>1:4 dilution</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>1:8 dilution</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>1:16 dilution</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>1:32 dilution</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>1:64 dilution</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>1:128 dilution</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>&gt;1:256 dilution</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>450</strong></td>
<td><strong>450</strong></td>
</tr>
</tbody>
</table>
Table 1. Distribution of different RPR reactivity in the WHO syphilis specimen reference panel.

<table>
<thead>
<tr>
<th>RPR titre range</th>
<th>Sensitivity</th>
<th>RPR titre range</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPR titre &lt;1:4</td>
<td>90%</td>
<td>RPR titre &lt;1:4</td>
<td>75%</td>
<td>98%</td>
</tr>
<tr>
<td>RPR titre &gt;1:4</td>
<td>98%</td>
<td>RPR titre &gt;1:4</td>
<td>98%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Performance criteria of Treponemal and non treponemal RDTs

6.2. Performance criteria of TP and non-TP RDTs

Table 2 summarize the expected performance of the TP and non-TP RDTs at different levels of RPR reactivity as agreed in the WHO Technical Consultation on WHO prequalification requirements for Syphilis Rapid Diagnostic Tests.

6.3. TP lot-to-lot variation panels

Lot-to-lot variation is assessed by testing 16 member dilution series (comprised of 2-fold dilutions of 10 stock anti-TP positive or 10 stock anti-non-TP positive specimens in commercially available normal human serum) on two separate production lots of the assay under evaluation in the same testing session.

6.4. TP seroconversion panels

A seroconversion panel is a series of specimens, sequentially collected over a period of time, from an individual developing antibody in response to acute infection. The assay under evaluation is tested using the commercial TP seroconversion panel; PSS901 AccuVert™ Syphilis Seroconversion Panel 0615-0017 [sourced from SeraCare Life Sciences Inc] in singular on one lot.

6.5. TP performance panel

One AccuSet Syphilis Performance Panel 0820-0300 containing 17 members, (PSS202) [sourced from SeraCare, Life Sciences Inc] is tested using the assay under evaluation in singular on one lot.

6.6. WHO international reference preparations

The WHO 1st Internal Standard for human syphilitic plasma IgG and IgM [NIBSC code: 05/132] is tested using the assay under evaluation in singular on one lot.

Table 3 summarizes the panels which will be used in the performance evaluation

<table>
<thead>
<tr>
<th>Panel name</th>
<th>Number of specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO syphilis (Treponemal and non-treponemal) performance evaluation panel</td>
<td>450 TP positive</td>
</tr>
<tr>
<td></td>
<td>750 TP negative</td>
</tr>
<tr>
<td>TP lot-to-lot variation panels</td>
<td>TP positive specimens, (160 in total)</td>
</tr>
<tr>
<td>Commercial TP seroconversion panel</td>
<td>1 panel comprising 9 specimens in total</td>
</tr>
<tr>
<td>Commercial TP performance panel</td>
<td>1 panel comprising 17 specimens in total</td>
</tr>
<tr>
<td>WHO 1st Internal Standard for human syphilitic plasma IgG and IgM</td>
<td>1 panel comprising 5 specimens in total</td>
</tr>
</tbody>
</table>
7. Laboratory testing

7.1. Review of instructions of use
Each assay under evaluation is used strictly in accordance with the instructions for use (IFU) issued by the manufacturer. WHO PEL will send a copy of the IFU to WHO upon delivery of the test kits and prior to the commencement of the performance evaluation.

The IFU must be reviewed against the IFU submitted to WHO as part of the dossier assessment for the prequalification assessment. If the IFU has been updated since dossier submission, WHO technical officer-in-charge of performance evaluation will inform WHO technical officer-in-charge of dossier assessment of any ramifications for dossier assessment prior to the performance evaluation commencing.

Any specific procedural aspects of the IFU that should be reinforced or clarified, such as use of specimen transfer device included within the test kit, will be communicated by email to WHO PEL, prior to commencement of the evaluation.

7.2. Sequence of testing
The WHO syphilis specimen reference panel is run such that approximately one half of the specimen panel will be run with the one lot and the other half of the panel with the other lot. The specimens of the WHO TP or non-TP specimen reference panel should initially be tested in singular and in a blinded manner.

Lot-to-lot variation is assessed by testing the same set of dilution series (comprised 2-fold dilutions of 10 stock TP or non-TP positive specimens) on two separate production lots of the assay under evaluation.

For the additional TP seroconversion panels, performance panels, WHO Anti-syphilitic plasma IgG IgM panel, are then tested in singular on a single lot.

For the purpose of evaluating the rapid diagnostic tests, a ‘test run’ is defined as a consecutive run of simple/rapid tests of the same production lot performed during the same ‘session’. A ‘testing session’ might be considered to be a morning or afternoon.

7.3. Recording test results
All test results are recorded on standardized test result forms and then entered in a Microsoft Excel spreadsheet for further data analysis. For subjectively read assays such as RDTs or line immunoassay, the intensity of band/line/spot is additionally entered into the data collection sheet. The intensity rating read outs is described in Table 4.

<table>
<thead>
<tr>
<th>Scoring index</th>
<th>Intensity reading scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Negative</td>
</tr>
<tr>
<td>1</td>
<td>Very Weak, but Definitely Reactive</td>
</tr>
<tr>
<td>2</td>
<td>Medium to Strong Reactivity</td>
</tr>
<tr>
<td>7</td>
<td>Debris (control line present but unable to read the test results clearly)</td>
</tr>
</tbody>
</table>

Table 4 - Results legend for data collection sheets for subjectively read assays
Visual interpretation of results of subjectively read assays is made independently by three readers (without the knowledge of the other two sets of results and blinded to the reference result for the specimen) and entered onto the data collection forms. After all results are recorded the outcome results are compared by the operator carrying out the assay so that any mistakes may be identified and rectified immediately. Should recording errors be identified, both the original and corrected result are recorded and initialed by the reader. When the three readers interpret the results differently from each other, the consensus is recorded as that interpretation which occurs two out of three times. In cases where all three interpretations are different, the result is recorded as indeterminate and excluded from the data analysis.

A technician's appraisal is made of each assay under evaluation and is completed by the operator performing the testing. It includes questions about the ease of the procedure, reading of results, clarity of IFU, as well as room to record any specific difficulties encountered during the evaluation.

8. Quality control and interpretation of test results

8.1. Test kit controls
If available, manufacturer-supplied positive and negative test kit controls will be run as indicated in the IFU for all test formats included at the commencement of each testing session for rapid diagnostic tests and other formats. Where positive and negative test kit controls are not supplied by the manufacturer, as will be the case for many RDTs, the external quality control specimen will act at the control specimen, see later section 9.3.

8.2. Internal control lines for rapid diagnostic tests
Generally, RDTs contain a control band, line or spot to determine that the test device is operating correctly. Most control bands/lines/spots will become visible with the addition of reagent (e.g. buffer) only. However, some RDTs will contain a control band/line/spot that becomes visible with addition of specimen (i.e. presence of IgG). It is imperative that the exact nature of the control band/line/spot is ascertained and recorded in the report. An experiment is performed to verify this point, if not explicitly mentioned in the IFU.

8.3. External quality control specimen
WHO PEL supplies external quality control (QC) specimens which are tested in singular at the beginning of each test session for rapid diagnostic tests. The QC specimens represent a lowly reactive anti-TP or non TP positive specimen, and thus may be different for different assays and different assay formats. The QC specimens are supplied by WHO PEL or acquired commercially, depending on the assay under evaluation.

8.4. Proficiency panels
A proficiency panel must be run successfully for each assay by each operator before the evaluation commences. This may be the same panel as that used at the time of assay demonstration by the manufacturer or for training purposes.

8.5. Limits of acceptability
All results on test kits controls and QC specimens are entered on the data collection forms. Should the test kit controls or the QC specimen not give results within the expected time frame, testing on that assay is suspended until the cause has been identified and a satisfactory solution identified. Such problems must be communicated immediately to WHO and recorded on the data forms. The PI is responsible for checking all data entry forms for legibility, accuracy and completeness.
8.6. Interpretation of results
The interpretation of results for each assay under evaluation is made strictly according to the manufacturers IFU. Invalid test results are recorded on the data collection forms including where the control line does not appear or in any other way the test result is invalid as defined by the IFU.

9. Analysis of data

9.1. Invalid devices
The number of invalid devices (if RDT or other subjectively read format) is recorded as the number of invalid test results as a percentage of the total number of devices used for the performance evaluation using clinical specimens (excluding commercially acquired panels, reference preparations, lot to lot variation panels).

Invalid results may mean invalid test results as defined by the instructions for use such as where the control line/band/spot does not appear or invalid due to obviously defective test device or defective transfer pipette.

9.2. Inter-reader variability
The inter-reader variability is calculated when assay readings must be performed without any objective reading instruments i.e. rapid diagnostic tests. Three persons independently interpret each test result. The inter-reader variability is expressed as the percentage of specimens for which initial test results are interpreted differently (i.e. reactive or non-reactive) by the independent readers for the clinical specimens (excluding commercially acquired panels, reference preparations, lot to lot variation panels).

Inter-reader variability is assessed for each test band i.e. TP or non-TP test bands/lines/dots.

9.3. Performance characteristics from WHO specimen reference panel
The following methods are used to calculate the performance characteristics for each assay under evaluation and is closely linked to the reference testing results gained during specimen panel characterization.

Table 4 - 2 x 2 table for calculation of performance characteristics

<table>
<thead>
<tr>
<th>Results of assay under evaluation</th>
<th>Results of reference testing</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>–</td>
<td>a+b</td>
<td></td>
</tr>
<tr>
<td>reactive</td>
<td></td>
<td></td>
<td>False positives</td>
<td></td>
</tr>
<tr>
<td>True positives</td>
<td>a</td>
<td>b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-reactive</td>
<td></td>
<td></td>
<td>True negatives</td>
<td></td>
</tr>
<tr>
<td>False negatives</td>
<td>c</td>
<td>d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9.3.1. Sensitivity
Sensitivity is the ability of the assay under evaluation to detect correctly specimens that contain the analyte of interest i.e. TP or non-TP antibodies (reference results positive). Thus sensitivity is the number of true positive specimens identified by the assay under evaluation as positive (a), divided by the number of
specimens identified by the reference assays as positive (a+c), expressed as a percentage. Sensitivity will be calculated for each analyte separately.

\[
\text{Sensitivity} = \frac{a}{a + c}
\]

9.3.2. Specificity
Specificity is the ability of the assay under evaluation to detect correctly specimens that do not contain the analyte of interest i.e. TP or non-TP antibodies (reference results negative). Thus specificity is the number of true negative specimens identified by the assay under evaluation as negative (d), divided by the number of specimens identified by the reference assays as negative (b+d), expressed as a percentage. Specificity will be calculated for each analyte separately.

\[
\text{Specificity} = \frac{d}{b + d}
\]

9.3.3. Confidence intervals
The 95% confidence intervals are calculated for both sensitivity and specificity in order to assess the level of uncertainty introduced by sample size, etc. Exact 95% confidence intervals for binomial proportions will be calculated from the F-distribution. [Armitage, 2002; Kirkwood, 2003]

9.3.4. Positive predictive value (PPV)
The probability that when the test is reactive that the specimen does contain the analyte of interest TP antibodies. PPVs at various prevalence rates will be calculated for each analyte separately, according to the following formula.

\[
\text{PPV} = \frac{(\text{prevalence})(\text{sensitivity})}{(\text{prevalence})(\text{sensitivity}) + (1 - \text{prevalence})(1 - \text{specificity})}
\]

9.3.5. Negative predictive value (NPV)
The probability that when the test is negative that a specimen does not contain the analyte of interest TP antibodies. NPVs at various prevalence rates will be calculated for each analyte separately, according to the following formula.

\[
\text{NPV} = \frac{(1 - \text{prevalence})(\text{specificity})}{(1 - \text{prevalence})(\text{specificity}) + (\text{prevalence})(1 - \text{sensitivity})}
\]

The probability that a test result will accurately determine the true infection status of a person being tested varies with the prevalence of syphilis infection in the population from which the person comes. In general, the higher the prevalence of syphilis infection in the population, the greater the probability that a person testing positive is truly infected (i.e., the greater the positive predictive value [PPV]). Thus, with increasing prevalence, the proportion of individuals testing false-positive decreases; conversely, the likelihood that a person whose test result is negative is truly uninfected (i.e., the negative predictive value [NPV]), decreases as prevalence increases. Therefore, as prevalence increases, so does the proportion of individuals testing false-negative.

The PPV and NPV are calculated at a prevalence of 0.1%, 1% and 5%.
9.4. Discrepant results

Those specimens with results that are consistent with the reference results i.e. the characterized specimen results, undergo no further testing. Those specimens with results discrepant from the reference result are retested in duplicate using the same lot number by the same operator. The results that occur two out of three times are recorded as the final result. If the result is again discrepant, the specimen is retested on a second lot number, if available. If the result on the second lot is concordant with the reference result, no further testing is required. If the result is still discrepant from the reference results, the result is recorded as is.

An initial sensitivity and specificity are calculated based on the initial results obtained for the assay under evaluation on the first lot. The final sensitivity and specificity are calculated taking into consideration the results of repeat testing performed on a second lot of the assay under evaluation i.e. if found to be concordant on the second lot will be recorded as such, and if found to be discrepant on the second lot will be recorded as such.

9.5. Interpretation of results from lot-to-lot variation panels

The results of both the TP or non-TP lot-to-lot variation panels for the two production lots are compared.

9.6. Interpretation of TP seroconversion sensitivity

The results obtained from seroconversion panels using the assay under evaluation are compared with those obtained using <validated Chemiluminescence Treponemal immunoassay>; the assay arbitrarily designated the reference for determination of relative sensitivity in these panel. For each seroconversion series (panel) the first specimen in the sequence to become reactive with <validated Chemiluminescence Treponemal immunoassay> is assigned the value “0”. Results from the assay under evaluation are compared with <validated Chemiluminescence Treponemal immunoassay > by determining the difference between the specimen assigned value “0” and the relative position in the sequence of the first specimen showing a reactive result for the assay under evaluation. For example, if an assay becomes reactive two specimens earlier in a series than <validated Chemiluminescence Treponemal immunoassay>, the value assigned for that series in that assay would be -2. Similarly, if an assay becomes reactive one specimen later than <validated Chemiluminescence Treponemal immunoassay>, the value assigned would be +1. The assigned values over the seroconversion panel (series) are averaged to determine a mean relative seroconversion sensitivity index for each assay and the 95% confidence interval is determined.

9.7. Interpretation of results from TP performance panels

The number of specimens correctly identified by the assay under evaluation in the anti-Treponema pallidum mixed titer performance panel is determined by comparison with the combined reference results generated by PEL <validated Chemiluminescence Treponemal immunoassay>.

9.8. Interpretation of results from WHO TP international reference preparations

The number of specimens correctly identified by the assay under evaluation for the WHO international reference preparation are determined by comparison with the combined reference results generated by WHO PEL for <validated Chemiluminescence Treponemal immunoassay >

9.9. Technician's appraisal

The technical aspects of the assay under evaluation is assessed by the technician who performed the testing. These assessments, along with other selected assay characteristics, contribute to an overall appraisal of each assay's suitability for use in small laboratories. To enable comparison between assays, a scoring system is used to rate specified operational characteristics.
9.10. **Report preparation**

The data analysis and report drafting is carried out by WHO PEL and sent to WHO in a timely manner. WHO verifies the draft report and sends to the authorized contact designated by the manufacturer for comment. The company has one month to provide comments. The final report is prepared after one month has elapsed. WHO ensures that the comments of the authorized contact are reviewed and any outstanding issues are resolved before final publication. The final report is prepared and disseminated by WHO. A copy of the final report is sent to the authorized contact designated by the manufacturer and to WHO PEL.

10. **Materials and supplies**

10.1. **Data collection sheets**

All data will be reported to WHO on the following forms:

- Data collection Microsoft Excel spreadsheet for the rapid diagnostic tests
- Technician's appraisal worksheet
- Findings of supplementary testing on discrepant specimens

10.2. **Supplies**

The manufacturers of products will provide the products and any equipment necessary for the evaluation free of charge.

11. **Roles and responsibilities**

11.1. **Responsibilities of WHO PEL**

a) Act as repository for the WHO TP and non TP specimen reference panels, lot-to-lot variation, seroconversion and performance panels;

b) Conducting the performance evaluation in accordance with internationally recognized best practice;

c) Preparation of QC specimens and proficiency panels;

d) Preparation of draft report on performance evaluation;

e) Advising WHO on operational characteristics of assays evaluated.

All source data, data analysis records and all correspondence are retained and archived for a period of at least ten years.

11.2. **Responsibilities of WHO**

a) Technical advice to the PI;

b) Technical and administrative management of the performance evaluation;

c) Procurement and delivery of supplies/assays;

d) Verification of the draft report, seeking of comments from manufacturer;

e) Preparation and dissemination of the final report;

f) Formal contacts with authorized contacts of the manufacturers.

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 WHO PEL, do not warrant or represent that the evaluations conducted with the TP test kits referred to in this document are accurate, complete and/or error-free. WHO and the WHO PEL e 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. References


World Health Organization Guidelines on HIV Safety Precautions, and Guidelines for the Safe Transport of Specimens (WHO/EMC/97.3)

International Standards
EN 13612:2002 Performance evaluation of in vitro diagnostic medical devices
ISO 17025:2017 (General requirements for the competence of testing and calibration laboratories)
ISO15189:2012 (Medical laboratories — Particular requirements for quality and competence)
13. Other documents required

Evaluation Protocols
SOP_PQDx_326 Protocol for performance evaluation of Treponemal anad non-Treponemal simple rapid diagnostic tests

Work Instructions
SOP_PQDx_072 PQDx work instruction for performance testing at WHO Collaborating Centre
SOP_PQDx_073 PQDx work instruction for specimen panels storage and characterization at WHO Collaborating Centre
SOP_PQDx_074 PQDx work instruction for data entry and analysis
SOP_PQDx_075 PQDx work instruction for report preparation and dissemination
SOP_PQDx_076 PQDx work instruction for specimen acquisition for performance evaluation (serum/plasma specimens)

Master Templates
PQDx_xxx PQDx report template for discriminatory anti-TP simple/rapid

Standard Letters
SOP_PQDx_045 PQDx standard letter to accompany evaluation protocol
SOP_PQDx_077 PQDx WHO PEL standard letter to request test kits:
SOP_PQDx_081 Standard letter for draft report
SOP_PQDx_083 Standard letter for final report (simple/rapid assays)