

What Do We Know About the Reliability and Validity of Physical Examination Tests Used to Examine the Upper Extremity?

Robert G. Marx, MD, Claire Bombardier, MD, James G. Wright, MD,
Toronto, Ontario, Canada

The literature regarding the reliability and validity of commonly used clinical tests for disorders of the upper extremity was reviewed. Formal literature search, standard texts, and experts in the field of upper extremity were consulted to locate relevant articles. Range of motion and strength testing of the upper limb have been shown to be reliable, while various tests used for the diagnosis of conditions in the upper limb, such as carpal tunnel syndrome and rotator cuff tendinopathy, have been shown to have varying degrees of validity. Overall, however, we determined that there is little evidence regarding the reliability and validity of physical examination for the upper extremity and specifically less information available regarding the reliability of diagnostic physical examination tests and the validity of impairment measures used for the upper limb. Further studies in this area are warranted in view of the impact of these findings on the treatment of patients. (*J Hand Surg* 1999;24A:185-193. Copyright © 1999 by the American Society for Surgery of the Hand.)

Key words: Reliability, validity, upper limb, physical examination, diagnosis.

Physical examination is used in conjunction with history, laboratory investigation, and imaging studies to evaluate the upper extremity.¹ Physical examination is used primarily to document the severity of impairments or to diagnose clinical conditions. Impairment is defined as a limitation or abnormality of physiologic or anatomic structure or function at the organ level.² For

example, a limitation in the range of motion of the elbow is an impairment. Diagnostic physical examination tests are used to identify or rule out pathologic conditions. For example, impingement tests are used to make the diagnosis of rotator cuff tendinopathy.

The elements of the physical examination should be reliable and valid.³ Reliability is the extent to which repeated measurements of a relatively stable phenomenon are close to each other.⁴ Validity is the degree to which the results of measurement correspond to the true state of the phenomenon being measured.^{5,6} Accuracy is defined as the proportion of all test results, both positive and negative, that are correct.⁵ The results of an examination are of limited value if they are not consistently repeatable in the hands of the same (intraobserver reliability) or different (interobserver reliability) clinicians.^{5,6} Measurement reliability also may be affected by procedural variability due to the technician's performance or equipment.⁴ An examination that does not accurately measure the phenomenon of interest is not a useful tool.^{4,5}

From the University of Toronto, Toronto, Ontario, Canada.

Dr Marx is supported by the Arthritis Society, the Institute for Work and Health, and the Surgeon Scientist Program of the University of Toronto. Dr Wright is supported by an MRC scientist award.

Received for publication August 6, 1997; accepted in revised form July 27, 1998.

Although the author or authors have not received or will not receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article, benefits have been or will be received but are directed solely to a research fund, foundation, educational institution, or other non-profit organization with which one or more of the authors are associated.

Reprint requests: James G. Wright, MD, The Hospital for Sick Children, 555 University Ave, S-107, Toronto, ON, Canada M5G 1X8.
Copyright © 1999 by the American Society for Surgery of the Hand
0363-5023/99/24A01-0028\$3.00/0

The reliability of a test can be determined without a gold standard for diagnosis. In many cases in which a gold standard does not exist, the reliability of the test may be of greater importance than the validity. Despite the complexity of the determination of validity in the measurement of clinical phenomena, it is incumbent on the clinician to strive to use valid tools.

Validity, or accuracy, is difficult to define in many clinical situations. There is no easily measurable gold standard for many medical and/or upper limb conditions. For example, the ultimate diagnosis of osteoarthritis of the glenohumeral joint would entail biopsy of the articular cartilage, which is not feasible in most clinical situations. If an inaccurate standard of validity is used, a new test can perform no better than that standard and may seem inferior even though it approximates the truth more closely.⁶ Furthermore, many factors affect the assessment of validity. Although the validity of a test is theoretically independent of the prevalence of the condition, patient characteristics such as stage and severity of disease, as well as clinician variables such as experience and expertise, may affect the performance of the test.⁴

Physical examination is an important part of patient assessment that is used by all upper extremity surgeons. The history of present illness, diagnostic imaging studies, and electrodiagnostic tests also play an important role in patient evaluation. For the physical examination to yield useful information, each test performed should be both reliable and valid.

The purpose of this study was to review the literature specific to the reliability and validity of commonly used physical examination tests for disorders of the upper extremity. This information will inform clinicians who use these tests of their properties and focus future research in this area.

Materials and Methods

A formal literature search was performed to find studies evaluating the reliability and validity of physical examination for the upper extremity. Medline (1966 to 1996) was used as the database. Several specific searches were conducted. Arm (exploded) or shoulder (exploded) were crossed with anatomy (exploded), kinematics (text word), physiology (exploded), sensitivity and specificity (exploded), reliability (text word), validity (text word), physical exam (exploded), epidemiologic methods (exploded), and reproducibility of results (exploded).

Three other strategies in addition to the literature

search were used to locate important articles. The reference sections of retrieved articles were used to locate other pertinent literature, the reference lists of standard texts were reviewed,⁷⁻¹³ and experts in the field of upper extremity were asked to identify articles relating to the reliability and validity of physical examination for the upper limb. No articles were excluded from this search because of the paucity of literature in this area. Rather, papers with notable weaknesses were included for completeness, with the relevant flaws identified. Rather than reviewing all aspects of the upper extremity physical examination, however, we restricted our attention to the most commonly performed components of the physical examination. Seven standard texts relating to physical examination of the upper limb were identified by polling a group of experienced upper extremity surgeons. These texts were used to define the examination techniques for the upper limb reviewed in this article.⁷⁻¹³ We reviewed tests used to measure impairments and those used for diagnosis separately. Reports published before 1966 were not systematically reviewed. Moreover, the literature review presented is not purported to be exhaustive, but rather as complete as possible.

Results

Measures of Impairment for the Hand and Wrist

Only one study that examined the reliability of the range of motion of the metacarpophalangeal and interphalangeal joints in the hand was located¹⁴ (Table 1). This study demonstrated that the intraobserver variations both for clinical observation and the use of a goniometer were not significant and that the interobserver variation was greater. Three studies examined the reliability of the measurement of wrist range of motion and demonstrated similar findings¹⁵⁻¹⁷ (see Table 1).

The two methods that have been used to test muscle power include manual muscle testing²¹ and muscle dynamometry.¹² Manual muscle testing is described as semiquantitative because it is an ordinal level of measurement.²² This method of quantifying muscle strength has been shown to be reliable both within and among observers, particularly when the scale is expanded to plus or minus a half or a full grade.^{23,24}

One study found the reliability of pinch strength measurements, including palmar (3-jaw chuck), key (lateral), and tip (2 point) pinch, to be high²⁰ (see

Table 1. Reliability and Validity of Impairment Measures for the Wrist and Hand

<i>Physical Examination Test</i>	<i>Reliability/Validity</i>	<i>Significance</i>
Range of motion of the MCP and IP joints ¹⁴	Intraobserver variation, $p < .01$ (ANOVA)	Reliability acceptable for clinical practice
Range of motion of the wrist ¹⁵⁻¹⁷	Intraobserver variation, 5° to 8° ; interobserver variation, 6° to 10°	Reliability acceptable for clinical practice
Manual and dynametric muscle testing ^{18,19}	Reliability coefficient = 0.88	Reliability acceptable for clinical practice
Pinch strength measurement ²⁰	Pearson $r = .98$ between observers; $r = .98$ for test-retest	Reliability acceptable for clinical practice

MCP, metacarpophalangeal; IP, interphalangeal.

Table 1). Manual muscle testing for wrist extension¹⁸ and dynametric muscle testing for wrist flexion and extension also had high reliability coefficients within individuals ($r = .88$)^{18,19} (see Table 1).

Measures of Impairment for the Shoulder and Elbow

Goniometry has been used to measure both active and passive range of motion of the shoulder. Several studies have shown the intraobserver variation to be satisfactory (intraclass correlation coefficients ≥ 0.89).^{15,25-29} Williams and Callaghan³⁰ found that visual estimation by experienced and skilled therapists was equal to measurements by goniometry (Table 2).

Hand-held dynamometry has adequate test-retest reliability for the measurement of shoulder abduction, flexion, internal rotation, and external rotation power ($r \geq .85$)^{20,31-34} (see Table 2).

We were unable to locate any studies validating measures of range of motion or strength of the upper extremity.

Diagnostic Physical Examination Tests for the Hand and Wrist

The reliability and validity (sensitivity and specificity) of diagnostic physical examination tests have been reported for three conditions: carpal tunnel syn-

drome, peripheral nerve injuries, and traumatic wrist instability. We were unable to locate any studies documenting the reliability and validity of diagnostic physical examination tests for other conditions of the hand and wrist.

Many clinical diagnostic physical examination tests have been described to aid in the diagnosis of carpal tunnel syndrome, including thenar muscle atrophy,³⁵ the wrist flexion (Phalen's) test,³⁶ the Tinel test,⁸ the tourniquet test,³⁷ vibrometry,³⁸ and Semmes-Weinstein monofilaments.³⁸ The assessment of validity of these tests is difficult because there is no generally accepted gold standard for the diagnosis of carpal tunnel syndrome.^{39,40} Although not a true gold standard, most studies use electrodiagnostic testing as the benchmark for validity testing.⁴¹ The validity of physical examination tests for carpal tunnel syndrome varies for the different tests (Table 3).

Although there have been many studies on the validity of physical examination for carpal tunnel syndrome, relatively little attention has been directed toward documenting the reliability of these diagnostic physical examination tests. We recently performed a study to evaluate the reliability of physical examination for carpal tunnel syndrome⁴² (see Table 3).

Table 2. Reliability and Validity of Impairment Measures for the Shoulder and Elbow

<i>Physical Examination Test</i>	<i>Reliability/Validity</i>	<i>Significance</i>
Active and passive shoulder range of motion by goniometry and visual estimation ^{15,25,26,28-30}	High intraobserver and interobserver reliability (ICC = 0.89)	Reliability acceptable for clinical practice
Hand-held dynamometry for shoulder strength ^{18,31-34}	High test-retest reliability ($r \geq .85$)	Reliability acceptable for clinical practice

ICC, intraclass correlation coefficient.

Table 3. Reliability and Validity of Diagnostic Physical Examination Tests for the Wrist and Hand

<i>Physical Examination Test</i>	<i>Reliability/Validity</i>	<i>Significance</i>
Phalen's test for carpal tunnel syndrome ^{35,38,42-47}	Intraobserver reliability, kappa = 0.53; interobserver reliability, kappa = 0.65; sensitivity, 67% to 88%; specificity, 32% to 86%	Reliability and validity moderately acceptable for use in clinical practice
Tinel's test for carpal tunnel syndrome ^{35,38,42-49}	Intraobserver reliability, kappa = 0.80; interobserver reliability, kappa = 0.79; sensitivity, 26% to 73%; specificity, 55% to 94%	Reliability and validity moderately acceptable for use in clinical practice
Thenar eminence wasting for carpal tunnel syndrome ^{35,43}	Sensitivity, 3% to 19%; specificity, 93% to 100%	Validity acceptable to aid in diagnosis if positive
Tourniquet test for carpal tunnel syndrome ^{35,43,44}	Sensitivity, 16% to 65%; specificity, 60% to 87%	Insufficient evidence for use in clinical practice
Carpal compression test for carpal tunnel syndrome ⁵⁰	Sensitivity, 87%; specificity, 90% to 95%	Validity acceptable for use in clinical practice
Two-point discrimination for carpal tunnel syndrome ^{38,42,44,46}	Intraobserver reliability, ICC = 0.45 (static), ICC = 0.66 (moving); interobserver reliability, ICC = 0.58 (static), ICC = 0.77 (moving); sensitivity, 22% to 33%; specificity, 81% to 100%	Reliability acceptable for use in clinical practice; insufficient validity for use in clinical practice
Semmes-Weinstein monofilament testing for carpal tunnel syndrome ^{38,42,44}	Intraobserver reliability, ICC = 0.15; interobserver reliability, ICC = 0.71; sensitivity, 80% to 91%; specificity, 80%	Reliability not acceptable for use in clinical practice; validity acceptable for use in clinical practice if the expertise is available
Weber static 2-point discrimination for nerve-injured patients ⁵¹	Pearson $r = 0.88-0.92$ between observers	Reliability acceptable for use in clinical practice
Dellon moving 2-point discrimination for nerve-injured patients ⁵¹	Pearson $r = 0.92-0.96$ between observers	Reliability acceptable for use in clinical practice
Scaphoid shift test for traumatic wrist instability ⁵²⁻⁵⁴	Sensitivity, 69%; specificity, 64% to 68%	Insufficient evidence for use in clinical practice
Ballottement test for traumatic wrist instability ⁵²	Sensitivity, 64%; specificity, 44%	Insufficient evidence for use in clinical practice
Ulnomeniscotriquetral glide test for traumatic wrist instability ⁵²	Sensitivity, 66%; specificity, 64%	Insufficient evidence for use in clinical practice

ICC, intraclass correlation coefficient.

Static and moving 2-point discrimination were found to be reliable in nerve-injured patients⁵¹ (see Table 3). A hand surgeon and a hand therapist measured within 1 mm of each other 93% and 87% of the time for the moving and static 2-point discrimination, respectively. There were no cases in which the examiners differed by more than 2 mm.⁵¹

Tests to evaluate patients with traumatic wrist instability include the scaphoid shift test,⁵⁵ the ballottement test,⁵⁶ and the ulnomeniscotriquetral dorsal glide test.⁵² These tests were found to have moderate to low sensitivity and specificity, with arthroscopy used as the gold standard (see Table 3).

Diagnostic Physical Examination Tests for the Shoulder and Elbow

The sensitivity and specificity of diagnostic physical examination tests for thoracic outlet syndrome,

shoulder instability, and rotator cuff tendinopathy have been studied.

In addition to a typical history and physical findings, the diagnosis of thoracic outlet syndrome relies on various provocative maneuvers, including the Adson maneuver, the costoclavicular maneuver, and the Allen hyperabduction test. The Adson maneuver involves the detection of a diminished radial pulse when the shoulder is abducted, extended, and externally rotated with the head turned toward the ipsilateral side.⁸ The Allen maneuver is positive when a decreased radial pulse is noted with the shoulder abducted to 90° in 90° of external rotation.¹⁰ The costoclavicular maneuver is positive when the radial pulse is decreased with the arm in extension and external rotation but not abduction.¹⁰ No studies documenting the reliability of these findings have been performed. The specificity, determined in

asymptomatic patients, has been reported to be between 18% and 87%.⁵⁷⁻⁶⁰ The sensitivity has been documented at 94%.⁵⁹ (Table 4).

The classic method of diagnosing anterior instability of the shoulder is the apprehension test.⁸ A study of 103 patients with arthroscopically proven anterior instability showed this test to have a sensitivity of 63%.⁶¹ The anterior drawer test for the shoulder involves sliding the humeral head in an anterior direction on the glenoid to detect anterior instability. This test has been reported to be 100% sensitive for the detection of instability in patients with recurrent dislocation but not in cases of recurrent subluxation.⁶² Similarly, examination under anesthesia was found to be 100% sensitive in 66 patients about to undergo anterior repair for instability.⁶⁶ In both cases, the investigators failed to describe the false-positive rates of the tests (see Table 4).

The relocation test⁶⁷ for diagnosing anterior instability was evaluated by Speer et al.⁶³ The diagnosis was confirmed in all patients at the time of surgery. In 54 patients with diagnoses other than anterior instability (8 posterior instability, 34 rotator cuff tendinopathy, 5 acromioclavicular arthritis, 5 glenohumeral arthritis, and 2 biceps tendon instability), none had a positive relocation test (specificity = 100%). The sensitivity of the test in 46 patients with anterior instability was 74% (see Table 4).

A knee laxity tester has been used to attempt to measure instability of the shoulder in an objective fashion.⁶⁸ The amount of shoulder translation (mm) was measured reliably when tested on 40 normal and unstable shoulders (intraclass correlation coefficient, .996). The laxity tester also found statistically significant differences between the normal and unstable

shoulders of patients with unilateral, recurrent episodes of dislocation and overhead athletes who felt a sensation of instability during overhead activity and had a positive apprehension test. No statistically significant difference was found, however, between the shoulders of patients without shoulder pathology or in patients with atraumatic multidirectional instability due to generalized joint hyperlaxity.⁶⁸

We were only able to locate one study evaluating the reliability and validity of tests for posterior instability of the shoulder.⁶¹ In 13 patients who were suspected to have traumatic posterior instability and who had subsequent arthroscopic confirmation, only 4 had a positive apprehension test in adduction and internal rotation.⁶¹ The method used to perform the test, however, was not described in detail.

Lyons and Tomlinson⁶⁴ reported the accuracy of the diagnosis of the presence and size of rotator cuff tears in patients referred to a special shoulder unit. In this study, physical examination consisted of palpation and strength testing of the rotator cuff. The sensitivity of physical examination for detecting the presence of a cuff tear was 91% and the specificity was 75%, with open surgery used as the gold standard.⁶⁴ Unfortunately, the investigators neglected to mention who examined the patients (see Table 4).

Another investigation compared multiple physical examination tests to arthroscopy for the diagnosis of impingement syndrome,⁶⁵ as described by Neer.⁶⁹ The investigators did not feel that it was appropriate to use a single test to diagnose impingement syndrome and that a combination of tests used by an experienced clinician was more accurate. In 45 patients, the sensitivity and speci-

Table 4. Reliability and Validity of Diagnostic Physical Examination Tests for the Shoulder and Elbow

<i>Physical Examination Test</i>	<i>Reliability/Validity</i>	<i>Significance</i>
Adson maneuver for thoracic outlet syndrome ^{58,60}	Specificity, 32% to 87%	Insufficient evidence for use in clinical practice
Costoclavicular maneuver for thoracic outlet syndrome ^{59,60}	Sensitivity, 94%; specificity, 53% to 85%	Insufficient evidence for use in clinical practice
Allen hyperabduction test for thoracic outlet syndrome ^{57,60}	Specificity, 18% to 43%	Insufficient evidence for use in clinical practice
Apprehension, relocation, and anterior drawer tests for anterior instability of the shoulder ⁶¹⁻⁶³	Sensitivity, 63% to 100%	Validity acceptable for use in clinical practice
Palpation and strength testing for rotator cuff tears ⁶⁴	Sensitivity, 91%; specificity, 75%	Validity acceptable for use in clinical practice
Multiple physical examination tests for the diagnosis of impingement syndrome ⁶⁵	Sensitivity, 86%; specificity, 74%	Validity acceptable for use in clinical practice

ficity of the multiple tests was 86% and 74%, respectively⁶⁵ (see Table 4).

Discussion

In virtually all patients with upper extremity disorders, the physical examination is used to grade the level of impairment and to make a diagnosis. This study has shown that not all aspects of the physical examination have been shown to be reliable and valid. The determination of the reliability and validity of physical examination tests is further complicated by the fact that there is no gold standard for many of the conditions that are routinely diagnosed by clinicians. In these cases, the reliability of the tests used assumes a greater importance to ensure that clinicians are measuring the same phenomenon when the true entity is difficult to define. Furthermore, the determination of the reliability must be performed using both diseased and nondiseased individuals (if it is possible to make this distinction) to avoid bias and to ensure that the test is useful in both situations.⁴

Measurements of joint motion in the upper extremity have been shown to be relatively reliable both between examiners and in terms of intraobserver reliability; interobserver and intraobserver variability is in the range of 5° to 10°. A study comparing the standard techniques, such as visual estimation or goniometric measurement, with a gold standard, however, would be useful to further validate these methods. Strength testing using a dynamometer also has been shown to be reliable both within and between observers.^{18,31-34,70} Intraobserver¹⁸ and interobserver^{23,24} reliability have been shown to be satisfactory for manual muscle testing. Thus, measurements of joint motion and muscle power (manual and with dynamometer) in the upper extremity have adequate intraobserver and interobserver reliability and we recommend their use for clinical practice.

The diagnosis of carpal tunnel syndrome often relies on aspects of the physical examination. However, the diagnostic accuracy of these tests has often used electrodiagnostic studies as a gold standard. The results of electrodiagnostic testing may vary between operators or testing centers due to technical factors.^{41,71,72} The disparate results in the studies that have been performed to evaluate the sensitivity and specificity of physical examination are likely due to several factors. First, the criteria selected for the electrodiagnostic diagnosis of carpal tunnel syndrome differ among investigators. Second, the clin-

ical criteria used to determine a positive response for the physical examination tests vary from investigator to investigator and in some cases they are not described in sufficient detail to allow reproducibility. Finally, various methods were used to locate patients for the studies. Investigators recruited patients by survey,³⁵ in electrodiagnostic clinics,⁴³ and before surgery.⁵⁰ Thus, the studies evaluated different samples of patients, which also could have contributed to differences in the results.

Tinel's and Phalen's tests for the diagnosis of carpal tunnel syndrome have been shown in multiple studies to have varying sensitivities and specificities. Taking the methodologic strengths and weaknesses of these studies into account, these tests should be used in clinical practice with the understanding that a positive or negative result may not bear a direct relationship with the truth. The carpal compression test is a recently described test that has been shown to be valid in 2 studies and is suitable for clinical practice. Another study described a similar test, the pressure provocative test, which involves the use of a sphygmomanometer over the carpal tunnel rather than the examiner's finger.⁷³ These investigators also found the test to be very sensitive and specific. Semmes-Weinstein monofilaments also have been shown to be useful for the diagnosis of this condition; however, recent work has documented poor interobserver reliability,⁴² limiting the usefulness of this test in patients with suspected carpal tunnel syndrome. Thenar muscle wasting has a very low sensitivity and a very high specificity for this diagnosis. This test should thus be used to help diagnose carpal tunnel syndrome when positive in a patient in whom the diagnosis is suspected. There is insufficient evidence to support the use of the tourniquet test, sensibility testing, or 2-point discrimination for this diagnosis. The reliability of tests for the diagnosis of carpal tunnel syndrome has been shown to be adequate overall.⁴² Two-point discrimination testing was found to be highly reliable and may be useful in the examination of nerve-injured patients. There is insufficient evidence to advocate the use of provocative tests for carpal instability.

There have been several studies of the sensitivity and specificity of diagnostic maneuvers for thoracic outlet syndrome.^{57-60,74} The results were conflicting, which likely is due to the fact that the investigators chose varying methods to define the presence of thoracic outlet syndrome. Thus, the true sensitivity or specificity of these maneuvers cannot be accurately determined based on the available literature and

should be used with caution, if at all, in clinical practice.

We were unable to identify any studies that adequately documented the reliability and validity of physical examination for patients with either posterior or multidirectional instability. Anterior instability of the shoulder is a more common condition in which physical examination plays an important role in the evaluation of the patient. In view of this fact, it is surprising that there has been only one formal test of validity for the standard tests. Speer et al⁶³ found a sensitivity of 74% and a specificity of 100% for the relocation test. These figures may have been higher and lower, respectively, if the examiners had used nonverbal signs of apprehension. By not recording such signs, the investigators likely increased their rate of false-negative results and decreased their rate of false-positive results.

Rotator cuff tendinopathy is a very common condition, with tears having been identified in 5% to 80% of cases in cadaveric studies.⁷⁵ There have been relatively few studies documenting the measurement properties of physical examination for this condition. Two studies documented the accuracy of multiple tests performed by an experienced orthopedic surgeon for the diagnosis of impingement syndrome⁶⁵ and rotator cuff tears.⁶⁴ The tests performed well; however, further studies would be useful to confirm these findings and to document the reliability and validity of the tests on an individual basis.

The reliability of range of motion and strength testing in the upper extremity has been well-documented. There is a lack of documentation of the reliability and validity of diagnostic physical examination tests for conditions of the upper extremity, particularly reliability. Many tests have no studies documenting their properties, such as deep tendon reflexes,⁸ Cozen's test for tennis elbow,¹⁰ O'Driscoll's test for posterolateral rotatory instability of the elbow,⁷⁶ Froment's sign,¹³ Finkelstein's test,⁸ Neer's sign,⁶⁹ Hawkins' test,⁷⁷ and Jobe's test.⁷⁸ Further studies in this area are warranted in view of the impact of these findings on the treatment of patients.

The tests for physical examination discussed above are used daily by upper extremity surgeons. Although these tests are not used in isolation, it is important to document the properties of each test to allow clinicians to reliably and accurately examine patients. There is little evidence regarding the reliability of diagnostic physical examination tests for the upper limb. As in all fields of clinical medicine,

the determination of the validity of physical examination tests is complicated by the lack of an easily measurable gold standard for many conditions. Furthermore, the many factors that may affect the reliability or validity of these tests, such as clinician experience or patient variables, have not been studied. The reliability and validity of physical examination for the upper extremity should be an area of future research.

References

1. Cofield RH. Physical examination of the shoulder: effectiveness in assessing shoulder stability. In: Matsen FA III, Fu FH, Hawkins RJ, eds. *The shoulder: a balance of mobility and stability*. Rosemont, IL: American Academy of Orthopaedic Surgeons, 1993:331-343.
2. Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med* 1994;38:1-14.
3. Fess EE. The need for reliability and validity in hand assessment instruments. *J Hand Surg* 1986;11A:621-623 (editorial).
4. Wright JG, Feinstein AR. Improving the reliability of orthopaedic measurements. *J Bone Joint Surg* 1992;74B:287-291.
5. Feinstein AR. *Clinometrics*. Westford, MA: Murray Printing Company, 1987.
6. Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: the essentials*. 2nd ed. Baltimore: Williams & Wilkins, 1988.
7. *Manual of orthopaedic surgery*. American Orthopaedic Association, 1972.
8. Hoppenfeld S. *Physical examination of the spine and extremities*. East Norwalk, CT: Appleton-Century-Crofts, 1976.
9. Lister G. *The hand: diagnosis and indications*. 3rd ed. Edinburgh: Churchill-Livingstone, 1993.
10. Magee DJ. *Orthopedic physical assessment*. Philadelphia: WB Saunders, 1987.
11. Matsen FA III, Fu FH, Hawkins RJ. *The shoulder: a balance of mobility and stability*. Rosemont, IL: American Academy of Orthopaedic Surgeons, 1993.
12. Rothstein JM. *Measurement in physical therapy*. New York: Churchill Livingstone, 1985.
13. Tubiana R, Thomine J-M, Mackin E. *Examination of the hand and wrist*. London: Martin Dunitz, 1996.
14. Hamilton GF, Lachenbruch PA. Reliability of goniometers in assessing finger joint angle. *Phys Ther* 1969;49:465-469.
15. Boone DC, Azen SP, Lin C-M, Spence C, Baron C, Lee L. Reliability of goniometric measurements. *Phys Ther* 1978;58:1355-1360.
16. Low JL. The reliability of joint measurement. *Physiotherapy* 1976;62:227-229.
17. Solgaard S, Carlsen A, Kramhoft M, Petersen VS. Reproducibility of goniometry of the wrist. *Scand J Rehabil Med* 1986;18:5-7.
18. Wadsworth CT, Krishnan R, Sear M, Harrold J, Nielsen DH. Intrarater reliability of manual muscle testing and

- hand-held dynamometric muscle testing. *Phys Ther* 1987;67:1342-1346.
19. Rheault W, Beal JL, Kubik KR, Nowak TA, Shepley JA. Intertester reliability of the hand-held dynamometer for wrist flexion and extension. *Arch Phys Med Rehabil* 1989;70:907-910.
 20. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. *J Hand Surg* 1984;9A:222-226.
 21. Kendall FP. Manual muscle testing: there is no substitute. *J Hand Ther* 1991;4:159-161.
 22. Sapega AA. Muscle performance in orthopaedic practice. *J Bone Joint Surg* 1990;72A:1562-1574.
 23. Iddings DM, Smith LK, Spencer WA. Muscle testing: part 2. Reliability in clinical use. *Phys Ther Rev* 1961;41:249-256.
 24. Silver M, McElroy A, Morrow L, Heafner BK. Further standardization of manual muscle test for clinical study: applied in chronic renal disease. *Phys Ther* 1970;50:1456-1465.
 25. Greene BL, Wolf SL. Upper extremity joint movement: comparison of two measurement devices. *Arch Phys Med Rehabil* 1989;70:288-290.
 26. Harris SR, Smith LH, Krukowski L. Goniometric reliability for a child with spastic quadriplegia. *J Pediatr Orthop* 1985;5:348-351.
 27. Laupattarakasem W, Sirichativapee W, Kowsuwon W, Sribunditkul S, Suibnugarn C. Axial rotation gravity goniometer: a simple design of instrument and a controlled reliability study. *Clin Orthop* 1990;251:271-274.
 28. Mayerson NH, Milano RA. Goniometric measurement reliability in physical medicine. *Arch Phys Med Rehabil* 1984;65:92-94.
 29. Riddle DL, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting: shoulder measurements. *Phys Ther* 1987;67:668-673.
 30. Williams JG, Callaghan M. Comparison of visual estimation and goniometry in determination of a shoulder joint angle. *Physiotherapy* 1990;76:655-657.
 31. Agre JC, Magness JL, Hull SZ, et al. Strength testing with a portable dynamometer: reliability for upper and lower extremities. *Arch Phys Med Rehabil* 1987;68:454-458.
 32. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. *Phys Ther* 1986;66:206-209.
 33. Frisiello S, Gazaille A, O'Halloran J, Palmer ML, Waugh D. Test-retest reliability of eccentric peak torque values for shoulder medial and lateral rotation using the Biodex isokinetic dynamometer. *J Orthop Sports Phys Ther* 1994;19:341-344.
 34. Stratford PW. Summarizing the results of multiple strength trials: truth or consequence. *Physiother Can* 1992;44:14-18.
 35. De Krom MCTFM, Knipschild PG, Kester ADM, Spaans F. Efficacy of provocative tests for diagnosis of carpal tunnel syndrome. *Lancet* 1990;335:393-395.
 36. Phalen GS. The carpal-tunnel syndrome: seventeen years' experience in diagnosis and treatment of six hundred fifty-four hands. *J Bone Joint Surg* 1966;48A:211-228.
 37. Gilliatt RW, Wilson TG. A pneumatic-tourniquet test in the carpal-tunnel syndrome. *Lancet* 1953;2:595-597.
 38. Szabo RM, Gelberman RH, Dimick MP. Sensibility testing in patients with carpal tunnel syndrome. *J Bone Joint Surg* 1984;66A:60-64.
 39. Feierstein MS. The performance and usefulness of nerve conduction studies in the orthopedic office. *Orthop Clin North Am* 1988;19:859-866.
 40. Grundberg AB. Carpal tunnel decompression in spite of normal electromyography. *J Hand Surg* 1983;8:348-349.
 41. Jablecki CK, Andary MT, So YT, Wilkins DE, Williams FH. Literature review of the usefulness of nerve conduction studies and electromyography for the evaluation of patients with carpal tunnel syndrome. *Muscle Nerve* 1993;16:1392-1414.
 42. Marx RG, Hudak PL, Bombardier C, Graham B, Goldsmith C, Wright JG. The reliability of physical examination for carpal tunnel syndrome. *J Hand Surg* 1998;23B:499-502.
 43. Golding DN, Rose DM, Selvarajah K. Clinical tests for carpal tunnel syndrome: an evaluation. *Br J Rheumatol* 1986;25:388-390.
 44. Gellman H, Gelberman RH, Tan AM, Botte MJ. Carpal tunnel syndrome: an evaluation of the provocative diagnostic tests. *J Bone Joint Surg* 1986;68A:735-737.
 45. Heller L, Ring H, Costeff H, Solzi P. Evaluation of Tinel's and Phalen's signs in diagnosis of the carpal tunnel syndrome. *Eur Neurol* 1986;25:40-42.
 46. Katz JN, Larson MG, Sabra A, et al. The carpal tunnel syndrome: diagnostic utility of the history and physical examination findings. *Ann Intern Med* 1990;112:321-327.
 47. Kuschner SH, Ebrahimzadeh E, Johnson D, Brian WW, Sherman R. Tinel's sign and Phalen's test in carpal tunnel syndrome. *Orthopaedics* 1992;15:1297-1302.
 48. Stewart JD, Eisen A. Tinel's sign and the carpal tunnel syndrome. *BMJ* 1978;2:1125-1126.
 49. Rietz K-A, Ötne L. Analysis of sixty-five operated cases of carpal tunnel syndrome. *Acta Chir Scand* 1967;133:443-447.
 50. Durkan JA. A new diagnostic test for carpal tunnel syndrome. *J Bone Joint Surg* 1991;73A:535-538.
 51. Dellon AL, Mackinnon SE, Crosby PM. Reliability of two-point discrimination measurements. *J Hand Surg* 1987;12A:693-696.
 52. LaStayo P, Howell J. Clinical provocative tests used in evaluating wrist pain: a descriptive study. *J Hand Ther* 1995;8:10-17.
 53. Easterling KJ, Wolfe SW. Scaphoid shift in the uninjured wrist. *J Hand Surg* 1994;19A:604-606.
 54. Wolfe SW, Gupta A, Crisco JJ III. Kinematics of the scaphoid shift test. *J Hand Surg* 1997;22A:801-806.
 55. Watson HK, Ashmead D, Makhlof MV. Examination of the scaphoid. *J Hand Surg* 1988;13A:657-660.
 56. Regan DS, Linscheid RL, Dobyns JH. Lunotriquetral sprains. *J Hand Surg* 1984;9A:502-514.
 57. Wright IS. The neurovascular syndrome produced by hyperabduction of the arms. *Am Heart J* 1945;29:1-19.
 58. Telford ED, Mottershead S. Pressure at the cervico-brachial junction: an operative and anatomical study. *J Bone Joint Surg* 1948;30B:249-265.

59. Winsor T, Brow R. Costoclavicular syndrome: its diagnosis and treatment. *JAMA* 1966;196:697-699.
60. Rayan GM, Jensen C. Thoracic outlet syndrome: provocative examination maneuvers in a typical population. *J Shoulder Elbow Surg* 1995;4:113-117.
61. Mok DWH, Fogg AJB, Hohan R, Bayley JIL. The diagnostic value of arthroscopy in glenohumeral instability. *J Bone Joint Surg* 1990;72B:698-700.
62. Gerber C, Ganz R. Clinical assessment of instability of the shoulder with special reference to anterior and posterior drawer tests. *J Bone Joint Surg* 1984;66B:551-556.
63. Speer KP, Hannafin JA, Altchek DW, Warren RF. An evaluation of the shoulder relocation test. *Am J Sports Med* 1994;22:177-183.
64. Lyons AR, Tomlinson JE. Clinical diagnosis of tears of the rotator cuff. *J Bone Joint Surg* 1992;74B:414-415.
65. Ure BM, Tiling T, Kirchner R, Rixen D. Zuverlässigkeit der klinischen Untersuchung der Schulter im Vergleich zur Arthroskopie: eine prospektive Studie. *Unfallchirurg* 1993;96:382-386.
66. Cofield RH, Irving JF. Evaluation and classification of shoulder instability with special reference to examination under anaesthesia. *Clin Orthop* 1987;223:32-43.
67. Jobe FW, Tibone JE, Jobe CM. The shoulder in sports. In: Rockwood CA Jr, Matsen FA III, eds. *The shoulder*. Vol. 1. Philadelphia: WB Saunders, 1990:961-990.
68. Jørgensen U, Bak K. Shoulder instability: assessment of anterior-posterior translation with a knee laxity tester. *Acta Orthop Scand* 1995;66:398-400.
69. Neer CS II. Impingement lesions. *Clin Orthop* 1983;173:70-77.
70. Byl NN, Richards S, Asturias J. Intrarater and interrater reliability of strength measurements of the biceps and deltoid using a hand held dynamometer. *J Orthop Sports Phys Ther* 1988;9:399-405.
71. Cioni R, Passero S, Paradiso C, Giannini F, Battistini N, Rushworth G. Diagnostic specificity of sensory and motor nerve conduction variables in early detection of carpal tunnel syndrome. *J Neurol* 1989;236:208-213.
72. Redmond MD, Rivner MH. False positive electrodiagnostic tests in carpal tunnel syndrome. *Muscle Nerve* 1988;11:511-517.
73. Williams TM, Mackinnon SE, Novak CB, McCabe S, Kelly L. Verification of the pressure provocative test in carpal tunnel syndrome. *Ann Plast Surg* 1992;29:8-11.
74. Waris P. Occupational cervicobrachial syndromes: a review. *Scand J Work Environ Health* 1980;6:3-14 (suppl 3).
75. Iannotti JP. Lesions of the rotator cuff: pathology and pathogenesis. In: Matsen FA III, Fu FH, Hawkins RJ, eds. *The shoulder: a balance of mobility and stability*. Rosemont, IL: American Academy of Orthopaedic Surgeons, 1993:239-251.
76. O'Driscoll SW, Bell DF, Morrey BF. Posterolateral rotatory instability of the elbow. *J Bone Joint Surg* 1991;73A:440-446.
77. Hawkins RJ, Kennedy JC. Impingement syndrome in athletes. *Am J Sports Med* 1980;8:151-158.
78. Jobe FW, Jobe CM. Painful athletic injuries of the shoulder. *Clin Orthop* 1983;173:117-124.