

Multi-centre evaluation of repeatability and reproducibility of the direct agglutination test for visceral leishmaniasis

M. Boelaert¹, S. El Safi³, H. Mousa³, J. Githure⁴, P. Mbatia⁴, V. L. Gurubacharya⁵, J. Shrestha⁵, D. Jacquet², A. De Muynck¹, D. Le Ray² and P. Van der Stuyft¹

¹ Department of Public Health, Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium

² Department of Parasitology, Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium

³ Department of Microbiology and Parasitology, University of Khartoum, Sudan

⁴ Kenya Medical Research Institute, Nairobi, Kenya

⁵ Nepal Health Research Council, Kathmandu, Nepal

Summary

OBJECTIVE To evaluate the repeatability and reproducibility of the serological direct agglutination test (DAT) for visceral leishmaniasis (VL) with aqueous antigen in a multi-centre study in VL-endemic areas in Sudan, Kenya and Nepal.

METHODS Repeatability within each centre and reproducibility between the centres' results and an external reference laboratory (Belgium) was assessed on 1596 triplicate plain blood samples collected on filter paper.

RESULTS High kappa values (range 0.86–0.97) indicated excellent DAT repeatability within the centres. The means of the titre differences between the reference laboratory and the centres in Sudan, Kenya and Nepal (2.3, 2.4 and 1.1, respectively, all significantly different from 0) showed weak reproducibility across centres. 95% of the titre differences between the reference laboratory and the respective centres were accounted for by large intervals: 0.6–9 fold titre variation for Sudan, 0.7–8 fold for Kenya and 0.26–4 fold for Nepal.

CONCLUSION High repeatability of DAT confirms its potential, but reproducibility problems remain an obstacle to its routine use in the field. Reproducibility was hindered by alteration of the antigen through temperature and shaking, especially in Kenya and Sudan, and by nonstandardization of the test reading. DAT handling procedures and antigen quality must be carefully standardized and monitored when introducing this test into routine practice.

keywords reproducibility, direct agglutination test, visceral leishmaniasis, diagnosis, repeatability, agreement

correspondence Marleen Boelaert, Department of Public Health, Prince Leopold Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerpen, Belgium. E-mail: boelaert@itg.be

Introduction

The value of a diagnostic test depends on the validity of the results, its repeatability and reproducibility, its cost and on operational considerations. Very often, only the first determinant, validity, is taken into consideration (Begg 1987). This can lead to erroneous expectations about the performance of the test in clinical practice. While the direct agglutination test (DAT) for visceral leishmaniasis (VL) has been consistently reported as very promising for field use (El Harith *et al.* 1986; Singla *et al.* 1993; Shiddo *et al.* 1995), those reports were predominantly based on validity assessment. Repeatability

data were only reported on a small series ($n = 48$) (el Safi & Evans 1989). Prior to any recommendation on the inclusion of DAT in diagnostic/therapeutic algorithms for VL, WHO/TDR initiated a thorough evaluation of both repeatability and reproducibility, including an assessment of validity under field conditions. We studied within-centre-repeatability of DAT in endemic areas in Sudan, Kenya and Nepal and the agreement (reproducibility) with results obtained in the laboratory where the DAT antigen had been produced. We report the results of this multicentric study and discuss the operational consequences for DAT use as a diagnostic test in health districts.

Materials and methods

DAT antigen

DAT antigen was prepared at the Laboratory of Protozoology, Prince Leopold Institute of Tropical Medicine, Antwerp (ITMA) as a modification of the method described by El Harith (El Harith *et al.* 1986; 1988). Logarithmic growth phase promastigotes of *L. (L.) donovani* 1-S at $1-2 \times 10^7$ parasites/ml in GLSH (Le Ray 1975) were harvested at 1590 g and washed twice in phosphate-glucose-saccharose buffer solution (PGS), pH 8.0, before cell surface digestion with 0.4% trypsin for 60 min at 37.0 °C under gentle orbital stirring (90 r.p.m.). A 1%-promastigote suspension in cold PGS medium was fixed with 4% formaldehyde at 0 °C for 20 h under stirring. Fixative was removed through repeated washing in phosphate-buffered-saline citrate (PBS), and promastigotes were stained with 0.025% Coomassie Brilliant Blue in PBS for 20 h at 4 °C under stirring. A final suspension with 5×10^7 parasites/ml was made in formaldehyde-citrate saline. Each batch of antigen was checked against a panel of reference sera, and inadequately reacting batches were discarded. The ready-to-use antigen batches were stored in 5-ml vials at 4–8 °C until use. Half the volume of a batch was dispatched to the overseas centre, the other half being kept at ITMA to test the corresponding samples. However, due to delay between antigen production and return of samples, a different batch of antigen had to be used at ITMA on part of the samples from Sudan. Only freshly produced antigen of proven quality was sent to the overseas centre. The antigen expiry date was set at 6 months.

DAT kit

A DAT diagnosis kit providing all necessary equipment and supplies for the implementation of 660 tests was designed according to previous experience with the CATT-test, a card agglutination test for trypanosomiasis also sold in a comprehensive kit permitting field use independently from a laboratory (Magnus *et al.* 1978). ITMA sent DAT kits and antigen to the three participating centres: the Department of Medical Microbiology, University of Khartoum (Sudan); the Kenya Medical Research Institute, Nairobi (Kenya); and the Nepal Health Research Council, Kathmandu (Nepal). The kit included an instruction manual on implementation and reading of the DAT test. No training of the staff of collaborating centres was provided prior to the study. After local data collection, a team from ITMA visited the collaborating centres in Sudan and Kenya to review data collection and to evaluate the DAT kit jointly with the local investigators.

Data collection

During the study period (1993–96), cross-sectional field surveys were carried out about every 6 months by the three participating centres in the VL-endemic areas of Gedaref (Sudan), Baringo (Kenya) and Kathmandu (Nepal). A clinical suspect was defined as a person with fever of more than 15 days' duration and either splenomegaly or lymphadenopathy. A healthy control was defined as 'a person living in the same endemic area who did not present any of the above symptoms or signs at the time of recruitment.' Individuals with antecedents of VL were excluded from the series of healthy controls; patients with a blood film positive for malaria or with signs of postkala azar dermal leishmaniasis were excluded from the group of clinical suspects. Thus 807 clinical suspects participated in the study: 148 in Sudan, 305 in Kenya, and 354 in Nepal. 789 healthy controls were enrolled: 176 in Sudan, 304 in Kenya and 309 in Nepal.

Blood was collected by finger prick in heparinized haematocrit tubes after disinfecting the fingertip. Each haematocrit tube contained sufficient blood to deliver two drops of 30 µl each on preprinted filter paper. Three filter papers labelled A, B and C, each loaded with two blood drops, were collected from each person. The filter papers were allowed to dry for a minimum of 2 h in ambient temperature, protected from sunshine. Filter papers A and B were examined by the local centre; filter papers C were sent to ITMA wrapped in plastic bags containing silica gel beads. All samples were received at ITMA in dry condition.

The parasitological examination used Giemsa-stained bone marrow smears from clinical suspects and from healthy controls with a DAT positive at 1 : 500. All parasitologically positive cases and all clinical suspects with a DAT agglutination at 1 : 500 were treated with sodium stibogluconate, 20 mg/kg for 30 days (Pentostam, Wellcome, Manchester, UK).

DAT serology

To test the samples, the 30 µl blood stains were cut from the filter paper with scissors and eluted overnight at room temperature in a test tube filled with 7.5 ml saline DAT-VL dilutant (making up a 1 : 250 plain blood dilution). 50 µl of the eluate were dispensed in one of the 96 V-shaped wells of a microtitre plate. Necessary positive and negative controls were added. According to the protocol, either single 1 : 500 dilutions, or two fold serial dilutions in the range 1 : 500 – 1 : 512 000 were prepared for each specimen. After gentle mixing of the antigen, 50 µl of DAT-VL antigen was added to each well. The plate was sealed and incubated overnight at room temperature. The test was read visually against a white background. The endpoint titre was estimated as the last dilution before appearance of a clear sharp-edged blue spot identical

Table 1 Correspondence between blood and serum dilutions, expressed with and without added DAT-antigen volume

DAT dilution for blood and antigen (this study)	DAT dilution for blood only (El Harith <i>et al.</i> 1988)	DAT dilution for serum only (El Harith <i>et al.</i> 1986)	DAT dilution for serum and antigen
1:200	1:100	1:150	1:300
1:500	1:250	1:375	1:750
1:1000	1:500	1:750	1:1,500
1:2000	1:1000	1:1500	1:3,000
1:4000	1:2000	1:3000	1:6,000
1:8000	1:4000	1:6000	1:12,000
1:16 000	1:8000	1:12 000	1:24,000
1:32 000	1:16 000	1:24 000	1:48,000
1:64 000	1:32 000	1:48 000	1:96,000
1:128 000	1:64 000	1:96 000	1:192,000
1:256 000	1:128 000	1:192 000	1:384,000
1:512 000	1:256 000	1:384 000	1:768,000

to the one observed in the antigen control well. In contrast to El Harith *et al.* (1986), the expression of the dilution in this study took into account the volume of antigen added in the well. As DAT dilutions mentioned in the literature depend further on whether serum or plain blood was used, Table 1 lists conversions.

The Sudan centre performed a screening at a single dilution, 1:500, at the field site where patients were enrolled. Since the field site is 800 km from the capital, DAT was repeated at 1:500 in the Khartoum laboratory and if positive, full serial titration (1:500–1:512 000) was performed. The two other centres implemented all the above test procedures in one location: in Kenya a temporary laboratory was set up in a school at Baringo for the duration of the study, and in Nepal all tests were performed in the Kathmandu laboratory. At ITMA, filter papers were screened at 1:500, positive samples were re-screened at 1:500–1:2000 and serial titration (1:2000–1:512 000) was performed if agglutination was observed at 1:2000.

Agreement

Data were analysed with EPI-INFO and SPSS-PC. Agreement was studied on both a categorical (positive-negative at 1:500 and at 1:4000 dilution) and a discrete numerical scale. Numerical DAT results were expressed as 'end titres': on a scale of 0–12, a transformation from the dilution to the corresponding number of twofold serial dilutions. A negative DAT result was coded 0, the first dilution (1:500) was coded as 1, and so on to code 12 for the dilution 1:512 000.

Repeatability refers to within-laboratory agreement between replicate observations of the same test performed by the same observer under similar conditions (Bland & Altman 1986). Repeatability of DAT was assessed through the com-

parison of the within-centre duplicate results of DAT at the screening dilution 1:500. Cohen's Kappa coefficient κ was calculated and interpreted according to Landis and Koch (Landis & Koch 1977): 1.00–0.81 excellent, 0.80–0.61 good, 0.60–0.41 moderate, 0.40–0.21 weak, 0.20–0.00 negligible agreement. Kappa is dependent on sensitivity and specificity of the test but also on the prevalence of the condition in the study population, and this renders comparison of κ across different levels of prevalence difficult (Thompson & Walter 1988). Under conditions of repeatability assessment, the sensitivity and specificity of the test can be considered fixed in the comparison, and to allow comparison across different prevalence levels, the maximal kappa κ_m was computed according to Grenier (1990). This gives the maximum attainable chance-corrected agreement for a given data set. The ratio of the observed κ over κ_m allows thus for a more accurate comparison between centres than the mere comparison of κ .

Reproducibility refers to between-laboratory agreement, i.e. agreement between separate observations by different laboratories under different conditions. We compared each participating centre's results with corresponding ITMA results. Reproducibility was first assessed on a categorical scale at two cut-off points: 1:500, the screening dilution, and 1:4000, the DAT serological positive threshold (El Harith *et al.* 1986). The kappa coefficient was calculated and interpreted as above. Sensitivity and specificity of the test performed in-country were calculated and compared to the results obtained at ITMA.

As the DAT gives quantitative results on a discrete numerical scale (titres coded 0–12), agreement was also examined for the subset of samples with at least one positive agglutination at 1:500. Duplicate zeros were excluded (207 from the Sudan data, 480 from the Kenya data, 288 from the Nepal data),

mainly because the distribution of the differences in titres was too skewed to allow for an estimation of limits of agreement (see below). To assess agreement on a numerical scale, the data were first plotted as a scatter diagram with the ITMA titre on the *x*-axis and the other centre's titre on the *y*-axis, the line of identity allowing for visual assessment of agreement. Subsequently, data were plotted as (*x*-*y*) against (*x* + *y*)/2 according to Altman & Bland (1983). The difference is plotted against the average of both titres, as the average of both observations is the best estimate of the true value.

Systematic error (bias) between measurements at ITMA compared to the centres' observations was estimated by the mean difference of all the paired titre differences $x_i - y_i$. A paired *t*-test examined the null-hypothesis of a zero mean difference. Random error in the measurement differences, or the variability in the differences between the paired observations, was estimated by the limits of agreement. They indicate the range containing 95% of all interobserver differences. The limits of agreement are computed as the mean difference ± 2 standard deviations, assuming differences are normally distributed. If a trend emerges between interobserver difference and size of the measurements, transformation of the data to a log scale is recommended (Altman & Bland 1983). Logarithmic transformation followed the formula:

$$\ln \text{ titre} = \ln (\text{end titre} + 1).$$

After computation of the mean difference and the limits of agreement, back transformation permitted expression of both parameters in the original titres.

Results

Table 2 shows the repeatability of DAT performed at one single dilution (1 : 500) within each participating centre. The respective kappa coefficients were excellent in all four centres. Maximal kappa coefficients (κ_m) varied slightly, depending on the marginal totals in the 2×2 table, but ratios of kappa over maximal kappa (κ/κ_m) again showed excellent agreement within each centre. However, 6/609 samples in Kenya, 0/647 in Nepal, 2/322 in Sudan and 49/1582 in Antwerp were read as 'doubtful' in the single dilution reading and included as positive readings for the analysis in Table 2. Their higher numbers explain the slightly lower κ observed in Antwerp. Categorical analysis of reproducibility between centres at two cut-off titres (1 : 500 and 1 : 4000) is shown in Table 3. Kappa-values showed only moderate to good agreement. The sensitivity of DAT performed in Sudan and Kenya is low compared to the ITMA result.

Figure 1 plots the titres in each centre compared to the titres obtained in Antwerp, with the line of identity drawn for visual assessment of agreement. Clearly Sudan and Kenya have consistently lower titres than ITMA, whereas the Nepalese data do not exhibit this systematic deviation. To analyse the systematic deviation and the random variability in titre differences in more detail, we plotted the differences of the log titres against their average, after exclusion of duplicate zero values (Figure 2). Table 4 lists the mean paired difference between ln titres, the corresponding mean paired difference in titre and the limits of agreement.

Table 2 Repeatability of DAT at 1 : 500 dilution (Kappa)

	N (missing)	Cell entries in the contingency tables				+ + /n	κ (95% CI)	κ_m	κ/κ_m
		- -	- +	+ -	+ +				
Sudan	322 (2)	239	0	4	78	0.24	0.97 (0.86-1.0)	0.97	1
Kenya	609 (0)	562	2	5	40	0.07	0.91(0.83-0.99)	0.98	0.92
Nepal	647 (16)	304	4	13	326	0.50	0.95 (0.87-1)	0.97	0.97
ITMA	1582 (14)	984	82	19	497	0.31	0.86 (0.81-0.91)	0.91	0.94

	<i>n</i> (missing)	κ (95% C.I.)	Sensitivity	Specificity
Sudan at 1 : 500	317 (7)	0.74 (0.64-0.84)	0.71	0.98
Sudan at 1 : 4000	317 (7)	0.61 (0.51-0.71)	0.52	1
Kenya at 1 : 500	607 (2)	0.41 (0.30-0.50)	0.32	0.99
Kenya at 1 : 4000	607 (2)	0.73 (0.63-0.83)	0.64	0.99
Nepal at 1 : 500	655 (8)	0.73 (0.65-0.81)	0.93	0.82
Nepal at 1 : 4000	655 (8)	0.86 (0.84-0.88)	0.96	0.91

Table 3 Reproducibility, sensitivity and specificity of DAT at 1 : 500 and 1 : 4000 compared to reference laboratory results (ITMA)

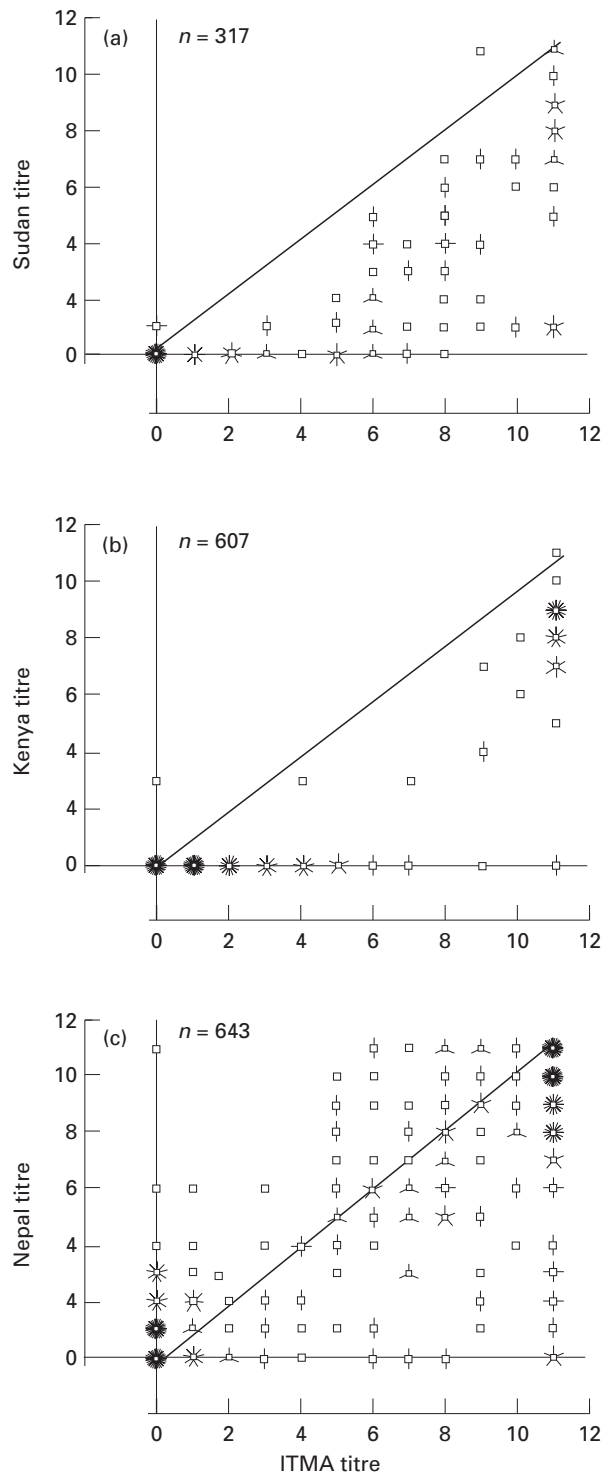


Figure 1 Comparison of titre at ITMA with corresponding titre in Sudan, Kenya and Nepal. First observation is represented by \circ , every petal representing 1 additional observation. Thus: \square is 1 observation, \square is 2 observations, etc.

There was a very significant mean difference exceeding 2 titres between the results from ITMA and those from Sudan and Kenya. The random variability in the titre differences was appreciable to high everywhere, as 95% of the differences in titres between ITMA and the respective centre

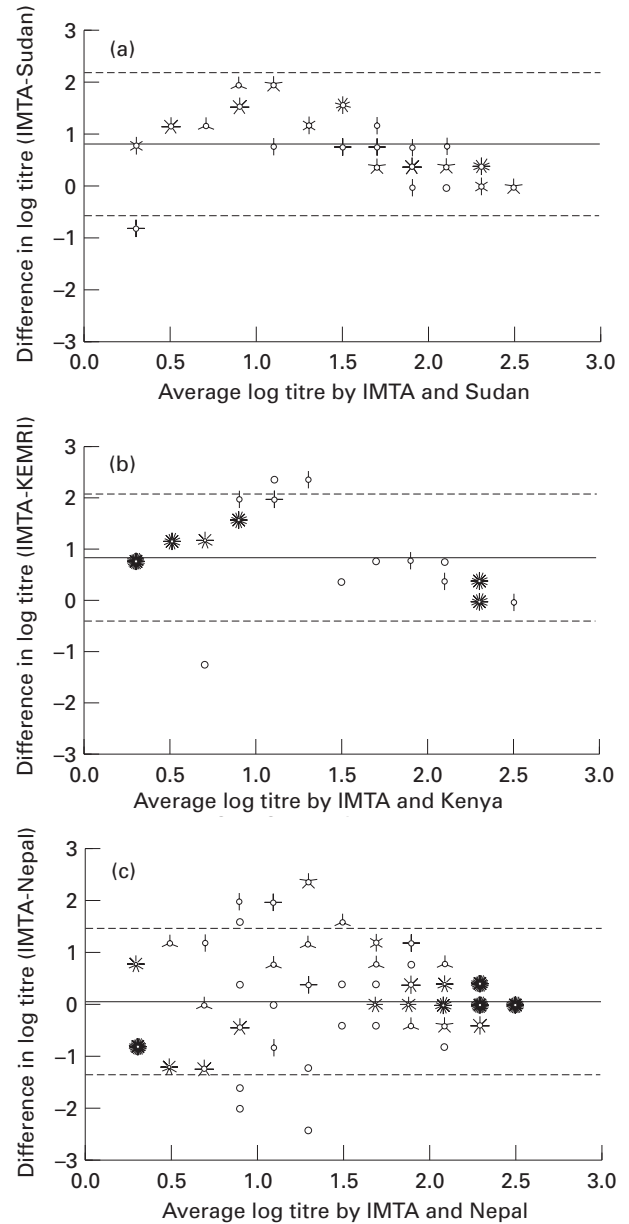


Figure 2 Differences of the log titres against their average, after exclusion of duplicate zero values: (a) Sudan; (b) Kenya; (c) Nepal. First observation is represented by \circ , every petal representing 1 additional observation. Thus: \circ is 1 observation, \circ is 2 observations, etc. (a) $n = 110$; 207 duplicate zeros excluded; (b) $n = 127$, 480 duplicate zeros excluded; (c) $n = 355$, 288 duplicate zeros excluded. — mean; - - - - + or - 2 SD.

	Mean of the differences ln titre (ITMA-centre)	Mean of the differences in titre (ITMA - centre)	'95% Limits of agreement' in original titres
Sudan (<i>n</i> = 110)	0.83 (<i>P</i> < 0.001)	2.3	0.6-9 fold
Kenya (<i>n</i> = 127)	0.86 (<i>P</i> < 0.001)	2.4	0.68-8 fold
Nepal (<i>n</i> = 355)	0.075 (<i>P</i> = 0.046)	1.1	0.26-4 fold

Table 4 Mean paired difference between ln titres, the corresponding mean paired difference in titre and the limits of agreement

could be expected to be found within 9.6 dilutions for Sudan, 8.6 dilutions for Kenya and 4.3 dilutions for Nepal. Titres at ITMA were significantly higher than in the three centres, especially in Sudan and Kenya.

Discussion

Reproducibility and repeatability of a laboratory test are crucial characteristics but, unlike sensitivity and specificity, they are seldomly taken into account when evaluating the usefulness of a test in a particular setting (Begg 1987). Twelve years after its introduction the question whether DAT gives the same result when applied twice on the same sample by the same laboratory or by two different laboratories remained largely unanswered. This is worrisome since the test has to be used under difficult field conditions and in laboratories with suboptimal infrastructure and not always highly skilled staff.

We found that DAT had almost perfect repeatability in the four participating centres, which confirms the previous findings on 48 samples by (el Safi & Evans 1989). We express caution about repeatability being assessed at 1 : 500 dilution only, since end point differences may have gone unnoticed. Moreover, the proportion of VL in the three centres varied, as indicated by the proportion of samples with duplicate positive results ranging from 7 to 50% (Table 2). This implies that their kappas are not comparable and was circumvented by computing the ratio of κ to the theoretical maximum κ for each centre (Grenier 1990). Again, performance was excellent, as centres achieved between 92% and 100% of the attainable agreement. This corroborates earlier claims on the ease of implementation of the DAT technique with skilled personnel (El Harith *et al.* 1986; Sinha & Sehgal 1994). Nevertheless, given its performance at a wide prevalence range in different centres, we may safely conclude that the repeatability of DAT is excellent and that the test has a high potential for good reproducibility.

Surprisingly, reproducibility was only moderate in a categorical analysis, with κ ranging from 0.41 to 0.86 in function of dilution and centre. We found low sensitivity in Sudan and Kenya if we used the DAT results obtained at ITMA as a reference test. Closer analysis of the agreement between titre readings for samples with at least one non-zero titre reading

pointed to the reasons for the low reproducibility. Random variability is high for all centres and the Sudan/ITMA and Kenya/ITMA comparisons were clearly biased with average underestimates of over 2 dilutions.

Quality decay of the antigen seems the most plausible explanation for these systematic lower readings. Deterioration has occurred during transportation and storage. The antigen is heat-sensitive and should be kept at 4 °C; when stored correctly, stability of DAT is guaranteed for at least 10 months (Mengistu *et al.* 1990; ITMA, unpublished). Securing the cold chain during shipment from the European production centre to the intertropical laboratories participating in the study was, however, a problem. Local research teams reported that important cold-chain interruptions occurred during handling of the freight at airports and customs. Transport by car from airports to the field study areas in Baringo and Gedaref was a very 'shaky' experience for both staff and equipment and may have affected antigen quality. Differences between batches may also account for part of the poor reproducibility.

The 95% limits of agreement for titre differences, excluding the duplicate zeros, exceed substantially the acceptable range of ± 1 titre in routine DAT serology. No formal training or reading standardization had taken place before the study and the kit manual instructions, deemed sufficient to enable the centres to implement the technique, may have been interpreted in different ways. This was confirmed in an *ad-hoc*, small-scale rater agreement study organized between the Kenian and ITMA lab technicians. The reading problems it revealed, particularly with regard to ascertainment of end point titres, may have contributed to the bias described above. The implications of such poor reproducibility are dramatic. In Sudan, DAT detected only 57.1% of a series of 49 parasitologically positive cases at the recommended 1 : 4000 diagnostic cut-off, whereas the DAT performed on the same samples at ITMA detected 95.9% (Boelaert *et al.* 1999).

In conclusion, DAT has high repeatability, but its reproducibility was poor in our study. More elaborate and clearer instructions in the accompanying manual, preimplementation training, tight supervision and continuous quality control seem to be critical. DAT antigen in suspension is heat and shock-sensitive. Continuous temperature monitoring of the

M. Boelaert *et al.* **Evaluation of direct agglutination test for visceral leishmaniasis**

cold chain, as is customary for vaccine storage, seems mandatory. Local calibration would control for antigen quality decay but render test execution more complex. Local antigen production is an alternative, but it will not completely solve the transport and storage-related problems which occur mainly in the periphery. Improved DAT stability (El Harith *et al.* 1988) would provide a viable solution and current efforts to produce a freeze-dried DAT antigen are encouraging (Meredith *et al.* 1995; Zijlstra *et al.* 1997).

Given the comparative advantages of DAT (El Harith *et al.* 1987; Chowdhury *et al.* 1993; Sinha & Sehgal 1994; Kar 1995; Shiddo *et al.* 1995), we should call for urgent improvement of the available technology. Meanwhile, DAT should only be introduced in laboratories staffed with well-trained biotechnicians who practice continuous quality control and secure an uninterrupted cold chain. Unfortunately, the above conditions may, today, still be too stringent for many district laboratories in VL-endemic areas.

Acknowledgements

This investigation received support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Disease (TDR), Project ID 960315 and the Fonds voor Wetenschappelijk Onderzoek-Vlaanderen, Project # 1.5.480.98. We are indebted to Dr D. Payne, Dr F. Modaber and Dr P. Desjeux for their support throughout the study.

References

- Altman DG and Bland JM (1983) Measurement in Medicine: the Analysis of Method Comparison Studies. *The Statistician* **32**, 307–317.
- Begg CB (1987) Biases in the Assessment of Diagnostic Tests. *Statistics in Medicine* **6**, 411–423.
- Bland JM and Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.
- Boelaert M, El Safi S, Jacquet D, De Muynck M, Van der Stuyt P & Le Ray D (1999) Operational validation of the Direct Agglutination Test for diagnosis of visceral leishmaniasis. *American Journal of Tropical Medicine and Hygiene* **60** (4), in press.
- Chowdhury MS, El Harith A, Al Massum A, al Karim E and al Rahman A (1993) Prevalence of agglutinating anti-Leishmania antibodies in two multi-thousand Bengali communities. *Parasitology Research* **79**, 444–450.
- El Harith A, Kolk AH, Kager PA *et al.* (1986) A simple and economical direct agglutination test for serodiagnosis and sero-epidemiological studies of visceral leishmaniasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **80**, 583–587.
- El Harith A, Kolk AH, Kager PA *et al.* (1987) Evaluation of a newly developed direct agglutination test (DAT) for serodiagnosis and sero-epidemiological studies of visceral leishmaniasis: comparison with IFAT and ELISA. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **81**, 603–606.
- El Harith A, Kolk AH, Muigai R, Huigen E, Jelsma T and Kager PA (1988) Improvement of a Direct Agglutination Test for Field Studies of Visceral Leishmaniasis. *Journal of Clinical Microbiology* **26**, 1321–1325.
- Grenier B (1998) Concordances. In: *Décision Médicale*. Masson, Paris, p. 137.
- Kar K (1995) Serodiagnosis of leishmaniasis. *Critical Review of Microbiology* **21**, 123–152.
- Le Ray D (1975) Structures Antigéniques de *Trypanosoma brucei* (Protozoa, Kinetoplastidae). *Annales de la Société Belge de Médecinotropical* **55**, 132–292.
- Landis JR and Koch JJ (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics* **33**, 159–174.
- Magnus E, Vervoort T and Van Meirvenne N (1978) A card-agglutination test with stained trypanosomes (CATT) for the serological diagnosis of *Trypanosoma brucei gambiense* trypanosomiasis. *Annales de la Société Belge de Médecinotropical* **58**, 169–176.
- Mengistu G, Kiessling R and Akuffo H (1990) The value of a direct agglutination test in the diagnosis of cutaneous and visceral leishmaniasis in Ethiopia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **84**, 359–362.
- Meredith SE, Kroon NC, Sondorp E *et al.* (1995) Leish-KIT, a stable direct agglutination test based on freeze-dried antigen for serodiagnosis of visceral leishmaniasis. *Journal of Clinical Microbiology* **33**, 1742–1745.
- el Safi SH and Evans DA (1989) A comparison of the direct agglutination test and enzyme-linked immunosorbent assay in the serodiagnosis of leishmaniasis in the Sudan. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **83**, 334–337.
- Shiddo SA, Aden MA, Akuffo H *et al.* (1995) Visceral leishmaniasis in Somalia. Circulating antibodies as measured by DAT, immunofluorescence and ELISA. *Tropical Geographical Medicine* **47**, 68–73.
- Singla N, Singh GS, Sundar S and Vinayak VK (1993) Evaluation of the direct agglutination test as an immunodiagnostic tool for kala-azar in India. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **87**, 276–278.
- Sinha R and Sehgal S (1994) Comparative evaluation of serological tests in Indian kala-azar. *Journal of Tropical Medicine and Hygiene* **97**, 333–340.
- Thompson WD and Walter SD (1988) A reappraisal of the Kappa coefficient. *Journal of Clinical Epidemiology* **41**, 949–958.
- Zijlstra EE, Osman OF, Hofland HW *et al.* (1997) The direct agglutination test for diagnosis of visceral leishmaniasis under field conditions in Sudan: comparison of aqueous and freeze-dried antigens. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **91**, 671–673.