THE RELIABILITY OF PHYSICAL EXAMINATION FOR CARPAL TUNNEL SYNDROME

R. G. MARX, P. L. HUDAK, C. BOMBARDIER, B. GRAHAM, C. GOLDSMITH and J. G. WRIGHT

From the University of Toronto, Toronto and McMaster University, Hamilton, Canada

The goal of this study was to determine the interobserver and intraobserver reliability of static and moving two-point discrimination, Semmes-Weinstein monofilament testing, Tinel's test, manual motor testing of abductor pollicis brevis, vibration and Phalen's test in the diagnosis of carpal tunnel syndrome. Twelve patients with suspected carpal tunnel syndrome were examined in an outpatient setting. The interobserver reliability was satisfactory for all tests except for Semmes-Weinstein monofilament testing. Intraobserver reliability was also satisfactory for all tests. Static two point discrimination had higher reliability than moving two-point discrimination. Seven tests for the diagnosis of carpal tunnel syndrome were reliable in the hands of skilled health care professionals. Hand surgeons and hand therapists examined patients more reliably than occupational health workers.

Journal of Hand Surgery (British and European Volume, 1998) 23B: 4: 499–502

Physical examination is used to evaluate patients who may have carpal tunnel syndrome. The diagnosis of carpal tunnel syndrome is generally based on history, physical examination, electromyography and nerve conduction studies. Physical examination is used both to diagnose the condition and to grade the severity of any associated nerve dysfunction.

Measurements used to evaluate patients should be reliable and valid (Feinstein, 1987). Reliability is the extent to which repeated measurements of a relatively stable phenomenon fall closely to each other (Feinstein, 1987; Fletcher et al, 1988; Wright and Feinstein, 1992). Validity is the degree to which the results of measurement correspond to the true state of the phenomenon being measured (Fletcher et al, 1988). The reliability and validity directly influence the quality of the information gathered (Fess, 1986). A test for physical examination that is not reliable both for a given observer and between observers, is of limited or no use. Conversely, a test that is documented to have high intraobserver and interobserver reliability will yield useful information providing the test is valid.

The validity of the physical examination for carpal tunnel syndrome has been extensively investigated (De Krom et al, 1990; Durkan, 1991; Gellman et al, 1986; Gilliatt and Wilson, 1953; Golding et al, 1986; Gonzalez Del Pino et al, 1997; Gunnarsson et al, 1997; Katz et al, 1990; Kuschner et al, 1992; Phalen, 1972; Pryse-Phillips, 1984; Stewart and Eisen, 1978; Szabo et al, 1984). However, the reliability of physical examination in the evaluation of patients with carpal tunnel syndrome is unknown. The purpose of this study was to determine the reliability of physical examination for carpal tunnel syndrome when done by occupational health workers, hand therapists and hand surgeons in an outpatient setting.

PATIENTS AND METHODS

Twelve patients referred to a tertiary care centre for possible carpal tunnel syndrome were asked to return for one afternoon of testing. Patients were examined at the time of study enrolment by a physician with experience in carpal tunnel syndrome assessment to ensure that patients with and without positive findings on the tests being studied were included in the study group. All patients underwent electrophysiological testing which revealed six patients with carpal tunnel syndrome, four patients with generalized sensory neuropathy, one patient with cubital tunnel syndrome and one patient with generalized diabetic neuropathy. Informed consent was obtained for all patients and patients were reimbursed for their travel expenses. The mean age of the 12 patients (six men and six women) was 55 years (range, 31–77).

Examiners with varying backgrounds were selected to determine the effect of clinical experience and training on reliability of physical examination. The six examiners consisted of two hand surgeons, two hand therapists (working in tertiary care hand centres) and two occupational health workers with experience in interviewing patients with work-related musculoskeletal disorders but with no experience of examining patients. Examiners were blinded to the diagnoses of the patients and had had no previous contact with them.

One week before the study, all examiners met to review the examination protocol. The seven tests studied included moving and static two-point discrimination, Semmes-Weinstein monofilament testing, vibration sense, motor power, Phalen's test and Tinel's test. A description of each test was provided and all examiners were instructed in order to standardize the protocol. The tests were also re-reviewed with the examiners on the day of the study.

Each examiner performed all seven tests on each of the 12 patients according to an extra period Latin square design (Hoaglin et al, 1991). This study design allows the determination of the effect of the patient, the examiner and the order in which the test was performed on the reliability of the test. Each patient was evaluated a total of seven times (once by six of the examiners and twice by one examiner). The final examination was performed by

the examiner who first examined the patient to determine intraobserver variation. The patients were randomly assigned to the order of tests and each test was done an equal number of times first, second, third and so on. Tests were done one after another, with a 1 min rest interval following Phalen's tests in an attempt to avoid a potential "carry-over" provocative effect secondary to transiently impaired nerve function.

Patients were instructed to answer only questions that pertained directly to the physical examination and not to comment on their previous examinations. The examiners did not reveal their area of expertise to the patients and did not make reference to previous examiners.

The intraobserver and interobserver variability for two-point discrimination (moving and static) and Semmes-Weinstein monofilament testing was determined using the random-effects intraclass correlation coefficient (Fleiss, 1986). The intraclass correlation is an index of agreement for continuous data. An intraclass correlation coefficient of < 0.40 indicates poor reproducibility, 0.41 to 0.75 fair to good reproducibility and > 0.75 excellent reproducibility (Fleiss, 1986). The intraobserver and interobserver variability for Tinel's test, motor power, vibration and Phalen's test were determined using the Kappa statistic. The Kappa statistic is an index of agreement for ordinal data. A Kappa between 0 and 0.20 is considered poor agreement, 0.21 to 0.40 fair, 0.41 and 0.60 moderate, 0.61 and 0.80 substantial and > 0.80 almost perfect (Landis and Koch, 1977). All reliability calculations were made for the whole group of examiners combined, as well as separately for the surgeons, therapists and occupational health workers (OHW).

RESULTS

All patients were examined by all examiners. There were no missing values as all patients were examined the appropriate number of times. The overall intraclass correlation coefficient (ICC) for interobserver reliability was "fair" (0.45) for moving two-point discrimination and "good" (0.66) for static two-point discrimination (Tables 1 and 2). In both instances the occupational health workers had a lower ICC than the therapists and surgeons (Tables 1 and 2). The interobserver reliability for Semmes-Weinstein monofilament testing was "poor" (ICC = 0.15, overall; Table 3). The ICC was lowest for the therapists. Interobserver reliability was "substantial" for Tinel's test and Phalen's test, "moderate" for vibration (mean pairwise Kappa 0.40) and "fair" for motor power (mean pairwise Kappa 0.25; Table 4). The mean pairwise Kappa was higher for all four tests when calculated for only the surgeons and therapists (mean pairwise Kappa varying from 0.50 to 0.83; Table 4).

The intraobserver reliability was "good to excellent" for moving and static two-point discrimination as well as for Semmes-Weinstein monofilaments, with all three groups of examiners showing acceptable reliability (Table 5). Intraobserver reliability of Tinel's test, Phalen's test, motor power and vibration were also "moderate" to "almost perfect" with Cohen's Kappa ranging from 0.53 to 1.0 (Table 6). The therapists had the highest agreement as their Kappa was 1.0 for all four tests (Table 7). For the three tests with a continuous outcome where the intraclass correlation coefficient was used, the order in which the tests were performed did not have an effect on the reliability.

DISCUSSION

This study evaluated the reliability of seven standard tests in the assessment of patients with carpal tunnel syndrome and investigated the reliability of different kinds of examiners. Moving and static two point discrimination, performed using the discriminator (Mackinnon and Dellon, 1985), had fair to good interobserver and intraobserver reliability (Tables 1, 2 and 5). Interobserver reliability was higher for health care professionals than for the occupational health workers, who had had no previous experience examining patients, while intraobserver reliability was similar in all three groups of

Table 1—Interobserver reliability for moving two-point discrimination

Examiners	*% variation due to patients*	*% variation due to order*	*% variation due to raters*	*% variation due to error*	*Intra-class correlation coefficient*
Overall	45.3	0.6	5.9	48.2	0.45
Therapists	53.1	11.6	9.2	26.1	0.53
Surgeons	44.3	6.7	10.7	38.3	0.44
OHW	37.0	16.0	9.8	37.2	0.37

Table 2—Interobserver reliability for static two-point discrimination

Examiners	*% variation due to patients*	*% variation due to order*	*% variation due to raters*	*% variation due to error*	*Intra-class correlation coefficient*
Overall	66.2	0.5	3.2	30.1	0.66
Therapists	70.3	4.0	2.9	22.8	0.70
Surgeons	85.1	3.4	0.0	11.5	0.85
OHW	49.5	5.9	17.8	26.7	0.50

Table 3—Interobserver reliability for Semmes-Weinstein monofilament testing

Examiners	*% variation due to patients*	*% variation due to order*	*% variation due to raters*	*% variation due to error*	*Intra-class correlation coefficient*
Overall	14.6	0.0	30.0	55.4	0.15
Therapists	0.0	0.0	4.2	95.8	0.00
Surgeons	11.4	0.0	32.4	56.2	0.11
OHW	43.0	1.2	42.1	13.7	0.43

Table 4—Interobserver reliability for Tinel's test, motor power, vibration and Phalen's test

Test	*Mean pairwise Kappa, Overall (asymptotic standard error)*	*Mean pairwise Kappa, Surgeons and therapists only (asymptotic standard error)*
Tinel's test	0.77 (0.15)	0.79 (0.18)
Motor power	0.25 (0.24)	0.50 (0.23)
Vibration	0.40 (0.14)	0.51 (0.24)
Phalen's test	0.65 (0.19)	0.83 (0.19)

Table 5—Intraobserver reliability for moving two-point discrimination, static two-point discrimination and Semmes-Weinstein monofilament testing

Test	*% variation due to patients*	*% variation due to order*	*% variation due to error*	*Intra-class correlation coefficient*
2 point moving Overall	57.6	0.0	42.4	0.58
2 point static Overall	76.9	0.6	22.5	0.77
SW monofilaments Overall	71.1	0.0	28.9	0.71

observers. Both interobserver and intraobserver reliability were higher for static two-point discrimination compared with moving two point discrimination (Tables 1, 2 and 5).

Intraobserver reliability for Semmes-Weinstein monofilament testing was good to excellent for all three groups. However, Semmes-Weinstein monofilament testing had very low overall interobserver reliability (Table 3). Thus, although the same examiner can be expected to obtain the same results on re-examination, the results vary substantially between examiners, raising concerns about the usefulness of this test. Interestingly, this was the one test where the occupational health workers substantially out-performed the therapists and the surgeons. This difference between observers likely arose due to chance, as the proportion of variance due to error was only 13.7% for the occupational health workers whereas it was 56.2% and 95.8% for the surgeons and therapists, respectively (Table 3).

Interobserver reliability was substantial for both Tinel's and Phalen's tests (Table 4). Intraobserver reliability was substantial for Tinel's test and moderate for Phalen's test (Table 6). Interobserver reliability was moderate for vibration whereas it was fair for motor power (Table 4). There was perfect intraobserver reliability for vibration and motor power in all cases (Table 6). For Tinel's test, Phalen's test, motor power and vibration the interobserver reliability increased when calculated for the surgeons and therapists only. This is likely to be due to greater experience and expertise in performing the tests. The occupational health workers also had a lower intraobserver reliability for two of the four tests compared with the health care professionals (Table 7); however this effect was small and the number of examinations was limited.

Table 6—Intraobserver reliability for Tinel's test, motor power, vibration and Phalen's test

Test	*Estimated Cohen's Kappa, (asymptotic standard error) (Overall)*
Tinel's test	0.80 (0.19)
Motor power	1.00 (0.0)
Vibration	1.00 (0.0)
Phalen's test	0.526 (0.209)

Table 7—Intraobserver reliability broken down by observer group

Test	*Therapists*	*Surgeons*	*OHW*
2-point moving (ICC)	0.80	0.73	0.67
2-point static (ICC)	0.60	0.42	0.78
SW monofilaments (ICC)	0.73	0.75	0.80
Tinel's test (Kappa)	1.0	1.0	0.50
Motor power (Kappa)	1.0	1.0	1.0
Vibration (Kappa)	1.0	1.0	1.0
Phalen's test (Kappa)	1.0	0.50	0.20

This study had three potential limitations. First, it was done on a single day in an experimental setting, leading to potential recall on the parts of both the patients and the examiners. However, with seven tests, 12 patients, and six examiners, the ability to recall previous examinations was minimized. Second, before conducting the study, we were concerned about the possibility of median nerve irritation from multiple examinations, since this could alter the findings during the day long examination. One patient who had a negative Phalen's test when first examined, had converted to a positive at the time of re-examination. This may have been due to a change in the patient's condition from nerve irritation, or may have been due to patient variation, examiner variation in performing the test, or measurement error. However, for the remainder of the patients, the order in which the tests were done did not have a significant impact on the three tests for which we were able to calculate the effect of this variable (Tables 1, 2 and 3). The final limitation concerns the general applicability of the results. Patients were referred to a tertiary care centre and were evaluated by sub-specialists. However, the fact that the patients had several different diagnoses on neurophysiological testing allows application of the results to patients referred to a hand unit for possible carpal tunnel syndrome, which is

the clinical context in which these tests are often used.

In conclusion, the interobserver reliability of all tests, except for Semmes-Weinstein monofilaments, was satisfactory. Furthermore, the interobserver reliability was increased when the tests were performed by surgeons or therapists with expertise in the physical examination of the hand. Although Semmes-Weinstein monofilament testing had a high intraobserver reliability, it should not be used in the diagnosis of carpal tunnel syndrome because of poor interobserver variability.

Acknowledgements

The authors gratefully acknowledge C.V.A. Bowen MD FRCS (C), Lucy Winston BSc OT, Louise Kelly CHT, Jonathan Smith and Sue Ferrier for participating in the study.

Dr Wright is supported by a MRC scholarship.

Dr Marx is supported by The Arthritis Society, the Institute for Work and Health and the Surgeon Scientist Program of the University of Toronto.

A copy of the protocol used for the tests can be obtained from the corresponding author.

References

De Krom MCTFM, Knipschild PG, Kester ADM, Spaans F (1990). Efficacy of provocative tests for diagnosis of carpal tunnel syndrome. Lancet, 335: 393–395.

Durkan JA (1991). A new diagnostic test for carpal tunnel syndrome. Journal of Bone and Joint Surgery, 73A: 535–538.

Feinstein AR. *Clinimetrics*. Westford, Mass., Murray Printing Company, 1987.

Fess EE (1986). The need for reliability and validity in hand assessment instruments. Journal of Hand Surgery, 11A: 621–623.

Fleiss JL. *The design and analysis of clinical experiments*. New York, John Wiley & Sons, Inc. 1986.

Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology*. The essentials, 2nd edn. Baltimore, Williams and Wilkins, 1988.

Gellman H, Gelberman RH, Tan AM, Botte MJ (1986). Carpal tunnel syndrome. An evaluation of the provocative diagnostic tests. Journal of Bone and Joint Surgery, 68A: 735–737.

Gilliatt RW, Wilson TG (1953). A pneumatic-tourniquet test in the carpal-tunnel syndrome. Lancet, ii: 595–597.

Golding DN, Rose DM, Selvarajah K (1986). Clinical tests for carpal tunnel syndrome: an evaluation. British Journal of Rheumatology, 25: 388–390.

Gonzalez Del Pino J, Delgado-Martinez AD, Gonzalez Gonzalez I, Lovic A (1997). Value of the carpal compression test in the diagnosis of carpal tunnel syndrome. Journal of Hand Surgery, 22B: 38–41.

Gunnarsson LG, Amilon A, Hellstrand P, Leissner P, Philipson L (1997). The diagnosis of carpal tunnel syndrome: sensitivity and specificity of some clinical and electrophysiological tests. Journal of Hand Surgery, 22B: 34–37.

Hoaglin DC, Mosteller F, Tukey J W. *Fundamentals of exploratory analysis of variance*. New York, John Wiley & Sons, Inc. 1991.

Katz JN, Larson MG, Sabra A, et al (1990). The carpal tunnel syndrome: diagnostic utility of the history and physical examination findings. Annals of Internal Medicine, 112: 321–327.

Kuschner SH, Ebramzadeh E, Johnson D, Brien WW, Sherman R (1992). Tinel's sign and Phalen's test in carpal tunnel syndrome. Orthopedics, 15: 1297–1302.

Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. Biometrics, 33: 159–174.

Mackinnon SE, Dellon L (1985). Two-point discrimination tester. Journal of Hand Surgery, 10A: 906–907.

Phalen GS (1972). The carpal tunnel syndrome. Clinical evaluation of 598 hands. Clinical Orthopaedics and Related Research, 83: 29–40.

Pryse-Phillips W (1984). Validation of a diagnostic sign in carpal tunnel syndrome. Journal of Neurology, Neurosurgery and Psychiatry, 47: 870–872.

Stewart JD, Eisen A (1978). Tinel's sign and the carpal tunnel syndrome. British Medical Journal, 2: 1125–1126.

Szabo RM, Gelberman RH, Dimick MP (1984). Sensibility testing in patients with carpal tunnel syndrome. Journal of Bone and Joint Surgery, 66A: 60–64.

Wright JG, Feinstein AR (1992). Improving the reliability of orthopaedic measurements. Journal of Bone and Joint Surgery, 74B: 287–291.

Received: 17 December 1997
Accepted after revision: 18 March 1998
Dr J. G. Wright, The Hospital for Sick Children, 555 University Avenue, S-107, Toronto, Ontario M5G 1X8, Canada.