Orthopaedic Division

The mission of CPA's Orthopaedic Division is the advancement of Orthopaedic physiotherapy through the promotion of excellence in education, practice and research. Having developed a standardized system of post graduate orthopaedic education, members have access to discounted courses and the opportunity to achieve a diploma in manipulative therapy, as well as access to the national orthopaedic symposium held annually. The Division also supports research in the field of orthopaedic physiotherapy through the Alan Morgan Fund (administered by the Physiotherapy Foundation of Canada), has an active web site, and produces five issues of the Orthopaedic Division Review Annually.

Scott Anderson, Chair
orthopaedics@physiotherapy.ca
Further information available at
http://www.physiotherapy.ca/divisions.htm

Scaphoid fracture: A case report illustrating evidence-based diagnosis and discussing measures of reliability and concurrent criterion-related validity

Peter Huijbregts, PT, MSc, MHSc, DPT, OCS, CGIMS, FAAOMPT, FCAMT; Michelle Hobby, BA, MPT; Penny Salmas, BPE, BSc PT, RCAMT, CGIMS

Peter Huijbregts, PT, MSc, MHSc, DPT, OCS, CGIMS, FAAOMPT, FCAMT
* Assistant Professor Online Education, University of St. Augustine for Health Sciences; * Residency Coordinator/Mentor, OMPT Residency Program Shelbourne Physiotherapy Clinic; * Consultant Physiotherapist, Shelbourne Physiotherapy Clinic Michelle Hobby, BA, MPT
* Resident, OMPT Residency Program Shelbourne Physiotherapy Clinic; * DPT Student, University of St. Augustine for Health Sciences Penny Salmas, BPE, BSc PT, RCAMT, CGIMS
* Mentor, OMPT Residency Program Shelbourne Physiotherapy Clinic; * Owner/Physiotherapist, Shelbourne Physiotherapy Clinic

Introduction

Physiotherapists establish a diagnosis and a prognosis by way of history taking, systems review, and tests and measures[1]. The data collected in these processes are used to determine whether the patient will benefit from physiotherapy intervention or needs to be referred for (co-) management with another health care provider. A patient presenting to physiotherapy with an undiagnosed fracture is not only a prime example illustrating the need for referral to a medical physician, but also a plausible scenario in jurisdictions where physiotherapists are allowed to act as direct access health care providers. The British Columbia Health Professions Act[2] clearly limits physiotherapy scope of practice disallowing treatment of a recent fracture, except when under physician direction and, therefore, signs and symptoms indicative of a scaphoid fracture constitute a clear indication for referral to a physician in this prvince (and likely also in other jurisdictions).

Evidence-based practice (EBP) represents a recent and major paradigm shift within medicine and allied health education and clinical practice from a reliance on authority-based knowledge and anecdotal evidence to the use of research-based evidence. However, research is not the sole component of EBP: Sackett et al[3] defined evidence-based medicine as the process of integrating the best research evidence available with both clinician expertise and patient values. For history items and tests and measures to be clinically useful for diagnosis, prognosis, and treatment planning within the EBP paradigm, the data they yield need to be reliable, valid, and responsive to clinically relevant change[1]. In the clinical scenario introduced above of a patient presenting to physiotherapy with signs and symptoms indicative of a scaphoid fracture, the question the physiotherapist needs to answer is, when is the probability of a scaphoid fracture high enough to warrant referral to a physician? Knowledge of the statistical measures associated with reliability and concurrent criterion-related validity of available diagnostic tests is a prerequisite to answer this question based on research evidence.

The goal of this article is threefold:

- To discuss the process of evidence-based diagnosis
- To discuss the statistical measures associated with reliability and concurrent criterion-related validity of diagnostic tests and measures
- To illustrate evidence-based diagnosis (and the use of statistical measures of reliability and validity) using the example of a patient with a suspected scaphoid fracture

Statistical measures of reliability

Reliability of a diagnostic test has two aspects. Intra-rater reliability refers to the stability of measurements taken by one rater across two or more trials; inter-rater reliability is concerned with the level of agreement between findings of two or more raters measuring the same subject or group of subjects[4]. Statistical measures used in research to establish

reliability are percentage agreement, variations of the (κ-statistic, intra-class correlation coefficients, measures of correlation, and measures of clinical significance[5].

Huijbregts[5] has provided an in-depth discussion of reliability and its statistical measures and noted that percentage agreement values, measures of correlation, and measures of clinical significance are invalid statistical measures for expressing agreement. Percentage agreement values do not correct for agreement based on chance, measures of correlation express the level of covariance rather than actual agreement, and statistical significance is influenced more by sample size than by actual agreement[5].

Both the intra-class correlation coefficients and variations of the (κ-statistic are chance-corrected indicators of agreement. Both statistics have many variants. Study methodology determines the appropriate type of these statistics to be used and both indices produce invalid results when limited variation exists within the data set[5]. An evaluation of statistical conclusion validity is needed when discussing the quality of evidence provided by reliability studies[5]. For the purpose of this case report, it is sufficient to know that (κ-values of 0.60-0.80 are agreed upon as indicating substantial agreement and are commonly assumed sufficient for clinical decision making[4].

Statistical measures of concurrent criterion-related validity

The validity of a measurement is the degree to which a meaningful interpretation can be inferred from this measurement[1]. Pretty and Maupome[6] defined the validity of a diagnostic procedure at the basic conceptual level as the extent to which it measures what it claims to measure. Of all aspects of validity, most relevant to history items and tests and measures collected for diagnostic purposes is concurrent criterion-related validity. Criterion-related validity indicates the extent to which a test can be used as a substitute measure for an established gold standard or criterion test[4]. Concurrent criterion-related validity involves two tests performed at approximately the same time; the research then evaluates whether the test studied could be used as a clinical alternative to the gold standard test[4]. Research into concurrent criterion-related validity yields data on the statistical measures of accuracy, sensitivity, specificity, and positive and negative predictive values of a history item, test, or measure[7]. Positive and negative likelihood ratios can be calculated from values for sensitivity and specificity[7].

Often, diagnostic tests and measures yield a dichotomous result: either the patient has or does not have the disease or dysfunction. When we compare a clinical test or measure to a gold standard (or criterion) test, there are four possible outcomes[6]:

- True positive (TP): the test indicates that the patient has the disease or dysfunction and this is confirmed by the criterion test
- False positive (FP): the clinical test indicates that the disease or dysfunction is present, but this is not confirmed by the criterion test
- False negative (FN): the clinical test indicates absence of the disorder, but the criterion test shows that the disease or dysfunction is present
- True negative (TN): both the clinical and the criterion test agree that a disease or dysfunction is absent

We can map out these four outcomes in a two-by-two contingency table[6,7] as illustrated in Table 1. From this table we can calculate the statistical measures of accuracy, sensitivity, specificity, and positive and negative predictive values as indicated in Table 2[7].

Data on diagnostic accuracy can give us an overall idea of the value of a diagnostic test, but accuracy has little further value in making actual diagnostic decisions, as it does not distinguish between the diagnostic value of positive and negative test results. Negative predictive values tell us how likely it is that patients who test negative do not have the disease or dysfunction, whereas positive predictive values indicate how likely it is for a person who tests positive to in fact have the disease or dysfunction. The usefulness of predictive values seems great, but is limited by the fact that for predictive values to apply, the prevalence in the clinical population we are examining has to be identical to the prevalence in the study population from which the predictive values were derived[7]. Davidson[7] noted that consequently we could almost disregard positive and negative predictive values in the diagnostic process.

Sensitivity and specificity are easy to interpret when they are high[6,7]. In case of a highly sensitive test, negative test results will likely rule out the disease or dysfunction, as there are very few false negatives when sensitivity is high[7]. The higher the specificity of a test, the more likely that a positive test result will rule in a disease or dysfunction: in tests with high specificity, the number of false positives is low[7]. So we can use highly sensitive tests for ruling out disease or dysfunction and highly specific tests for ruling in disease or dysfunction. Davidson[7] used the mnemonics **SnOUT** and **SpIN**:

- **SnOUT**: with highly **S**ensitive tests, a **N**egative result wil rule a disorder OUT
- **SpIN**: with highly **S**pecific tests, a **P**ositive result will rule a disorder **IN**

Of course, the ideal diagnostic test would possess both high sensitivity and high specificity, but for most diagnostic procedures these statistical measures are inversely related: high sensitivity usually means low specificity, and vice versa[6]. In fact, a diagnostic test can only be 100% sensitive and 100% specific if there is no overlap between the population that has the disease or dysfunction and the population that does not[6]. However, as Pretty and Maupome[6] noted, in this case the presence of the disease or disorder is often so obvious that no diagnostic testing is needed.

Another problem with sensitivity and specificity is that these measures tell us how often a test will be positive or negative in patients that we already know have or do not have the condition[7]. This, of course, does not correspond with the clinical situation, where we do not know whether our patient has or does not have a certain condition. Likelihood ratios (LR) summarize the information

Result clinical test	Result criterion test		
	Present	*Absent*	*Totals*
Positive	TP	FP	TP + FP
Negative	FN	TN	FN + TN
Totals	TP + FN	FP + TN	TP + FP + FN + TN

Table 1: Two-by-two contingency table[6,7].

Statistical measure	Definition	Calculation
Accuracy	The proportion of people who were correctly identified as either having or not having the disease or dysfunction	(TP + TN) / (TP + FP + FN + TN)
Sensitivity	The proportion of people who have the disease or dysfunction who test positive	TP / (TP + FN)
Specificity	The proportion of people who do not have the disease or dysfunction who test negative	TN / (FP + TN)
Positive predictive value	The proportion of people who test positive who have the disease or dysfunction	TP / (TP + FP)
Negative predictive value	The proportion of people who test negative who do not have the disease or dysfunction	TN / (FN + TN)

Table 2: Definition and calculation of statistical measures of concurrent criterion-related validity[7].

contained in the statistical measures of sensitivity and specificity[7]. We can calculate likelihood ratios as follows:

- Positive LR = sensitivity / (1-specificity)
- Negative LR = (1-sensitivity) / specificity

A positive LR tells us how likely a positive test result is in patients that have a certain disease or dysfunction as compared to how likely it is in patients who do not have the disease or dysfunction. A negative LR provides information on how likely a negative test result is in patients with the disorder as compared to how likely it is in patients without the disorder.

Davidson[7] provides simple guidelines on interpreting likelihood ratios:

- The higher a positive LR, the more certain one can be that a patient with a positive test has the disorder for which you are testing
- The lower a negative LR, the greater the chance that a person with a negative test result does not have the disorder
- A positive or negative LR close to 1.0 provides little change in the probability that a patient has or does not have a disease or dysfunction, i.e., this test is of little diagnostic value

A LR can be used qualitatively, but it can also be used quantitatively to express the effect a test result has on post-test probability of a certain disorder. We can use a nomogram or a mathematical solution involving the calculation of pre- and post-test odds to determine the effect on post-test probability. Further information on this quantitative approach is provided in the reference by Davidson[7], accessible full-text on the Internet.

Evidence-based diagnosis of scaphoid fractures

Evidence-based diagnosis makes use of the best available research evidence into properties of demographic data, history items, and tests and measures used for diagnostic purposes. As an example of the use of data on reliability and validity for the OMPT residency program at his clinical workplace, the first author did a search for full-text, peer-reviewed references using the key word scaphoid on the Infotrac and Proquest databases available to students and faculty through the University of St. Augustine for Health Sciences.

Demographic data

Scaphoid fractures account for 75% of all carpal injuries in men aged 15-30[8,9] and for eight percent of all sports injuries[8]. The fracture is rare in children and elderly people: children more frequently fracture the distal radial epiphysis with a fall on the outstretched wrist and older people will commonly fracture the distal radius[10].

History items

The mechanism of injury for a scaphoid fracture is frequently a fall on the extended wrist. For a fracture to occur, a wrist extension of greater than 95° is required[8]. In the controlled laboratory situation, scaphoid fractures have been produced with wrists in more than 90° of extension and more than 10° of radial deviation with in excess of 400(kg) of force[9]. However, flexion and compression mechanisms are described in the literature[10]. A description of a rapid onset of swelling is indicative of haemarthrosis,

possibly due to fracture[8]. Patients complaining of pain localized to the dorsoradial wrist, pain on gripping, and decreased wrist active range of motion (AROM) constitutes the classical group of symptoms indicative of a scaphoid fracture[8]. Data regarding reliability and validity of these history items is lacking.

Tests and measures

Tests with established inter-rater reliability and concurrent criterion-related validity for the diagnosis of scaphoid fractures include AROM tests and palpation tests:

- Scaphoid fractures will result in an inability to extend the wrist with frequently more normal wrist flexion[8,9]. Tenderness on thumb movement is 100% sensitive for diagnosing scaphoid fractures and 48% specific[10].
- Palpation of the anatomic snuffbox may be positive for pain, fullness, or swelling. Palpation with the wrist in ulnar deviation exposes the waist of the scaphoid, the location of 70% of scaphoid fractures[8]. Exquisite tenderness is common with fracture[8]. Inter-rater agreement for anatomical snuffbox tenderness yielded a (κ-value in patients with suspected scaphoid injury of 0.66411. Tenderness and swelling upon palpation of the anatomical snuffbox is 100% sensitive, but has a specificity of only 9%[10].
- Tenderness of the scaphoid tubercle at the radio-palmar aspect of the wrist, immediately radial to flexor carpi radialis tendon (which is more accessible in radial deviation) is more reliable and more specific for scaphoid fracture. Scaphoid tubercle tenderness yielded an inter-rater agreement of κ=0.739 in patients with suspected scaphoid injury[11]. In addition, this test has been shown to be 100% sensitive and 30% specific[10].
- A combination of tenderness with palpation of the anatomic snuffbox and the scaphoid tubercle and pain on thumb movement is 74% specific 24 hours post-injury[10]. Assuming for the sake of this example of diagnostic clinical reasoning conform the EBP paradigm that all three, rather than two, tests have an individual (and therefore also combined) sensitivity of 100%, we can calculate a positive LR for this combination of tests as 100/(100-74) = 3.85.

Pain on compression through the thumb is another test commonly done to diagnose scaphoid lesions. However, this test yielded only an inter-rater agreement of κ=0.289 in patients with a suspected scaphoid injury[11].

Patient presentation

A 66-year-old woman presented to physiotherapy with a complaint of left-sided bicipital region pain. Evaluation showed a tendinopathy of the tendon of the long head of the biceps, a supraspinatus insertion tendinopathy, decreased rotator cuff strength and endurance, and a decreased length of the posterior glenohumeral capsuloligamentous structures. Treatment consisted of transverse friction massage, glenohumeral joint mobilization, education, a home exercise program, and modalities (ultrasound, interferential current, and cold application).

On the third visit, the patient reported a fall onto the extended right wrist. There was a temporary inability to grip objects, but the wrist pain was already decreasing. On a fourth visit, two weeks later, the patient reported continued pain in the dorsoradial right wrist with axial pressure and contraction of the wrist and finger flexors. On AROM and PROM testing, especially wrist extension was painful and limited. Pain was present with palpation of the right scaphoid tubercle, the anatomic snuffbox, and on AROM of the thumb. The patient was referred to her physician with a clinical suspicion of scaphoid fracture. A subsequent series of six radiographic views showed a radiolucent line across the proximal ventro-ulnar aspect of the right scaphoid. The patient was immobilized in a short-arm thumb spica splint due to suspected undisplaced scaphoid fracture.

Discussion

There is no clear consensus, let alone an evidence-based consensus, on the management of undisplaced scaphoid fractures: authors disagree on length of immobilization and immobilization position[8,10]. What is evident is the adverse effect an undiagnosed and, therefore, inappropriately treated scaphoid fracture can have: Hooper et al[8] noted the risk of non-union, rotatory instability of the wrist, avascular necrosis, degenerative radiocarpal arthritis, and carpal collapse with subsequent impairments of wrist and hand function. One could argue that the physiotherapist should refer any patient with a suspected scaphoid fracture due to the great risk of adverse effects in a missed fracture as noted above. However, pain in the dorsoradial aspect of the wrist can be the result of many different pathologies and dysfunctions: Table 3 provides a list of differential diagnostic options. It is the physiotherapist's responsibility to establish a physiotherapy diagnosis prior to treatment. It is, of course, not the responsibility of the physiotherapist to establish a medical diagnosis for every patient presenting with dorsoradial wrist pain. However, the therapist needs to be able to identify signs and symptoms indicating that a patient is not appropriate for (sole) management by physiotherapy and, when needed, refer a patient for medical management.

This case report demonstrates the process of evidence-based diagnosis for a patient with a suspected scaphoid fracture. A literature search provided information on

Osteoarthritis CMC I
Scapholunate dissociation
Radial styloid fracture
Trapezium fracture
Trapezoid fracture
Lunate fracture
De Quervain's disease
Flexor carpi radialis tendonitis
Tendonitis extensores carpi radialis
Tendonitis extensor pollicis longus
Tendonitis extensor digitorum communis
Kienbock's disease
Colles fracture

Table 3: Differential diagnosis for dorsoradial wrist pain[8].

reliability, sensitivity, and specificity of diagnostic tests and measures: a combination of tenderness with palpation of the anatomic snuffbox and the scaphoid tubercle and pain on thumb movement was shown to have a specificity of 74% and an individual test sensitivity of 100% (resulting in a calculated positive LR of 3.85) for diagnosing a scaphoid fracture[10]. Inter-rater agreement for two of these tests was determined to be substantial and thus sufficient for clinical decision making[11]. A positive response for a test or regimen of tests with a high specificity and a high positive LR helps to rule in a particular condition. Both specificity and positive LR in this case were moderately, but not extremely high. This brings into play the other two components of EBP: clinician expertise and patient values. Dorsoradial wrist pain after a fall onto the extended wrist with a painful restriction of mainly wrist extension and pain with gripping in combination with positive results on the above combination of tests, despite the demographics not fitting, sufficiently raised the suspicion of the clinician to warrant referral to a physician. The risk of adverse effects in case of a missed diagnosis was good reason for the patient to follow up with the physiotherapist's request to go see the physician. As noted above a referral for further medical diagnosis was appropriate and justified by imaging findings.

Conclusion

We presented a case report illustrating the use of data collected from a literature search on test reliability and concurrent criterion-related validity for identifying a patient with signs and symptoms of dorsoradial wrist pain not amenable to physiotherapy management. Referral to the medical physician and subsequent imaging confirmed our suspicion of an undisplaced scaphoid fracture. Using data on test reliability and validity is of benefit to our diagnostic abilities and a necessity in this era of EBP, but also strengthens our case when communicating the reason for referral to the physician thereby facilitating appropriate patient care. We advocate for a continued and increased emphasis on diagnostic test properties and their effect on clinical decision-making in entry-level and post-graduate medical and allied health education.

Address correspondence to:
Dr. Peter Huijbregts, PT, Shelbourne Physiotherapy Clinic,
100B-3200 Shelbourne Street, Victoria, BC V8P 5G8, Canada
(250) 598-9828 (phone); (250) 598-9588 (fax);
shelbournephysio@telus.net (e-mail)

References

1. Guide to Physical Therapist Practice, 2nd Ed. Phys Ther 2001; 81: 9-744.

2. Health Professions Act, Physical Therapists Regulation (BC Reg. 485/94: OC 1563/94). Available at: http://www.qp.gov.bc.ca/statreg/reg/H/HealthProf/485_94.htm. Accessed on June 2, 2004.

3. Sackett DL, et al. Evidence-Based Medicine. How to Practice & Teach EBM. New York, NY: Churchill Livingstone, 1997.

4. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Norwalk, CT: Appleton & Lange, 1993.

5. Huijbregts PA. Spinal motion palpation: A review of reliability studies. J Manual Manipulative Ther 2002;10:24-39.

6. Pretty IA, Maupome G. A closer look at diagnosis in clinical dental practice: Part 1. Reliability, validity, specificity and sensitivity of diagnostic procedures. J Can Dent Assoc 2004;70:251-255. Available at: http://www.cda-adc.ca/jcda/vol-70/issue-4/251.pdf. Accessed July 4, 2004.

7. Davidson M. The interpretation of diagnostic tests: A primer for physiotherapists. Aust J Physiother 2002;48:227-233. Available at: http://www.physiotherapy.asn.au/AJP/48-3/0902Davidson.pdf. Accessed July 4, 2004.

8. Hooper RM, Sonzogni JJ, Taleisnik J, Weber ER. Wrist: Beware the scaphoid fracture. Patient Care 1992;26(4):74-93.

9. Rettig AC. Athletic injuries of the wrist and hand: Part I: Traumatic injuries of the wrist. Am J Sports Med 2003;31:1038-1048.

10. Krasin E, Goldwirth M, Gold A, Goodwin DR. Review of current methods in the diagnosis and treatment of scaphoid fractures. Postgrad Med J 2001;77:235-237.

11. Gow K, Carley S, Mackway-Jones K. Interobserver variability in examination of the injured scaphoid. Emergency Medicine Journal 2002;19:280.