Diagnostics Evaluation Series
No.4

Visceral Leishmaniasis Rapid Diagnostic Test Performance

World Health Organization

TDR For research on diseases of poverty
UNICEF • UNDP • World Bank • WHO

WHO Library Cataloguing-in-Publication Data

Visceral leishmaniasis rapid diagnostic test performance.

1.Leishmaniasis, Visceral - diagnosis. 2.Leishmaniasis, Visceral - prevention and control. 3.Antibodies, Protozoan - analysis. 4.Reagent strips - diagnostic use. 5.Immunologic tests -utilization. 6.Sensitivity and specificity. I.World Health Organization. II.UNICEF/UNDP/World Bank/WHO Special
Programme for Research and Training in Tropical Diseases.

ISBN 978 92 4 150223 8 **(NLM classification: WC 175)**

Web version only
Cover design: Lisa Schwarb

Acknowledgements

The evaluation described in this report was a joint project of WHO/TDR and nine test evaluation centres around the world. The project was financed by TDR with technical support from the Prince-Leopold Institute of Tropical Medicine (ITM), Belgium and the Royal Tropical Institute (KIT), Netherlands. The report was prepared by Emily Adams (KIT), Epco Hasker (ITM) and Jane Cunningham (WHO/TDR).

WHO/TDR wishes to acknowledge the efforts of lead investigators (*) and the staff at all the participating evaluation centres:

Rajendra Memorial Research Institute of Medical Sciences, India	Pradeep Das* S. Bimal A.K. Gupta
Laboratório de Soroepidemiologia e Imunobiologia Instituto de Medicina Tropical de São Paulo, Brazil	Hiro Goto* Beatriz Julieta Celeste Jose Angelo Lauletta Lindoso Anamaria Mello Miranda Paniago Maria Carmen Arroyo Sanchez
Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Fiocruz, Brazil	Ana Rabello* Tália Machado de Assis Edward Oliveira Dorcas Lamounier Costa[a]
B P Koirala Institute of Health Sciences, Nepal	Suman Rijal* Basudha Khanal Murari L Das
Kenya Medical Research Institute (KEMRI), Kenya	Margaret Mbuchi* Monique Wasunna* Charles G. Magiri Alfred M. Muia
Parasitology Laboratory, International Centre for Diarrhoeal Disease Research, Bangladesh	Dinesh Mondal* Khondaker Rifat Hasan Bhaskar Mamun Huda Mukidul Hassan Afrad
University of Khartoum, Sudan	Sayda El Safi* Al Farazdag Saad Nuha Galal Osman Salih Osman Awad Hammad Ahmed Al Mustafa Bashir
Institute Endemic Diseases, University of Khartoum, Sudan	Maowia Mukhtar* Sahar M. Bakhiet Mohamed Osman Hussien Aymen Awad
Institute of Medical Sciences, Banaras Hindu University, India	Shyam Sundar* Dinesh Kumar Prajapati Kamlesh Gidwani Puja Tiwary

[a] Maternal and Childhood Department, Federal University of Piauí
WHO/TDR is grateful for the cooperation exhibited by the manufacturers of VL rapid diagnostic tests who agreed to the external and independent evaluation of their products.

WHO/TDR also wishes to acknowledge all those who contributed to protocol development, site assessments, site monitoring, conduct of the evaluation, data analysis and/or preparation of this report:

Olivia Achonduh	University of Yaoundé , Cameroon
Emily Adams	Royal Tropical Institute (KIT), The Netherlands
Marleen Boelaert	Institute of Tropical Medicine (ITM), Belgium
Carmen Canavate	WHO Collaborating Centre for Leishmaniasis, Spain
Jane Cunningham	Special Programme for Research and Training in Tropical Diseases (TDR), Switzerland
Varalakshmi Elango	Consultant (GCLP, Monitoring)
Hashim Ghalib	Special Programme for Research and Training in Tropical Diseases (TDR), Switzerland
Kamlesh Gidwani	Banaras Hindu University (BHU), India
Epco Hasker	Institute of Tropical Medicine (ITM), Belgium
Diane Jacquet	Institute of Tropical Medicine (ITM), Belgium
Joris Menten	Institute of Tropical Medicine (ITM), Belgium
Francis Moussy	Special Programme for Research and Training in Tropical Diseases (TDR), Switzerland
Rosanna Peeling	London School of Hygiene and Tropical Medicine, United Kingdom
Nidia Rizzo	Universidad del Valle de Guatemala, Guatemala
Gerard Schoone	Royal Tropical Institute (KIT), The Netherlands

Abbreviations

BHU	Banaras Hindu University
CL	Cutaneous leishmaniasis
DAT	Direct Agglutination Test
EA	East Africa
GCLP	Good Clinical Laboratory Practice
GCP	Good Clinical Practice
HEC	Healthy endemic control
HIV	Human immunodeficiency virus
ICT	Immunochromatographic tests
ID	Identification number
ISC	Indian subcontinent
ITM	Institute Tropical Medicine
KIT	Royal Tropical Institute
LN	Lymph node
QA	Quality assurance
RDT	Rapid diagnostic test
rK 39	Recombinant antigen 39
rKE 26	Recombinant antigen 26
RFA	Request for applications
SA	South America
SOP	Standard Operating Procedure
TDR	Special Programme for Research and Training in Tropical Diseases
VL	Visceral Leishmaniasis
VL-LN	Visceral Leishmaniasis Laboratory Network
WHO	World Health Organization

Contents

Annexes

Executive Summary

Visceral Leishmaniasis (VL) is one of the world's most neglected diseases, affecting the poorest people in developing countries. Some 500,000 new cases occur annually with 90% of all VL cases in the Indian subcontinent (Bangladesh, India, Nepal), Sudan, Ethiopia and Brazil. Leishmaniasis is a poverty-related disease associated with malnutrition, weakness of the immune system, displacement, poor housing, illiteracy, gender discrimination, and lack of resources.

Up until the 1990s, accurate VL diagnosis necessitated parasitological confirmation by microscopy or culture of the blood, bone-marrow, lymph nodes or spleen. The invasiveness and potentially fatal complications associated with splenic aspiration motivated the development of non-invasive serological tests such as direct agglutination test (DAT) and immunochromatographic lateral-flow assays, commonly referred to as rapid diagnostic tests (RDT).

Enthusiasm and rapid uptake of RDTs for VL, particularly in the Asian region, has translated into the emergence of several commercialized tests targeting serum antibodies to rK39 and other antigens, e.g. rKE 16. However, there are few head-to-head evaluations of diagnostic accuracy for these tests. Therefore, in collaboration with clinicians and laboratory scientists in endemic regions, TDR coordinated a multicentre and multiregional comparative evaluation of commercially available antibody-detecting RDTs to inform country policy. This report describes the performance of four commercially available rK39 and rKE16 antibody detecting RDTs in three global regions of VL endemicity (Indian subcontinent (ISC), East Africa (EA), South America (SA)) using well characterized panels of human sera; a fifth RDT was evaluated in the ISC.

VL RDT Evaluation Programme

All companies manufacturing RDTs for VL under license ISO-13485 Quality System Standard or US FDA 21 CFR were invited to submit tests for evaluation; 4 responded to the call contributing 5 products in total. This report describes the performance of these five commercially available rK39 and rKE16 antibody detecting RDTs. Nine laboratories were chosen in three global endemic regions; namely ISC (n=4), South America (n=2) and Eastern Africa (n=3) (Table 1). Each region collected archived sera from 250 confirmed VL cases, 210 healthy endemic controls and 40 cases of other regionally relevant disease conditions for the evaluation panel.

Before commencing product testing, a proficiency panel of 20 well-characterized sera was assembled and sent to each evaluation laboratory. The proficiency panel was tested blindly in evaluation laboratories using the DAT assay. The DAT was subsequently used to re-validate all samples prior to their inclusion in the evaluation panel.

RDTs were tested against the evaluation panel for clinical accuracy, ease of use, thermal stability and reproducibility.

Results of the evaluation

The results of the evaluation are summarized in Figures 2-4 and Table 4. Accuracy of RDTs between centres was highly comparable; however, results between global regions were significantly different. In general, against ISC panel samples, commercially available RDTs have high sensitivity and specificity, good reproducibility and most are heat stable. However, RDT sensitivity is more variable in sample panels from East Africa and Brazil. In Brazil and in two sites in East Africa, the rKE16 based products appeared to perform less well than rK39 products. Specificity in Brazil and East Africa of most tests remains high.

In all regions, agreement between lots or batches of the same product was good to excellent. Further, operator-to-operator and run-to-run reproducibility was very good. Limitations of this study include the use of retrospectively collected serum samples from a series of case and control patients which do not necessarily represent the true 'VL suspect' population. Potentially, this may lead to an overestimation of clinical accuracy of the RDTs. Furthermore, results cannot be extrapolated to RDT performance with whole blood or in field conditions, where ease of use may significantly influence test performance.

Use of these results

This evaluation was designed to provide comparative data on the performance of submitted lots of each RDT tested. Ultimately, the results indicate that in clinical practice in the East African and South American region used alone VL RDT positive results may be adequate to direct treatment but inadequate to rule out a diagnosis of VL. This highlights the need to establish limits of acceptable performance and to implement RDTs within a diagnostic algorithm as appropriate for each global endemic area. In the ISC, all brands performed well.

This performance data can be used to guide priority setting for field trials and/or procurement decisions. The final decision on product selection needs to be taken in a rational way, considering not only the minimal performance limits, but also the global endemic region, patient characteristics, experience of the intended users, climate and cost.

1. Introduction

Visceral Leishmaniasis (VL), also known as Kala azar, is a vector-borne parasitic disease which is nearly always fatal if left untreated. Protozoa of the leishmania complex cause an obligate intra-macrophage infection. The clinical syndrome is characterized by fever, weight loss, splenomegaly, lymphadenopathy and hepatomegaly. VL is one of the world's most neglected diseases, largely affecting the poorest people, mainly in developing countries. VL is endemic in 60 countries and some 500,000 new cases occur annually, though 90% of all reported cases occur in just 6 countries: Bangladesh, Brazil, Ethiopia, India, Nepal and Sudan [*1*]. The disease is transmitted through the bite of an infected *phlebotomine* sandfly. The host reservoir varies in different parts of the world, reflecting a difference in parasite strains. Whereas in Europe and South America the causative organism agent is *Leishmania infantum*, which has the domestic dog as its main reservoir, in the Indian subcontinent (ISC) (Bangladesh, India, Nepal) and in East Africa, the causative organism is *L. donovani*, which is essentially a parasite of humans with anthroponotic transmission patterns. In South America, the causative agent is *L. infantum* (previously referred to as *L. chagasi*) [*2*].

Up until the 1990s, accurate VL diagnosis necessitated parasitological confirmation by microscopy or culture of the blood, bone-marrow, lymph nodes or spleen. Microscopic detection of parasites in clinical material from the spleen is still considered the reference standard; however, splenic aspirates are associated with risks of serious bleeding and should only be carried out in settings with access to blood transfusion and surgical services. The invasiveness and potentially fatal complications associated with splenic aspiration has motivated the development of non-invasive serological tests such as direct agglutination test (DAT) [*3*] and lateral flow immuno-chromatographic tests (ICT), commonly referred to as rapid diagnostic tests (RDT). The first ICT was based on a 39-amino-acid-repeat recombinant leishmanial antigen from *L. chagasi* (rK39) and was seen as a potential breakthrough allowing user and patient friendly, rapid diagnosis in peripheral health care settings. The initial validation study of an rK39 ICT (Arista Biologicals, Allentown, PA, USA) reported 100% sensitivity and 98% specificity in an Indian setting when combined with a strict clinical case definition [*4*]. However, an RDT from the same manufacturer evaluated in Sudan showed only 67% sensitivity [*5*]. In 2006 Chappuis et al. published a meta analysis on the performance of the DAT and the rK39 RDTs for VL [*6*]. This meta analysis, when combined with the results of a TDR coordinated multicentre evaluation [*7*], formed the evidence base that corroborated the diagnostic accuracy of the rK39 RDTs when combined with the WHO clinical case definition for VL; this led to its adoption as a first-line diagnostic test in the VL Elimination Initiative of the

ISC. Due to several reports of lower test sensitivity in other endemic regions such as East Africa [*5,8,9*], a negative rK39 RDT result may not reliably exclude a diagnosis of VL in this region and additional serological or parasitological tests are necessary. Further-more, in all settings rK39 RDTs must only be used for patients without a history of past VL and who meet the WHO clinical case definition, as antibodies will persist after treatment and can also be found in asymptomatic individuals living in endemic areas.

Overall, the enthusiasm and rapid uptake of RDTs for VL, particularly in the Asian region, has translated into the emergence of several commercialized tests targeting serum antibodies to rK39 and other antigens, e.g. rKE16[1]. However, there are few, if any, reports of RDT performance in the peer reviewed literature for tests other than Kalazar Detect™ Inbios International, Inc and DiaMed-IT LEISH (Bio-Rad Laboratories) and equally few head-to-head comparisons of performance. Therefore, in collaboration with clinicians and laboratory scientists in endemic regions, TDR decided to coordinate a multicentre and multi-regional comparative evaluation of commercially available antibody-detecting RDT.

2 Objectives

- To perform a head-to-head evaluation of commercially available antibody-detecting immuno-chromatographic tests (ICTs) or rapid diagnostic tests (RDTs) for the diagnosis of visceral leishmaniasis (VL) using archived serum samples in three global endemic regions.
- To assess the operational characteristics of the commercially available RDTs including ease of use and thermal stability.
- To produce performance data to guide UN agency, NGO and national government procurement of VL RDTs.
- To create an international inventory of well-characterized clinical materials derived from VL patients and non-VL endemic controls to support diagnostic development, evaluation and quality control.

[1] rKE16 tests are based on a recombinant antigen (*L.d.*-rKE16) from a newly isolated Indian strain of *L. donovani* (MHOM/IN/KE16/1998).

3 Evaluation methodology

3.1 Selection of evaluation centres

TDR issued a public request for applications (RFA) for any laboratory interested in becoming a member of a laboratory network to perform the VL-RDT evaluation. Twenty-two responses were received and applicants were ranked by an independent review committee according to several criteria including: access to patients, geographical location, laboratory facilities, expertise and experience with RDTs. Nine laboratories were chosen in three global endemic regions; namely Indian subcontinent (n=4), South America (n=2) and Eastern Africa (n=3), (for full details of participating laboratories see Table 1).

The laboratory at the Prince-Leopold Institute of Tropical Medicine, Antwerp, Belgium (ITM) was contracted as an independent partner to coordinate logistics of RDT supplies, as well as proficiency testing in the direct agglutination test (DAT) for VL at all laboratories prior to implementation of the study protocol. When required, independent consultants from The Royal Tropical Institute, The Netherlands and Banaras Hindu University (BHU), India facilitated the standardization of laboratory procedures (mainly DAT) and subsequent repeat proficiency testing and re-validation of panel samples at several evaluation laboratories. The 9 participating laboratories, together with experts from WHO/TDR, ITM and KIT formed the TDR Visceral Leishmaniasis Laboratory Network (VL-LN).

Table 1: Evaluation centres

Region	Country	Institution
East Africa	Sudan	University of Khartoum
East Africa	Sudan	Institute Endemic Diseases, University of Khartoum
East Africa	Kenya	Kenya Medical Research Institute
South America	Brazil	Instituto de Medicina Tropical de São Paulo
South America	Brazil	Centro de Pesquisas Réné Rachou , Fiocruz
Indian subcontinent	India	Rajendra Memorial Research Institute of Medical Sciences
Indian subcontinent	India	Institute of Medical Sciences, Banaras Hindu University
Indian subcontinent	Nepal	B P Koirala Institute of Health Sciences
Indian subcontinent	Bangladesh	International Centre for Diarrhoeal Disease Research

3.2 Test Selection

The TDR VL-LN developed a set of RDT operational characteristics required for inclusion of an RDT in the evaluation. To be considered as an RDT, the test should be:

- **rapid**: with a test result available in 15 minutes
- **simple**: test can be performed after minimal training and equipment
- **easy to interpret**: cassette or strip format with visual readout.

Requirements also included evidence of quality manufacturing either ISO 13485:2003 or US FDA 21 CFR certification, supply of sufficient quantities of products (Lot 1: 1,849; Lot2: 1,794 [2]) and a signed Confidentiality Agreement with the WHO permitting the publication of results in the public domain.

In March 2009, the evaluation process was initiated with an open call for expression of interest to companies that manufacture and sell tests that fit the above inclusion criteria. The expression of interest was advertised on the TDR website and distributed via email to European, North American, Indian and South American manufacturers associations, companies advertising commercially available tests, a TDR scientist mailing list and TDR Steering Committee members. Three manufacturers with a total of four commercially available products responded to the letter of interest prior to the deadline. A fourth manufacturer expressed interest after the deadline and a decision was taken by the TDR Diagnostics Steering Committee to allow participation of this test in the Indian subcontinent evaluation centres as results could immediately inform procurement practices in the region.

Three of the five products targeted detection of antibodies to rK39 and two products detect antibodies to rKE16. The manufacturer, product and catalogue number are described in Table 2, and Table A1.1 provides a comprehensive overview of test characteristics [3].

Table 2: Manufacturers and products accepted for evaluation

Product Name	Manufacturer	Catalogue Number	Bound Antigen
DiaMed-IT LEISH	Bio-Rad Laboratories	46240	rK39
Crystal®KA	Span Diagnostics Ltd.	56IC102-25[a]	rKE16
Signal®-KA[b]	Span Diagnostics Ltd.	56FT100-050 [a]	rKE16
Kalazar Detect™	InBios International Inc.	INS025	rK39
Onsite Leishmania Ab Rapid Test[c]	CTK Biotech,Inc.	R0122S	rK39

[a] These products may include different catalogue numbers for different box sizes, contact manufacturer for details.
[b] In the ISC Signal® KA, which is recommended for storage between 2-8°C, was excluded from evaluation in the two Indian centres because of prolonged exposure to temperatures exceeding 30°C during shipping.
[c] Evaluated only in Indian subcontinent evaluation centres.

[2] Exception for *Onsite* Leishmania Ab Rapid Test-Strip (CTK Biotech, Inc.) who provided Lot 1: 623 and Lot 2: 598 tests.

[3] In 2007, DiaMed GmbH was acquired by Bio-Rad Laboratories'

In accordance with the confidentiality agreement, companies were able to review their product(s) data for a period of 60 days prior to publication of results. However, they were not permitted to make any changes to the data or to modify the conclusions of the report.

3.3 Sample size calculations

Each rapid test was evaluated at all nine laboratories using a locally assembled evaluation panel (Section 4.4) and results were pooled regionally, except for the Onsite Leishmania AB Rapid test –Strip (CTK Biotech, Inc.) that was only evaluated in the ISC. To demonstrate with a power of 80% that the test under evaluation is at least 90% sensitive, assuming that its true sensitivity is 95%, a sample size of 250 cases was required. As specificity was expected to be ~98% based upon Boelaert, et al. (2007) a healthy endemic control (HEC) size of 210 was required to achieve a power of 90% [7]. Additionally, 40 controls with other disease conditions that can mimic VL were selected appropriately for each region, for example patients with confirmed *Plasmodium* sp., *M. tuberculosis*, Chagas disease or cutaneous leishmaniasis (CL).

3.4 Sample selection

In total, archived sera from 250 confirmed VL cases, 210 healthy endemic controls and 40 cases of other regionally relevant disease conditions were assembled per region[4].

Confirmed VL cases were defined as symptomatic individuals with positive microscopy on splenic, lymph node **or** bone marrow aspiration. Samples were prioritized based on the following factors:

- A minimum volume of 220µl was required to complete all tests under evaluation.
- Distribution of DAT titres within samples (20% medium to low <6,400, 80% high >6,400).
- Healthy controls from endemic area and DAT titres < 3,200.
- Most recent date of sample collection and least freeze/thaw cycles.
- Known HIV status.

3.5 Blinding

Selected samples were added to an Epi Info database and assigned random identification numbers.

In order that samples were tested blindly in the centres, evaluation samples, after inclusion by DAT, were assigned random ID's by a laboratory technician who was not involved in the study.

[4] Except in Brazil, where the number of confirmed sera was between 237-250 and controls between 2006-209. Signal ®-KA (Span Diagnostics Ltd.) was not tested in India, therefore, cases tested were 175 and HEC totaling 170. In the ISC 249 controls sera were tested.

4 Evaluation Site Preparation

4.1 Ethical considerations

Each laboratory participating in the evaluation obtained approval from the local ethics board and the WHO Research Ethics Committee. If informed consents for previous research projects, through which the archive samples were derived, did not explicitly request permission for future use of left-over samples, then a specific clearance from the local ethics committee was requested. This letter granted special permission for these biological samples to be used for the evaluation of RDTs, as is normal procedure for left over samples in the respective institute. Samples were unlinked to unique patient identifiers so sera could not be traced to individual patients.

4.2 Laboratory efficiency

All personnel involved in the evaluation took part in TDR sponsored Good Clinical Laboratory Practice (GCLP) regional training courses. The evaluation complied with national workplace bio-safety guidelines, including those related to the safety of laboratory personnel and to the disposal of infectious waste. A TDR clinical monitor visited each laboratory for assessment of compliance with GCLP.

4.3 Proficiency testing

A proficiency panel of 20 well-characterized sera was assembled, coded and aliquots were sent to each evaluation laboratory. The proficiency panel was tested blindly in evaluation laboratories using the DAT assay according to the Standard Operating Procedure (KIT, Netherlands) and using the same batch of DAT antigen and microtitre plates. Results were returned to ITM and decoded. All evaluation sites reporting titre discrepancies >1 titre received on-site refresher training. When all sites read within 1 titre of the reference laboratories they permitted to proceed with study preparations. Subsequently, DAT was used to revalidate all potential evaluation panel samples prior to their inclusion (see section 4.4).

4.4 Evaluation panel composition

Each laboratory was asked to inventory their archived clinical materials and enter data in a common database. Patient and sample characteristics[5] were recorded and the evaluation panel was derived based on a set of prioritized criteria (Section 3.4).

HECs were defined as healthy subjects living in a transmission area of leishmania infection, with DAT serology <1:3,200. Samples from patients with diseases that could potentially cross-react with VL serology were included in the evaluation panel to assess specificity of the RDTs. Serum samples of cases with malaria, Chagas disease, tuberculosis and CL were included in each endemic area as geographically relevant.

All evaluation panel samples were re-validated with the DAT to ensure storage conditions had not affected sample quality and to accurately record the DAT titre as a semi-quantitative measure of total antibody response.

Evaluation panel samples were stored at -70°C until the time of RDT testing.

[5] Sample type, country of origin, history of past VL, duration of fever, baseline (and repeat) laboratory investigations, e.g. type and results of biopsy (lymph node, bone marrow, spleen) microscopy, culture, DAT, rk39 RDT, rk39ELISA, HIV, sample volume.

5 The Evaluation

5.1 Preparatory phase

5.1.1 RDT shipments

All manufacturers shipped tests from two lots to ITM, Antwerp. In Antwerp, the RDTs were repackaged in the required amounts for courier shipment to each of the nine laboratories participating in the evaluation. Crates of RDTs were shipped by courier with at least one temperature monitoring device and RDTs requiring refrigeration were shipped with cooling agents and insulating packaging.

For reasons beyond the control of TDR, shipments destined for two evaluation centres in India were delayed and temperature log data revealed that all RDTs [6] in the shipment were subjected to temperatures exceeding manufacturers' recommendations (>30°C) for a period of several weeks. An ad-hoc meeting of the VL-LN advisory group recommended the exclusion of Signal®-KA (Span Diagnostics) from the two Indian evaluation centres, because of its requirement for storage between 2-8°C. All other tests were included at both centres.

5.1.2 Registration and storage at testing sites

Upon arrival, tests were immediately unpacked and stored according to the manufacturers' instructions. Date of arrival and condition of shipment were noted; temperature log data from transport of the RDTs was forwarded to TDR for downloading. Daily temperatures were taken from the storage area of the RDTs with electronic temperature recorders until the day of testing.

5.2 Performing Rapid Diagnostic Tests

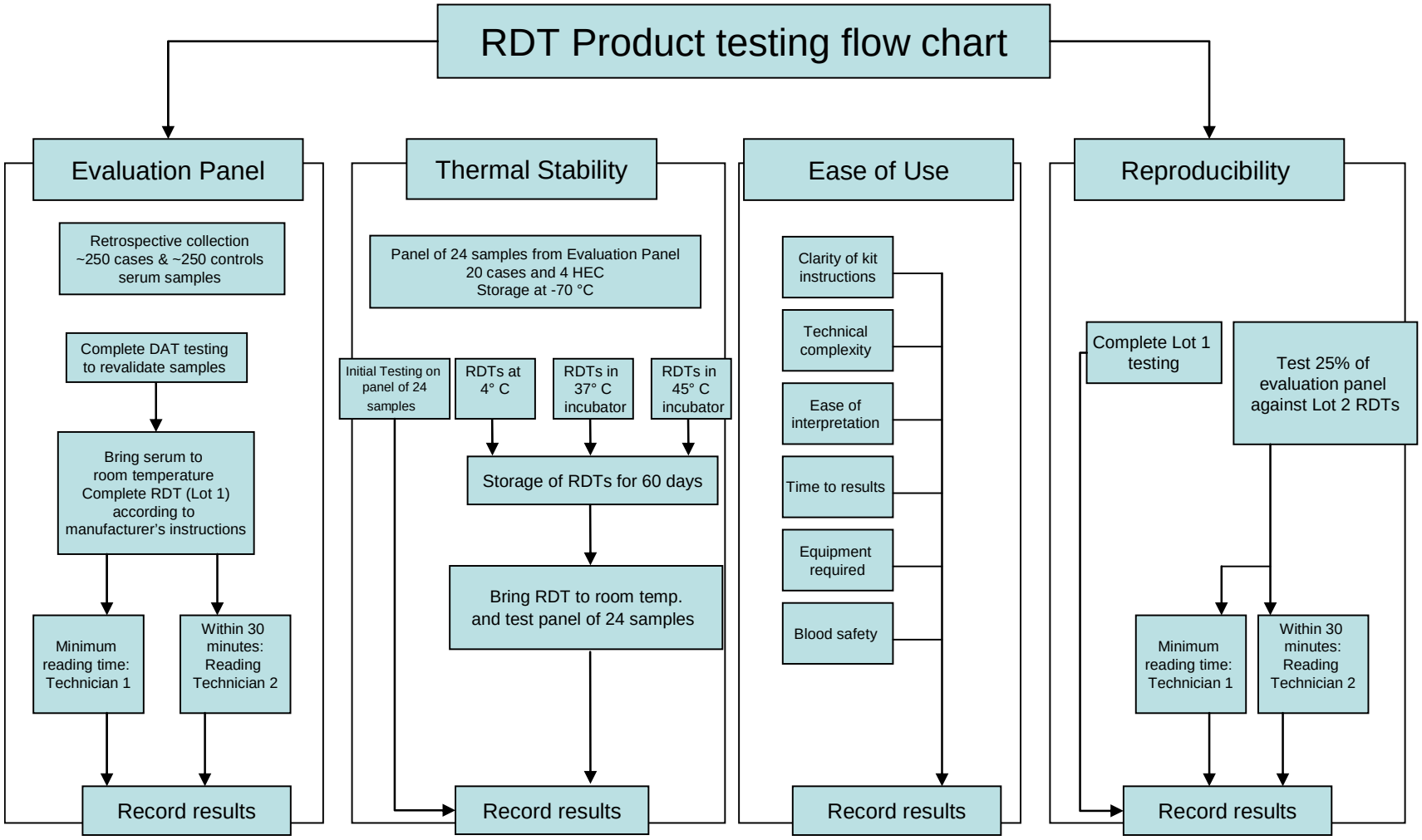
Figure 1 provides an overview of the VL RDT evaluation process.

5.2.1 Evaluation panel

RDTs were tested against an evaluation panel comprising approximately 250 cases and 250 controls (see section 4.4). RDTs from 'Lot 1' were brought to room temperature prior to use and labelled with the random sample code. Serum samples were taken out of the -70 °C freezer and brought to room temperature prior to testing. RDT envelopes were only opened immediately before use. Using a micropipette, the volume of serum specified by the manufacturer was dispensed onto the RDT. The buffer was applied according to the manufacturers' instructions, using the dropper provided. Results were read by a first technician at the minimum reading time and within 30 minutes the test was read by a second technician, blinded to the reading of the first. Results of test and control lines were recorded as positive or negative by each technician. More specifically, band intensity was recorded for each test, based on standardized charts: 0-negative; 1- faint/weak positive line; 2 -positive line; 3- strong positive line. If the control line was recorded as absent by either technician, the same sample was tested once against a new RDT. If the control line was still absent, the test result was recorded as invalid.

[6] Except *Onsite* Leishmania Ab Rapid Test-Strip (CTK Biotech Inc.) which was shipped separately on different dates.

Figure 1: Visceral leishmaniasis RDT product testing overview



5.2.2 Thermal stability

In one laboratory per region, RDTs were evaluated for thermal stability on a panel of 24 serum samples (20 VL cases and 4 negative controls). RDTs were tested at day = 0 as a baseline measure; then stored in original packaging in calibrated incubators of 4°C, 37°C and 45°C and tested again on the same sample panel on day = 60. The panel included serum samples from 20 cases, including low DAT titre cases if available (<1:6,400) and 4 healthy endemic controls. Aliquots of sera used for the thermal stability panel were kept at -70°C until the day of testing. Thermal stability evaluation was performed only for those tests not requiring storage between 2-8°C; i.e. Crystal® KA, DiaMed-IT LEISH and Kalazar Detect™.

5.2.3 Ease of use description

Tests were assessed for ease of use by technicians who had become familiar with the devices.

Tests were assessed according to the following parameters:

- Clarity of kit instructions
- Technical complexity (number of steps)
- Ease of interpretation of results
- Time to results
- Equipment required (for example micropipettes and tips)
- Blood safety

The maximum score in this schedule is 18 which would indicate that the RDT kit has operational characteristics most suitable for use in the primary health care settings of resource-limited settings.

5.2.4 Reproducibility

Lot-to-lot reproducibility was assessed by re-testing 25% of the evaluation panel against a second lot (Lot 2) of RDTs as provided by the manufacturers. The smaller evaluation panel contained all groups of samples including cases of high and low DAT titre, healthy endemic controls and other disease conditions. Assessment took place after Lot 1 testing had been completed. Operator-to-operator and run-to-run variability was tested by two technicians independently performing the RDT using the same 8 samples and negative control on three consecutive days [7].

5.2.5 Modification to RDT procedures

Performance of RDTs was according to manufacturer's instructions with the exception that frozen sera instead of fresh sera were used. Furthermore, the result was recorded by a technician at the minimum specified reading time and a second technician read within 30 minutes.

5.3 Data management and analysis

Sensitivity and specificity estimates are based on compiled regional data. Test results were considered positive if a Lot 1 RDT was recorded as positive against a VL positive sample by the first technician. Double data entry was completed at each site and exported to TDR for subsequent data cleaning procedures and analysis using STATA software. All source documents and electronic records of study data are maintained in secure storage until conclusion of the evaluation, data analysis and report publication.

[7] Signal®-KA (Span Diagnostics Ltd.) was not assessed in ISC and *Onsite* Leishmania Ab Rapid Test-Strip (CTK Biotech Inc.) was only tested in ISC.

The diagnostic accuracy of each diagnostic test was assessed in classical 2x2 contingency tables. Sensitivity was computed as the number of specimens positive by Test X out of the total number of specimens of confirmed VL cases. Specificity was calculated as the number of specimens negative in Test X out of the total number of controls. Accuracy is presented as receiver operating characteristic (ROC) plots. Agreement was assessed on a binary scale using Cohen's Kappa coefficient [10]. Kappa coefficients were interpreted following Landis and Koch [11]: 1.00–0.81 excellent, 0.80–0.61 good, 0.60–0.41 moderate, 0.40–0.21 weak and 0.20–0.00 negligible agreement.

5.4 Quality Assurance

All assays were performed according to the manufacturer's package insert; any deviations are reported in section 5.2.5.

A TDR clinical monitor visited every laboratory at the time of initiating the evaluation phase to ensure study specific SOPs, principles of GCP and GCLP were being followed correctly.

Temperature monitoring of all incubators, fridges, freezers and during transport was performed by calibrated minimum/ maximum thermometers.

6 Results

6.1 Summary of results

Overall, 750 cases and 754 controls were evaluated across 3 endemic regions[8] (Table 3). All RDTs performed with high specificity (>95%) in all regions; however, sensitivity varied between tests (range: 36.8% -100%) and between regions. In ISC all tests had a high sensitivity (> 93%); in East Africa and Brazil sensitivity results were variable, but no test exceeded 92% sensitivity (95% confidence interval, 87.8-94.8%:). The main results are summarized in Table 4 and Figures 2-4.

6.2 Sample characteristics

6.2.1 Direct Agglutination Test

DAT titres among cases ranged from serum dilution 1:100 to >1:204,800. In Brazil, 12% of sera were ≤1:6,400 and 88% were >1:6,400. In East Africa, 9% of sera were ≤1:6,400 and 91% were >1:6,400; and in ISC only 3% of cases were ≤1:6,400 and 97% were >1:6,400.

DAT titres among healthy endemic control samples ranged from serum dilution ≤1:100 to 1:3200. In Brazil all DAT titres of controls were ≤1:100; in East Africa 99% of all DAT titres of controls were ≤1:800; whereas on the Indian subcontinent 82% of DAT titres of controls were ≤1:800.

6.2.2 HIV status

The distribution of known HIV status is presented in Table 3. Overall HIV status was known in less than half of the patients in ISC and East Africa and less than 2% in Brazil. There were no samples from HIV infected patients included in the panels from East Africa and Brazil, and only two patient samples included in the ISC.

6.2.3 Age of patients

The average age of cases and controls is presented in Table 3. In general cases and controls were older in the ISC compared to East Africa with Brazil between the two.

6.2.4 RDT performance: Sensitivity and specificity

Sensitivity and specificity of the commercial RDTs evaluated were variable between regions but were highly comparable intra-regionally. This strongly suggests that real differences in test performance exist between the three major geographical regions of VL. In the Indian subcontinent all tests performed well, however, only one test had an average regional sensitivity that exceeded 85% in both East Africa and Brazil 87.2% (82.5%- 90.8%) and 92% (87.8%- 94.8%), respectively.

[8] With notable exceptions: Signal®-KA (Span Diagnostics Ltd.) was not tested in India, therefore 675 cases and 675 controls were tested. *Onsite* Leishmania Ab Rapid Test-Strip (CTK Biotech Inc.) was tested exclusively in the ISC against 250 cases and 249 controls samples. In Brazil, the case number in the two evaluation centres ranged from 237-250 and 206-209 controls.

Sensitivity and specificity were calculated based upon pooled data per region per test. Specificity results based on all control samples have been combined as there were no major differences in specificity between the samples from healthy endemic controls and those from patients with disease conditions that clinically mimic VL, in any of the regions. Table 4 and Figures 2-4 illustrate average sensitivity and specificity for each of the tests in the three regions. Data show that in the ISC all tests performed well with four out of five tests having a sensitivity and specificity above 95%. In East Africa, three out of the four tests had a specificity >95%; however several of the tests had poor sensitivity ranging from: 36.8% (31.1-42.9%) to 87.2% (82.5%- 90.8%). Similarly, in Brazil all tests showed high specificity >95% but variable sensitivity 61.5% (55.1%- 67.4%) - 92% (87.8%- 94.8%). Sensitivity was notably poorer with the anti-rKE16 antibody based tests in Brazil and East Africa.

The performance of products that were exposed to temperatures exceeding 30°C during several weeks of delayed transport did not appear to be significantly affected by the heat stress in the Indian testing centres (See Tables A2.3a and A2.33b, Figures A2.3a and A2.3b).

Table 3: Performance evaluation panel characteristics

	Sample size			DAT titre – cases (%)			Average age of patients (yrs)		HIV status [a] (%)		
									unknown	known	
	Cases	HEC	Non-malaria disease condition controls	Low (≤100)	Medium (1:200 - 1:6,400	High (>1:6,400)	Case	controls		positive	negative
East Africa	250	210	40	0.4	8.8	90.8	14	20	47.2	0	52.8
Brazil	237; 250	206; 209	42; 45	0.0	11.6	88.4	19	27	98.6	0	1.4
ISC	250	210	39	0.0	3.2	96.8	24	31	50.8	0.5	48.7

[a] Includes cases and controls combined

Table 4: Sensitivity and specificity of RDTs, all regions

		East Africa		Brazil		Indian subcontinent	
Product	Manufacturer	Sensitivity (95% CI) n=250	Specificity (95% CI) n=250	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI) n=250	Specificity (95% CI) n=249
Crystal® KA	Span Diagnostics Ltd.	36.8% (31.1-42.9%)	98.0% (95.4-99.1%)	61.5% (55.2-67.4%)[a]	98.4% (95.9-99.4%)[b]	92.8% (88.9-95.4%)	99.2% (97.1-99.8%)
DiaMed-IT LEISH	Bio-Rad Laboratories	87.2% (82.5-90.8%)	96.4% (93.3-98.1%)	92.0% (87.8-94.8%)[c]	95.6% (92.2-97.5%)[d]	98.8% (96.5-99.6%)	97.6% (94.8-98.9%)
Kalazar Detect™	InBios International, Inc.	67.6% (61.6-73.1%)	90.8% (86.6-93.8%)	84.7% (79.7-88.7%)[e]	96.8% (93.9-98.4%)[f]	99.6% (97.8-99.9%)	96.0% (92.8-97.8%)
Signal® – KA	Span Diagnostics Ltd.	73.2% (67.4-78.3%)	96.4% (93.3-98.1%)	79.2% (73.7-83.8%)[g]	98.8% (96.6-99.6%)[h]	100% (97.9-100%)[i]	100% (97.8-100%)[j]
Onsite Leishmania Ab Rapid	CTK Biotech. Inc.	na	na	na	na	99.6% (97.8-99.9%)	96.8% (93.8-98.4%)

CI - confidence interval; na - not applicable

[a] n-244; [b] n=249; [c] n=237; [d] n=248; [e] n=249; [f] n=252; [g] n=250; [h] n=254; [i] n=175 (see section 5.1.1); [j] n=170 (see section 5.1.1)

Figure 2: ROC plot, East Africa

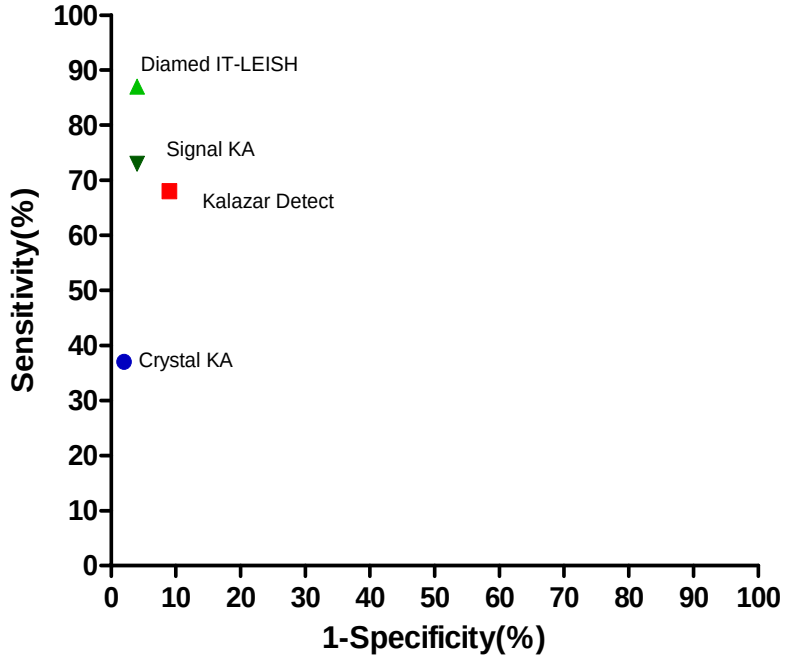


Figure 3: ROC plot, Brazil

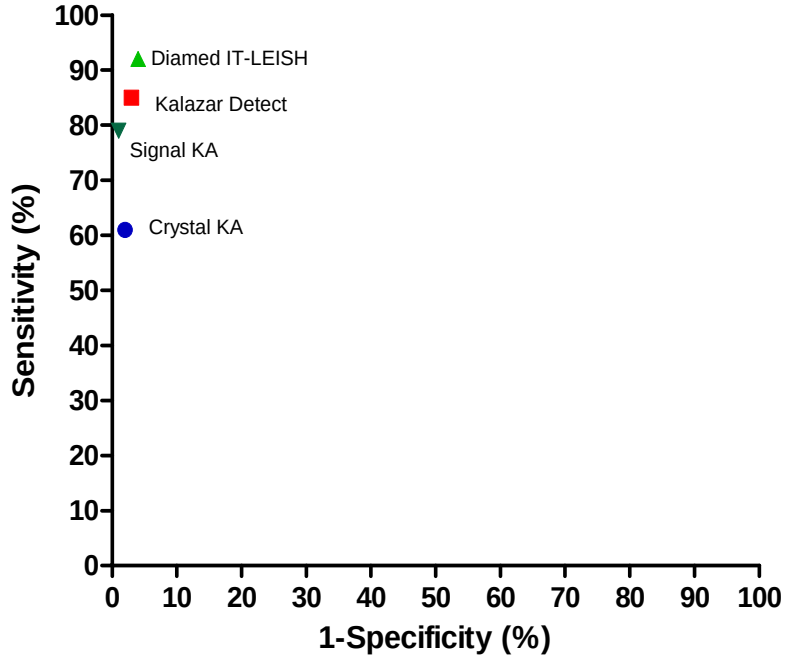
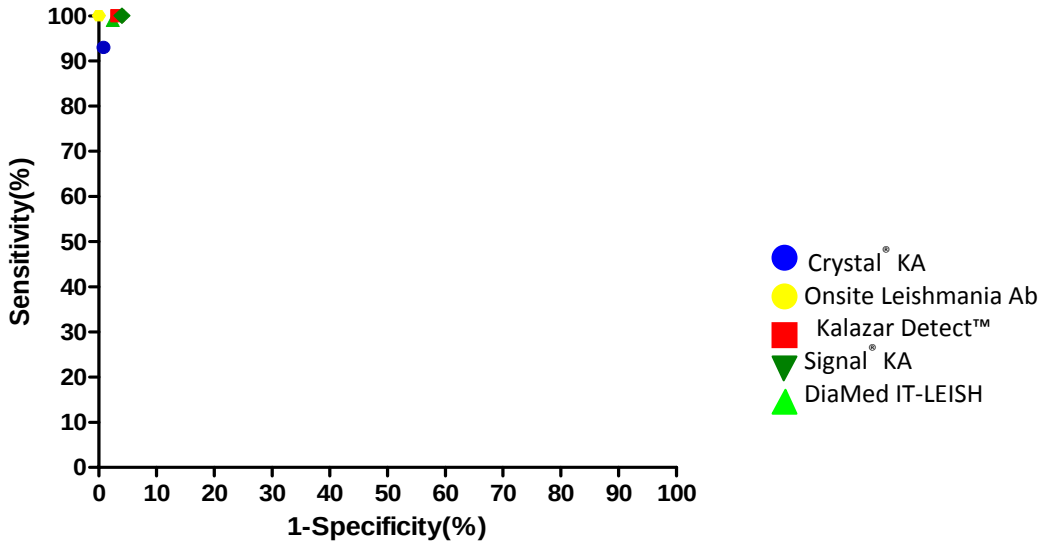


Figure 4: ROC plot, Indian subcontinent



	1-Specificity	Sensitivity
Crystal KA	0.8	93
Diamed IT-LEISH	2.4	99
Onsite Leishmania Ab	3.2	100
Kalazar Detect	4.0	100
Signal KA	0.0	100

6.2.5 Variation in test performance in centres within the same region

As each laboratory assembled their own evaluation panel, results between centres within each region were carefully scrutinized for differences prior to pooling of data regionally. Figures and Tables in Annex 2 reveal that with perhaps only one exception (Kalazar Detect™ in East Africa), there were negligible differences in intra-regional performance.

6.2.6 Band intensity

Although RDTs are not quantitative, technicians did grade the positive results according to a standard colour chart and mean band intensity was calculated. There was a positive correlation between sensitivity and band intensity (See Annex A3.1a,b,c).

6.3 Thermal stability

Tables 5-7 present the results of the thermal stability assessment which was limited to tests not requiring refrigerated storage temperatures (Annex A1.1).

In the ISC all tests assessed showed excellent baseline performance which was maintained post incubation for 60 days incubated at 4°C and 37°C. However, one product's performance was seriously affected by post 60 day-incubation at 45°C, returning 100% invalid results (n=24).

In East Africa and Brazil suboptimal baseline performance of two tests makes it challenging to draw any conclusions regarding test line stability. Samples may be at the limit of detection of some products and therefore variations at baseline reflect inter-test variation. Nonetheless, in Brazil two products appear to be stable and the other, despite the best baseline performance, showed reductions in performance at 35°C and 45°C but no invalid results as reported in both ISC and East Africa. In East Africa test performance actually appeared to improve for one product (80% detection at baseline to 100% detection after 60 days at 45°C).

6.4 Ease of use description

Here, "sample type" was not specifically assigned a score contributing to the ease of use score, it is a critical consideration in RDT selection. Only one of the five RDTs evaluated is currently recommended for use with whole blood (Annex A1.1).

Each test kit was scored out of a possible 18 points for ease of use (see section 5.2.3 and Figure A3.1). An average total score per region was calculated. In general all tests performed well with scores ranging between 9.9-13.5 (See Table 8). These results constitute a description of the product with emphasis on aspects considered of importance to ease of use in field settings.

Table 5: Thermal stability of RDTs in East Africa (post 60 days incubation)

Product	Manufacturer	Proportion positive results among cases (n=20)			Proportion negative results among controls (n=4)			Proportion invalid results (n=24)		
		4°C	37°C	45°C	4°C	37°C	45°C	4°C	37°C	45°C
Crystal® KA	Span Diagnostics Ltd.	65%	40%	20%	100%	100%	100%	0%	0%	0%
DiaMed-IT LEISH	Bio-Rad Laboratories	100%	100%	85%	100%	100%	100%	0%	0%	8%
Kalazar Detect™	InBios International, Inc.	80%	100%	100%	100%	100%	100%	0%	0%	0%

Table 6: Thermal stability of RDTs in Brazil (post 60 days incubation)

Product	Manufacturer	Proportion positive results among cases (n=20)			Proportion negative results among controls (n=4)			Proportion invalid results (n=24)		
		4°C	37°C	45°C	4°C	37°C	45°C	4°C	37°C	45°C
Crystal® KA	Span Diagnostics Ltd.	45%	50%	50%	100%	100%	100%	0%	0%	0%
DiaMed-IT LEISH	Bio-Rad Laboratories	95%	85%	35%	100%	100%	75%	0%	0%	0%
Kalazar Detect™	InBios International, Inc.	65%	65%	65%	100%	100%	100%	0%	0%	0%

Table 7: Thermal stability of RDTs in the Indian subcontinent (post 60 days incubation)

Product	Manufacturer	Proportion positive results among cases (n=20)			Proportion negative results among controls (n=4)			Proportion invalid results (n=24)		
		4°C	37°C	45°C	4°C	37C	45°C	4°C	37°C	45°C
Crystal® KA	Span Diagnostics Ltd.	95%	90%	90%	100%	100%	100%	0%	0%	0%
DiaMed-IT LEISH	Bio-Rad Laboratories	100%	100%	0%	100%	100%	100%	0%	0%	100%
Kalazar Detect™	InBios International, Inc	100%	100%	100%	100%	100%	100%	0%	0%	0%
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	100%	100%	100%	100%	100%	100%	0%	0%	0%

Table 8 Ease of use description - all regions combined

Product	Manufacturer	Average score (n= 9 evaluation centres)
Crystal® KA	Span Diagnostics Ltd.	9.9
DiaMed-IT LEISH	Bio-Rad Laboratories	11.0
Kalazar Detect™	InBios International, Inc.	11.8
Signal® – KA	Span Diagnostics Ltd.	10.9 [a]
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	13.5[b]

[a] n = 8 centres
[b] n = 4 centres

6.5 Reproducibility

In general, agreement of test results between lots and readers was high; with a few exceptions, points ranging from even the lower margins the confidence intervals can still be considered 'good' agreement.

Lot testing was completed by testing ~25% of all samples with Lot 2 RDTs provided by the manufacturer. 185 cases (81 Brazil, 39 East Africa and 65 ISC) and 158 controls (43 Brazil, 53 East Africa and 62 ISC) were tested with both Lot 1 and Lot 2 RDTs [9]. Lot-to-lot results were more variable than reader-to-reader with most Kappa values ranging from 0.82-0.98 (Tables 9-10). Two tests in two different regions fell outside of this range with values of 0.73 and 0.79. Even considering these lower values the lot-to-lot variability can be considered low, with good agreement of test results between the two lots.

Variability between readers was tested by the reading of each result by two technicians, blinded to each other's reading on all samples. Results show good agreement between readings with Kappa values ranging from 0.87-0.99 (Table 10).

Eight samples from VL cases and one negative control, chosen randomly, were tested over 3 consecutive days to assess run-to-run and operator-to-operator variability. Between 63-100% (0.63-1.00) samples give the same results on each day; there is one exception to this, with consistent results in only 37.5% (0.38) (Annex A4.1.).

Operator-to-operator results can be considered good, with 75-100% (0.63-1.00) samples positive between operators and consecutive days testing, indicating good agreement between operators and tests over time (Annex A5.1).

[9] Exceptions: *Onsite* Leishmania Ab Rapid Test-Strip (CTK Biotech Inc.) which was tested on 102 cases and 100 controls in ISC only. Signal®-KA (Span Diagnostics Ltd.) was tested on 45 cases and 42 controls in ISC, and DiaMed-IT LEISH (DiaMed AG) was tested on 80 samples in Brazil.

Table 9: Lot-to-lot agreement, all regions

Product	Manufacturer	Kappa (95% CI)		
		East Africa	Brazil	Indian subcontinent
Crystal®KA	Span Diagnostics Ltd.	0.73 (0.53-0.94)	0.84 (0.66-1.00)	0.90 (0.73-1.00)
DiaMed-IT LEISH	Bio-Rad Laboratories	0.95 (0.75-1.00)	0.93 (0.75-1.00)	0.97 (0.79-1.00)
Kalazar Detect™	InBios International, Inc.	0.82 (0.62-1.00)	0.79 (0.61-0.96)	0.91 (0.73-1.00)
Signal® – KA	Span Diagnostics Ltd.	0.88 (0.68-1.00)	0.92 (0.74-1.00)	0.98 (0.77-1.00)
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	na	na	0.92 (0.78-1.00)

na - not applicable

Table 10: Reader to reader agreement, all regions

Product	Manufacturer	Kappa (95% CI)		
		East Africa	Brazil	Indian subcontinent
Crystal®KA	Span Diagnostics Ltd.	0.87 (0.78-0.96)	0.89 (0.80-0.98)	0.98(0.89-1.00)
DiaMed-IT LEISH	Bio-Rad Laboratories	0.98 (0.89-1.00)	0.95 (0.86-1.00)	0.99(0.90-1.00)
Kalazar Detect™	InBios International, Inc.	0.96 (0.88-1.00)	0.93(.85-1.00)	0.98(0.89-1.00)
Signal® – KA	Span Diagnostics Ltd.	0.94 (0.86-1.00)	0.91(0.82-1.00)	0.99(0.90-1.00)
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	na	na	0.98(0.88-1.00)

na - not applicable

7 Discussion

VL RDTs offer the potential to improve access to diagnosis and link this to prompt treatment in peripheral health care settings. To be useful, VL RDTs must have adequate i) sensitivity to detect a high proportion of clinically significant cases; ii) specificity to accurately discriminate VL from other regionally relevant disease conditions iii) thermal stability for accuracy to be maintained after transport and storage in ambient conditions and iv) ease of use and safety procedures to allow the correct interpretation of results.

Over the past decade many reports have been published on the accuracy of both prototypes and commercially available rK39 and less commonly, other antigen based lateral flow assays. When these tests were adopted into the ISC VL Elimination Programme in 2005 it prompted a surge in commercially available VL RDTs. In order to assist VL control programmes and other procurement agencies in the selection of products appropriate to their needs, TDR coordinated a multiregional head-to-head laboratory based evaluation of VL RDTs to determine how well they meet major requirements. This report describes the performance of four commercially available rK39 and rKE16 antibody detecting RDTs in three global regions of VL endemicity using well characterized panels of human sera; a fifth RDT was also included in the Indian subcontinent evaluation centres.

Archived samples were used to avoid the time and expense associated with prospective sample collection and the complexities of comparing several products simultaneously. This is a common approach for evaluation of serological assays. Samples with the fewest freeze thaw cycles were preferentially selected; and retested by DAT as an estimate of total antibody reactivity. Due to the limited volumes of sera and restrictions in shipping of biological specimens internationally, each evaluation centre assembled its own panel following proficiency testing, and sample revalidation using study-specific standard operating procedures and materials. The absence of significant differences in test performance within regions supports the pooling of data [10] and suggests that samples from each laboratory were representative of the patient population in each global region.

In the ISC, our results reflect reports in the published literature of high accuracy of rK39 RDTs; however, our study extends to rKE16 based products and an rK39 commercial test not previously independently evaluated. Overall, our findings illustrate that in the ISC several RDTs demonstrated high sensitivity and specificity (Figure 4). This included products targeting antibodies to both rK39 and rKE16 antigens. Furthermore, performance of these products [11] did not seem to be significantly affected by the heat stress induced, inadvertently, during

[10] The sensitivity of Kalazar Detect™ (InBios International, Inc.) varied between the 3 centres in East Africa from 48.8% (95% CI 38.1-59.5%) to 86.3% (95% CI 77.0-92.1%).

[11] Kalazar Detect™ (InBios International, Inc.), DiaMed IT-Leish (Bio-Rad Laboratories), Crystal® KA (Span Diagnostics).

transport of tests to evaluation centres in India. This finding is reassuring as these circumstances can be expected to recur regularly in routine settings. On the other hand, in East Africa and Brazil, the tests performed with variable sensitivity but high specificity (3 out of four tests >95%). In Brazil and in two sites in East Africa, the rKE16 based products appeared to perform less well than rK39 products. This may be partially explained by the fact that rKE16 antibody detecting tests are based on a recombinant antigen (L.d.-rKE16) from a newly isolated Indian strain of *L. donovani* (MHOM/IN/KE16/1998), whereas rK39 is based on *L. infantum* [*12*]. Ultimately, in clinical practice in these regions, used alone VL RDT positive results may be adequate to direct treatment (when combined with the clinical case definition) but inadequate to rule out a diagnosis of VL.

Differences in product performance between regions is likely attributable to differences in antibody concentrations which may in turn be linked to different age patterns, immune and/or nutritional status of patient and/or parasite diversity. The average patient age of VL cases in each region was 24 years, 14 years and 19 years in ISC, East Africa and Brazil, respectively. Furthermore, the proportion of DAT titres >1:102,400, representing very high antibody response in ISC was 58%, this drops to 40% and 27% in East African and Brazil respectively. Samples sizes were too small to do sub-group analysis by DAT titre but when RDT performance was examined on low DAT titre samples (<1:6,400) there was a trend of reduced performance, at these lower titres.

In general, the ISC population appear to have a higher total antibody response to VL infection which corresponds with good RDT performance.

In addition, sub-group analysis based on HIV status could not be completed due to limited availability of information on HIV positive patients. HIV and VL co-infections are important to consider in VL endemic areas and test reliability is known to fall in some of these cases [*13*]. Further evaluation of RDTs against a well defined population of VL and HIV co-infected cases is needed in order to gain precise clinical accuracy data on RDTs in this special group of patients.

In all regions, agreement between lots or batches of the same product was good to excellent (Kappa 0.73-0.99) according to Landis and Koch [*11*] and agreement between readers (second reading within 30mins of the first) was excellent (Kappa >0.9). Furthermore, operator to operator and run to run reproducibility was very good. All manufacturers entered in this evaluation have current ISO 13485:2003 certification, a standard designed to give assurance of consistency of quality of final product, if correctly implemented; however, it cannot be guaranteed that the results here will predict results from different RDT lots. Therefore, ideally quality control materials should be developed so that manufacturers and procurers alike can assess test lots prior to purchase and/or field deployment to ensure that expected performance is maintained. Furthermore, training and supervision of operators must be implemented on a programmatic level to ensure the quality of testing (preparation and interpretation) at the point of care.

VL is endemic in rural regions where daytime temperatures can regularly exceed 30°C or 40°C so it is very likely that RDTs will be exposed to temperatures above the manufacturers recommendations (usually 30°C) during transport, storage or use in field settings. Heat is known to diminish the performance of some malaria RDTs [*14*] but there are no published reports on the thermal stability of VL RDTs. Our assessment does not mimic the fluctuating heat and humidity conditions in real-life settings, nor does it necessarily predict long-term stability in field conditions so it should not be considered conclusive. However, the results are useful to highlight potential losses to test sensitivity should similar conditions be encountered e.g. during transport. It is quite clear from the assessment in each region, that one of the products was less stable after 60 days at 45°C (Tables 5-7).

The sensitivity and specificity of RDT results are dependent on the quality of preparation and interpretation of the test. In general a simpler format with fewer steps or fewer required materials is likely to be conducted more reliably. For example, cassette-format RDTs are generally more reliably performed than dipstick format [*15*]. Field settings are ideal for assessing ease of use and are encouraged prior to RDT procurement; nonetheless, our study incorporated an assessment based on relatively objective criteria and all tests scored similarly well. The importance of format and simplicity of test design will depend on the intended end-users: laboratory technicians versus community health workers. In all cases, training and supervision should be an integral part of programmes implementing VL RDTs.

Samples were selected for inclusion in the evaluations if they had parasitological confirmation of VL according to microscopy or culture of clinical material (spleen, lymph node, bone marrow) and positive DAT titre, a surrogate for total antibody response. Unfortunately, neither of these diagnostics are 100% sensitive and patients with high parasite burden and elevated antibody response may have been selected for preferentially which may have artificially enhanced the clinical accuracy of the RDTs. Furthermore, the majority of controls were healthy individuals from endemic areas (with negative DAT results) and 10% were from patients with other disease conditions that mimic VL. This distribution does not represent the 'VL suspect' population and therefore may overestimate RDT specificity. This underlines the importance in clinical practice of combining RDTs with the WHO clinical case definition (or atypical presentations in high risk groups e.g. HIV infected) to avoid false positive results.

Laboratory vs field-based evaluations

Despite the strengths of this evaluation, it was laboratory based, and does not replace proper evaluation in the clinical, point-of-care, setting where the test is most likely going to be used. This evaluation was designed to provide comparative data on the performance of submitted lots of each product, mainly to inform purchasing policies of countries and agencies. RDTs were performed using sera and results cannot be extrapolated to whole blood. Furthermore, only one of

the five RDTs is recommended for use with whole blood and this may impact feasibility of performing RDTs at the point-of-care.

Ultimately, the final decision on product selection needs to be taken in a rational way taking in consideration the minimal performance limits, the patient characteristics, experience of the intended users and health care settings, climate and cost.

During the establishment of this programme of evaluation for VL RDTs, investigators came together to harmonize the details of their biological sample archives from VL cases and controls to create a common database. This international repository of information constitutes a virtual global specimen bank for VL and will hopefully serve the international community in future research and development and evaluations, in support of improved diagnostics for VL.

8 Conclusions

This report describes the performance of four commercially available rK39 and rKE16 antibody detecting rapid diagnostic tests (RDTs) in three global regions endemic for visceral leishmaniasis (VL) using well characterized panels of human sera; a fifth RDT was also included in the evaluation in the Indian sub-continent (ISC).

In all regions, agreement between lots or batches of the same product was good to excellent. Further, operator to operator and run to run reproducibility was very good. Diagnostic accuracy of RDTs between centres was highly comparable; however, results between global regions were substantially different. In general, the 5 RDTs tested in the ISC show high sensitivity and specificity, good reproducibility and most are heat stable. RDT sensitivity is more variable in East Africa and Brazil (Figure 2 & 3); in Brazil and in two sites in East Africa, the rKE16 based products appeared to perform less well than rK39 products. Ultimately, outside the ISC, in clinical practice VL RDT positive results may be adequate to direct treatment (when combined with the clinical case definition) but inadequate to rule out a diagnosis of VL. In all settings, RDTs should be implemented according to pre-defined acceptable limits of performance and within an appropriate diagnostic algorithm.

The results of this evaluation may be used to guide procurement decisions but also highlight the need for additional research into test performance amongst HIV-VL co-infected patients and in determining performance of RDTs using whole blood (which would be programmatically more convenient), rather than serum. Furthermore, our results should be combined with a detailed ease of use assessment performed in clinical settings to best inform procurement decisions.

Reference List

1. Matlashewski, G., Arana, B., Kroeger, A., Battacharya, S., Sundar, S., Das, P., Sinha, P. K., Rijal, S., Mondal, D., Zilberstein, D., and Alvar, J. 2011 Visceral leishmaniasis: elimination with existing interventions. Lancet Infect.Dis. 11,322-325

2. Schonian, G., Mauricio, I., and Cupolillo, E. 2010 Is it time to revise the nomenclature of Leishmania? Trends Parasitol. 26,466-469

3. Harith, A. E., Kolk, A. H., Kager, P. A., Leeuwenburg, J., Muigai, R., Kiugu, S., Kiugu, S., and Laarman, J. J. 1986 A simple and economical direct agglutination test for serodiagnosis and sero-epidemiological studies of visceral leishmaniasis. Trans.R.Soc.Trop.Med.Hyg. 80,583-36

4. Sundar, S., Reed, S. G., Singh, V. P., Kumar, P. C., and Murray, H. W. 1998 Rapid accurate field diagnosis of Indian visceral leishmaniasis. Lancet 351,563-565

5. Zijlstra, E. E., Nur, Y., Desjeux, P., Khalil, E. A., el-Hassan, A. M., and Groen, J. 2001 Diagnosing visceral leishmaniasis with the recombinant K39 strip test: experience from the Sudan. Trop.Med.Int.Health 6,108-113

6. Chappuis, F., Rijal, S., Soto, A., Menten, J., and Boelaert, M. 2006 A meta-analysis of the diagnostic performance of the direct agglutination test and rK39 dipstick for visceral leishmaniasis. BMJ 333,723-

7. Boelaert, M., El-Safi, S., Hailu, A., Mukhtar, M., Rijal, S., Sundar, S., Wasunna, M., Aseffa, A., Mbui, J., Menten, J., Desjeux, P., and Peeling, R. W. 2008 Diagnostic tests for kala-azar: a multi-centre study of the freeze-dried DAT, rK39 strip test and KAtex in East Africa and the Indian subcontinent. Trans.R.Soc.Trop.Med.Hyg. 102,32-40

8. Chappuis, F., Mueller, Y., Nguimfack, A., Rwakimari, J. B., Couffignal, S., Boelaert, M., Cavailler, P., Loutan, L., and Piola, P. 2005 Diagnostic accuracy of two rK39 antigen-based dipsticks and the formol gel test for rapid diagnosis of visceral leishmaniasis in northeastern Uganda. J.Clin.Microbiol. 43,5973-5977

9. Ritmeijer, K., Melaku, Y., Mueller, M., Kipngetich, S., O'keeffe, C., and Davidson, R. N. 2006 Evaluation of a new recombinant K39 rapid diagnostic test for Sudanese visceral leishmaniasis. Am.J.Trop.Med.Hyg. 74,76-80

10. Cohen, J. 1960 A coefficients of agreement for nominal scales. Edu and Psych Meas 20,37-46

11. Landis, J. R. and Koch, G. G. 1977 An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 33,363-374

12. Sivakumar, R., Dey, A., Sharma, P., and Singh, S. 2008 Expression and characterization of a recombinant kinesin antigen from an old Indian strain (DD8) of *Leishmania donovani* and comparing it with a commercially available antigen from a newly isolated (KE16) strain of *L. donovani*. Infect.Genet.Evol. 8,313-322

13. ter, Horst R., Tefera, T., Assefa, G., Ebrahim, A. Z., Davidson, R. N., and Ritmeijer, K. 2009 Field evaluation of rK39 test and direct agglutination test for diagnosis of visceral leishmaniasis in a population with high prevalence of human immunodeficiency virus in Ethiopia. Am.J.Trop.Med.Hyg. 80,929-934

14. Chiodini, P. L., Bowers, K., Jorgensen, P., Barnwell, J. W., Grady, K. K., Luchavez, J., Moody, A. H., Cenizal, A., and Bell, D. 2007 The heat stability of Plasmodium lactate dehydrogenase-based and histidine-rich protein 2-based malaria rapid diagnostic tests. Trans.R.Soc.Trop.Med.Hyg. 101,331-337

15. Rennie, W., Phetsouvanh, R., Lupisan, S., Vanisaveth, V., Hongvanthong, B., Phompida, S., Alday, P., Fulache, M., Lumagui, R., Jorgensen, P., Bell, D., and Harvey, S. 2007 Minimising human error in malaria rapid diagnosis: clarity of written instructions and health worker performance. Trans.R.Soc.Trop.Med.Hyg. 101,9-18

ANNEXES:

A1.1 Characteristics of Visceral leishmania rapid diagnostic tests included in the evaluation

Product	Manufacturer	Catalogue number	Format	Storage temperature	Shelf life	Control line	Line 1	Sample volume (µl)	Buffer quantity (drops)	Minimum reading time	Maximum reading time
Crystal® KA	Span Diagnostics Ltd.	56FT101	Dipstick	4°- 30°	18 months	Yes	rKE16	20	5	15	30
DiaMed-IT LEISH	Bio-Rad Laboratories	46240	Cassette	2°- 30°	16 months	Yes	rK39	8 to 12 (10ul mark on pipette)	Step 1: 1 (conjugate well) Step 2: 4 (wash well)	5	Not specified
Kalazar Detect™	InBios International, Inc.	INS105	Dipstick	RT (20°-28°C)	24 months	Yes	rK39	20	3	10	10
Signal® – KA	Span Diagnostics Ltd.	56FT100	Cassette	2°- 8°	12 months	Yes	rKE16	Mix 1 (20ul) part serum with 4 parts NS. Then use 50µl.	Step 1: 2 Step 2: 2 Step 3: 3	2	10
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	RO122S	Dipstick	2°- 30°	18 months	Yes	rK39	5µl	2 drops	15 mins	15 mins

RT: room temperature; NS: normal saline

A2.0 Sample size per evaluation centre

Evaluation Centre	Abbreviation	Cases	Controls
University of Khartoum	SDUOK	80	50
Kenya Medical Research Institute	KEKEM	90	50
Institute Endemic Diseases, University of Khartoum	SDIED	80	150
Instituto de Medicina Tropical de São Paulo	BRIMT	112-125[a]	124-129[b]
Pesquisadora Titular, Fiocruz	BROCG	125	125
International Centre for Diarrhoeal Disease Research	BDICD	50	50
Institute of Medical Sciences, Banaras Hindu University	INBHU	75	79
Rajendra Memorial Research Institute of Medical Sciences	INRMR	50	50
B P Koirala Institute of Health Sciences	NPBPK	75	70

[a] Signal®-KA n=125, Kalazar Detect™ n=124, DiaMed IT-LEISH n=112, Crystal®-KA n= 119
[b] Signal®-KA n=129, Kalazar Detect™ n= 127, DiaMed IT-LEISH n=123, Crystal®-KA n=124

A2.1a Sensitivity (with 95% confidence intervals) of RDTs by centre in East Africa

Evaluation Centre	Sensitivity (with 95% confidence intervals)			
	Crystal® KA	DiaMed-IT LEISH	Kalazar Detect™	Signal® – KA
SDUOK	36.3% (26.6-47.2%)	93.8% (86.2-97.3%)	48.8% (38.1-59.5%)	66.3% (55.4-75.7%)
KEKEM	38.9% (29.5-49.2%)	85.6% (76.8-91.4%)	67.8% (57.6-76.5%)	76.7% (66.9-84.2%)
SDIED	35% (25.5-45.9%)	82.5% (72.7-89.3%)	86.3% (77-92.1%)	76.3% (65.9-84.2%)

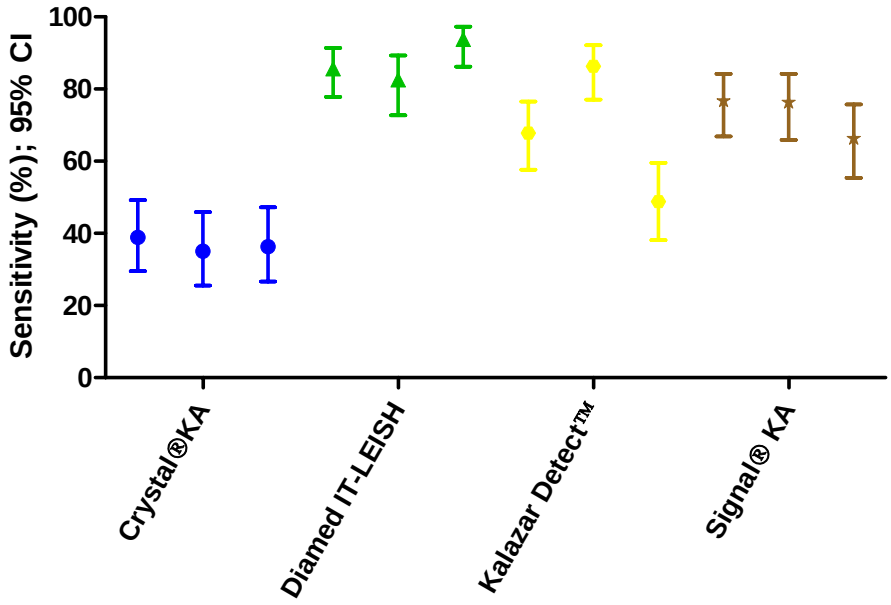
SDUOK – Sudan, University of Khartoum; KEKEM - Kenya, Kenyan Medical Research Institute; SDIED – Sudan, Institute of Endemic Diseases (University of Khartoum).

A2.1b Specificity (with 95% confidence intervals) of RDTs by centre in East Africa

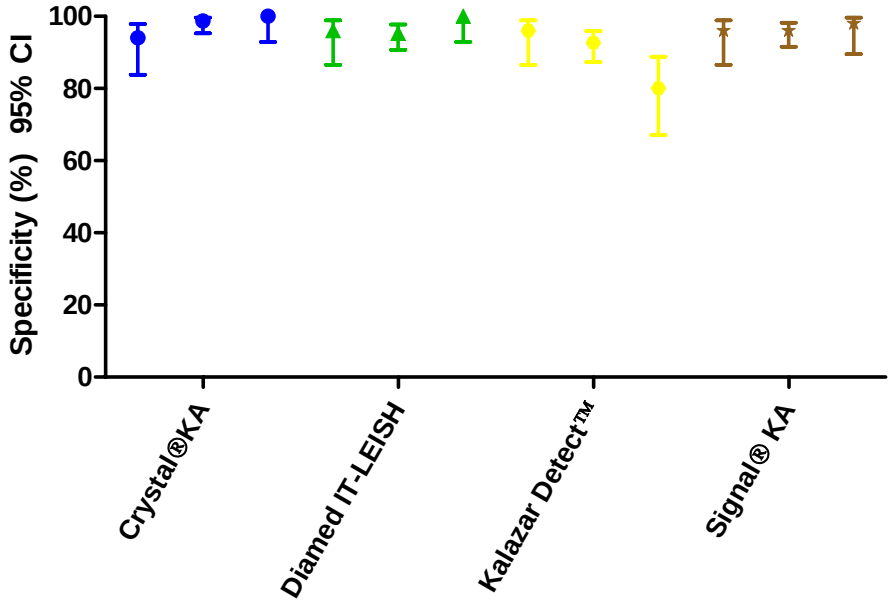
Evaluation centre	Specificity (with 95% confidence intervals)			
	Crystal® KA	DiaMed-IT LEISH	Kalazar Detect™	Signal® – KA
SDUOK	100% (92.9-100%)	100% (92.9-100%)	80% (67-88.8%)	98% (89.5-99.6%)
KEKEM	94% (83.8-97.9%)	96% (86.5-98.9%)	96% (86.5-98.9%)	96% (86.5-98.9%)
SDIED	98.7% (95.3-99.6%)	95.3% (90.7-97.7%)	92.7% (87.3-95.9%)	96% (91.5-98.2%)

SDUOK – Sudan, University of Khartoum; KEKEM - Kenya, Kenyan Medical Research Institute; SDIED – Sudan, Institute of Endemic Diseases (University of Khartoum).

A2.1a Sensitivity (with 95% confidence intervals) of RDTs by centre in East Africa



A2.1b Specificity (with 95% confidence intervals) of RDTs by centre in East Africa



A2.2a Sensitivity (with 95% confidence intervals) of RDTs by centre in Brazil

Evaluation Centre	Sensitivity (with 95% confidence intervals)			
	Crystal® KA	DiaMed-IT LEISH	Kalazar Detect™	Signal® – KA
BRIMT	63% (54.1-71.2%)	92% (85.4-95.7%)	85.5% (78.2-90.6%)	78.4% (70.4-84.7%)
BROCG	60% (51.2-68.2%)	92% (85.9-95.6%)	84% (76.6-89.4%)	80% (72.1-86.1%)

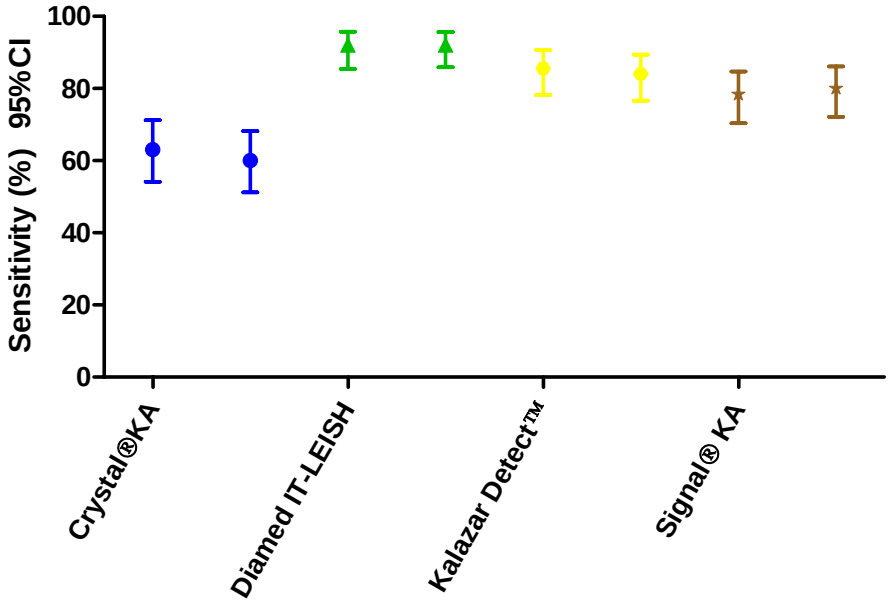
BRIMT – Brazil, Laboratório de Soroepidemiologia e Imunobiologia Instituto de Medicina Tropical de São Paulo; BROCG – Brazil, Centro de Pesquisas Réné Rachou , Fiocruz

A2.2b Specificity (with 95% confidence intervals) of RDTs by centre in Brazil

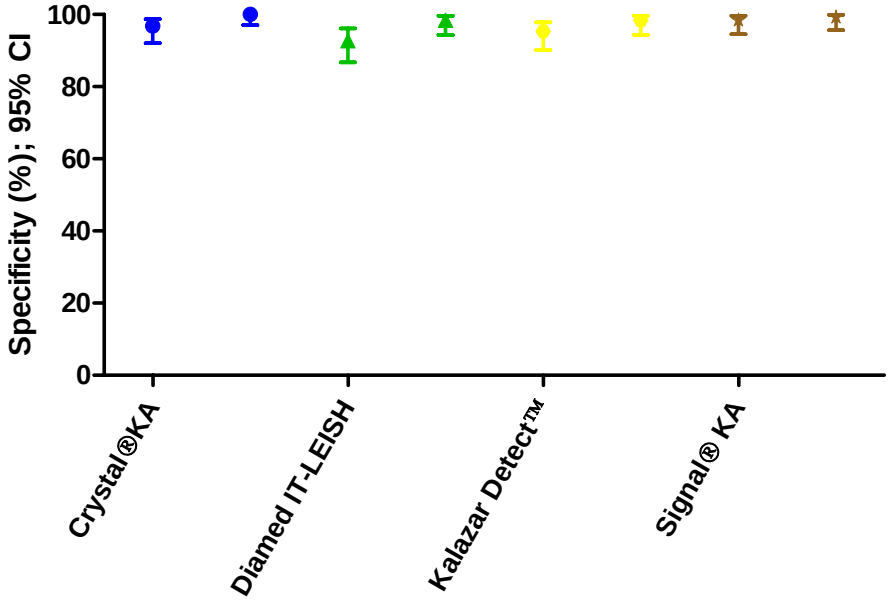
Evaluation Centre	Specificity (with 95% confidence intervals)			
	Crystal®KA	DiaMed-IT LEISH	Kalazar Detect™	Signal® – KA
SDUOK	96.8% (92-98.7%)	92.7% (86.7-96.1%)	95.3% (90.1-97.8%)	98.5% (94.5-99.6%)
BROCG	100% (97-100%)	98.4% (94.4-99.6%)	98.4% (94.4-99.6%)	99.2% (95.6-99.9%)

BRIMT – Brazil, Laboratório de Soroepidemiologia e Imunobiologia Instituto de Medicina Tropical de São Paulo; BROCG – Brazil, Centro de Pesquisas Réné Rachou , Fiocruz

A2.2a Sensitivity (with 95% confidence intervals) of RDTs by centre in Brazil



A2.2b Specificity (with 95% confidence intervals) of RDTs by centre in Brazil



A2. 3a Sensitivity (with 95% confidence intervals) of RDTs by centre on the Indian subcontinent

	Sensitivity (with 95% confidence intervals)				
Evaluation Centre	**Crystal®KA**	**DiaMed-IT LEISH**	**Kalazar Detect™**	**Onsite Leishmania Ab Rapid Test**	**Signal® – KA**
BDICD	100% (92.9-100%)	98% (89.5-99.6%)	100% (92.9-100%)	100% (92.9-100%)	100% (92.9-100%)
INBHU	89.3% (80.3-94.5%)	100% (95.1-100%)	100% (95.1-100%)	100% (95.1-100%)	na
INRMR	84% (71.5-91.7%)	98% (89.5-99.6%)	98% (89.5-99.6%)	98% (89.5-99.6%)	na
NPBPK	97.3% (90.8-99.3%)	98.7% (92.8-99.8%)	100% (95.1-100%)	100% (95.1-100%)	100% (95.1-100%)

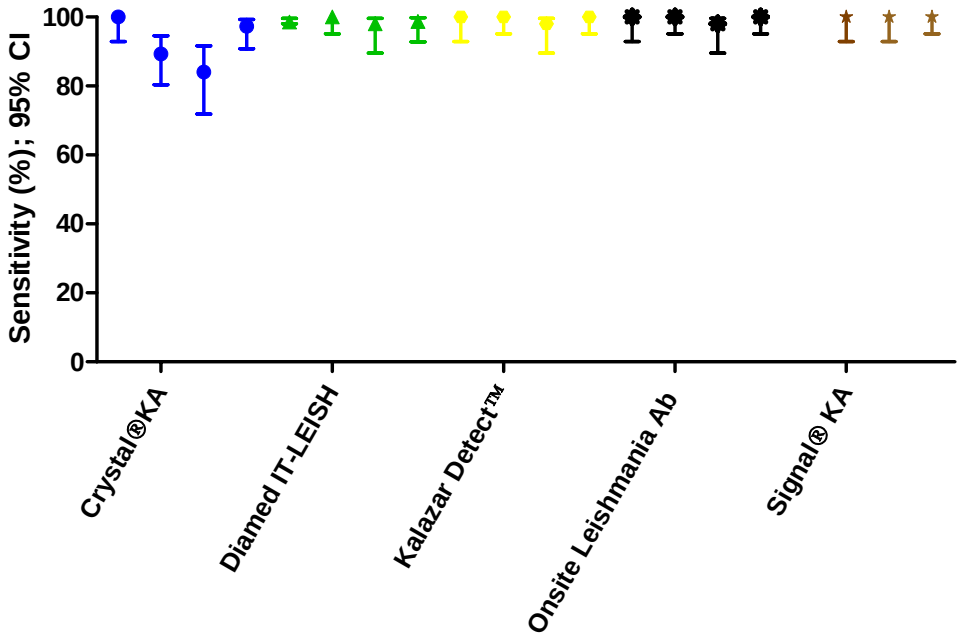
na- not applicable

A2.3b Specificity (with 95% confidence intervals) of RDTs by centre on the Indian subcontinent

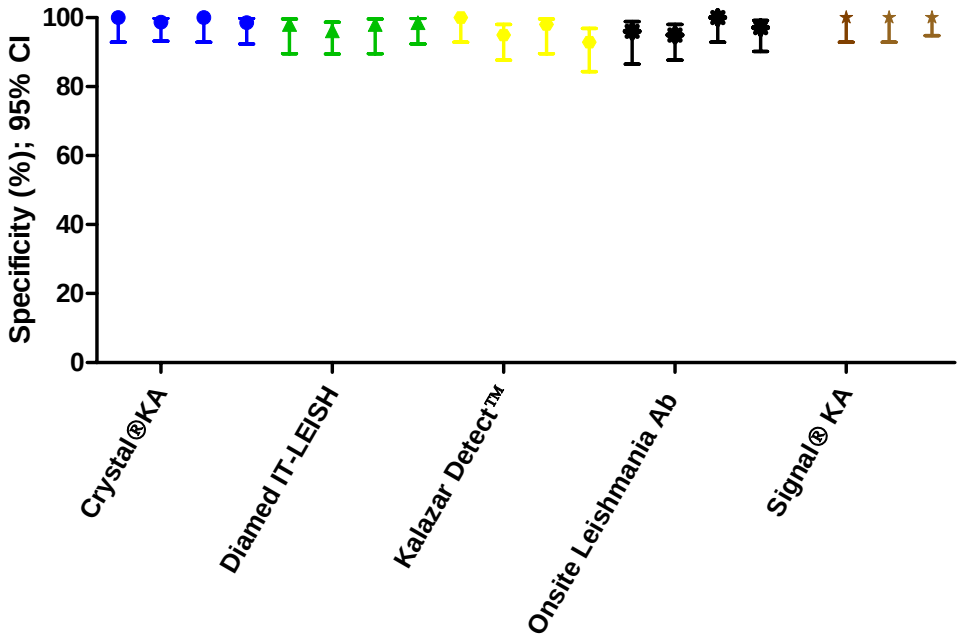
	Specificity (with 95% confidence intervals)				
Evaluation centre	**Crystal®KA**	**DiaMed-IT LEISH**	**Kalazar Detect™**	**Onsite Leishmania Ab Rapid Test**	**Signal® – KA**
BDICD	100% (92.9-100%)	98% (89.5-99.6%)	100% (92.9-100%)	96% (86.5-98.9%)	100% (92.9-100%)
INBHU	98.7% (93.1-99.8%)	96.2% (89.4-98.7%)	94.9%.2 (87.7-98.0%)	94.9%.2 (87.7-98.0%)	na
INRMR	100% (92.9-100%)	98% (89.5-99.6%)	98% (89.5-99.6%)	100% (92.9-100%)	na
NPBPK	98.6% (92.3-99.7%)	98.6% (92.3-99.7%)	92.9%.2 (84.3-96.9%)	97.1% (90.2-99.2%)	100% (94.8-100%)

na- not applicable

A2. 3a Sensitivity (with 95% confidence intervals) of RDTs by centre on the Indian subcontinent



A2.3b Specificity (with 95% confidence intervals) of RDTs by centre on the Indian subcontinent



BDICD – Bangladesh, International Centre for Diarrhoeal Disease Research; INBHU – India, Banaras Hindu University; INRMR – India, Rajendra Memorial Research Institute; NPBPK - Nepal, B P Koirala Institute of Health Sciences.

A3.1a Distribution of band intensity scores against VL positive samples from East Africa

		Band Intensity (%) [a]			
	Manufacturer	**0**	**1**	**2**	**3**
Crystal® KA	Span Diagnostics Ltd.	63.2	25.2	8.8	2.8
DiaMed-IT LEISH	Bio-Rad Laboratories	0.8	6.4	37.6	55.2
Kalazar Detect™	InBios International, Inc.	32.4	31.6	24.4	11.6
Signal® – KA	Span Diagnostics Ltd.	26.8	35.6	27.2	10.4

[a] Grading based on standard colour charts: 0-negative; 1- faint/weak positive line; 2 -positive line; 3- strong positive line.

A3.1b Distribution of band intensity scores against VL positive samples from Brazil

		Band Intensity (%)[a]			
	Manufacturer	**0**	**1**	**2**	**3**
Crystal® KA	Span Diagnostics Ltd.	38.5	35.7	17.2	8.6
DiaMed-IT LEISH	DiaMed AG	8	16.5	28.7	46.8
Kalazar Detect™	InBios International, Inc.	15.3	21.3	13.7	49.8
Signal® – KA	Span Diagnostics Ltd.	20.8	24.4	35.2	19.6

[a] Grading based on standard colour charts: 0-negative; 1- faint/weak positive line; 2-positive line; 3-strong positive line.

A3.1c Distribution of band intensity scores against VL positive samples from the Indian subcontinent

		Band Intensity (%)[a]			
	Manufacturer	**0**	**1**	**2**	**3**
Crystal® KA	Span Diagnostics Ltd.	7.2	37.2	34.4	21.2
DiaMed-IT LEISH	DiaMed AG	0.8	6.4	37.6	55.2
Kalazar Detect™	InBios International, Inc.	0.4	7.6	23.6	68.4
Signal® – KA	Span Diagnostics Ltd.	0	10.3	40.6	49.1
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	0.4	11.2	30.4	58

[a] Grading based on standard colour charts: 0-negative; 1-faint/weak positive line; 2-positive line; 3-strong positive line.

A4.1 Run to run variability scores for RDTs by region

			Mean RRV score[a]		
Product	Manufacturer	Tech	East Africa	Brazil	ISC
Crystal® KA	Span Diagnostics Ltd.	1	0.63	1	0.88
		2	0.38	1	1
DiaMed-IT LEISH	Bio-Rad Laboratories	1	1	1	1
		2	1	1	1
Kalazar Detect™	InBios International, Inc.	1	0.63	1	1
		2	0.88	1	1
Signal® – KA	Span Diagnostics Ltd.	1	0.88	1	na
		2	1	1	na
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	1	na	na	1
		2	na	na	1

RRV- run to run variability; na - not applicable

[a] Two technicians in one laboratory per region perform one RDT against the same set of 8 positive/VL case samples and one negative control on three consecutive days; RRV = 1 if sample returns all positive or all negative results on 3 consecutive days; RRV=0 if sample returned < 3 positive results. Scores are summed for each sample and a mean reported.

A5.1 Operator-to-operator variability scores for RDTs per region

		East Africa Mean OOV score[a]			Brazil Mean OOV score[a]			ISC Mean OOV score[a]		
Product	Manufacturer	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
Crystal® KA	Span Diagnostics Ltd.	0.75	0.75	1	1	1	1	0.875	1	1
DiaMed-IT LEISH	DiaMed AG	1	1	1	1	1	1	1	1	1
Kalazar Detect™	InBios International, Inc.	0.75	0.63	1	1	1	1	1	1	1
Signal® – KA	Span Diagnostics Ltd.	0.88	0.88	1	1	1	1	na	na	Na
Onsite Leishmania Ab Rapid Test	CTK Biotech. Inc.	na	na	na	na	na	na	1	1	1

a Two technicians in one laboratory per region perform one RDT against the same set of 8 positive/VL case samples and one negative control on three consecutive days; OOV=1 if there are no discrepancies between technician 1 and 2. OOV=0 if there were discrepancies between tech 1 and 2. A mean daily OOV score is reported.

A6.1Ease of use description form

FORM 07: Ease of use assessment

Date (dd/mm/yyyy):	
Name of the test:	
Manufacturer:	
Evaluation site:	
Technician name:	

1a. Clarity of kit instructions (Qualitative)

Difficult to follow	0	◯
Fairly clear	1	◯
Very clear	2	◯

1b. Clarity of kit instructions (Objective)

No diagrams	0	◯
Diagrams of results	1	◯
Diagrams of method and results	2	◯

2a. Technical complexity (Qualitative)

Complex	0	◯
Fairly easy	1	◯
Very easy	2	◯

2b. Technical complexity (Objective)

Total number of steps		
Total number of timed steps		
Steps requiring measurement or exact volumes	Yes (0)	No (1)
Need for manipulation of tubes	Yes (0)	No (1)
Need for manual interpretation of results	Yes (0)	No (1)

3. Ease of interpretation of results

Difficult	0	◯
Fairly easy	1	◯
Very easy	2	◯

4. Time to results

<15 mins	3	◯
15- 20 mins	2	◯
21- 30 mins	1	◯
> 30 mins	0	◯

5. Equipment required but not provided. E.g. micropipette

Yes	0	◯
No	1	◯

6. Biosafety (check all applicable)

Mixing wells involved	Yes(0)	No (1)
Retractable needle	Yes (1)	No (0)
Strip exposed (not within cassette)	Yes (0)	No (1)

7. Comments Highest possible score: 18

TDR For research on diseases of poverty
UNICEF • UNDP • World Bank • WHO

TDR/World Health Organization
20, Avenue Appia
1211 Geneva 27
Switzerland

Fax: (+41) 22 791-4854
tdr@who.int
www.who.int/tdr

ISBN 978 92 4 150223 8

The Special Programme for Research and Training in Tropical Diseases (TDR) is a global programme of scientific collaboration established in 1975. Its focus is research into neglected diseases of the poor, with the goal of improving existing approaches and developing new ways to prevent, diagnose, treat and control these diseases. TDR is sponsored by the following organizations:

unicef UNDP World Bank World Health Organization