

Postprint Version	1.0
Journal website	http://dx.doi.org
Pubmed link	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12161838&query_hl=14&itool=pubmed_docsum
DOI	10.1053/apmr.2002.33728

From the Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam (Smidt, van der Windt, Mourits, Devillé, Bouter); Department of General Practice, Division of Public Health, Academic Medical Center, University of Amsterdam, Amsterdam (Assendelft); and Department of Social Psychiatry, University of Groningen, Groningen (de Winter), The Netherlands.

Supported by the Netherlands Organization for Scientific Research and the Health Insurance Council's Fund for Investigative Medicine.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.

Reprint requests to Nynke Smidt, PhD, Institute for Research in Extramural Medicine, VU University Medical Center, Van der Boechorststr 7, 1081 BT Amsterdam, The Netherlands, e-mail: n.smidt.emgo@med.vu.nl.

Interobserver Reproducibility of the Assessment of Severity of Complaints, Grip Strength, and Pressure Pain Threshold in Patients With Lateral Epicondylitis

NYNKE SMIDT, PHD, DANIËLLE A. VAN DER WINDT, PHD, WILLEM J. ASSENDELFT, MD, PHD, ANNEKE J. MOURITS, MSC, WALTER L. DEVILLE, MD, ANDREA F. DE WINTER, PHD, LEX M. BOUTER, PHD

Objective: To evaluate the interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis in primary care.

Design: Two physiotherapists assessed independently, and in randomized order, the severity of complaints scored on an 11-point numeric rating scale, pain-free grip strength, maximum grip strength, and pressure pain threshold.

Setting: Primary care center in the Netherlands.

Participants: Fifty patients were assessed by both physiotherapists.

Interventions: Not applicable.

Main Outcome Measures: Intraclass correlation coefficients (ICCs) and 95% confidence intervals (CIs) were calculated to assess interobserver reliability. The Bland and Altman method was used to assess interobserver agreement, which included calculation of the mean difference between the observers (d), the 95% CI for d, the standard deviation of the differences, and the 95% limits of agreement. Finally, the smallest detectable difference (SDD) was calculated.

Results: The ICC for the severity of complaints was .90. The ICCs for the pain-free grip strength and maximum grip strength were .97 and .98, respectively. The pressure pain threshold showed a lower reliability (ICC=.77). The interobserver agreement for all outcome measures was good, but systematic differences in assessment between the

physiotherapists were found for the maximum grip strength and pressure pain threshold. For pressure pain threshold, the SDD was clearly larger than the predefined acceptable difference of 10% of the total range of measurement.

Conclusions: The interobserver reliability of severity of complaints and grip strength was excellent, whereas the pressure pain threshold showed unsatisfactory reliability. Grip strength and overall assessment of the severity of complaints are useful and reliable measures for the assessment of lateral epicondylitis. Pain-free grip strength, in particular, is relatively easy to perform and has been shown to be associated with other measures of functional disability in patients with lateral epicondylitis. We, therefore, recommend the use of pain-free grip strength in both research and clinical practice.

Lateral epicondylitis is characterized by pain at the lateral side of the elbow, which is increased by pressure on the lateral epicondyle and during resisted dorsiflexion of the wrist.¹ Pain provoked by palpation or by specific manual tests at the lateral epicondyle are the most important diagnostic findings. In clinical practice, these tests are also used to describe patient status or to evaluate treatment efficacy.

A recent systematic review² including 23 randomized clinical trials (RCTs) on the effectiveness of physiotherapy for lateral epicondylitis showed that most RCTs included grip strength (19/23) and pain (15/23) to evaluate the efficacy of treatment. However, specification of grip strength in terms of pain-free grip strength or maximum grip strength was not properly reported in 10 of 19 RCTs. Only 9 RCTs described their measurement protocol, and the large majority of studies did not address the reliability of the methods used. Pain was, in most studies, assessed according to a visual analog scale or a Likert scale. Only 1 study³ measured the pressure pain threshold.

Many RCTs on musculoskeletal disorders included assessment of the severity of complaints when evaluating the efficacy of treatment. However, none of these has investigated the interobserver reproducibility of this assessment.^{4,5} Although the literature contains several reports on the reproducibility of grip strength in healthy subjects, the reproducibility of grip strength in patients with lateral epicondylitis has received little attention. Stratford et al⁶ investigated the intraobserver reliability of grip strength in 35 patients with tennis elbow and reported high intraobserver reliability coefficients for pain-free grip strength (.97) and maximum grip strength (.98). However, information on the interobserver reliability of grip strength is still lacking. In their study, Stratford⁶ used the Smedley hand dynamometer, although the Jamar hand dynamometer is more frequently used by hand therapy clinics.

Several authors⁷⁻¹³ have recommended the use of the pressure pain threshold meter in clinical practice and research. Interobserver reliability of pressure pain threshold measurements has been reported as satisfactory in healthy subjects¹⁴⁻¹⁶ and in patients with myofascial trigger points.¹⁷ However, some studies^{14,18} showed that reaction time of the observer and variation in the rate of pressure increase may affect the reliability. There is, as yet, no information about the reproducibility of measurements of pressure pain threshold in patients with lateral epicondylitis.

Good reproducibility is a prerequisite for a valid measurement. In addition, during long-term follow-up or prolonged treatment of patients, different examiners will often assess the same patient. Therefore, interobserver reproducibility should be known and acceptable. Given the importance of clinical outcomes such as grip strength and pressure pain threshold in clinical practice and research and the scarcity of information on reproducibility of these measures, we decided to determine the interobserver reproducibility of the severity of complaints, grip strength, and the pressure pain threshold in patients with lateral epicondylitis. The study was performed within the framework of an RCT on the effectiveness of primary care interventions for patients with lateral epicondylitis.¹⁹

METHODS

Patients

The participants of the reproducibility study were recruited from the study population of an RCT, comparing the effectiveness of corticosteroid injections, physiotherapy, and a waitand- see policy for lateral epicondylitis.¹⁹ Patients were eligible for participation in the RCT if they had pain at the lateral side of the elbow, which increased with pressure on the lateral epicondyl and during resisted dorsiflexion of the wrist. Other inclusion criteria were elbow complaints ≥ 6 weeks, age between 18 and 70 years, ability to complete questionnaires (sufficient knowledge of the Dutch language), and informed consent. Exclusion criteria were treatment with physiotherapy or injection(s) during the preceding 6 months; bilateral symptoms; presence of signs and symptoms suggesting some other cause of lateral elbow pain (eg, neck problems); congenital or acquired deformities of the elbow; surgery of the elbow; dislocation, tendon ruptures, or fractures in the elbow area in the preceding 12 months; systemic disorders of the musculoskeletal system; neurologic disorders; and contraindications for corticosteroid injections. A total of 185 patients were enrolled in the RCT between September 1997 and October 1998.

During follow-up (3, 6, 12, 26, or 52wk after randomization), a random sample of 50 patients were invited to participate in the reproducibility study. To ensure inclusion of patients with a wide range of the severity of symptoms, patients were eligible for participation at any time of follow-up. Additional selection criteria for the reproducibility study were informed consent, no bilateral elbow complaints, no surgery of the elbow, and no cast.

Measurements

Between July 1998 and January 1999, 2 research physiotherapists (SK, AJM, see Acknowledgments) independently, and in random order, evaluated each of the 50 patients. The research physiotherapists were registered as physiotherapists with the Royal Dutch College of Physiotherapists and had several years of experience in clinical practice and research. The measurement protocol was discussed and the physiotherapists were trained thoroughly before the start of patient selection. A standardized history was taken by one of the research physiotherapists, in the presence of the other. Subsequently, both research physiotherapists independently performed a physical examination of the elbow and wrist of each patient consisting of inspection of the elbow, active and passive examination, and resisted movements. The severity of complaints was subsequently scored on an 11-point scale (0=no complaints, 10=very severe complaints).

Pain-free grip strength and maximum grip strength was measured with the Jamar hand dynamometer,^a while the patient was standing with the elbow in complete extension and the shoulder and radioulnar joints in neutral rotation. Pain-free grip strength of the involved arm was measured first. Patients were instructed to squeeze the dynamometer slowly until they began to feel discomfort. Both physiotherapist and patient were unable to read the dial of the dynamometer during testing. The pain-free grip strength was measured 3 times, with a 20-second rest interval between each measurement. The mean value (in kg) of 3 efforts was calculated and included in the analysis. Pain-free grip strength was always assessed before maximum grip strength. Following a similar procedure, 3 maximum grip strength measurements were performed for the uninvolved arm followed by measurements for the involved arm. The intraobserver reliability of the 3 measurements was excellent for both pain-free and maximum grip strength (intraclass correlation coefficient [ICC] range, .95-.97). Verbal instructions were given to each subject: "Are you ready? Squeeze as hard as you can . . . harder . . . harder . . . relax!" The Jamar hand dynamometer was calibrated according to a standard procedure by using standardized test weights. Less than 4% error has been reported.²⁰

Pressure pain threshold on the lateral epicondyle was measured with an algometer^b while the patient was sitting with the arm in 30° of abduction; the elbow in 90° of flexion; and with the forearm, wrist, and hand supported. The research physiotherapist lightly palpated the involved common extensor tendon to identify the most sensitive area. Subsequently, beginning with the uninvolved side, the head of the algometer (hard rubber tip, surface area 1cm²) was applied perpendicular to the common extensor tendon at the lateral epicondyle until patients indicated the moment at which the sensation of pressure changed to pain. The pressure pain threshold was measured 3 times on both sides, with a 20-second rest interval between each measurement. The intraobserver reliability of these 3 measurements

was excellent (ICC range, .91–.96). For the analysis, the mean value (in kg/cm²) of the 3 efforts was calculated. The maximum applied pressure was 6kg/cm².²¹

The first physiotherapist left the room after taking all measurements, and the same procedure was repeated by the second physiotherapist. To prevent sequence effects between the physiotherapists, the sequence of evaluation by the physiotherapists was randomized and stratified according to the timing of the evaluation (≤ 6 wk or > 6 wk after the start of the RCT). The physiotherapists independently recorded the results of their assessment (history taking, physical examination, overall assessment of severity, grip strength, pressure pain threshold) on standardized forms. Their findings were not discussed until completion of the study in all 50 patients. In the framework of the randomized trial, the physiotherapists were also blinded regarding the allocated treatment of patients because this information might have influenced their assessment.

Analysis of Reproducibility

For the quantification of interobserver reproducibility, 2 types of analysis were applied: the Bland and Altman method for assessing agreement²² and the ICC²³ for the evaluation of reliability.

The Bland and Altman method provides insight into the distribution of differences between observers.²⁴ It presents the size, direction, and range of differences between observers in the same units as the original scale. Interobserver agreement (Bland and Altman) was quantified by calculating the mean difference (d) between the 2 observers and the standard deviation (SD) for this difference. The closer d is to zero and the smaller the SD of this difference, the better the interobserver agreement. Differences between the 2 observers were plotted against the mean of the measurements made by the 2 observers. This graph shows the size, direction, and range of the differences and shows whether differences between observers are consistent across the range of measurements. The 95% limits of agreement were defined as the mean difference between the observers ± 1.96 SD of the differences, indicating the total error (bias and random error together). The presence of bias between the observers is estimated by calculating the 95% confidence interval (CI) for d . The 95% CI for d was calculated as $d \pm t_{n-1} SE(d)$, where n is the number of subjects and SE the standard error of the mean equal to SD/\sqrt{n} . If zero lies outside the 95% CI, systematic differences (bias) between the observers exist.²²

When trying to determine whether a real change in outcome has occurred in clinical practice and research, a change must be at least the smallest detectable difference (SDD) of a measurement procedure. The SDD was calculated as $1.96 \cdot \sqrt{2} \cdot SEM^2$ where the standard error of measurement was $SD \cdot \sqrt{1-r}$ and r the reliability coefficient.²⁵ Although no clear criteria for an acceptable value of interobserver agreement are available for these specific outcome measures, we defined, a priori, that a difference in measurements between the observers (d) of 10% of the total range of measurements would be acceptable. The decision regarding this cutoff point is more or less arbitrary. The reader may apply his/her own cutoff point in the interpretation of our findings.

ICCs provide information on the ability of 2 observers to differentiate between subjects (reliability).²⁶ The ICC was defined as the ratio of variance among patients (subject variability) over the total variance (subject variability, observer variability, random error variability), which ranges between 0 (no reliability) and 1 (perfect reliability). An ICC of less than .75 was considered unsatisfactory.^{27,28} Two-way random effect models were used to calculate ICCs according to Fleiss.²⁹

RESULTS

Patients

The 50 participants of the reproducibility study were enrolled between July 1998 and January 1999. The mean age \pm SD of the 50 patients was 47 ± 11 years, and 40% of the patients were women (table 1). In 35 cases (70%), the dominant arm was involved. Nine (18%) of the 50 participants indicated much improvement of their elbow complaints, but most patients still reported symptoms. Fifteen (30%) patients reported moderate or severe pain scores (pain during the day: numeric rating scale [NRS] score, ≥ 3).

[TABLE 1]

Agreement

Table 2 summarizes the results of interobserver agreement on the severity of complaints, pain-free grip strength, maximum grip strength, and pressure pain threshold on the lateral epicondyl. The results of the assessments made by the 2 observers were quite similar for all measurements. Pain intensity did not appear to influence the degree of interobserver agreement (data not shown). However, the mean level of pain was low (2.2 ± 2.7) according to most patients.

[TABLE 2]

Figure 1 shows the differences between observers, plotted against the mean score of both observers for pain-free grip strength, and the maximum grip strength of the involved arm. There was a wide variation in grip strength between patients. Although the limits of agreement were wider for the pain-free grip strength than for the maximum grip strength, systematic differences were found for the latter ($d[95\% \text{ CI}], -.81 [-1.37 \text{ to } -.25]$). The maximum grip strength for the involved arm reported by observer B was higher than reported by observer A in approximately 72% of the patients.

[FIGURE 1]

The interobserver differences against the mean score of both observers for pressure pain threshold are presented in figure 2. Observer B recorded higher scores than observer A for approximately 74% of the patients. As the mean pressure pain threshold increased, observer B reported higher scores more frequently than observer A. The interobserver differences also increased with increasing mean scores for the involved arm. Bias was found for the pressure pain threshold of the involved arm ($d[95\% \text{ CI}], -.66 [-.96 \text{ to } -.36]$).

[FIGURE 2]

For grip strength, the SDD was clearly lower than the predefined acceptable difference of 10% of the total range of measurement. For the severity of complaints, the SDD was equal to the predefined acceptable difference of 10% (0.7 point vs 10% of 7=0.7). The SDD for the pressure pain threshold was much greater than the predefined acceptable difference (1.3 vs 10% of 5=0.5).

Reliability

The results of the analysis of variance (ANOVA) are presented in table 3. The systematic differences between the observers for maximum grip strength and the pressure pain threshold of the uninvolved arm were confirmed in the ANOVA by low *P* values for observer variance. The ICC values ranged from .72 to .98. The reliability of the pressure pain threshold was relatively poor and unsatisfactory for the uninvolved arm ($\text{ICC} < .75$). Almost perfect reliability was found for the pain-free grip strength and maximum grip strength ($\text{ICCs} \geq .97$).

[TABLE 3]

DISCUSSION

This study investigated the interobserver reproducibility of severity of complaints, grip strength, and pressure pain threshold in 50 patients with lateral epicondylitis. The results showed excellent reliability for the severity of complaints, pain-free grip strength, and maximum grip strength ($\text{ICC} \geq .90$). The pressure pain threshold was clearly less reliable (ICC range, .72–.77). Systematic differences (bias) between the physiotherapists were found for the assessment of maximum grip strength and pressure pain threshold of the involved arm. This bias could be partly caused by differences in the verbal encouragement given by the physiotherapists during the measurement of maximum grip strength. The systematic differences in the assessment of pressure pain threshold may also be explained by variation in the instructions given to the patient, placement of the algometer on

the lateral epicondyl, rate of pressure increase, and reaction time of the physiotherapists when patients indicated that application of pressure should be stopped.

Standardized positioning of patients for grip strength measurements, according to the American Society of Hand Therapists⁸ (ASHT) indicates that “the patient should be seated with his shoulder adducted and neutrally rotated, elbow flexed at 90° and the forearm and wrist in neutral position.” Measuring grip strength with extended elbow, as we did, may reflect a more sensitive outcome for patients with lateral epicondylitis,^{30,31} although it limits comparability with the recommended standardized positioning method for grip strength. Furthermore, ASHT recommended that “the second handle position of the dynamometer should be used when evaluating grip strength” and “the mean of three successive trials should be used as the outcome measure of grip strength.”⁸ In our study, the handle position of the Jamar dynamometer was adjusted to the hand size of the patient and standardized for each measurement.

Earlier studies^{32,33} showed excellent interobserver reliability of grip strength measurements with the Jamar dynamometer in healthy subjects and patients with cumulative trauma disorders (ICC range, .93–.98). Our study confirms these findings in patients with lateral epicondylitis and showed that the mean difference between observers was only 0.8kg for maximum grip strength and pain-free grip strength of the involved arm.

To determine the effectiveness of an intervention, knowledge of factors contributing to the variation in observed measurement scores is necessary. When trying to detect a real change between 2 observations with statistical significance, this change must be at least the SDD of the measurement procedure. The SDD for grip strength fell within the predefined criteria for interobserver agreement (<10% of the total range). For the general assessment of severity, the SDD was just equal to this criterion (0.7 points equals 10% of 7). The SDD for the pressure pain threshold was clearly higher than 10% (1.3 > 10% of 5), indicating that a true difference between measurements may be difficult to detect when using different observers. This finding, again, confirms the relatively unfavorable clinimetric performances of pressure pain threshold in patients with lateral epicondylitis.

We limited our study to the evaluation of the reproducibility of outcome measures for lateral epicondylitis. The validity of these measures is difficult to evaluate and has been less often subjected to further investigation. Because of a lack of gold standards in this area, no data on criterion validity are available. Yet, a few studies provide useful information about the construct validity of pressure pain threshold and grip strength in lateral epicondylitis. The pressure pain threshold has been reported to be significantly lower in patients with lateral epicondylitis compared with healthy controls³⁴ or compared with the unaffected elbow.³⁵ However, Pienimäki et al³⁴ could not find an association with pain drawings or with scores on a pain and disability questionnaire, which seems to question the validity of pressure pain threshold in lateral epicondylitis. Data about the construct validity of grip strength in lateral epicondylitis are also scarce, but previous studies have shown that pain-free grip strength is associated with working disability³⁶ and functional disability (as measured by the pain-free function questionnaire and by a visual analog scale).³⁷ These associations were weaker for maximum grip strength. These findings seem plausible. Patients will try to avoid pain during daily activities and may rarely challenge their maximum grip strength. Functional ability may therefore mainly be determined by their pain-free capabilities. Furthermore, changes of pain-free grip strength over time were associated with changes of functional disability and with general perceived improvement. In contrast, changes in maximum grip strength were not associated with improvement of symptoms.^{30,31} These results and their clinical implications favor the use of pain-free grip strength in lateral epicondylitis.

The research physiotherapists in our study remained unaware of each other’s findings in all 50 patients. Therefore, we feel that blinding was successful and that our results on interobserver reproducibility were not biased by knowledge about the findings of the other observer. Our research physiotherapists were well trained, had considerable expertise in assessing patients with lateral epicondylitis, and had discussed the assessment protocol thoroughly before the start of patient entry. Consequently, not all results of our reproducibility study can be directly generalized to everyday clinical care. However, the procedure for the assessment of grip strength can be described very specifically, and the assessment can be performed without much training. Therefore, we are quite confident about the external validity of our findings on grip strength. This may not hold true for the

overall assessment of severity. Our study showed that reproducibility was very good but not as high as that of grip strength, even with our trained observers. In usual care, the variation among examiners may be much larger for this outcome measure because it strongly depends on a subjective interpretation of symptoms and signs in patients with epicondylitis. These subjective interpretations of severity, however, are an important aspect of the assessment of patients in clinical practice, implying that the measure itself is clinically relevant. Readers should take potential interobserver variation into account when evaluating the severity of elbow problems and when discussing their findings with other physicians or therapists. Pain-free grip strength may be better suited to quantify the severity of lateral epicondylitis when precise findings are needed for research or clinical practice.

CONCLUSION

Interobserver reproducibility for all outcome measures, except for pressure pain threshold, showed excellent results. An overall assessment of severity of symptoms by a trained physiotherapist showed good agreement and reliability and is a relevant measure in day-to-day clinical care. Some interobserver variation should be taken into account, particularly when examiners have different backgrounds and training. Pain-free grip strength showed high reliability and agreement, is relatively easy to perform, and has been shown to be associated with other measures of functional disability in patients with lateral epicondylitis. We can therefore recommend the use of pain-free grip strength in both research and clinical practice.

Acknowledgments: We thank Samyra Keus for performing the assessments.

REFERENCES

1. Geoffroy P, Yaff MJ, Rohan I. Diagnosing and treating lateral epicondylitis. *Can Fam Physician* 1994;40:73-8.
2. Smidt N. Conservative treatments for tennis elbow in primary care [thesis]. Wageningen (Netherlands): Ponsen & Looijen BV; 2001. p 13-46.
3. Vicenzino B, Collins D, Wright A. The initial effects of a cervical spine manipulative physiotherapy treatment on the pain and dysfunction of lateral epicondylalgia. *Pain* 1996;68:69-74.
4. Van der Windt DA, Koes BW, Deville W, Boeke AJ, de Jong BA, Bouter LM. Effectiveness of corticosteroid injections versus physiotherapy for treatment of painful stiff shoulder in primary care: randomised trial. *BMJ* 1998;317:1292-6.
5. Koes BW, Bouter LM, van Mameren H, et al. Randomised clinical trial of manipulative therapy and physiotherapy for persistent back and neck complaints: results of one year follow up. *BMJ* 1992; 304:601-5.
6. Stratford PW, Norman GR, McIntosh JM. Generalizability of grip strength measurements in patients with tennis elbow. *Phys Ther* 1989;69:34-9.
7. Fischer AA. Documentation of myofascial trigger points. *Arch Phys Med Rehabil* 1988;69:286-91.
8. Fess EE, Moran CA, editors. *Clinical assessment recommendations*. Chicago: American Society of Hand Therapists; 1981.
9. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluation. *J Hand Surg [Am]* 1984;9:222-6.
10. Olson SL, O'Connor DP, Birmingham G, Broman P, Herrera L. Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. *J Orthop Sports Phys Ther* 2000;30: 13-20.
11. Buskila D, Fefer P, Harman-Boehm I, et al. Assessment of nonarticular tenderness and prevalence of fibromyalgia in hyperprolactinemic women. *J Rheumatol* 1993;20:2112-5.
12. Dondt W, Willaëys T, Verbruggen LA, Oostendorp RA, Duquet W. Pain threshold in patients with rheumatoid arthritis and effect of manual oscillations. *Scand J Rheumatol* 1999;28:88-93.
13. Granges G, Littlejohn G. Pressure pain threshold in pain-free subjects, in patients with chronic regional pain syndromes, and in patients with fibromyalgia syndrome. *Arthritis Rheum* 1993;36: 642-6.
14. Delaney GA, McKee AC. Inter- and intra-rater reliability of the pressure pain threshold meter in measurement of myofascial trigger point sensitivity. *Am J Phys Med Rehabil* 1993;72:136-9.
15. Nussbaum EL, Downes L. Reliability of clinical pressure pain algometric measurements obtained on consecutive days. *Phys Ther* 1998;78:160-9.

16. Antonaci F, Sand T, Lucas GA. Pressure algometry in healthy subjects: inter-examiner variability. *Scand J Rehabil Med* 1998; 30:3-8.
17. Reeves JL, Jaeger B, Graff-Redford SB. Reliability of the pressure algometer as a measure of myofascial trigger point sensitivity. *Pain* 1986;24:313-21.
18. Jensen K, Andersen HO, Olesen J, Lindblom U. Pressure pain threshold in human temporal region. Evaluation of a new pressure algometer. *Pain* 1986;25:313-23.
19. Smidt N, van der Windt DA, Assendelft WJ, Deville WL, Korthals-de Bos IB, Bouter LM. Corticosteroid injections, physiotherapy, or a wait-and-see policy for lateral epicondylitis: a randomised controlled trial. *Lancet* 2002;359:657-62.
20. Fess EE. A method for checking Jamar dynamometer calibration. *J Hand Ther* 1987;1:28-32.
21. Fischer AA. Pressure algometry over normal muscles. Standard values, validity and reproducibility of pressure threshold. *Pain* 1987;30:115-26.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307-10.
23. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998;12:187-99.
24. Tinsley HE, Weiss DJ. Interrater reliability and agreement of subjective judgments. *J Couns Psychol* 1975;22:358-76.
25. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Chronic Dis* 1987;40:171-8.
26. Stratford P. Reliability: consistency or differentiating between subjects? [letter]. *Phys Ther* 1989;69:299-300.
27. Kramer MS, Feinstein AR. Clinical biostatistics LIV. The biostatistics of concordance. *Clin Pharmacol Ther* 1981;29:11-23.
28. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000;81 Suppl 2:S15-20.
29. Fleiss JF. The design and analysis of clinical experiments. London: John Wiley; 1986. p 1-33.
30. Stratford PW, Levy DR, Gowland C. Evaluative properties of measures used to assess patients with lateral epicondylitis at the elbow. *Physiother Can* 1993;45:160-4.
31. Stratford PW, Levy DR. Assessing valid change over time in patients with lateral epicondylitis at the elbow. *Clin J Sport Med* 1994;4:88-91.
32. MacDermid JC, Kramer JF, Woodbury MG, MacFarlane RM, Roth JH. Interrater reliability of pinch and grip strength measurements in patients with cumulative trauma disorders. *J Hand Ther* 1994;7:10-4.
33. Bellace JV, Healy D, Besser MP, Byron T, Hohmann L. Validity of the Dexter Evaluation System's Jamar dynamometer attachment for assessment of hand grip strength in a normal population. *J Hand Ther* 2000;13:46-51.
34. Pienimäki T, Vanharanta H. Pain questionnaire, pain drawing, and pressure pain thresholds in chronic lateral epicondylitis. *Eur J Phys Med Rehabil* 1998;8:3-9.
35. Wright A, Thurnwald P, O'Callaghan J, Smith J, Vicenzino B. Hyperalgesia in tennis elbow patients. *J Musculoskeletal Pain* 1994;2:83-97.
36. Pienimäki T, Siira P, Vanharanta H. Muscle function of the hand, wrist and forearm in chronic lateral epicondylitis. *Eur J Phys Med Rehabil* 1997;7:171-8.
37. Stratford P, Levy DR, Gaudie S, Levy K, Miferi D. Extensor carpi radialis tendonitis: a validation of selected outcome measurements. *Physiother Can* 1987;39:250-5.

Suppliers

- a. Asimow Engineering Co, 1818 Franklin St, Santa Monica, CA 90404.
- b. Pain Diagnostics and Thermography, 17 Wooley Ln E, Great Neck, NY 11021.

TABLES AND FIGURES

Table 2: Interobserver Agreement and SDD for Elbow Severity, Grip Strength, and Pressure Pain Threshold (N=50)

Outcome Measures	Observer A		Observer B		Mean ± SD Observer A and B	Mean ± SD Difference A – B	95% Limits of Agreement*	SDD
	Mean ± SD	Range	Mean ± SD	Range				
Severity [†]	2.02±1.7	0–7	1.98±1.6	0–6	2.0±1.6	.04±0.8	–1.53 to 1.97	0.7
Pain-free grip strength (kg) [‡]	24±12	3–48	23±11	3–48	23.4±11.4	.79±3.0	–5.09 to 6.6	1.4
Maximum grip strength (kg)								
Involved arm [‡]	30±11	6–50	30±11	9–53	30.1±10.8	–.81±2.0	–4.73 to 3.11	0.8
Uninvolved arm [‡]	34±10	16–54	34±10	17–52	34.3±9.8	–.17±1.9	–3.89 to 3.55	0.7
Pressure pain threshold (kg/m ²)								
Involved arm [§]	3.3±1.5	1.2–6	3.9±1.6	1–6	3.6±1.5	–.66±1.1	–2.82 to 1.50	1.5
Uninvolved arm [§]	5.0±1.3	1.8–6	5.1±1.2	1.6–6	5.0±1.2	–.17±0.9	–1.93 to 1.59	1.3

* 95% limits of agreement: mean (A–B)±1.96 SD.

[†] Rated on an NRS: scores range from 0 to 10; with 10 indicating very severe complaints.

[‡] Mean value of 3 efforts.

[§] Scores range from 1 to 6kg/m², with 1 indicating low pain threshold; mean value of 3 efforts.

Table 1: Main Characteristics of the Participants

	Participants (N=50)
Female (%)	20 (40)
Mean age (SD) in years	47 (11)
Dominant side affected (%)	35 (70)
Elbow disability (median, IQR)*	5.5 (1.5–13.3)
Pain during the day (median, IQR) [†]	1 (0–4)

Abbreviation: IQR, interquartile range.

* Modified Pain Free Function Questionnaire: scores range from 0 to 40, with 40 indicating severe disability.³⁷

[†] Scored by the patient on an NRS. Scores range from 0 to 10, with 10 indicating very severe pain.

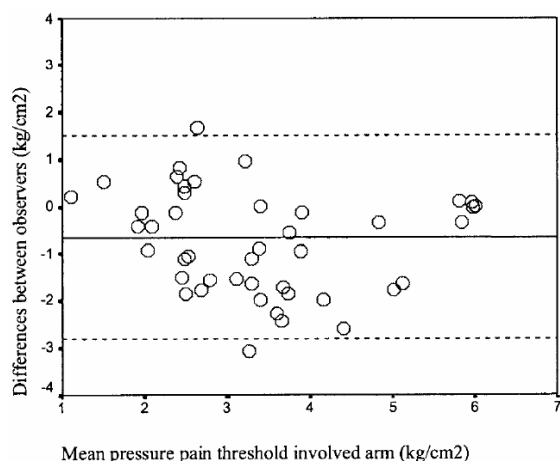


Fig 2. The differences between observers plotted against the mean value of both observers for pressure pain threshold (involved arm only) (N = 50). —, mean differences; ---, limits of agreement.

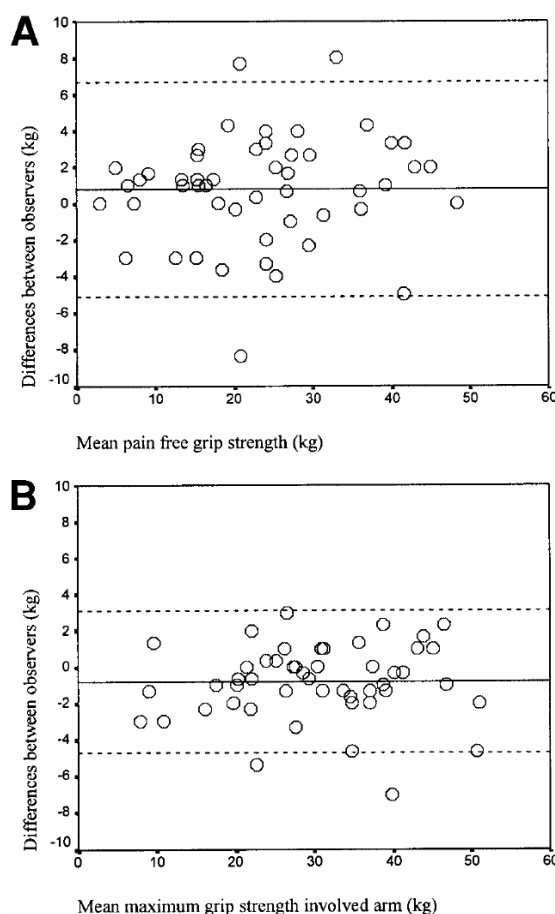


Fig 1. The differences between observers plotted against the mean value of both observers for (A) pain-free grip strength and (B) maximum grip strength (involved arm only) (N = 50). —, mean differences; ---, limits of agreement.

Table 3: Interobserver Reliability: ANOVA and ICC for Measurements of Elbow Severity, Grip Strength, and Pressure Pain Threshold by 2 Observers (N=50)

	Source of Variation							ICC	95% CI
	Between Patients			Between Observers					
	Mean square	F	P	Mean square	F	P			
Severity	5.31	18.62	<.01	0.04	0.14	NS	.90	.83–.94	
Pain-free grip strength	258.35	57.37	<.01	15.73	3.49	NS	.97	.94–.98	
Maximum grip strength									
Involved arm	233.60	115.34	<.01	16.54	8.17	<.01	.98	.97–.99	
Uninvolved arm	192.39	104.97	<.01	0.69	0.38	NS	.98	.97–.99	
Pressure pain threshold									
Involved arm	4.25	7.56	<.01	10.91	19.39	<.0001	.77	.62–.86	
Uninvolved arm	2.72	6.16	<.01	0.71	1.60	NS	.72	.55–.83	

Abbreviation: NS, not statistically significant (P>.05).