Lepr Rev (1999) **70,** 305–313

Reliability of Semmes Weinstein monofilament and ballpoint sensory testing, and voluntary muscle testing in Bangladesh

ALISON M. ANDERSON* & RICHARD P. CROFT**
**INF RELEASE Project, PO Box 5, Pokhara, Nepal*
***Danish Bangladesh Leprosy Mission, Nilphamari 5300, Bangladesh*

Accepted for publication 5 July 1999

Summary The reliability of methods of testing nerve function is important, since diagnostic decision making is a direct function of the quality of the test. Three methods of nerve function testing were investigated at the Danish Bangladesh Leprosy Mission (DBLM) in north Bangladesh, and assessed for inter-observer reliability. The three methods were 1) ballpoint pen test (BPT) for sensory function; 2) graded Semmes Weinstein monofilament test (SWM) for sensory function and 3) voluntary muscle testing (VMT) for motor function. The weighted kappa (κ_w) statistic was used to express inter-observer reliability. Using this statistic, 0 represents agreement no better than random, and 1·0 complete agreement. κ_w values of $\geq$0·80 are reckoned to be adequate for monitoring and research. Fifty-three patients were tested, a Senior physiotechnician acting as 'gold standard' against whom four other staff physiotechnicians were assessed. All three testing methods were found to have minimal inter-observer variation, with the κ_w for inter-observer agreement using BPT being 0·86, the SWM 0·92, and VMT 0·94. It is concluded that in trained and experienced hands, all three methods are reliable and repeatable to a level allowing confident use of results obtained in monitoring and research.

Introduction

Routine assessment of nerve function in leprosy patients is essential for the early detection of nerve function impairment and its treatment.[1,2] Recently, van Brakel has drawn attention to the need for the measurements used to test nerve function to themselves be the subject of reliability testing.[3] It is empirically true that the results of any measurement cannot be better than the measure itself.

While there is general agreement about the use of the modified MRC scale for measuring muscle strength,[4,5] there is less agreement about the preferred method for sensory testing. Owen and Stratford[6] reviewed several commonly used methods and concluded that the WHO (ballpoint) test, cotton wool and pinprick were all cheap and easy to use but were not

Correspondence to: R. P. Croft, 70 Culver Lane, Reading, RG6 1DY, UK, e-mail: richard@crofts32.freeserve.co.uk

0305-7518/99/070305+09 $1.00

sensitive enough to be of practical value. They found monofilaments and the biothesiometer to be reliable pieces of equipment, but the latter are expensive and dependent on electricity.

Perhaps the two methods that are most commonly used for sensory testing are the Ballpoint Pen test (BPT) described by Jean Watson[7] and the Semmes Weinstein Monofilaments (SWM) test described by Judith Bell-Krotoski.[8] Both methods have their advocates and critics. The BPT is advocated on the grounds that is cheap and readily available, and criticised on the grounds that the force applied may vary considerably and therefore the results are likely to be unreliable.[6] Also, since it is a threshold (yes/no only) test, it relies on a count of the number of sites at which gross sensation is lost to provide an indication of the level of sensory loss. On the other hand, SWM testing is advocated on the grounds that the results are reliable, since the force required to bend the accurately manufactured monofilaments is relatively constant and repeatable,[9] and since they are a graded test they provide a quasi-quantitative estimate of sensory loss. The SWM test is sometimes criticized on the grounds that the monofilaments used less easily available, and too 'technical' and time-consuming for widespread use.

Lienhardt, Currie and Wheeler carried out inter-observer testing using BPT, SWM and voluntary muscle testing (VMT) in Ethiopia.[10] They found a 32–58% agreement using SWM with a weighted kappa (κ_w) statistic of 0·736–0·814, indicating good agreement ($\kappa_w \geq 0{\cdot}60$ indicates good agreement, see explanation in the Methods section of this paper[11]) but with wide confidence intervals. With the BPT the agreement was 71–84%, κ_w 0·604–0·793; and 79–98% agreement for the VMT (κ_w could not be determined for all tests). Van Brakel[3] in a similar kind of study assessed the reliability of SWM testing, moving touch sensibility and pinprick testing in Nepal. He found the intra-observer κ_w for SWM to be very good at 0·83–0·92, inter-tester κ_w agreement for moving touch sensibility 0·75–0·82, and the pinprick 0·54–0·82. He concluded that the SWM and moving touch sensibility tests were both suitable for reliable sensory testing.

This study was performed to check the reliability of a core team of physiotechnicians at the Danish Bangladesh Leprosy Mission (DBLM) in the use of the most common diagnostic tests for leprosy related nerve impairment. The reliability study was performed after training in the one new test, SWM, and after refresher training in the other two (BPT and VMT).

The results of the reliability testing were to be used as follows:

1. To help the project decide whether to change from BPT and SWM for routine use in sensory testing.
2. To give the project confidence in the use of these tests as outcome measures in research running at the project (Bangladesh Acute Nerve Damage Study BANDS, and Trials of Prevention of Disability, TRIPOD).

Method

The percentage of occasions on which direct agreement is obtained can be measured. This provides a simple useful indicator of the reliability of the test. The statistic of choice for measurement of reliability is the weighted kappa (κ_w). The use of weighting ensures that where operators disagree on the result, bigger disagreements have a bigger effect on the κ_w. κ_w ranges from 0 (agreement no better than random), to 1 (perfect agreement). A target of a κ_w of at least 0·60 (Altman – good agreement[11]) was believed at the outset to be achievable

immediately post-training for the SWM test. 'Good agreement' would be required for any test to be implementable. A κ_w higher than this ($\geq 0{\cdot}80$) would be desirable for monitoring and research work.

The DBLM project at Nilphamari used the BPT and VMT as components of their testing for nerve function impairment (see Appendix). Patients were treated with corticosteroids on the basis of a locally developed scoring method. This composite score was made up of one point for every point of lost sensation (as determined by the BPT), plus one point for every grade lost on the MRC scale of muscle strength. The composite score obtained was therefore made up of both sensory and motor loss in any of the nerves tested. Subjects whose composite score increased by 2 or more points within a 6-month period were treated for nerve function impairment.[12] Thus patients were treated on the basis of 2 or more points of sensory loss only, or 2 or more points of lost muscle strength only, or a combination of at least one point of sensory loss and 1 point of lost muscle strength. The Appendix gives details of testing procedures used.

For this project, a knowledge of the comparative reliability of the two types of sensitivity testing was important in choosing the best test for the project. For tests like the SWM and BPT, the reliability depends on the patient, the operator and on the equipment. This study examined inter-tester reliability only, assuming this to be the major source of test variability. The other potential sources of test variability are intra-tester and equipment reliability, but these were not tested in this study, since it was assumed that they would be at least as good as the inter-tester reliability.

SAMPLE SIZE

Fifty-three patients were each tested twice by a pair of operators. Four physiotechnicians were used in the study, as well as the senior physiotechnician against whom, as the gold standard, each operator's results were checked. Each pair of operators tested a minimum of 10 patients. Patients were tested with the two types of sensory test (SWM and BPT) in random order, and the VMT always performed during or after the sensory testing.

PATIENT SELECTION

Patients selected were known to have at least one nerve trunk impaired. The target was to test a selection of all grades of impairment within the sensory and motor testing scales. The patients were selected from among hospital inpatients at the DBLM hospitals in Nilphamari and Rangpur, and at some outpatient clinics. In order to maximize the number of patients available, some patients with missing limbs were included, but a total of 304 nerves were tested. One patient declined further testing after BPT, resulting in only 300 nerves being tested using SWMs.

TRAINING

For the SWM test, $2\frac{1}{2}$ days of training were given by one author, AMA, to all testers. The procedure was written in training notes. A set of six graded Semmes Weinstein monofilaments was used. These filaments were obtained from Carville, USA and were designed to bend slightly when forces of 70 mg, 200 mg, 2 g, 4 g, 10 g and 300 g are respectively applied. The filaments were individually touched on the skin until they bent slightly, and then withdrawn. The patients were asked to point to where they felt the stimulus applied. If the

filament was not felt, the next thickest one was used until a level was found where a filament could be felt. Details of filaments used on the hands and feet are given in the Appendix.

For BPT and VMT a short 1-h refresher training was given by the senior physiotechnician to ensure that all staff were familiar with the test criteria. DBLM has used BPT and VMT for 5 years regularly on all patients both in clinic and inpatient situations, so staff are very experienced (see Appendix for details of tests). All testers participated in the testing of some pilot patients, to ensure they were confident in the study procedure and recording of results.

TESTING

Patients were selected by the testers themselves, asked for their consent and cooperation and brought to the test site. They were allocated a study number and tested by the testers in the order given in the randomization. Pairs of testers took turns to test and observe the study, each tester testing three to five patients on one occasion, watching for some sessions, then testing a further block of patients later on.

Each test was performed with the tester blind to previous results. At the end of each block of testing (morning, afternoon) any results showing unusually large discrepancies were discussed, along with possible methods for improving testing technique. Although such feedback was given the data obtained were still included, unchanged.

DATA RECORDING

Test results were recorded on forms separate from the patient record. A new form was used for each tester for each patient, to ensure that blinding occurred. Patient details (age, sex, registration number etc.) were also recorded.

DATA HANDLING

Since the data for the several points of the hand and foot are not independent, only two points for each hand and one for each foot (corresponding to the three nerves being tested) were used in calculations. The chosen sites were the first and fifth metacarpal heads and the first metatarsal head. All data was entered into an Epi Info database and exported to Stata software for the calculation of κ_w.

PATIENT TREATMENT

Where sensory loss was found by the BPT, or motor loss which had not previously been recorded on the patient card, the patient was asked to describe the duration of the impairment. The standard DBLM criteria for prednisolone prescription were used, and patients who met these criteria were referred for treatment.

Results

SENSORY AGREEMENT

Tables 1 and 2 show the agreement between the senior physiotherapist and all staff members for BPT and SWM tests. Table 3 shows the absolute agreement and κ_w for these two tests, and Table 4 the absolute agreement and κ_w by testing pair.

Table 1. Agreement between senior physiotechnician and other physiotechnicians using the ballpen test (BPT)

	Other Physios			
Senior physio	0	1	2	Total
0	85	3	14	102
1	1	2	3	6
2	1	4	191	196
Total	87	9	208	304

0 = Ballpen not felt (complete anaesthesia).
1 = Ballpen felt uncertainty (partial anaesthesia).
2 = Ballpen felt normally (normal sensation).

Table 2. Agreement between senior physiotechnician and other physiotechnicians using the Semmes Weinstein monofilaments

	Other physiotechnicians						
Senior physio	0	1	2	3	4	5	Total
0	81	2	4	1	1	0	89
1	5	6	7	1	0	1	20
2	0	4	7	2	3	2	18
3	1	1	4	16	10	3	35
4	0	1	1	7	20	8	37
5	0	0	0	5	11	85	101
Total	87	14	23	32	45	99	300

The numbers 0–5 indicate levels that the at which the different monofilaments were felt.

Different sets of filaments were used for hands and feet. Details are given in the Appendix.

Table 3. Absolute agreement and weighted kappa (κ_w) with confidence intervals, for the ballpoint and Semmes Weinstein tests

Test	*n*	Absolute agreement	Agreement within 1 grade	κ_w	95% CI
BPT	304	91%	95%	0·86	0·75–0·97
SWM	300	72%	92%	0·92	≥0·80

Table 4. Absolute agreement and κ_w by testing pair for the ballpoint and Semmes Weinstein tests

Test	Testers	*n*	Absolute agreement	Agreement within 1 grade	κ_w	95% CI
BPT	A	69	93%	94%	0·89	≥0·65
	B	68	91%	97%	0·89	≥0·65
	C	87	95%	98%	0·92	≥0·70
	D	80	86%	90%	0·77	0·66–0·88
SWM	A	69	78%	96%	0·96	≥0·70
	B	68	71%	96%	0·95	≥0·70
	C	83	65%	83%	0·87	≥0·65
	D	80	74%	88%	0·90	≥0·65

Table 5. Agreement and κ_w for the individual muscles in the VMT

Test	*n*	Absolute agreement	Agreement within 1 grade	κ_w	95% CI
Tight eye closure	105	90%	96%	0·80	0·62–0·98
Little finger abduction	101	86%	98%	0·94	≥0·72
Thumb abduction	102	90%	98%	0·94	≥0·72
Wrist extension	103	98%	100%	–	–
Dorsiflexion of foot	95	98%	99%	0·89	≥0·67
Eversion of foot	94	98%	100%	0·88	≥0·67
All tests except eye	495	93%	98%	0·94	≥0·84

MOTOR TESTING

Table 5 shows the percentage agreement and κ_w for the individual muscles tested in the VMT. The muscles tested did not cover the full range of the possible grades, therefore in some cases κ_w is not calculable. Table 6 shows the agreement by tester.

Discussion

These data suggest that both sensory tests and the VMT can be performed by this team of testers to an excellent standard of reliability, using a sample of patients similar to the

Table 6. Absolute agreement by tester for the VMT

Tester	*n*	Absolute agreement	Agreement within 1 grade	κ_w	95% CI
A	110	95%	98%	0·94	≥0·73
B	109	91%	99%	0·93	≥0·73
C	144	92%	98%	0·89	≥0·72
D	132	92%	99%	0·95	≥0·77

population they normally test. The BPT showed a κ_w for inter-observer agreement of 0·86, with the weakest operator having an individual κ_w of 0·77. This is an encouraging result, indicating that not only can the BPT be performed to a high degree of reliability, but that in DBLM itself it is a reliable and repeatable test. The reliability of the monofilament was found to be even higher, with a κ_w for inter-observer agreement of 0·92, the weakest tester's individual κ_w being 0·87. This result is excellent, and staff could be expected to improve still further in reliability and specificity with more experience. While both methods of testing were found to be reliable, it can be expected that SWM testing will be more consistently reliable since the force applied by each monofilament is limited by bending. However, whilst inter-observer reliability was good for both methods, the 5-point SWM testing method provides richer data than the BPT since it gives a semi-quantitative assessment of sensory loss.

The reliability of motor testing was also found to be very good. The κ_w for inter-observer agreement was overall 0·94, excluding eye strength testing, and the poorest individual tester's result was 0·89.

In terms of the first aim of testing as described in the introduction, the project decided to switch over to SWM testing as a routine for sensory testing. It also enables data from DBLM to be comparable with data from other projects using SWMs.

Secondly, the exercise boosted the project's confidence in all three methods of testing. This lends weight to the results from the Bangladesh Acute Nerve Damage Study.

Whilst this study indicates that high levels of reliability can be obtained from the three nerve function testing methods in common use, it must be emphasized that this followed a long period of experience with BPT sensory and VMT methods, and $2\frac{1}{2}$ days of training for SWM testing. Such levels of reliability may not be found amongst workers with less experience and training.

In conclusion, ballpoint pen and Semmes Weinstein Monofilament sensory testing techniques for assessing sensory function, and standard voluntary muscle testing for motor function testing are reliable test methods in trained hands. Results obtained may be used with confidence for monitoring and research.

References

[1] Becx-Bleumink M, Berhe D, Mannetje WT. The management of nerve damage in the leprosy control services. *Lepr Rev*, 1990; **61:** 1–11.

[2] Lienhardt C, Fine PEM. Type 1 reaction, neuritis and disability in leprosy. What is the current epidemiological situation? *Lepr Rev*, 1994; **65:** 9–33.

[3] Van Brakel WH, Khawas IB, Gurung KS, Kets CM, van Leerdam ME, Drever W. Intra- and Inter-tester reliability of sensibility testing in leprosy. *Int J Lepr*, 1996; **64:** 287–298.

[4] MRC. *Aids to the investigation of peripheral nerve injuries.* Memo no. 7, 2nd edition, London: HMSO 1962 (quoted in Lienhardt and Fine[2]).

[5] Brandsma W. Basic nerve function assessment in leprosy patients. *Lepr Rev*, 1981; **52:** 161–170.

[6] Owen BM, Stratford CJ. Assessment of the methods available for testing sensation in leprosy patients in a rural setting. *Lepr Rev*, 1995; **66:** 55–62.

[7] Watson JM. *Essential action to minimise disability in leprosy patients.* London: The Leprosy Mission International, 1986.

[8] Bell-Krotoski JA. 'Pocket' monofilaments and specifications for the Semmes-Weinstein monofilaments. *J Hand Ther*, 1990; **3:** 26–31.

[9] Bell-Krotoski JA, Tomancik E. The repeatability of testing with Semmes-Weinstein monofilaments. *J Hand Surg*, 1987; **12A:** 155–161.

[10] Lienhardt C, Currie H, Wheeler JG. Inter-observer variability in the assessment of nerve function in leprosy patients in Ethiopia. *Int J Lepr*, 1995; **63:** 62–76.
[11] Altman DG. *Statistics for medical research.* London: Chapman and Hall, 1991.
[12] Croft RP, Richardus JH, Smith WCS. The effectiveness of corticosteroids in the treatment of long-term nerve function impairment. *Lepr Rev*, 1997; **68:** 316–325.

Appendix

Details of motor and sensory testing used in the study.

a Modified 5-point MRC scale for muscle strength scoring[4,5]

Hands and feet	MRC grade	Eyes
Full ROM[1], full resistance	5	Normal muscle strength
Full ROM, reduced resistance	4	Closes, stays closed against some resistance
Full ROM, no resistance	3	Closes, no resistance[2] (may be gap)
Reduced ROM, some joint movement	2	Gap on strong closure[2]
Flicker only	1	Flicker only
Full paralysis	0	Complete paralysis

[1]ROM: Range Of Movement.
[2]In addition, lid gap in mm is measured and recorded.

b Movements/muscles tested

Nerve	Movement	Muscle/muscle group
Ulnar	Little finger abduction	Abductor digiti minimi
Median	Thumb abduction	Abductor pollicis brevis
Radial	Wrist extension	Wrist extensors
Lateral popliteal	Foot dorsiflexion	Foot dorsiflexors
Facial	Close eyes	Orbicularis oculi

c Ballpoint pen testing technique

The skin is gently dented using an upright, ordinary ballpoint pen to create a dimple of approximately 1 cm across. The subject is asked to point to the place where he feels the sensation, whenever he feels a touch. Accurate pointing to within approximately 3 cm of the point touched is taken to indicate *normal sensation*. Some sensation, but inability to point to within 3 cm is taken as *partial anaesthesia*. Inability to feel anything is taken as *complete anaesthesia.*

d Filaments used in sensory testing of hands and feet

Filaments used for palms of hands*	Filaments used for soles of feet*	Level of sensation (see Table 2)
(No filament felt)	(No filament felt)	0
300 g	300 g	1
4 g	10 g	2
2 g	4 g	3
200 mg	2 g	4
70 mg	200 mg	5

*The weights given indicate the force at which the monofilament will bend.