WHO PREQUALIFICATION TEAM:
DIAGNOSTICS

World Health Organization

Technical Specifications Series for submission to WHO Prequalification – Diagnostic Assessment

TSS-6 Syphilis rapid diagnostic tests

Technical Specifications Series for submission to WHO Prequalification – Diagnostic Assessment: syphilis rapid diagnostic tests

Published 20 December 2018

ISBN 978-92-4-151516-0

Suggested citation. Technical Specifications Series for submission to WHO Prequalification – Diagnostic Assessment: syphilis rapid diagnostic tests. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.

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

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

Table of contents

Acknowledgements

The document *"Technical Specifications Series for submission to WHO Prequalification – Diagnostic Assessment: Syphilis rapid diagnostic tests"* was developed with support from the Bill & Melinda Gates Foundation and UNITAID. The document was prepared in collaboration with T. Crucitti, Centre Pasteur du Cameroun, Yaoundé, Cameroon; R. J. S. Duncan, London, United Kingdom of Great Britain and Northern Ireland; D. Healy, M. Lanigan, U. Ströher, Prequalification Team – Diagnostic Assessment, WHO, Geneva and with technical and programmatic input from M. Taylor and I. Toskin, Department of Reproductive Health and Research, Human Reproduction team, WHO, Geneva.

This document was produced under the coordination and supervision of D. Healy, U. Ströher and I. Prat, Prequalification Team – Diagnostic Assessment, WHO, Geneva, Switzerland.

List of contributors

A technical consultation on WHO prequalification requirements for syphilis rapid diagnostic tests was held in Geneva, Switzerland from 20 to 21 August 2018.

Meeting participants: R. Ballard, Atlanta, United States of America (USA); K. Bialas, Scientific Reviewer, Division of Microbiology Devices, Office of In Vitro Diagnostics and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland, USA[1]; S. Braniff, National Serology Reference Laboratory (NRL), Victoria, Melbourne, Australia; C. Bristow, UC San Diego School of Medicine, San Diego, California, USA; A. Chua, Leader, Diagnostics Network Working Group, Médecins Sans Frontières International ,Geneva, Switzerland; T. Crucitti, Centre Pasteur du Cameroun, Yaoundé, Cameroon; R. J. S. Duncan, London, United Kingdom of Great Britain and Northern Ireland; N. Gerald, Scientific Reviewer, Division of Microbiology Devices, Office of In Vitro Diagnostics and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland, USA[2]; A. Kijo, Tanzania Food and Drugs Authority (TFDA), Dar Es Salaam, Tanzania; A. L. Page, Epidemiologist, Epicentre, Paris, France; B. Vetter, Foundation for Innovative New Diagnostics (FIND), Geneva, Switzerland.
WHO Secretariat: D. Healy; M. Lanigan; I. Prat; U. Ströher; W. Urassa, Prequalification Team – Diagnostic Assessment Group, Regulation of Medicines and other Health Technologies; M. Taylor and I. Toskin, Department of Reproductive Health and Research, Human Reproduction team.

The draft technical specifications document was posted on the WHO website for public consultation on 19 September 2018. Various stakeholders, including manufacturers submitting to WHO Prequalification of IVDs, IVD manufacturing industry associations, various national and international regulatory bodies, and IVD standards organizations were informed of the consultation in order to solicit feedback. A two- month response period was provided.

Public comments were received for consideration from the Asian Harmonization Working Party (AHWP) Working group 2, China, Hong Kong SAR; S. Best, Melbourne, Australia; Division of Microbiology Devices, Office of In Vitro Diagnostics and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland, USA; C. Hayden, R&D Manager, bioLytical® Laboratories Inc, Richmond, British Columbia, Canada; In Vitro Diagnostic Devices Evaluation Division, Medical Devices Bureau, Therapeutic Products Directorate, Health Canada, Government of Canada, Canada; ISO/TC 212 Secretary on behalf of ISO/TC 212, International Standards Organization, Geneva, Switzerland; MedTech Europe, Brussels, Belgium; W. Schmandt, Siemens Healthcare Diagnostics Products GmbH, Marburg, Germany.

[1] Participated via web conferencing
[2] Participated via web conferencing

Abbreviations

ANA	anti-nuclear antibodies
CLIA	chemiluminescence Immunoassays
EDTA	ethylenediaminetetraacetic acid
EIA	enzyme immunoassay
IFU	instructions for use
IgG, IgM	Immunoglobulin G, M
IVD	in vitro diagnostic
LOD	limit of detection
POC	point of care
RDT	rapid diagnostic test
RPR	rapid plasma reagin
STI	sexual transmitted infection
TPHA	Treponema pallidum haemagglutination assay
TPPA	Treponema pallidum particle agglutination assay
TpN	*T. pallidum* Nichols (polypeptide)
TSS	Technical Specification Series
WHO	World Health Organization

A. Introduction

The purpose of this document is to provide technical guidance to in vitro diagnostic (IVD) medical device manufacturers that intend to seek WHO prequalification of rapid diagnostic tests (RDTs) used to detect syphilis infection. For the purpose of this document, RDTs are lateral-flow or flow-through immunochromatographic antigen or antibody detection tests, which rely on the capture of dye-labelled antibodies or antigens to produce a visible band or dot on a strip of nitrocellulose, often encased in plastic housing, referred to as cassettes.

The document is relevant to qualitative RDTs that detect antibodies *to T. pallidum* alone or in combination with antibody detection to non-treponemal antigens which are used as part of a testing algorithm.

- A *T. pallidum* antibody RDT submitted for WHO prequalification is expected to detect all disease stages with the exception of congenital syphilis or neurosyphilis, unless claimed;
- A non-treponemal antibody RDT is expected to detect syphilis related reagin antibodies.

The requirements outlined in this document do not include those necessary to demonstrate that the IVD could be used for confirmatory testing, nor the requirements for any accompanying quality control material. However, if quality control material is provided with the assay, it should demonstrate that the IVD is functional and performs as claimed. (ISO 15198) The requirements are not intended for self-testing.

Minimum performance requirements for WHO prequalification are summarized in this document and apply equally to RDTs intended solely for the detection of syphilis and to those tests manufactured where syphilis detection comprises one component of a multidetection system (e.g. a HIV/syphilis dual-detection RDT).

For the purpose of this document, the verbal forms identified below are defined as follows:

- “shall” indicates that the manufacturer is required to comply with the technical specifications. A documented justification and rationale shall be provided by the manufacturer when the WHO prequalification submission does not comply with the required technical specifications outlined in this document.

- "should" indicates that the manufacturer is recommended to comply with the technical specifications, but it is not a requirement.
- "may" indicates that the technical specifications are a suggested method to undertake the testing, but it is not a requirement.

Where possible, WHO performance conditions are aligned with published guidance, standards and/or regulatory documents. Although references to source documents are provided, in some cases WHO prequalification has additional requirements. Detailed numbers for specimens to be tested are provided for each study in Part 1 and 2 of this document. These are the minimum numbers that are necessary to meet WHO prequalification requirements. The final study numbers chosen by the manufacturer need to be evaluated based on the risk assessment of the RDT under evaluation.

For WHO prequalification purposes, manufacturers shall provide evidence in support of the clinical performance of an IVD to demonstrate that reasonable steps have been taken to ensure that a properly manufactured IVD, being correctly operated in the hands of the intended user, will detect the target analyte consistently and fulfil its indications for use.

WHO prequalification requirements summarized in this document do not extend to the demonstration of clinical utility, i.e. the effectiveness and/or benefits of an IVD, relative to and/or in combination with other measures, as a tool to inform clinical intervention in a given population or healthcare setting. To demonstrate clinical utility, a separate set of studies is required. Clinical utility studies usually inform programmatic strategy and are thus the responsibility of programme managers, ministries of health and other related bodies in individual WHO Member States. Such studies do not fall under the scope of WHO prequalification.

B. Other guidance documents

This document should be read in conjunction with other relevant WHO guidance documentation, including:

WHO prequalification documents[3]

- Technical Guidance Series for WHO Prequalification – Diagnostic Assessment;
- Sample Product Dossiers for WHO Prequalification – Diagnostic Assessment;
- Instructions for Compilation of a Product Dossier, WHO document PQDx_018.

WHO syphilis guidelines

- Laboratory diagnosis of sexually transmitted infections, including human immunodeficiency virus;[4]
- WHO guideline on syphilis screening and treatment for pregnant women[5].

C. Performance principles for WHO prequalification

C.1 Intended use

An IVD intended for WHO prequalification shall be accompanied by a sufficiently detailed intended use statement. This should allow an understanding of at least the following:

[3] http://www.who.int/diagnostics_laboratory/evaluations/en/
[4] https://www.who.int/reproductivehealth/publications/rtis/9789241505840/en/
[5] https://www.who.int/reproductivehealth/publications/rtis/syphilis-ANC-screenandtreat-guidelines/en/

- what is detected (e.g. detection of antibodies to *T. pallidum*, and/or detection of non-treponemal antibodies related to syphilis infection);
- the function of the IVD (e.g. screening for surveillance or case management among the sexually active population for symptomatic or asymptomatic *T. pallidum* infections, as an aid to diagnosis of syphilis infection by detection of antibodies to *T. pallidum*; as an aid for detection of non-treponemal antibodies related to syphilis infection);
- the clinical indication of the IVD (e.g. aid in the diagnosis syphilis), and that the result is qualitative;
- the intended testing population e.g. sexually active population, special populations (e.g. pregnant women);
- the intended user;
- the intended operational setting (e.g. for professional use in a laboratory setting, or point of care[6] (POC));
- any limitation to the intended use e.g. that the antibody test cannot differentiate between active disease and treated infection; exclusion of blood donor screening, neonatal screening, testing of cerebrospinal fluid (neurosyphilis), self-testing;
- the type of specimen required.

C.2 Diversity of specimen types, users and testing environments and impact on required studies

Depending on the intended use of the RDT, analytical and clinical performance studies shall be designed to consider the diversity of knowledge and skills of potential RDT users, and the operational settings in which testing is likely to occur. It is a manufacturer's responsibility to ensure that the risk assessment and subsequent validation studies for an RDT reflect the intended operational settings, including service delivery complexity and involve the user population expected to conduct the test. Prequalified syphilis RDTs in low- and middle-income countries are likely to be used by laboratory professionals[7] either in centralized testing laboratories (although access may be limited) or at POC, or lay providers[8] trained in the use of the test at POC.

For WHO prequalification submission, clinical performance studies shall be conducted using each specimen type (e.g. capillary whole blood, venous whole blood, serum, plasma) claimed in the instructions for use (IFU). Note that the specimen type that is most likely to be used in resource limited WHO Member States at POC is capillary whole blood. Testing of cerebral spinal fluid is not included in the scope of this document however; the performance of all specimen types claimed by the manufacturer shall be demonstrated. Laboratory demonstration of equivalence between specimen types without evidence of clinical validation is insufficient (with exception of anticoagulants). For example, studies that comprise testing of left-over or repository specimens by research and development staff at a manufacturer's facility shall not, on their own, be considered sufficient to meet most of the performance requirements for WHO prequalification.

[6] Point-of-care (POC) in-vitro diagnostic testing refers to decentralized testing that is performed by a minimally trained healthcare professional near a patient and outside of central laboratory testing facilities. It does not refer just to sample collection procedures.
[7] Medical technologists, medical laboratory technicians or similar, who have received a formal professional or paraprofessional certification or tertiary education degree.
[8] Any person who performs functions related to healthcare delivery and has not received a formal professional or paraprofessional certification or tertiary education degree.

C.3 Applicability of supporting evidence to RDT under review

Performance studies shall be undertaken using the specific, final (locked-down design) version of the RDT intended to be submitted for WHO prequalification. For WHO prequalification, design lock-down is the date that final documentation, including quality control and quality assurance specifications, is signed off and the finalized method is stated in the IFU. Where this is not possible, a justification shall be provided; additional supporting evidence may also be required. This may occur in the case of minor variations to the design where no negative impact on performance has been demonstrated (see WHO document PQDx_121 Reportable Changes to a WHO Prequalified In Vitro Diagnostic Medical Device[9]).

The version of the IFU used for performance evaluations submitted to WHO prequalification shall be stated. If the test procedure in the IFU is changed in any way after completing performance verification and validation studies the change shall be reported to WHO.

Detailed information regarding the numbers of lots required for certain performance studies is provided in Parts 1 and 2 of this document. Each lot should comprise different production (or manufacturing, purification, etc.) runs of critical reagents, representative of routine manufacture. It is a manufacturer's responsibility to ensure, via risk analysis of the RDT that the minimum numbers of lots chosen for estimating performance characteristics considers the variability in performance likely to arise from the diversity of key components and their formulation.

Differences found between lots during the analytical and clinical performance studies shall be reported.

The true *T. pallidum* status of a specimen shall be determined using suitable reference methods for which justification shall be provided; comparison with a similar device is insufficient for resolution of discrepant specimens (e.g. other method from the same manufacturer or other method using the same antigens provided by the same supplier). For WHO purposes the reference method should be to a level that is currently at a developed stage of technical capability based on the relevant consolidated findings of science, technology and experience (commonly referred to as state of the art).

Estimation (and reporting) of RDT performance shall include the rate of invalid test results and the 95% confidence interval around the estimated values for key performance metrics, as appropriate.

Analytical studies shall include testing for all specific characteristic factors (e.g. relevant epitopes) for which detection is claimed. For certain analytical studies it may be acceptable to use contrived specimens (e.g. where non-reactive human specimens have been spiked with those containing analyte specific antibodies). All reasonable attempts should be made to use clinical specimens (unless otherwise stated) and justification should be provided where contrived specimens are used in the submitted studies.

For quantitative assays, additional requirements may apply. Contact WHO prequalification for more information on these requirements.

[9] http://apps.who.int/iris/bitstream/handle/10665/251915/WHO-EMP-RHT-PQT-2016.01-eng.pdf;jsessionid=30D5BF0B09FFDA3B38A1698E65C8B496?sequence=1

D. Table of requirements

Part 1	Establishing non-clinical evidence (analytical performance characteristics)
1.1	**Stability of sample(s)**
1.1.1	Specimen collection, storage and transport
1.2	**Validation of specimens**
1.2.1	Specimen types
1.3	**Metrological traceability of calibrator and control material values**
1.3.1	Metrological traceability of control material values
1.4	**Precision (repeatability and reproducibility)**
1.4.1	Repeatability, reproducibility
1.5	**Analytical sensitivity**
1.5.1	Analytical sensitivity
1.6	**Analytical specificity**
1.6.1	Potentially interfering substances
1.6.1.1	Endogenous
1.6.1.2	Exogenous
1.6.2	Cross-reactivity
1.7	**High dose hook effect**
1.7.1	High dose hook effect
1.8	**Validation of the assay procedure**
1.8.1	Validation of reading times
1.8.2	Validation of controls
1.8.3	Establishment of reader cut-off
1.9	**Usability/human factors**
1.9.1	Flex studies
1.10	**Stability of the IVD**
1.10.1	Claimed shelf-life (including transport stability)
1.10.2	In-use stability (open pack or open vial stability)
Part 2	**Establishing clinical evidence (clinical performance characteristics)**
2.1	**Diagnostic sensitivity and specificity**
2.1.1	Mixed titre panels
2.1.2	Performance panels
2.1.3	Diagnostic sensitivity and specificity study general requirements
2.1.4	Diagnostic sensitivity
2.1.5	Diagnostic specificity
2.2	**Qualification of usability for RDT IVDs**
2.2.1	Labelling comprehension study (including IFU)
2.2.2	Results interpretation study

Part 1 Establishing non-clinical evidence (analytical performance characteristics)

Aspect	Testing requirements	Notes on testing requirements	Source documents
1.1 Stability of sample(s)			
1.1.1 Specimen collection, storage and transport	Real time studies accounting for: 1. storage conditions (duration at different temperatures, temperature limits, freeze/thaw cycles); 2. transport conditions, where applicable; 3. intended use (see note 1); 4. specimen collection and/or transfer devices intended to be used with the RDT.	1. Evidence shall be provided which validates the maximum allowable time between specimen collection, processing of the specimen and its addition to the RDT in the setting where testing takes place. 2. Unless all specimens are expected to be processed as fresh samples within a specified time frame, the RDT performance shall be established for each different storage condition and at the beginning and end of the stated period. 3. In case the use of archived specimens is considered for part 2, evidence of stability shall be demonstrated.	
1.2 Validation of specimens			
1.2.1 Specimen types	1. For each claimed specimen type, testing in paired specimens shall be undertaken in at least: • 50 treponemal positive specimens (see note 1 & note 3); • 50 treponemal negative specimens. 2. If equivalence is claimed between different anticoagulants, testing shall be conducted in at least: • 25 positive specimens of each claimed anticoagulant; • 25 negative specimens of each claimed	1. Specimens confirmed positive by either treponemal assays (enzyme immunoassay (EIA), chemiluminescence immunoassay (CLIA) and/or Treponema pallidum particle agglutination assay (TPPA) or non-treponemal assays (such as rapid plasma reagin (RPR) assays). 2. The relationship between RDT performance in claimed specimen types and materials used for analytical studies shall be established. The design of subsequent studies shall then take that relationship into account. 3. Positive specimens (undiluted) should be chosen so that	Technical Guidance Series for WHO Prequalification – Diagnostic Assessment TGS-3 (1)

Aspect	Testing requirements	Notes on testing requirements	Source documents
	anticoagulant. 3. The equivalence of specimen types shall be determined for all claimed analytes (e.g. non-treponemal antibodies, *T. pallidum* specific antibodies etc., as appropriate). 4. If there is no equivalence between claimed specimen types, then the impact that this will have on each subsequent performance claim shall be fully understood and described (see note 4). 5. If an RDT is intended for testing whole blood or capillary whole blood and some aspects of performance have been established using serum or plasma specimens, then • the relationship between analytical sensitivity in serum/plasma to that of the same characteristic in whole blood shall be understood (see note 4,5); • paired specimens shall be used for RDTs intended to test capillary and venous blood (see note 6).	the majority is weakly reactive for the respective analyte (near the RDT limit of detection (LOD)[10]) and that different stages of infection are included. 4. Where a significant difference in performance exists between specimen types, equivalence may need to be investigated as part of a larger clinical study (See Part 2). 5. Demonstration of the comparability of specimen types may be achieved by comparing RDT results between end-point dilution series of several positive whole blood specimens titrated into compatible (blood group type) whole blood and compared with the serum from those same specimens titrated into serum. 6. All reasonable attempts should be made to use clinical specimens giving responses close to the LOD for capillary and venous blood.	
1.3 Metrological traceability of calibrator and control material values			
1.3.1 Metrological traceability of control material values		1. If a control material has an assigned concentration value, the metrological- (not commercial- nor documentary-) traceability to an accepted international standard shall be demonstrated.	WHO Prequalification – Diagnostic Assessment PQDx_018 (2) ISO 15198 (3) ISO 17511 (4)
1.4 Precision (repeatability and reproducibility)			
1.4.1	Both repeatability (within-condition) and reproducibility	1. E.g. within- or between-run, -lot, -day, -site, etc.	EN 13612:2002 (5)

[10] measured quantity value, obtained by a given measurement procedure, for which the probability of falsely claiming the absence of a component in a material is β, given a probability α of falsely claiming its presence. NOTE 1 IUPAC recommends default values for α and β equal to 0,05. NOTE 2 The term analytical sensitivity is sometimes used to mean detection limit, but such usage is now discouraged. (ISO 18113-1:2009)

Aspect	Testing requirements	Notes on testing requirements	Source documents
Repeatability, reproducibility	(between-condition) shall be determined for each analyte for which detection is claimed (e.g. non-treponemal and/or *T. pallidum* antibodies as appropriate). The panel of spiked specimens shall include at least: 1. 1 non-reactive specimen; 2. 1 weak reactivity positive specimen (approx. 1-2 x RDT's LOD); 3. 1 medium reactivity positive specimen (approx. 2-3 x RDT's LOD); 4. the panel shall include whole blood specimens if claimed. Each panel member shall be tested: 1. in at least 5 replicates; 2. using 3 different lots (see note 4); 3. over 5 days (not necessarily consecutive) with 1 run per day (alternating morning/afternoon); 4. at each of 3 different testing sites. The effect of operator-to-operator variation on RDT performance should be included as part of the precision studies (see also note 8). Testing should be done: 1. by personnel representative of intended users. 2. unassisted; 3. using *only* those materials provided with the RDT (e.g. IFU, labels and other instructional materials).	2. Studies shall be statistically designed and analysed to identify and isolate the sources and extent of any variance. • The extent of variance (due to manufacturing, test procedures or environment) which could nullify any claim shall be identified and the power of testing shall be sufficient to identify any such variance. 3. Where possible, the testing panel should be the same for all operators, lots and sites. 4. Each lot shall comprise different production (or manufacturing, purification, etc.) runs of critical reagents. 5. The percentage of correctly identified, incorrectly-identified and invalid results shall be tabulated for each specimen and be separately stratified according to each site, lot, etc. This type of analysis is especially important for RDTs that may not have any numerical values. 6. To understand irregularities in results obtained, at least 2 lots should be tested at each of the 3 testing sites (3 different lots are required to be tested overall in the 3 testing sites). 7. The effect of operator-to-operator variation on RDT performance may also be considered as a human factor when designing flex studies (see 1.9.1 Flex studies) and may be addressed as part of clinical studies in representative populations (see Part 2). 8. Users shall be selected based on a pre-determined and contextually appropriate level of education, with literacy and auxiliary skills that will challenge the usability of the RDT and reflect the diversity of intended users and operational settings. These characteristics shall be detailed in the submission.	CLSI EP12-A2 (6)
1.5 Analytical sensitivity			
1.5.1 Analytical sensitivity	Analytical sensitivity shall be determined using the two approaches outlined below using a minimum of 2 lot.	1. For the international standards, the result shall be expressed in international units as analytical end-point	European Commission decision on CTS (7)

Aspect	Testing requirements	Notes on testing requirements	Source documents
	Analytical sensitivity shall be established by determining the lowest concentration for which the probability of detection is ≥95%. 1. Analytical sensitivity shall be determined relative to the available international standards or to secondary standards metrologically traceable to them (WHO International Standard: 1st IS for human syphilitic plasma IgG: NIBSC code: 05/122). 2. Analytical sensitivity shall be determined for at least 5 specimens collected from individual patients during the primary phase of syphilis infection in comparison with a widely used test system for that analyte (e.g. TPPA, RPR) (note 3 and 4).	sensitivity with its associated metrological uncertainty. 2. If the listed international standards are not available any more, then the version of the standard used needs to be stated. 3. A widely used test system that is used in large clinical laboratories and in reference laboratories in more than 1 country shall be used. 4. Patients may be selected by testing primary lesions for the presence of T. pallidum nucleic acids and confirming subsequent blood specimens for the presence of T. pallidum or non-treponemal antibodies. 5. The lots shall comprise different production (or manufacturing, purification, etc.) runs of critical reagents.	CLSI EP17 (8) WHO Technical Report Series, No. 1004, 2017 Annex 6. (9)
1.6 Analytical specificity			
1.6.1 Potentially interfering substances	The potential for false results (false-negatives and false-positives) arising from interference from, at least, the substances/conditions listed below shall be determined (see note 1). • a minimum of 100 specimens; • each substance represented by at least 5-10 specimens from different individuals. Testing shall be undertaken in both treponemal negative and -positive specimens spiked with each potentially interfering substance at physiologically relevant dosages.	1. The risk assessment conducted for the RDT should identify substances/conditions where the potential for interference can reasonably be expected for the analyte being detected in the areas of intended use and not simply rely on published lists of such compounds and conditions which might be of limited relevance in resource limited settings (and overlook those which might be of relevance). • By conducting appropriate risk assessment, testing can be performed on the substances or conditions identified as likely to be significant and testing of potentially irrelevant substances/conditions can be avoided. 2. Interference studies should be performed with specimens with an analyte response near the RDT LOD. • The methods and concentrations used must be validated so that any effect of clinical importance would be detected. 3. Any observed interference shall be investigated and	CLSI EP07-A3 (10) CLSI EP37-A (11) ISO 14971 (12)
1.6.1.1 Endogenous	1. Antibody interference • heterophile antibodies such as human antibodies to the expression system (for recombinants), e.g. anti-*Escherichia coli* (anti-*E.coli* positive); • human anti-animal antibodies e.g. anti-mouse; • autoantibodies including systemic lupus erythematosus, anti-nuclear antibodies (ANA),		

Aspect	Testing requirements	Notes on testing requirements	Source documents
	rheumatoid factor. 2. High titres of potentially interfering antibodies such as in patients with • recent infection; • immunization; • pregnant (including multiparous) women. 3. Biochemical interference • haemolysis or haemoglobin; • hyperglobulinaemia; • cholesterol, triglycerides and bilirubin; 4. sickle-cell disease; 5. thyroiditis.	performance limitations of the RDT reported in the IFU. 4. Results shall be reported with respect to each condition and not be reported as an aggregate of the total number of specimens tested in the study. • Any effect must be evaluated against the probability of that effect occurring and causing clinically significant issues in the population tested in resource limited settings. 5. Evaluation of endogenous interfering substances may be addressed as part of the clinical studies but the number of specimens of each type evaluated shall be in accord with the requirement in this section.	
1.6.1.2 Exogenous	Relevant medicines, including: 1. antiparasitic, antibacterial, antimalarial, antiretroviral, antiviral (including for hepatitis C, B, cytomegalovirus) and anti-tuberculosis medications; 2. common over-the-counter analgesic medications (aspirin, paracetamol).		
1.6.2 Cross-reactivity	The potential for false-positive results arising from cross-reactivity (see note 1) shall be determined for a minimum of 100 specimens, including, where possible, at least 5-10 of each (See **Annex 1** for the full list): 1. viral/bacterial/parasitic infections; 2. sexually transmitted infections; 3. infections by other spirochaetes (Borrelia, Leptospira, periodontal disease causing spirochaete); 4. immunization; 5. other unrelated conditions known to cause cross-reactivity in *T. pallidum* RDTs.	1. The types of conditions/disease tested for shall be risk-based, taking into consideration the operational setting as well as the intended users for the analyte being detected in the areas of intended use and not simply rely on published lists of such cross-reactivity which might be of limited relevance in resource limited settings. • See 1.6.1, note 1 2. Cross-reactivity with other *T. pallidum* subspecies causing nonvenereal treponematoses is known and is not required to be demonstrated unless specificity to *T. pallidum* is claimed. 3. Any observed cross-reactivity shall be investigated and performance limitations of the RDT reported in the IFU. • See 1.6.1 note 3	Anderson, M.D. (13)

Aspect	Testing requirements	Notes on testing requirements	Source documents
		4. For studies of interference by cytomegalovirus and Epstein-Barr-Virus it is most important to use specimens containing immunoglobulin M (IgM) as these are well known to cause clinically relevant interference while the corresponding immunoglobulin G (IgG) specimens do not.	
1.7 High dose hook effect			
1.7.1 High dose hook effect	For each claimed analyte the potential for a prozone/high dose hook effect shall be determined: 1. using multiple, highly-reactive specimens (minimum of 20); 2. using at least 2 different concentrations (diluted by at least a factor of 10).	1. Specimens shall be chosen that have a high antibody titre as determined using a method other than the RDT intended to be prequalified e.g. using an EIA. This second method shall be of a design not subject to high dose hook effect. 2. An increase in signal upon dilution of a specimen implies a hook effect. 3. At least 3 different lots should be tested.	Butch, A.W. (14)
1.8 Validation of the assay procedure			
1.8.1 Validation of reading times	For RDTs where a reading interval is specified (i.e. time when result can first be read; time beyond which result should not be read), validation of critical time points shall be provided. The study shall use panels of at least: 1. 1 non-reactive specimen; 2. 1 weak reactivity positive specimen (approx. 1-2 x RDT's LOD); 3. 1 medium reactivity positive specimen (approx. 2-3 x RDT's LOD); 4. the panel shall include whole blood and anticoagulated plasma (e.g. ethylenediamine tetraacetic acid (EDTA)) if claimed. Performance studies shall be conducted at the extremes of the intended operational temperature range; the effect of humidity on reading times shall also be investigated.	1. The ranges of temperature and humidity validated shall be risk-based, taking into consideration likely operational settings. 2. The intended operating temperature range within which reading time has been validated, shall be clearly stated in the IFU. 3. The studies should take into account possible differences between use of freshly made devices and those stored until near the end of their assigned shelf-lives under the conditions expected in resource limited settings and being used under those conditions. 4. Some of these aspects could be evaluated within the flex studies (1.9.1).	
1.8.2 Validation of controls		1. If control materials (positive controls) are provided, the control materials should be validated as showing that if	

Aspect	Testing requirements	Notes on testing requirements	Source documents
		the RDT would not meet the claims, that the positive control will indicate the failure.	
1.8.3 Establishment of reader cut-off	In *T. pallidum* assays provided with a reader, the way in which the reader has been designed to differentiate positive specimens from negative specimens shall be demonstrated.		
1.9 Usability/human factors			
1.9.1 Flex studies	The influence of the following factors on expected results (both reactive and non-reactive) should be considered as below (this list is not exhaustive): 1. Any numerical factor in the IFU method provided and/or identified by risk assessment such as: • specimen and/or reagent volume; • specimen dilution factor; • operating temperature, pressure and humidity; • time between opening packaging or preparing reagents and starting the assay; • any mixing, rotating or incubating times, temperatures; • reading time: both the time after starting any incubations and the time for which the result is stable. 2. Ruggedness: • RDT sturdiness including robustness of packaging and labelling. RDT in final packaging shall be subjected to drop-shock testing; • permanence of component labels: print legibility, adhesiveness; • effects of lighting and humidity (see note 4); • residual volumes and characteristics of liquids (potential evaporation, pH changes, microbial growth, antimicrobial efficacy). 3. Instrumentation (if applicable and based on a risk	1. Refer to WHO document PQDx_018 "Instructions for compilation of a product dossier" for other flex studies that may be relevant, taking into consideration the broad range of operational and environmental conditions consistent with intended use in resource limited settings. 2. The factors should be investigated using "designed experimentation" so that potential critical interactions between them can be understood e.g. the effect of low or high operating temperature with low or high volume of specimen at an incorrect reading time. 3. The factors listed opposite should be investigated in ways that not only reflect, but also exceed, likely operating conditions in lower- and middle-income countries so that the limitations of the RDT can be understood. For example, in addition to investigating deviations of temperature within those claimed in the IFU, temperature ranges should be investigated that exceed those of claimed operating conditions and which could cause test failure (incorrect/invalid results). 4. The impact of lighting can be twofold – i.e. the impact of lighting on packaging e.g. fading, and the sufficiency of lighting to read the test lines. 5. For the purposes of this document, ruggedness means the ability to resist environmental shocks of a variety of kinds.	WHO Prequalification – Diagnostic Assessment PQDx 018 (2) U.S. FDA CLIA Waiver guidance (15, 16)

Aspect	Testing requirements	Notes on testing requirements	Source documents
	assessment) including: • ruggedness (see note 5); • impact of dust and mould on componentry (e.g. optics).		
1.10 Stability of the IVD			
1.10.1 Claimed shelf-life (including transport stability)	Stability studies shall be conducted using the conditions expected in the environment of intended use (temperature, humidity, light); 1. Replicate testing shall be undertaken using a panel of spiked specimens consisting of at least: • a sufficient number of non-reactive specimens (note 1); • at least 1 specimen for each analyte and each epitope used or detected by the RDT (approx. 1 - 2x LOD). 2. A minimum of 3 lots in final packaging; 3. Lots shall be subjected to simulated "transport stress" before real time studies are undertaken on these lots. This mimics the real situation.	1. The testing panels should include whole blood (to verify correct functioning of the device, such as flow, clearance of debris, lack of autoagglutination, no background interference) and serum or plasma as the specimen matrix. 2. The testing panel shall include members to monitor all claimed critical epitopes, for example TpN47, TpN17 and TpN15. • Each of these epitopes play a role in detecting syphilis in different stages of the infection. It is necessary to have a panel member to monitor each epitope system present (and possibly each stage of infection), even if polyepitopic-fusion proteins are used. This may be avoided if the manufacturer can demonstrate that each epitope system is equally stable within the fusion protein. 3. Specimens to be diluted should represent a range of stages of infection (antibody maturation) to take into account the limitations of mimicking low RDT reactivity by dilution of high avidity specimens. 4. Each lot shall comprise different production (or manufacturing, purification, etc.) runs of critical reagents. 5. The numbers of invalid tests per lot shall be reported. 6. Claims for stability shall be based on the second-last successful data point from the least stable lot. For example: for testing conducted at 3, 6, 9, 12 and 15 months, if stability was observed at 15 months, then the maximum stability claim can be 12 months. 7. Accelerated studies do not replace the need for real time studies.	ISO 23640:2011 (17) CLSI EP25 (18) Technical Guidance Series for WHO Prequalification – Diagnostic Assessment: TGS-2 (19) ASTM D4169 (20)

Aspect	Testing requirements	Notes on testing requirements	Source documents
1.10.2 In-use stability (open pack or open vial stability)	1. There shall be evidence that once the RDT is removed from its primary packaging, it is stable at the expected temperature and humidity ranges for a defined period of time at the beginning and end of its assigned shelf-life; 2. Testing shall include all labile components (e.g. buffers vials, etc.) (see note 1); 3. Liquid components, once opened, shall have a validated life and number of stated uses under environmental (including microbial) conditions expected. 4. Minimum of 1 lot, using panel below: • a sufficient number of non-reactive (see 1.10.1 note 1); • at least 1 specimen for each analyte and each epitope used or detected by the RDT (approx. 1 - 2x RDT's LOD).	1. In-use stability of all components shall be conducted using components stored in their final configuration. 2. Statistically designed experiments should be involved to allow evaluation of any interactions between environmental conditions 3. Most aspects of in-use stability may be considered as part of "flex" studies (see 1.9.1 Flex studies).	

Part 2 Establishing clinical evidence (clinical performance characteristics)

Aspect	Testing requirements	Notes on testing requirements	Source documents
2.1 Diagnostic sensitivity and specificity			
2.1.1 Mixed titre panels	Testing of specimen panels with a range of analyte concentrations (e.g. non-treponemal and *T. pallidum* antibody 'mixed titre' panels).		
2.1.2 Performance panels	Testing of the RDT shall be undertaken using, as appropriate: • suitable performance panels which include all claimed critical epitopes, as available (see note 1); • with a minimum of 1 lot.	1. Each epitope plays a role in detecting syphilis in different stages of the infection. It is necessary to monitor each epitope system present (and each stage of infection), even if polyepitope-fusion proteins are used. • This may be done using specimens characterized by a line immunoassay.	
2.1.3 Diagnostic sensitivity and specificity study general requirements	Diagnostic sensitivity and specificity shall be determined for each claimed specimen type. Testing shall be conducted: • in different geographical settings (minimum of 2 regions) including high and low prevalence settings; • by a variety of intended users (see note 1) in the intended testing settings (e.g. decentralized at point of care use, laboratories in hospital setting); • using at least 2 different lots (see note 3).	1. Prequalified RDTs are generally used by lay providers. For WHO prequalification purposes, these should be considered as the intended user rather than a trained laboratory professional. 2. The performance shall be evaluated on the population in the environment of expected use in resource limited settings. 3. Each of the two lots shall comprise different production (or manufacturing, purification, etc.) runs of critical reagents. 4. Criteria for the selection of archived specimens shall be explained. 5. An effort should be made to include low RPR titre/early stage specimens. 6. Negative and positive archived specimens should be blinded to the user. 7. A separate specimen shall be collected prior to testing to establish the reference result. The laboratory where the clinical evaluation is occurring shall confirm all reactive specimens by the following methods: • Non-treponemal positive specimens confirmed with RPR; • Treponemal positive specimens screened with CLIA or EIA	European Commission decision on CTS (7)
2.1.4 Diagnostic sensitivity	Testing of at least 500 confirmed positive specimens: • consolidation of results from archived specimen collections and clinical evaluation studies is permissible (note 4); • however, at least 50% of the results from which the diagnostic sensitivity is calculated must be from freshly taken, unfrozen routine specimens of the types claimed (e.g. capillary blood, venous blood, serum, plasma); • at least 50 pregnant women; • at least 50 specimens from a low prevalence setting.		

Aspect	Testing requirements	Notes on testing requirements	Source documents
		and if positive, confirmed with TPPA.	
2.1.5 Diagnostic specificity	The specimens for specificity studies shall be: • as per requirements in 2.1.3; • unselected, other than as being syphilis negative; • archived specimens shall not exceed 20%; • if the RDT claim is for diagnostic use, blood bank specimens will be insufficient – the expected environment would be STI clinics or POC settings. At least 1000 specimens shall be tested. • At least 2 different lots (and ideally three) shall be used; • At least 100 pregnant women (to include at least 20 multiparous women).	8. Specimens with results discrepant between the confirmed laboratory result and the RDT under evaluation should be further evaluated: • For *T. pallidum* assays: a state of the art IgG and IgM syphilis immunoassay (LIA/EIA), with positive specimens further evaluated using TPPA/TPHA (treponema pallidum haemagglutination assay); • For non-treponemal assays: at least an IgM anti-treponemal immunoassay and a blot-type assay; • All discrepant specimens shall be repeated using the same lot of RDT, and then on all available lots and any variability noted; • Characterization of the donor of a specimen is acceptable evidence in the case of primary syphilis (e.g. detection of organisms by dark field microscopy). 9. Problematic specimens, those with unexpected results but which otherwise meet selection criteria for a study, shall not be excluded from analysis. 10. Performance characteristics shall be reported using initial results only. The results of further testing of specimens with discrepant results shall be reported separately as additional information about RDT performance. 11. All invalid test results and indeterminate results shall be recorded and reported as the invalid rate. 12. Estimates of diagnostic/clinical sensitivity and specificity shall be reported with 95% confidence intervals. 13. Results shall be expressed separately for each specimen type and for each specimen type per intended use (no aggregation of results).	
2.2 Qualification of usability for RDT IVDs			
2.2.1 Labelling	Testing shall be undertaken to assess the ability of intended users to correctly comprehend key messages from	1. If the labelling is available in different languages, the labelling comprehension study should be performed for	European Parliament IVD regulations

Aspect	Testing requirements	Notes on testing requirements	Source documents
comprehension study (including IFU)	packaging and labelling: • understanding key warnings, limitations and/or restrictions; • proper test procedure; • test result interpretation; • using only the information available to all users (IFU and any job aid). Studies shall include: • at least 15 intended users including those whose native language may not be the language of the IFU if necessary; • in their usual working environment, not employees of the manufacturer; • from 2 geographically diverse populations to demonstrate comprehension of key messages in each user group.	each language. 2. Requirements listed may be investigated as separate studies or included as part of clinical studies. 3. Testing may be conducted using questionnaire based surveys.	(21) U.S. FDA CLIA Waiver guidance (15, 16)
2.2.2 Results interpretation study	Intended users shall interpret the results of contrived RDTs (e.g. static/pre-made tests) to assess their ability to correctly interpret pre-determined test results. Contrived tests shall be made to demonstrate the following potential test results: • non-reactive; • range of invalid results; • reactive; • weak reactive. Testing subjects shall consist of: • at least 15 intended users, including those whose native language may not be the IFU language; • in their usual working environment, not employees of the manufacturer; • from 2 geographically diverse populations to demonstrate correct interpretation of simulated test results.		

E. Annex 1: List of conditions that should be evaluated for cross-reactivity.

Viral infections
- HIV
- hepatitis B infection
- hepatitis C infection
- acute hepatitis A infection
- acute cytomegalovirus (IgM)
- acute Epstein-Barr virus (IgM)

Sexually transmitted infections
- herpes simplex virus 2
- Chlamydia trachomatis
- human papillomavirus
- trichomoniasis

Infections caused by genus *Borrelia* (lymes disease)

Bacteria/parasites
- malaria
- visceral leishmaniasis
- tuberculosis
- brucellosis
- leptospirosis
- leprosy

Immunization
- influenza vaccine recipient
- vaccine-induced HIV seropositivity

F. Source documents

1. Technical Guidance Series for WHO Prequalification – Diagnostic Assessment. Principles for Performance studies, TGS–3. Geneva: World Health Organization; 2016 http://apps.who.int/iris/bitstream/10665/258985/1/WHO-EMP-RHT-PQT-TGS3-2017.03-eng.pdf?ua=1 accessed 03 December 2018.
2. WHO Prequalification – Diagnostic Assessment: Instructions for Compilation of a Product Dossier. WHO PQDx_018 v3, 27 August 2014. Geneva: World Health Organization; 2014 (http://www.who.int/entity/diagnostics_laboratory/evaluations/141015_pqdx_018_dossier_instructions_v4.pdf?ua=1 accessed 03 December 2018.
3. ISO 15198:2004 Clinical laboratory medicine -- In vitro diagnostic medical devices -- Validation of user quality control procedures by the manufacturer. Geneva: International Organization for Standardization; 2004.
4. ISO 17511:2003 In vitro diagnostic medical devices -- Measurement of quantities in biological samples -- Metrological traceability of values assigned to calibrators and control materials. Geneva: International Organization for Standardization; 2003.
5. EN 13612:2002. Performance evaluation of in vitro diagnostic medical devices. Brussels: European Committee for Standardization; 2002.
6. User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition. CLSI EP12-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
7. 2009/886/EC: Commission Decision of 27 November 2009 amending Decision 2002/364/EC on common technical specifications for in vitro diagnostic medical devices (notified under document C(2009) 9464) (Text with EEA relevance), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32009D0886;
8. User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition. CLSI EP17-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
9. WHO Technical Report Series, No. 1004, 2017. Annex 6. WHO manual for the preparation of secondary reference materials for in vitro diagnostic assays designed for infectious disease nucleic acid or antigen detection: calibration to WHO International Standards. Geneva, World Health Organization; 2016.
10. Interference Testing in Clinical Chemistry. Approved Guideline - Second Edition. CLSI EP07-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.
11. Supplemental Tables for Interference Testing in Clinical Chemistry, 1st Edition. CLSI EP37. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.
12. ISO 14971:2007 Medical devices -- Application of risk management to medical devices. Geneva: International Organization for Standardization; 2007.
13. Anderson MD, Kennedy CA, Lewis AW. Retrobulbar neuritis complicating acute Epstein-Barr virus infection. Clin Infect Dis. 1994 May;18(5):799-801.
14. Butch AW. Dilution protocols for detection of hook effects/prozone phenomenon. Clin Chem. 2000; 46:1719-21.
15. U.S. Food and Drug Administration Center for Devices and Radiological Health. Recommendations for Dual 510(k) and CLIA Waiver by Application Studies. Draft Guidance for Industry and Food and Drug Administration Staff. Issued November 29, 2018.
16. U.S. Food and Drug Administration Center for Devices and Radiological Health. Select Updates for Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices. Draft Guidance for Industry and Food and Drug Administration Staff. Issued November 29, 2018.
17. ISO 23640:2011. In vitro diagnostic medical devices - Evaluation of stability of in vitro diagnostic reagents. Geneva: International Organization for Standardization; 2011.

18. Evaluation of Stability of In Vitro Diagnostic Reagents; Approved Guideline. CLSI EP25-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2009.

19. Technical Guidance Series for WHO Prequalification – Diagnostic Assessment. Establishing stability of an in vitro diagnostic for WHO Prequalification, TGS–2. Geneva: World Health Organization; 2016 http://apps.who.int/iris/bitstream/handle/10665/259742/WHO-BS-2017.2304-eng.pdf?sequence=1 accessed 20 November 2018.

20. Standard Practice for Performance Testing of Shipping Containers and Systems. ASTM D4169-14. West Conshohocken, PA: ASTM International; 2014.

21. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in-vitro diagnostic medical devices. O. J. E. U. 2017 L 117/176 http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0746&from=DE accessed 30 November 2018.

List of related WHO Publications of related interest

WHO Prequalification Team- Diagnostic Assessment. Technical Guidance Series for WHO Prequalification – Diagnostic assessment (available online)

WHO Prequalification Team – Diagnostic Assessment. Instructions for Compilation of a Product Dossier. WHO document WHO/PQDx_18 (available online)

WHO Human Reproduction team. Point-Of-Care Diagnostic Tests (POCTs) for Sexually Transmitted Infections (STIs)Target product Profiles (available online)

WHO Human Reproduction team. Point-Of-Care Diagnostic Tests (POCTs) for Sexually Transmitted Infections (STIs)Clinic-based Evaluation (available online)

The Technical Specifications Series for submission to WHO Prequalification – Diagnostic Assessment set out appropriate performance evaluation criteria to meet prequalification requirements. Each Technical Specification provides information on the minimum performance requirements for WHO Prequalification that should be met by a manufacturer to ensure that their in vitro diagnostic medical device is safe and performs optimally.

Further information on these and other WHO publications can be obtained from WHO Press, World Health Organization, 1211Geneva 27, Switzerland

http:/www.who.int/bookorders

tel. : +41 22 791 3264; fax: +41 22791 4857; email bookorders@who.int

ISBN 978-92-4-151516-0