

Impact of Proficiency Testing on Results of the Microscopic Agglutination Test for Diagnosis of Leptospirosis

R. J. Chappel,^{1*} M. Goris,² M. F. Palmer,³ and R. A. Hartskeerl²

National Serology Reference Laboratory, Australia, Fitzroy, Victoria, Australia¹; Leptospirosis Reference Centre (WHO/FAO/OIE), Royal Tropical Institute, Amsterdam, The Netherlands²; and *Leptospira* Reference Unit, County Hospital, Hereford, United Kingdom³

Received 20 May 2004/Returned for modification 6 July 2004/Accepted 15 August 2004

A proficiency testing scheme for the leptospirosis microscopic agglutination test was provided to 37 laboratories in 23 countries in 2002 (round 1) and to 60 laboratories in 34 countries in 2003 (round 2). Thirty-four laboratories participated in both rounds. Each panel consisted of five rabbit serum samples, four of which were antisera raised against pathogenic serovars of *Leptospira*. One of these samples was a mixture of two different antisera. The rates of false-negative results, calculated on the basis of the assumption that serovars within a serogroup will cross-react, were 11% for round 1 and 14% for round 2. There were regional differences in the rates of false-negative results. The titers reported by laboratories testing for the same sample with the same serovar varied widely. Laboratories that had previously participated in round 1 reported fewer false-negative results in round 2 than new participants (10 and 21%, respectively [$P = 0.002$]) and reported 0.56 false-negative results per participant, whereas new participants reported 1.23 false-negative results per participant ($P = 0.041$). Laboratories that had previously participated also reported fewer false-negative results in round 2 than in round 1 when samples common to both rounds were tested (5 and 15%, respectively [$P = 0.028$]). The titers reported by the new participants were, on average, lower than those reported by the laboratories that had participated previously ($P = 0.019$) and were significantly more variable ($P = 0.001$). Analysis of these results suggests a positive impact of proficiency testing on the testing performance of the participating laboratories.

The microscopic agglutination test (MAT) for leptospirosis antibodies (2) has been used for the diagnosis of leptospirosis for many years in the human medical and the veterinary settings (1, 3). It is a simple technique and requires little expensive equipment, apart from a dark-field microscope, and it broadly differentiates between antibodies directed against different leptospiral serogroups. There are more than 200 pathogenic serovars of *Leptospira*, and these are grouped into 25 serogroups on the basis of antigenic similarity. Sometimes only one serovar from a particular serogroup is of interest to a diagnostic laboratory. The results of the MAT can thus give an indication of the infecting serovar, and this is important both for diagnosis and in epidemiological studies.

MAT uses live cultures of leptospiral bacteria as diagnostic antigens. As diagnostic reagents, these cultures are difficult to standardize. Eventually, newer methods may render MAT obsolete, but this is unlikely to happen for some years. MAT may remain in use the longest in countries with limited resources, some of which have wet tropical climates and, consequently, high prevalences of leptospirosis (1, 3).

Quality assurance is important for MAT testing. An international proficiency testing scheme for the leptospirosis MAT was developed because proficiency testing for this test was not available in many countries.

The aims of this study were to develop measures for the quality of MAT testing in different laboratories and to assess the impact of proficiency testing on the quality of testing.

* Corresponding author. Mailing address: National Serology Reference Laboratory, Australia, 4th Floor, Healy Building, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia. Phone: 61 3 9418 1115. Fax: 61 3 9418 1155. E-mail: roderick@nrl.gov.au.

MATERIALS AND METHODS

Proficiency testing rounds. Two proficiency testing panels were distributed to participating laboratories, round 1 in 2002 and round 2 in 2003.

Participants. Laboratories performing MAT for the diagnosis of leptospirosis in the medical or veterinary setting, or both, were invited to participate through an e-mail list administered by the International Leptospirosis Society and through approaches to scientists with an interest in leptospirosis. Each participating laboratory was given a confidential identifying number, known only to the participant and to the National Serology Reference Laboratory, Australia. Table 1 summarizes the geographic locations of the participants. Of 37 participants in round 1, 34 participated again in round 2, along with 26 new participants. The round 1 participants came from 23 countries, and the round 2 participants came from 34 countries.

Panels. Each proficiency testing panel consisted of five samples of rabbit serum. Individual samples were freeze-dried and distributed in International Air Transport Association-approved packaging. In round 1, each sample was provided as a 50- μ l volume, and participants were instructed to reconstitute each sample with 500 μ l of physiological saline. In round 2, the samples were diluted 1/10 with Sorensen buffer (pH 7.4) before they were freeze-dried and distributed. Preliminary testing showed that the dilution of 1/10 was likely to bring the titers in each of the positive serum samples into the range of titers that laboratories

TABLE 1. Geographic locations of participants in rounds 1 and 2 of the leptospirosis MAT proficiency testing scheme

Region	No. of participants			
	Round 1 only	Rounds 1 and 2	Round 2 only	Total
Australasia		2	1	3
Asia and Pacific	1	3	6	10
Central and South America		6	9	15
Europe	2	21	7	30
North America		2	3	5
Total	3	34	26	63

TABLE 2. Leptospiral serovars used to raise antisera constituting the panels distributed in rounds 1 and 2 of the leptospirosis MAT proficiency testing scheme

Sample	Round 1		Round 2	
	Serogroup	Serovar	Serogroup	Serovar
A	Australis	Australis	Canicola	Canicola
B	Canicola	Canicola	Grippotyphosa	Grippotyphosa
C	None	None	Sejroe	Sejroe
D	Tarassovi	Tarassovi	None	None
E	Icterohaemorrhagiae	Icterohaemorrhagiae	Icterohaemorrhagiae and Javanica	Icterohaemorrhagiae and Poi

were likely to measure. Participants were asked to test the samples on the day on which they were reconstituted.

Table 2 shows the compositions of the two panels. One sample in each panel was nonimmune rabbit serum. The remaining four samples in round 1 and three samples in round 2 were high-titer antisera raised against individual leptospiral serovars. One sample in round 2 was a mixture of two such antisera. The following strains were used to raise antisera: serovar Australis, strain Ballico; serovar Canicola, strain Hond Utrecht IV; serovar Grippotyphosa, strain Moskva V; serovar Icterohaemorrhagiae, strain Ictero 1; serovar Poi, strain Poi; serovar Sejroe, strain M84; and serovar Tarassovi, strain Perepelicin. The antisera that were used in both rounds (against serovars Canicola and Icterohaemorrhagiae) were identical.

Serovars used in MAT. Each participating laboratory was invited to test the samples with its normal panel of MAT antigens. As a consequence, the samples were tested with many different serogroups, serovars, and strains. Round 1 participants used a total of 60 pathogenic and 2 saprophytic serovars as test antigens. These came from 22 of 25 recognized pathogenic serogroups and from 2 saprophytic serogroups. Round 2 participants used a total of 66 pathogenic and 5 saprophytic serovars as test antigens. These came from all 25 recognized pathogenic serogroups and from 5 saprophytic serogroups. Some serogroups were used by far more participants than others, especially serogroups Australis, Canicola, Grippotyphosa, Icterohaemorrhagiae, Pomona, Sejroe, and Tarassovi.

Differences in MAT methods. MAT is often performed somewhat differently by different laboratories. One area of variation observed is in the dilution of test sera. Participants in this study used various dilution series. Most, but not all, involved twofold dilutions, usually based on a dilution of 1/25 but less commonly on a dilution of 1/20 or of 1/4.

In the collation of the results, the titer was expressed as the reciprocal of the highest dilution at which 50% agglutination was reported, including the volume of antigen added in this dilution. Results from participants who reported titers on the basis of dilutions that excluded added antigen were adjusted as required. Participants used different starting dilutions, and therefore, the lowest titer that could be detected varied from one participant to another. For example, participants who used doubling dilutions based on a dilution of 1/25 used dilution series that variously detected lowest titers of 25, 50, or 100. Those who used dilutions based on a dilution of 1/20 detected lowest titers of 20 or 40.

Because of the differences in dilution series, negative titers reported by different participants differed and were expressed as < 100, <50, <40, etc.

Arbitrary definition of a positive result. To compare the results reported by the different participants, it was necessary to use an arbitrary definition of a positive titer. For the purposes of this study, titers of 80, 100, or higher were defined as positive and titers of < 100, <80, or below were defined as negative. Positive titers, thus defined, were used for the purposes of comparison and analysis but were not assumed to be of diagnostic significance.

Definition of a false-negative result. A participant was considered to have obtained a false-negative result if it reported a negative result (a titer of <100 or <80) for any serovar from the serogroup against which a sample was raised.

Identification of serogroups. A participant was considered to have correctly identified the serogroup for which a sample was positive if it reported a positive titer for any serovar within the serogroup and if that titer was higher than all titers for the serovars in the other serogroups. This could be assessed only with the unmixed positive samples, which comprised four samples from the round 1 panel and three samples from the round 2 panel. If a laboratory did not test a sample with any serovar from the serogroup concerned, its result for that sample was not considered in this analysis.

Quantitative analysis of titers. The titers reported by participants for the same sample tested with the same serovar varied widely. In some cases, enough participants reported titers for the same serovar to allow statistical analysis after logarithmic transformation. So that all available data could be included in such an analysis, negative titers were taken as half the lowest titer that the participant

reported, and off-scale positive titers were taken as double the highest titer that the participant reported. For example, in the analysis a negative titer of <50 was considered 25 and a titer of $\geq 6,400$ was considered 12,800.

Statistical analysis. The SPSS package (version 12.1; SPSS Inc., Chicago, Ill.) was used for comparisons between groups by using *t* tests for independent samples and for comparison of variances by using Levene's test for equality of variances. Microsoft Excel 2000 software was used for comparisons by using Pearson's chi-square statistic. A *P* value of <0.05 was considered statistically significant.

RESULTS

False-negative results. Table 3 summarizes the false-negative results reported in round 1 and round 2. Overall, 13% of the reported results (73 of 552) were false negative, according to the definition given in the Materials and Methods. Participants obtained 37 of 320 (12%) negative results using the homologous serovars against which the samples tested were actually raised, but 36 of 232 (16%) negative results using heterologous serovars within the same serogroup. This apparent difference was not statistically significant ($\chi^2_{1\text{ df}} = 1.833$ [df, degrees of freedom]; *P* = 0.18).

Table 4 compares the false-negative results reported in round 2 for participants in the five geographic regions in relation to whether or not they had previously participated in round 1. Ongoing participants reported less than half the percentage of the false-negative results reported by the new participants ($\chi^2_{1\text{ df}} = 9.185$; *P* = 0.002). The difference between the two groups varied with the region; the greatest difference, for South and Central America, was individually significant ($\chi^2_{1\text{ df}} = 4.498$; *P* = 0.034). Ongoing participants also reported less than half the false-negative results per participant ($\chi^2_{1\text{ df}} = 4.164$; *P* = 0.041).

Two antisera raised against serovars Canicola and Icterohaemorrhagiae were included in the panels for both rounds (Table 2). A direct comparison of the false-negative results reported in the two rounds was made by using these samples

TABLE 3. False-negative results reported by participants in rounds 1 and 2 of the leptospirosis MAT proficiency testing scheme

Round	No. of false-negative results/total no. of results (%) with:		
	Homologous serovars	Heterologous serovars within the serogroup	Combined negative results
1	12/123 (10)	10/72 (14)	22/195 (11)
2	25/197 (13)	26/160 (16)	51/357 (14)
1 and 2 combined	37/320 (12)	36/232 ^a (16)	73/552 (13)

^a By comparison of results with homologous and heterologous serovars, $\chi^2_{1\text{ df}} = 1.833$ and *P* = 0.18.

TABLE 4. False-negative results reported by participants in round 2 of the leptospirosis MAT proficiency testing scheme related to geographic region and to participation in round 1 of the scheme

Region	No. of participants		No. of tests		No. of tests with negative results		% of tests with false-negative results		No. of false-negative results per participant	
	Ongoing	New	Ongoing	New	Ongoing	New	Ongoing	New	Ongoing	New
Australasia	2	1	10	2	0	0	0	0	0	0
North America	2	3	12	14	0	1	0	7	0	0.33
Europe	21	7	131	40	13	5	10	13	0.62	0.71
Asia and Pacific	3	6	11	30	2	6	18	20	0.67	1.00
South and Central America	6	9	35	66	4	20	11^a	30^b	0.67	2.22
Total	34	26	199	152	19	32	10	21^c	0.56	1.23^d

^a Significant differences between ongoing and new participants are shown in boldface.

^b $\chi^2_{1 \text{ df}} = 4.498$; $P = 0.034$.

^c $\chi^2_{1 \text{ df}} = 9.185$; $P = 0.002$.

^d $\chi^2_{1 \text{ df}} = 4.164$; $P = 0.041$.

(round 1, samples B and E; round 2, sample A; and round 2, sample E mixed with antiserum against serogroup Javanica, but including only tests for serogroup Icterohaemorrhagiae). Table 5 shows that when these sera were tested, participants in both rounds reported 15% false-negative results in round 1 but 5% false-negative results in round 2 ($\chi^2_{1 \text{ df}} = 4.837$; $P = 0.028$).

Identification of the correct serogroup. In round 1, the correct serogroup was identified by laboratories in 86% of cases. Twenty-one of 37 laboratories (57%) correctly identified all four serogroups. In round 2, the correct serogroup was identified by laboratories in 82% of cases, and 39 of 60 laboratories (65%) correctly identified all serogroups in the three unmixed samples.

Table 6 compares the success of serogroup identification in round 2 for participants in the five geographic regions in relation to whether or not they had previously participated in round 1. The serogroup was incorrectly identified in 19% of the cases by the new participants, whereas the serogroup was incorrectly identified in 9% of the cases by continuing participants, but this difference was not significant ($\chi^2_{1 \text{ df}} = 3.254$; $P = 0.071$).

Quantitative analysis of titers. Table 7 shows the titers reported for round 2 samples A, B, and C, tested with the corresponding homologous serovars (Canicola, Grippotyphosa, and Sejroe), after logarithmic transformation. For each sample, new participants reported titers that were overall lower

($t_{103 \text{ df}} = 2.37$; $P = 0.019$) and more variable ($F = 12.06$; $P = 0.001$) than those reported by continuing participants.

DISCUSSION

Definition of a positive titer. For the purposes of this study, a titer ≥ 80 was defined as positive. This arbitrary definition bears no relationship to the diagnostic significance of a MAT titer. Whether a particular titer indicates leptospirosis may depend on whether paired samples show that the titer is rising and may also depend on the magnitude of the current titer. The diagnostic significance of a positive titer may also vary with the infecting serovar, the infected host species, vaccination history, and the local prevalence of leptospirosis.

Definition of a false-negative result. The definition of a false-negative result likewise deserves discussion. For the purposes of this proficiency testing program, negative titers for any serovar within the serogroup for which the sample was positive were defined as false negative. Strictly, the term "false negative" relates only to results negative for the actual serovar against which the sample was raised. Under some circumstances, the MAT result may be genuinely negative for a heterologous serovar within the same serogroup.

Nevertheless, it is usual for positive results to be obtained when a sample containing antibodies directed against a particular serovar is tested by using another serovar in the same

TABLE 5. False-negative results reported by round 2 participants in rounds 1 and 2 of the leptospirosis MAT proficiency testing scheme for antisera tested in both rounds

Parameter	Ongoing participants in round 2	New participants in round 2
No. of participants	34	26
Round 1		
No. of participants reporting more than one false-negative result/total no. of participants (%)	8/34 (24)	
No. of false-negative reported/total no. of results (%)	13/87 (15)^{a,b}	
Round 2		
No. of participants reporting more than one false-negative result/total no. of participants (%)	4/34 (12)	7/26 (27)
No. of false-negative results reported/total no. of results (%)	4/83 (5)^c	10/67 (15)

^a Significant differences between round 1 and round 2 are shown in boldface.

^b Difference between rounds for ongoing participants $\chi^2_{1 \text{ df}} = 1.619$ and $P = 0.203$.

^c Difference between rounds for ongoing participants $\chi^2_{1 \text{ df}} = 4.837$ and $P = 0.028$.

TABLE 6. Identification of the incorrect serogroup in round 2 of the leptospirosis MAT proficiency testing scheme related to geographic region and to participation in round 1 of the scheme

Region	No. of participants		No. of determinations		No. of incorrect determinations		% Incorrect determinations		No. of incorrect determinations per participant	
	Ongoing	New	Ongoing	New	Ongoing	New	Ongoing	New	Ongoing	New
Australasia	2	1	6	2	0	0	0	0	0	0
North America	2	3	6	9	0	0	0	0	0	0
Europe	21	7	63	21	6	4	10	19	0.29	0.57
Asia and Pacific	3	6	6	15	2	2	33	13	0.67	0.33
South and Central America	6	9	15	27	1	8	7	30	0.17	0.11
Total	34	26	96	74	9	14	9	19 ^a	0.26	0.54

^a $\chi^2_{1\text{ df}} = 3.254; P = 0.071.$

serogroup as the diagnostic antigen (3, 4). Likewise, it is uncommon for a titer to be obtained when the sample is tested with a serovar in a different serogroup. The results reported in the first two rounds of this proficiency testing scheme support the general applicability of these two generalizations. Thus, broad comparisons based on the definition of false negative used in this study can be considered valid.

Impact of proficiency testing. This proficiency testing scheme has grown progressively over the first three rounds, with 75 participants in round 3 (2004), compared with 37 participants in round 1 and 60 participants in round 2. This growth and the feedback received indicate that many participants find it to be of value.

It can be seen from the results of the first two rounds of the proficiency testing scheme that laboratories performing the leptospirosis MAT quite often report incorrect results. Sometimes participating laboratories reported false-negative results, incorrectly identified the serogroup for which a sample was positive, or failed to identify a positive sample as positive for any serogroup. Participants also reported widely varying titers for a particular sample tested with a specific serovar. Contamination, misidentification, and deterioration of antigen cultures are among the factors that can contribute to diagnostic errors by MAT. Laboratories can reduce errors by carefully monitoring the identities of cultures and periodically seeking replacement cultures from reference laboratories.

Improved quality assurance needs to be applied to MAT in many or all laboratories. Proficiency testing is one of a number of ways in which the quality of serological testing can be monitored and improved.

It is not easy to demonstrate objectively that proficiency test-

ing has a positive impact. A proficiency testing scheme offered to a group of laboratories is not a controlled experiment.

An indication of the impact of this proficiency testing scheme can be seen when we compare two groups of participants who participated in round 2. Laboratories participating for the first time reported significantly more false-negative results than laboratories that had previously participated in round 1 of the scheme. They also reported titers that were overall significantly lower ($P = 0.019$) and significantly more variable ($P = 0.001$). New participants also apparently made more errors in serogroup identification, but this difference was not significant.

The key question is whether, in terms of false-negative results, the laboratories that had participated previously performed better overall because of their previous participation in this scheme or because they had a different profile of characteristics. One characteristic that can be analyzed is geographic location. There was some shift in the geographic locations of the participants in round 2 compared with those of the participants in round 1 (Table 1); for example, round 2 included a smaller proportion of laboratories from Europe and a larger proportion from the Americas and from Asia and the Pacific. Far more false-negative results were reported from some regions than others (Table 4). However, ongoing participants performed better across regional groupings. This difference was statistically significant within the laboratories from South and Central America, although only 15 laboratories were involved in this comparison.

The inclusion of two antisera in both rounds made it possible to compare the rates of false-negative results between the two rounds. A significant improvement was demonstrated by on-

TABLE 7. Analysis of titers reported for round 2 samples A, B, and C tested with the homologous serovar in relation to participation in round 1 of the scheme^a

Parameter	Sample A		Sample B		Sample C	
	Ongoing	New	Ongoing	New	Ongoing	New
No. of tests	33	26	29	23	20	12
Mean titer	3.48	3.11	3.10	2.80	2.76	2.43
SD of titer	0.48	0.86	0.61	0.77	0.57	0.90
Coefficient of variation (%)	13.8	27.5	19.7	27.6	20.7	37.0
95% confidence limits of mean	3.31–3.65	2.77–3.45	2.87–3.33	2.48–3.12	2.50–3.02	1.91–2.95

^a Titers are expressed as logarithms to base 10. Overall, for three samples, new participants reported titers that were significantly lower ($t_{103\text{ df}} = 2.37; P = 0.019$) and significantly more variable ($F = 12.06; P = 0.001$) than those reported by the ongoing participants.

going participants when these antisera were tested in round 2 (Table 5).

Although the comparisons reported here do not provide positive proof, they are consistent with the proposition that this proficiency testing scheme has had a definitive impact on testing quality.

ACKNOWLEDGMENTS

We thank Daniel W. Tholen, consultant expert of the Public Health Program and Practice Office, Division of Laboratory Systems, Centers for Disease Control and Prevention, and Barbara H. Francis of the

National Serology Reference Laboratory, Australia, for advice and assistance with statistical analysis.

REFERENCES

1. **Bharti, A. R., J. E. Nally, J. N. Ricaldi, M. A. Matthias, M. M. Diaz, M. A. Lovett, P. N. Levett, R. H. Gilman, M. R. Willig, E. Gotuzzo, and J. M. Vinetz.** 2003. Leptospirosis: a zoonotic disease of global importance. *Lancet Infect. Dis.* **3**:757-771.
2. **Cole, J. R., Jr., C. R. Sulzer, and A. R. Pursell.** 1973. Improved microtechnique for the leptospiral microscopic agglutination test. *Appl. Microbiol.* **25**:976-980.
3. **Faine, S., B. Adler, C. Bolin, and P. Perolat.** 1999. *Leptospira* and leptospirosis, 2nd ed. MediSci, Melbourne, Victoria, Australia.
4. **Levett, P. N.** 2003. Usefulness of serologic analysis as a predictor of the infecting serovar in patients with severe leptospirosis. *Clin. Infect. Dis.* **36**:447-452.