Diagnosis of Carpal Tunnel Syndrome: Interobserver Reliability of the Blinded Scratch-Collapse Test

Robin D. Blok · Stéphanie J. E. Becker · David C. Ring

Received: 14 April 2013 / Accepted: 17 September 2013 / Published online: 22 September 2013
© Society of the Hand & Microsurgeons of India 2013

Abstract The reliability of the scratch-collapse test for diagnosis of carpal tunnel syndrome (CTS) has not been tested by independent investigators. This study measured the reliability of the scratch-collapse test comparing the treating hand surgeon and blinded evaluators. We performed a prospective observational study of 41 patients with a provisional diagnosis of CTS or a combination of CTS and cubital tunnel syndrome and prescribed electrodiagnostic testing. The treating hand surgeon performed the scratch-collapse test. Next, the test was administered by one of the six observers, unaware of the patient's symptoms and the diagnosis made by treating hand surgeon. The kappa statistic (κ) was used to calculate the interrater reliability between the treating hand surgeons and blinded scratchers. The agreement between the blinded observers and the hand surgeons on the scratch-collapse test was substantial 0.63 (95 % CI, 0.34–0.87; $p < 0.001$). **The sensitivity of the blinded scratch test in our sample was 32 %.** In a small study with a spectrum bias favoring electrophysiologically confirmed CTS the reliability was lower than that reported by the inventors of the test, but was still substantial. We propose a larger study of patients with a greater variety of electrodiagnostic test results using fewer observers with more experience.

Keywords Scratch-collapse test · Interobserver reliability · Carpal tunnel syndrome

Introduction

Even though carpal tunnel syndrome (CTS) is the most common peripheral mononeuropathy [1, 2], there is still no consensus on a reference standard for diagnosis [3]. Currently doctors use a combination of interview, physical examination, and electrodiagnostic testing. The provocative tests that are most often used are the Phalen, Tinel and Durkan tests. The scratch-collapse test is said to be less subjective and works as follows: 1) The examiner applies bilateral shoulder adduction/internal rotation to the forearms and the patient tries to resist the examiner's force; 2) Next, the examiner 'scratches' or swipes with the fingertips over the area of the transverse carpal ligament, where the median nerve is potentially compressed; and 3) Immediately afterwards Step 1 is repeated. The scratch-collapse test is considered positive when momentary loss of the patient's external resistance tone is observed, in other words the patient's arm collapses against the resistance. It is said that the test can be repeated without signs of fatigue [4]. To date, there are a few studies that evaluated the accuracy, sensitivity and specificity of the scratch-collapse test and there is still a wide range of opinions about the usefulness of this test [4, 5].

This study measured the reliability of the scratch-collapse test between treating hand surgeons and trained blinded examiners among patients undergoing electrodiagnostic tests to confirm or exclude CTS.

Methods

The study protocol was approved by the Human Research Committee as part of an earlier study at our department [6]. Informed consent was obtained from each patient. Between January 2011 and April 2012, English-speaking adult patients (18 years or older) presenting to one of four hand surgeons with a provisional diagnosis of CTS or a combination of CTS

R. D. Blok · S. J. E. Becker · D. C. Ring (✉)
Orthopaedic Hand and Upper Extremity Service, Yawkey Center,
Massachusetts General Hospital, Suite 2100, 55 Fruit Street,
Boston, MA 02114, USA
e-mail: dring@partners.org

Table 1 Baseline characteristics

			<i>n</i> = 41
Parameter	Mean	SD	Range
Baseline			
Age (years)	58	14	27–82
Body mass index	27	5.2	21–45
Duration of symptoms (months)	26	48	0.8–240
Parameter	Number		%
Baseline			
Sex			
Female	28		68.0
Male	13		32.0
Numbness predominantly in median nerve distribution			
No	5		12.0
Yes	36		88.0
Phalen's test			
Negative	18		44.0
Positive	23		56.0
Durkan's test (<i>n</i> = 40)			
Negative	14		35.0
Positive	26		65.0
Tinel's test (<i>n</i> = 31)			
Negative	18		58.1
Positive	13		41.9
Thumb abduction weakness			
No	26		63.0
Yes	15		37.0
Thenar atrophy			
No	35		85.0
Yes	6		15.0
Severity of CTS on electrodiagnostic test (<i>n</i> = 37)			
None	6		16.2
Mild	4		10.8
Mild-moderate	3		8.1
Moderate	16		43.2
Moderate-severe	4		10.8
Severe	4		10.8

N number; *NS* not significant; *CTS* carpal tunnel syndrome

and cubital tunnel syndrome based on symptoms and signs and prescribed electrodiagnostic testing were asked to participate. Exclusion criteria were 1) a prior diagnosis of CTS by a specialist; 2) prior carpal tunnel release; 3) previous negative electrodiagnostic testing of the median nerve in the past 5 years; 4) previous positive electrodiagnostic testing of the median nerve at our institution at any point in time; 5) rheumatoid arthritis; and 6) pregnancy.

The treating surgeon performed the scratch-collapse test. Next, one of six researchers trained to perform the scratch-

Table 2 Kappa agreement of scratch test

			<i>n</i> = 41		
	Physicians		Total	Kappa (95 % CI)	<i>P</i> value
	Negative	Positive			
Blinded scratchers					
Negative	27	1	28	0.63 (0.34–0.87)	<0.001
Positive	5	8	13		
Total	32	9	41		

CI confidence interval

collapse test administered the test, but unaware of the patient's symptoms and the diagnosis that was made by the treating hand surgeon. Blinded observer 1 tested 27 patients, observer 2 tested 6 patients, observer 3 tested 3 patients, observer 4 and 5 each tested 2 patients, and observer 6 tested 1 patient.

Thirty-seven of the total of 41 patients underwent electrodiagnostic testing. The electrodiagnostic tests were performed by a small group of electrodiagnosticians in a single office at our institution using a standard technique. **Both motor and sensory conduction velocities were measured.** The severity of the CTS was rated using an agreed on, standardized rating scale. **Thirty-one of 37 patients had electrodiagnostically confirmed CTS. Four of these 31 patients were not assigned severity ratings.**

Statistical Analysis

We followed the general rule of thumb that 30 subjects is a reasonable minimum sample size for a reliability study as long as the hypothesis is to prove that kappa is different from a value of zero and a moderate level or better kappa value (>0.40) is expected [7].

Cohen's kappa (κ), a measure of agreement corrected for chance of nominally scaled data, was performed to calculate the interobserver reliability between the treating hand surgeons and blinded scratchers. The κ values were interpreted according to the proposed guidelines by Landis and Koch [8]: 0.01 to 0.20 represents a slight agreement; 0.21 to 0.40 a fair agreement; 0.41 to 0.60 a moderate agreement; 0.61 to 0.80 a substantial agreement; and greater than 0.81 an almost perfect agreement.

Results

Forty-one patients were enrolled in this study and had complete data for analysis. (Table 1) The sensitivity of the blinded scratch test in our sample was 32 % (10 patients had a true positive blinded scratch test among 31 patients with electrodiagnostically confirmed CTS). The agreement between the

blinded examiners and the surgeons on the scratch test was substantial 0.63 (95 % CI, 0.34–0.87; $p < 0.001$) (Table 2).

Discussion

We did not reject our null hypothesis that there would be no discrepancy in the result of the scratch-collapse test between treating hand surgeons and blinded observers among patients who would undergo electrodiagnostic testing to confirm or exclude CTS. The interobserver agreement was substantial ($\kappa = 0.63$; $p < 0.001$) between the blinded examiners and the hand surgeons. **The sensitivity of the scratch test for CTS among the blinded examiners was only 32 %.**

The strengths of this study included prospective evaluation; blinding of the second examiner from the patient's complaints and the diagnosis made by the hand surgeon, thereby limiting confirmation bias. There are several limitations as follows: First, for practical reasons we used ten observers (6 blinded examiners and 4 hand surgeons) and three blinded examiners performed the scratch test only once or twice. The number of observers might have negatively influenced the accuracy of the interobserver agreement. Another concern is the potential spectrum bias caused by the fact that most of the patients who underwent electrodiagnostic testing had CTS; thus the study population does not reflect the reported prevalence of 5–15 % of CTS in general population based studies [9, 10]. For most examiners, the scratch-collapse test was nearly always negative, which may be the only reason there was moderate reliability. Finally, because patients underwent the scratch test more than once, there may have been confirmation bias among the patients.

The inventors of the scratch-collapse test reported a reliability ($\kappa = 0.98$; $p < 0.001$) that seems almost too good to be true [4]. They did not provide enough detail to know how many raters there were and if this was actually intra- or interrater reliability. Intrarater reliability can be this high, but interrater reliability rarely is. In that study, the examiners knew the patient's history and current condition.

Another important consideration is performance bias. Cheng et al. [4] made clear that there is a learning curve associated with the scratch-collapse test. In our study, we had six blinded evaluators who were trained by one of the hand surgeons and by an instructional video made by the creator of the test. The raters in the study by Cheng et al. [4] were trained directly by the inventor of the test. There is a possibility that the six examiners were not trained properly or that they were not skilled adequately to administer the test.

To determine the reliability of the scratch-collapse test, a larger study of patients with a larger variety of electrophysiological test results (CTS, cubital tunnel syndrome, cervical radiculopathy, normal, and less common diseases) is needed, using fewer observers with more experience. The blinded observer should do the scratch-collapse test first. It is very difficult to study the accuracy of diagnostic tests where the performance of both the examinee and the examiner are subject to so many forms of bias.

Acknowledgment RDB is supported by a Dutch research grant from Spinoza Fonds – University of Amsterdam, Academic Medical Center, The Netherlands, for Scientific Research.

SJEB is supported by Dutch research grants from Anna Foundation/NOREF, GenootschapNoorthey, and StichtingVreedefonds, the Netherlands, for Scientific Research.

Conflict of Interest All named authors hereby declare that they have no conflicts of interest to disclose related to this study.

Ethical Statement All authors adhere to the ethical standards described by the Committee on Publication Ethics and the International Committee of Medical Journal Editors. The study was completed under an IRB approved protocol.

References

1. Aroori S, Spence RA (2008) Carpal tunnel syndrome. *Ulster Med J* 77:6–17
2. Atroshi I, Gummesson C, Johnsson R, Ornstein E, Ranstam J, Rosen I (1999) Prevalence of carpal tunnel syndrome in a general population. *JAMA* 282:153–158
3. Prime MS, Palmer J, Khan WS, Goddard NJ (2010) Is there light at the end of the tunnel? Controversies in the diagnosis and management of carpal tunnel syndrome. *Hand (N Y)* 5:354–360
4. Cheng CJ, Mackinnon-Patterson B, Beck JL, Mackinnon SE (2008) Scratch collapse test for evaluation of carpal and cubital tunnel syndrome. *J Hand Surg [Am]* 33:1518–1524
5. Gillenwater J, Cheng J, Mackinnon SE (2011) Evaluation of the scratch collapse test in peroneal nerve compression. *Plast Reconstr Surg* 128:933–939
6. Makanji HS, Becker SJ, Mudgal CS, Jupiter JB, Ring D (2013) Evaluation of the scratch collapse test for the diagnosis of carpal tunnel syndrome. *J Hand Surg Eur Vol.* doi:10.1177/1753193413497191
7. Crewson PE (2005) Reader agreement studies. *AJR Am J Roentgenol* 184:1391–1397
8. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
9. de Krom MC, Knipschild PG, Kester AD, Thijs CT, Boekkooi PF, Spaans F (1992) Carpal tunnel syndrome: prevalence in the general population. *J Clin Epidemiol* 45:373–376
10. Katz JN, Larson MG, Sabra A, Krarup C, Stirrat CR, Sethi R et al (1990) The carpal tunnel syndrome: diagnostic utility of the history and physical examination findings. *Ann Intern Med* 112:321–327