UvA-DARE (Digital Academic Repository)

Inter-observer reproducibility of measurements of range of motion in patients with shoulder pain using a digital inclinometer

de Winter, A.F.; Heemskerk, M.A.M.B.; Terwee, C.B.; Jans, M.P.; Devillé, W.; van Schaardenburg, D-J.; Scholten, R.J.P.M.; Bouter, L.M.

[Link to publication](#)

BioMed Central

Research article

Inter-observer reproducibility of measurements of range of motion in patients with shoulder pain using a digital inclinometer

Andrea F de Winter[1,2], Monique AMB Heemskerk[1,3], Caroline B Terwee*[1], Marielle P Jans[1,4], Walter Devillé[1,5], Dirk-Jan van Schaardenburg[6], Rob JPM Scholten[1,7] and Lex M Bouter[1]

Address: [1]Institute for Research in Extramural Medicine, VU University Medical Center, van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands, [2]Department of Social Psychiatry, University of Groningen, P.O. Box 30001, 9700 RB Groningen, the Netherlands, [3]Dutch National Institute for Allied Health Professions, P.O. Box 1161, 3800 BD Amersfoort, the Netherlands, [4]TNO Work and Employment, P.O. Box 718, 2130 AS Hoofddorp, the Netherlands, [5]Netherlands Institute for Health Services Research (NIVEL), P.O. Box 1568, 3500 BN Utrecht, the Netherlands, [6]Department of Rheumatology, VU University Medical Center, P.O. Box 7057, 1007 MB, Amsterdam, the Netherlands and [7]Dutch Cochrane Centre, J2-229, Academic Medical Center, P.O. Box 22660, 1100 DD Amsterdam, the Netherlands

Email: Andrea F de Winter - a.f.de.winter@med.rug.nl; Monique AMB Heemskerk - moniqueheemskerk@planet.nl; Caroline B Terwee* - cb.terwee@vumc.nl; Marielle P Jans - m.jans@arbeid.tno.nl; Walter Devillé - w.deville@nivel.nl; Dirk-Jan van Schaardenburg - d.vanschaardenburg@vumc.nl; Rob JPM Scholten - r.j.scholten@amc.uva.nl; Lex M Bouter - lm.bouter@vumc.nl

* Corresponding author

Abstract

Background: Reproducible measurements of the range of motion are an important prerequisite for the interpretation of study results. The digital inclinometer is considered to be a useful instrument because it is inexpensive and easy to use. No previous study assessed inter-observer reproducibility of range of motion measurements with a digital inclinometer by physical therapists in a large sample of patients.

Methods: Two physical therapists independently measured the passive range of motion of the glenohumeral abduction and the external rotation in 155 patients with shoulder pain. Agreement was quantified by calculation of the mean differences between the observers and the standard deviation (SD) of this difference and the limits of agreement, defined as the mean difference ± 1.96*SD of this difference. Reliability was quantified by means of the intraclass correlation coefficient (ICC).

Results: The limits of agreement were 0.8 ± 19.6 for glenohumeral abduction and -4.6 ± 18.8 for external rotation (affected side) and quite similar for the contralateral side and the differences between sides. The percentage agreement within 10° for these measurements were 72% and 70% respectively. The ICC ranged from 0.28 to 0.90 (0.83 and 0.90 for the affected side).

Conclusions: The inter-observer agreement was found to be poor. If individual patients are assessed by two different observers, differences in range of motion of less than 20–25 degrees can not be distuinguished from measurement error. In contrast, acceptable reliability was found for the inclinometric measurements of the affected side and the differences between the sides, indicating that the inclimeter can be used in studies in which groups are compared.

Background

Measurement of the range of motion of the shoulder joint plays a vital role in the understanding of the nature and the expected course of shoulder pain, as well as in the evaluation of treatment effects. Systematic reviews evaluating the efficacy of medication, steroid injection or physical therapy for shoulder disorders show that in most randomised clinical trials a measurement of range of motion was included [1-4]. The degree of external rotation and glenohumeral abduction is relevant for the evaluation of treatment effects, especially in patients with adhesive capsulitis [4].

Reproducible measurements of the range of motion are an important prerequisite for the interpretation of study results. Visual inspection, goniometric measurements, inclinometry and high-speed cinematography are examples of methods that have been used to quantify the range of motion. For this purpose the digital inclinometer is considered to be a useful instrument because it is inexpensive and easy to use [5]. A few studies have assessed the reproducibility of inclinometric assessment of the range of motion of the shoulder joint [6-10]. The first study showed that two trained physical therapists could obtain reproducible measurements for the assessment of external rotation and glenohumeral abduction of the shoulder joint [6]. However, only a small sample of healthy subjects was included. Three later studies that included patients reported poor reproducibility of range of shoulder motion [7,9,10]. One of these studies, however, was conducted in a very specific group of patients with reflex sympathetic dystrophy [7]. In addition, in all these studies the measurements were done by physicians (e.g. rheumatologists or surgeons) from a single practice. Hoving et al showed that physical therapists achieved higher reliability than rheumatologists, especially for external rotation, but the physical therapists from the study that Hoving referred to assessed only 6 patients [8,9].

Therefore, the purpose of our study was to evaluate the inter-observer reproducibility of the external rotation and glenohumeral abduction measurements by physical therapists in a large sample of patients from many different practices and with different degrees of shoulder pain, using the Cybex Electronic Digital Inclinometer-320 (EDI 320).

Methods

Patients

Within the framework of a study on inter-observer agreement on the diagnosis of shoulder disorders, which involved history taking and physical examination [11], an evaluation was made of the inter-observer reproducibility of external rotation and glenohumeral abduction measurements by physical therapists, using the EDI 320 inclinometer. During a 20-month period, consecutive patients with shoulder complaints who consulted one of the 20 participating general practitioners, one the 2 participating physicians in an orthopaedic practice, or one of the 20 participating rheumatologists in a secondary care rheumatology clinic, were considered for participation in the study. Patients were eligible for participation if they met the following inclusion criteria: aged between 18 and 75 years, ability to co-operate (no dementia, sufficient knowledge of the Dutch language) and informed consent given. Patients with shoulder problems due to neurological, vascular or internal disorders, systemic rheumatic diseases, prior dislocations or fractures were excluded. The study was approved by the local institutional review board of the VU University Medical Center.

Measurements

Two observers (MPJ & AFW), both experienced physical therapists, independently measured the range of motion of the shoulder joint using the Cybex Electronic Digital Inclinometer-320 (EDI 320) (Cybex Inc, Ronkonkoma, NY). This device is gravity dependent and indicates range of motion on a 360° scale. The EDI 320 consists of a hand-held unit and portable display unit with an integral rechargeable power source. The EDI 320 recorded gross movement and then calculated the differential range of motion by subtracting the initial position reading from the final position reading. The EDI 230 can be used to measure single joint motions of the elbow, forearm, wrist, thumb, fingers, shoulder, scapula, hip, knee and ankle, and combined motions of the spine and shoulder.

Each observer measured both shoulders of each patient once. Passive glenohumeral abduction was measured first, followed by measurement of the passive external rotation. Within one hour the second observer repeated the measurements of the first observer. In order to prevent the occurrence of systematic differences between the observers, due to repeated testing, the sequence of the observers was randomly allocated. The patients did not receive any therapy between the two measurements.

Prior to the study, the performance of all measurements were standardised, to make sure that the physiotherapists assessed the patients in the same way. For the measurement of passive glenohumeral abduction, the patients was seated upright, and the position of 0° was defined as the upper arm in a neutral position. While palpating the lower angle of the scapula with the thumb, the examiner elevated the upper arm of the patient until the scapula began to rotate or pain limited further motion. This range of motion was recorded in degrees.

For the measurement of passive external rotation, the patient was in a supine position, with the shoulder in 0°

of abduction and rotation, the elbow flexed at 90° and the forearm in a neutral position. This position was defined as the position of 0°. The observer then performed external rotation until pain limited the range of motion or the extreme of the range was reached. This range of motion was recorded in degrees.

Prior to the measurements, demographic characteristics (age, gender) and clinical characteristics (e.g. previous episodes of shoulder problems, duration of complaints, sleep disturbances) of the patients were recorded by means of a structured questionnaire. In addition, all patients recorded the severity of pain during the day and at night in the preceding week on a 100 mm visual analogue scale (VAS), ranging from 0 'no pain' to 100 ' very severe pain'.

Assessment of reproducibility
The reproducibility of the measurements of the affected side and the contralateral side, and the difference in range of motion between the sides was calculated. The difference between the sides was quantified by subtracting the results of the affected shoulder from those of the contralateral shoulder. The difference between the sides is an important outcome, since in clinical practice a conclusion on abnormal range of motion of the affected shoulder is usually drawn after comparison of the measurements of the affected shoulder with those of the contralateral shoulder. In this manner, differences in mobility between subjects due to age, gender or other factors [12] can be taken into account.

For the quantification of reproducibility, we distinguished two different types of measures of reproducibility with different interpretations: measures of agreement and measures of reliability. Measures of agreement refer to the absolute measurement error (presented in the units of measurement of the instrument) that is associated with one measurement taken from one individual patient [13]. Measures of agreement provide insight into the the ability of two or more observers to achieve the same value. Measures of reliability refer to the relative measurement error, i.e. the variation between patients in relation to the total variance of the measurements (see below). They provide information on the ability of two or more observers to differentiate between subjects in a group [13,14].

Agreement
The inter-observer agreement was quantified by calculating the mean difference between the two observers (A-B) and the standard deviation (SD) of this difference. Subsequently, the 95% limits of agreement were calculated according to the method of Bland & Altman [15], defined as the mean difference between the observers ± 1.96*SD of this difference. These limits represent the range in

which 95% of the differences between the two observers fall. If the values of observer A would be extracted from observer B (B-A instead of A-B), the limits of agreement would stay the same, but the signs (+ / -) of the mean differences and the upper and lower limits of agreement would be opposed. In this situation, the choice of extracting B-A or A-B is arbitrair. Therefore, the signs are irrelevant and should be ignored when interpreting the results. For the interpretation of the measurement error, the largest limit of agreement (either upper or lower limit) is most relevant.

Furthermore, plots of the differences between observers against the corresponding mean of the two observers for each patient were constructed to examine homoscedasticity, as proposed by Bland and Altman [15]. In addition, the frequency of agreement of the observers within 5° and 10° was calculated. Although no clear criteria for the acceptable degree of inter-observer agreement are available, based on our clinical experience, we decided prior to the study that differences exceeding 10° were determined as being unacceptable because they are likely to affect decisions on patient management.

Reliability
The intra-class correlation coefficient (ICC) was derived from a random-effects two-way analysis of variance. By means of analysis of variance the variation in measurements is partitioned into the potential sources of variation: observer differences, patient differences and random error. The ICC is defined as the ratio of the variance between patients over the total variance [16]. The values of the ICC can theoretically range from 0 to 1, with a higher value indicating that less variance is due to other factors such as differences between observers. An intraclass correlation coefficient of at least 0.70 is considered to be satisfactory for group comparisons, and a value of 0.90–0.95 for individual comparisons [17].

Results
Complete data on the inclinometric measurements were available for 155 of the 201 patients included. No inclinometric measurements were available for 46 patients, for the following reasons: no measurements were performed because of high pain intensity (30 patients), lack of time (3 patients), difficulties during the measurement procedure (e.g. difficulties with the test position or inability to relax)(8 patients), and errors in the registration of the measurement results (5 patients). The main characteristics of participants and non-participants are presented in Table 1. Diagnosis are not presented, because in our previous study we found only a moderate inter-observer agreement between the two observers (60% agreement, kappa 0.45) [11]. Compared to the participating patients,

Table 1: Main characteristics of the participants and the non-participants

	Participants	Non-participants
	(n = 155)	(n = 46)
Patients recruited by (%):		
- general practitioner	39	30
- physician in orthopaedic practice	18	11
- rheumatologist in secondary care rheumatology clinic	43	59
Female (%)	65	65
Mean age in years (SD)	47 (12.6)	48 (12.1)
Previous episode(s) of shoulder pain (%)	41	35
Duration of current episode (%):		
- < 3 months	26	30
- 3 – 6 months	17	13
- 6 – 12 months	24	13
- > 12 months	33	44
Mean pain score* (SD):		
- at night	48 (31)	63 (25)
- during the day	51 (26)	63 (26)

* Scores on a visual analogue scale ranging from 0 'no pain' to 100 'very severe pain' (in mm).

Table 2: Mean, standard deviation (SD) and range of the glenohumeral abduction and the external rotation according to observer A and B, followed by the mean differences between both observers, and the frequency of agreement within 5 and 10 degrees.

Tested movements	Observer A (in degrees) Mean (SD)	Observer B (in degrees) Mean (SD)	Observer A-B (in degrees) Mean (SD)	Upper and lower limit of agreement	Agreement (%) within 5° and 10°	
					5°	10°
Glenohumeral abduction						
- affected side	69.5 (17.6)	68.8 (16.3)	0.8 (10.0)	-18.8 – 20.4	43	72
- contralateral side	79.8 (7.6)	78.9 (8.4)	0.9 (9.6)	-17.9 – 19.7	43	72
- contralateral – affected	10.3 (16.3)	10.2 (14.2)	0.1 (10.1)	-19.6 – 19.8	48	75
External rotation						
- affected side	58.4 (24.1)	63.0 (24.0)	-4.6 (9.6)	-23.4 – 14.2	43	70
- contralateral side	72.8 (11.2)	79.4 (11.7)	-6.6 (9.5)	-25.2 – 12.0	39	63
- contralateral – affected	14.4 (20.5)	16.4 (20.8)	-2.0 (10.1)	-21.7 – 17.7	44	71

the severity of complaints of the non-participants was indeed higher, expressed by a higher mean pain intensity.

Reproducibility

Agreement
Table 2 summarises the results of the inter-observer agreement. The observers had quite similar measurements of glenohumeral abduction, but observer A measured a consistently smaller range of external rotation than observer B. For the affected side, the limits of agreement were 0.8 ± 19.6 for glenohumeral abduction and -4.6 ± 18.8 for external rotation. The percentage agreement within 10° for these measurements were 72% and 70% respectively.

Since the pain level of the non-participants was higher than that of the participants (Table 1), patients with a high pain intensity (pain score on the VAS during the day > 65; n = 54) were compared with patients with moderate pain intensity (pain score on the VAS during the day ≤ 65; n = 101). The inter-observer agreement was not different between these patient groups (data not shown).

Figures 1a and 1b show the differences between observers, plotted against the mean value of both observers for glenohumeral abduction and external rotation of the affected side, respectively (each point represents one patient). For both movements the error of measurement was found to be independent of the magnitude of the range of motion (homoscedasticity). This was also the case for the contralateral side and for the differences between the sides (data not shown).

Figure 1a Differences between observers, plotted against the mean value of both observers for each patient for glenohumeral abduction affected size
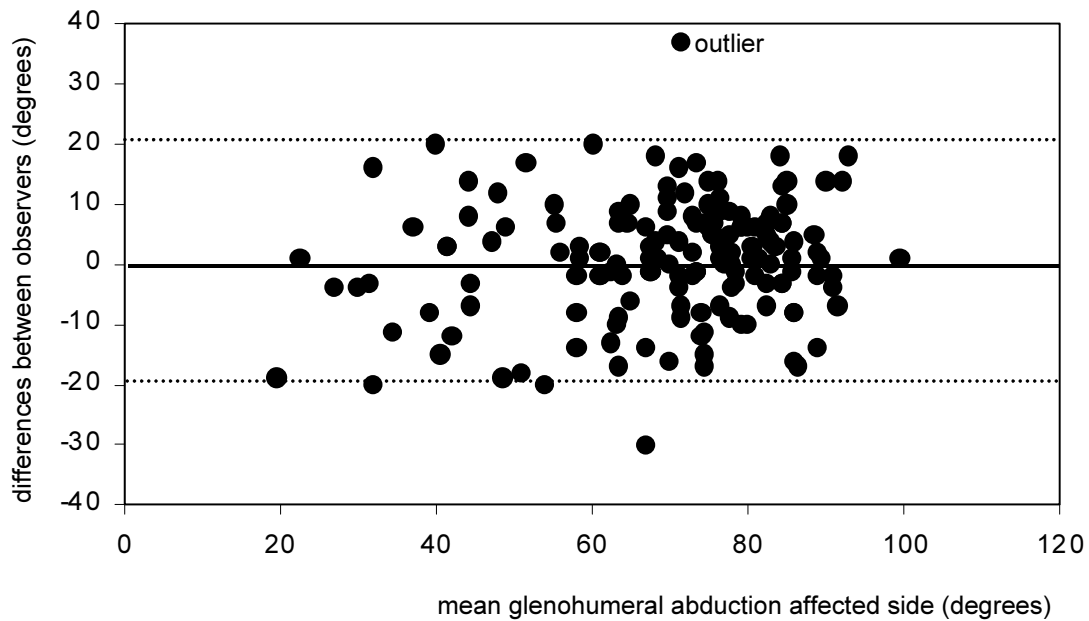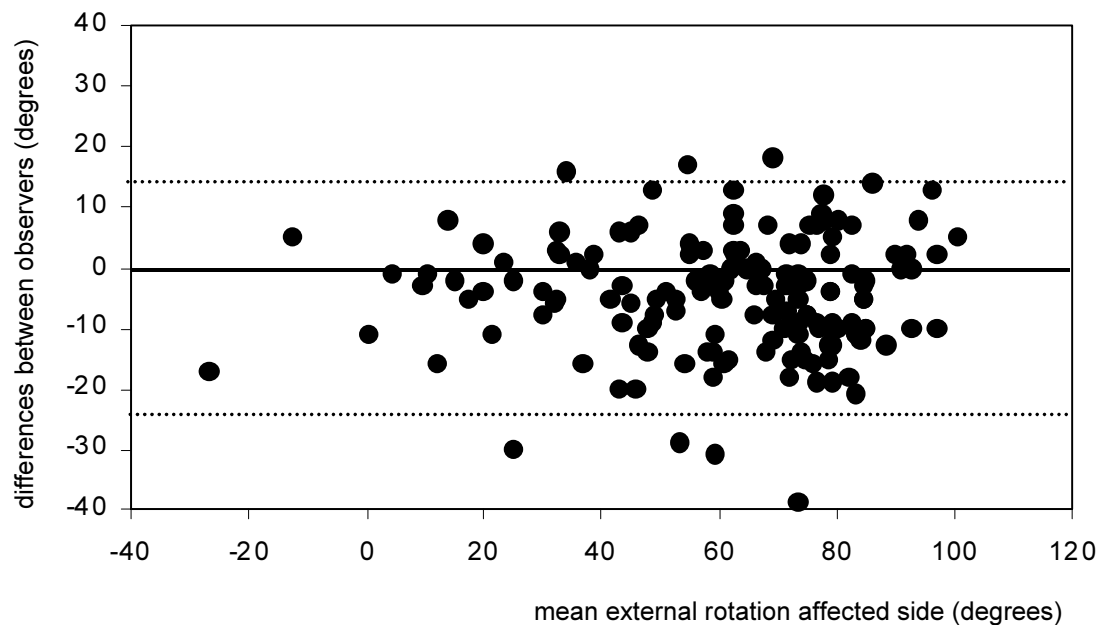

Figure 1b Differences between observers, plotted against the mean value of both observers for each patient for external rotation affected size

Solid lines: mean differences; dashed lines: limits of agreement

Figure 1
Differences between observers, plotted against the mean value of both obervers for each patient for glenohumeral abduction and external rotation of the affected side. Solid lines: mean differences; dashed lines: limits of agreement.

Table 3: Results of the analysis of variance

| | Source of variation | | | | |
| | Between patients | | Between observers | | |
Tested movements	F	P	F	P	ICC
Glenohumeral abduction					
- affected side	10.5	<0.01	0.99	N.S.	0.83
- contralateral side	1.8	<0.01	1.35	N.S.	0.28
- contralateral – affected	8.2	<0.01	0.02	N.S.	0.78
External rotation					
- affected side	24.0	<0.01	36.0	<0.01	0.90
- contralateral side	4.8	<0.01	77.5	<0.01	0.56
- contralateral – affected	15.7	<0.01	6.2	<0.05	0.88

Reliability

The results of the analysis of variance are presented in Table 3. The ICC-values ranged from 0.28 to 0.90. For both movements, the ICCs of the measurement of the affected shoulder were higher than ICCs of the contralateral shoulder.

Discussion

This study investigated the inter-observer reproducibility of the assessment of the passive range of motion of the glenohumeral abduction and the external rotation of the shoulder joint, using the EDI 320 digital inclinometer. A large number of patients from different clinics with different levels of mobility and varying severity of shoulder pain were examined. We chose to measure passive rather than active range of motion because according to diagnostic guidelines the degree of passive external rotation and glenohumeral abduction is important for the evaluation of adhesive capsulitis.

The results showed that there was considerable variation in measurement between the observers across the whole range of values of the tested movements. In a maximum of 75% of the various measurements, the differences between observers did not exceed 10°. Although it is a matter of clinical judgement, which other clinicians might not agree with, it was decided that differences between observers which exceed 10° are not acceptable for clinical purposes. The limits of agreement show that if patients, that are considered to be stable, are assessed by two different observers, the differences in the measured range of motion between the observers can be as large as 20–25 points (referring to the largest of the upper and lower limit of agreement). This means that if patients are assessed e.g. before and after therapy by two different observers, changes in range of motion of less than 20–25 degrees, can not be distinguished from measurement error.

In the present study, inclinometric measurements could often not be performed at all because of the high severity of shoulder pain, resulting in a large number of non-participants (n = 46). However, one could argue that if patients are not able to perform this kind of test because of their pain, there is no need to measure the range of motion anyway. In addition, in our study population no association was found between the level of pain and the inter-observer differences. We believe that our study provides a reasonably valid estimate of the reproducibility of inclinometric measurements of patients with shoulder pain, based on one measurement of each range of motion.

Contrary to those of the glenohumeral abduction, the measurements of the external rotation showed systematic differences between the observers, which is consistent with the findings of Croft et al. [18]. Although several factors might contribute to the systematic differences, differences in defining the limits of motion might explain the results. For glenohumeral abduction the limits of motion are determined by rotation of the scapula, whereas pain and reaching the extreme of the range of motion are the criteria for the limits of motion of external rotation. It was suggested that the amount of passive force applied is one of the reasons why passive movements are more difficult to reproduce than active movements [19]. However, Tousignant et al also found systematic differences in their study on reliability of the EDI-320 for measurement of active neck flexion and extension [20].

In contrast to the level of poor agreement, acceptable reliability was found for most inclinometric measurements for use for group comparisons (ICC above 0.70), but not for individual comparisons (ICC between 0.90–0.95). These findings are in accordance with the findings of Green et al. [8], who also reported acceptable reliability for the measurement of glenohumeral abduction and external rotation by physical therapists using an inclinom-

eter. Tousignant et al found comparable ICCs for measurement of active neck flexion and extension with the DI 320 [20].

Several other studies have evaluated observer variation in the measurement of range of motion of the shoulder joint based on various other measurement methods [18-30]. The results vary considerably across studies and estimates are difficult to compare due to differences in patient groups, raters, and measurement methods.

In general, is seems that most methods are reliable enough to use for group comparisons, but not for individual comparisons (ICCs between 0.70 and 0.90). This means that most instruments can be used in studies. Several authors suggested that visual estimation may be as reliable as measurement instruments, such as an inclinometer or a goniometer [7,21,24]. In our patient group reliability obtained with inclinometer measurements was higher than reliability obtained with visual estimation in a previous study on the same subjects (ICC for abduction was 0.83 compared with 0.71; ICC for external rotation was 0.90 compared with 0.78 for the affected side, data submitted for publication).

As in our study, most other studies that presented data on agreement found large measurement errors, especially for the assessment of external rotation. For example, large standard errors of measurement were found in the studies of Geertzen et al. [7] (approximately 25°), and Triffitt et al. [10] (approximately 25–30°).

Poor inter-observer agreement, but acceptable reliability of measurements may seem to be a puzzling result. ICCs, however, are strongly influenced by the heterogeneity of the population studied. In a patient group with large differences between patients, it is more easy to distinguish between patients than in a patient group with small differences between patients. Therefore, it is possible that an instrument is able to discriminate adequately between groups of patients despite a large measurement error in a heterogeneous patient population [13]. For the measurement of individual patients in clinical practice, or to assess intra-individual changes in range of motion over time, the measurement error of the observers, using the inclinometer, or most other instruments, seems to be too large. For the purpose of comparing groups in studies, the inclinometer seems to be a useful instrument, and is probably better than visual estimation. Finally, reproducibility is a function of the instrument that is used, the measurement conditions, the movements tested, the observers and the study population. Which method of assessment of range of motion is preferable should therefore be evaluated within one single study.

Investigators should quantify the reproducibility of their assessments before commencing a clinical trial, since the level of reproducibility has considerable impact on the power of a clinical trial. In general, intra-observer reproducibility is better than inter-observer reproducibility [18,19,25,29], so it is recommended that in clinical trials the same observer should be responsible for the measurement of treatment outcome for each patient. Reproducibility of measurements may also be improved by using the mean value of multiple measurements. Further psychometric studies should examine the validity of the EDI 320.

Conclusions
In conclusion, the inter-observer agreement was found to be poor. If patients are assessed by two different observers, differences in range of motion of less than 20–25 degrees, can not be distuinguished from measurement error. In contrast, acceptable reliability was found for the inclinometric measurements of the affected side and the differences between the sides, indicating that the inclimeter can be used in studies.

Since the measurements were already standardised and the observers trained prior to the study, the best way to reduce variation in measurements would seem to use the mean value of multiple measurements at each time point, preferably done by the same observer.

Authors' contributions
AFW designed and coordinated the reproducibility study, collected all data and carried out all measurements, analysed the data, and drafted the manuscript; MAMBH contributed to the analysis of the data and commented to various versions of the manscript; MPJ carried out all the measurements, contributed to data collection and commented to various versions of the manuscript; CBT contributed to the writing of the manuscript; WD contributed to the statistical analyses; DJS contributed to the design of the study ; RJPMS and LMB participated in the design of the study, supervised the performance of the study and commented on various versions of the manuscript.

References
1.   van der Windt DAWM, van der Heijden GJMG, Scholten RJPM, Koes BW, Bouter LM: **The efficacy of Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) for shoulder complaints. A systematic review.** *J Clin Epidemiol* 1995, **48:**691-704.
2.   van der Heijden GJMG, van der Windt DAWM, Kleijnen J, Koes BW, Bouter LM, Knipschild PG: **Steroid injections for shoulder disorders: a systematic review of randomised clinical trials.** *Br J Gen Pract* 1996, **46:**309-316.
3.   van der Heijden GJMG, van der Windt DAWM, de Winter AF: **Physical therapy for patients with soft tissue shoulder disorders: a systematic review of randomised clinical trials.** *BMJ* 1997, **315:**25-30.

4.   Green S, Buchbinder R, Glazier R, Forbes A: **Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy.** *BMJ* 1988, **316:**354-360.

5.   Chiarello CM, Savidge R: **Interrater reliability of the Cybex EDI-320 and fluid goniometer in normals and patients with low back pain.** *Arch Phys Med Rehabil* 1993, **74:**32-37.

6.   Heemskerk MAMB, van Aarst M, van der Windt DAWM: **De reproduceerbaarheid van het meten van de passieve beweeglijkheid van de schouder met de EDI-320 digitale hoekmeter.** *Ned T Fysiother* 1997, **107:**146-149.

7.   Geertzen JHB, Dijkstra PU, Stuwart RE, Groothoof JW, ten Duis HJ, Eisma WH: **Variation of measurements of range of motion: A study in reflex symathetic dystrophy patients.** *Clin Rehabil* 1998, **12:**254-264.

8.   Green S, Buchbinder R, Forbes A, Bellamy N: **A standardized protocol for measurement of range of movement of the shoulder using the Plurimeter-V inclinometer and assessment of its intrarater and interrater reliability.** *Arthritis Care Res* 1998, **11:**43-52.

9.   Hoving JL, Buchbinder R, Green S, Forbes A, Bellamy N, Brand C, Buchanan R, Hall S, Patrick M, Ryan P, Stockman A: **How reliable do rheumatologists measure shoulder movement?** *Ann Rheum Dis* 2002, **61:**612-616.

10. Triffitt PD, Wildin C, Hajioff D: **The reproducibility of measurement of shoulder movement.** *Acta Orthop Scand* 1999, **70:**322-324.

11. de Winter AF, Jans MP, Scholten RJPM, Devillé W, van Schaardenburg D, Bouter LM: **Diagnostic classification of shoulder disorders: Inter-observer agreement and determinants of disagreement.** *Ann Rheum Dis* 1999, **58:**272-277.

12. van Schaardenburg D, van den Brande KJS, Ligthart GJ, Breedveld FC, Hazes JMW: **Musculokeletal disorders in persons aged 85 and over, a communitiy survey.** *Ann Rheum Dis* 1994, **53:**807-811.

13. de Vet HCW: **Observer reliability and agreement.** In *Encyclopedia of Biostatistics Volume 4*. Edited by: Armitage P, Colton T. Boston University: John Wiley & Sons Ltd; 1998:3123-3128.

14. Stratford P: **Reliability: consistency or differentiating between subjects?** *Physical Therapy* 1989, **69:**299-300.

15. Bland JM, Altman DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, **1:**307-310.

16. Fleiss JF: *The design and analysis of clinical experiments. Chapter 1: Reliability of measurement* John Wiley & Sons; 1986:1-33.

17. Scientific Advisory Committee of the Medical Outcome Trust: **Assessing health status and quality-of-life instruments: Attributes and review criteria.** *Qual Life Res* 2002, **11:**193-205.

18. Croft P, Pope D, Boswell R, Rigby A, Silman A: **Observer variability in measuring elevation and external rotation of the shoulder.** *Br J Rheumatol* 1994, **33:**942-946.

19. Gajdosik RL, Bohannon RW: **Clinical measurement of range of motion; Review of goniometry emphasizing reliability and validity.** *Phys Ther* 1987, **67:**1867-1872.

20. Tousignant M, Boucher N, Bourbonnais J, Gravelle T, Quesnel M, Brosseau L: **Intratester and intertester erliability of the Cybex electronic digital inclinometer (EDI-320) for measurement of active neck flexion and extension in healthy subjects.** *Man Ther* 2001, **6:**235-241.

21. Awan R, Smith J, Boon AJ: **Measuring shoulder internal rotation range of motion: A comparison of 3 techniques.** *Arch Phys Med Rehab* 2002, **83:**1229-1234.

22. Boone DC, Azen SP, Lin C, Spence C, Baron C, Lee L: **Reliability of goniometric measurements.** *Phys Ther* 1978, **58:**1355-1360.

23. Bostrom C, Harms-Ringdahl K, Nordemar R: **Clinical reliability of shoulder function assessment in patients with rheumatoid artritis.** *Scand J Rheumatol* 1991, **20:**36-48.

24. Hayes K, Walton JR, Szomor ZL, Murrell GAC: **Reliability of five methods for assessing shoulder range of motion.** *Austr J Physiother* 2001, **47:**289-294.

25. Johnson GR, Fyfe NCM, Heward M: **Ranges of movement at the shoulder complex using an electromagnetic movement sensor.** *Ann Rheum Dis* 1991, **50:**824-827.

26. Jordan K, Dziedzic K, Jones PW, Ong BN, Dawes PT: **The reliability of the three-dimensional FASRAK measurement system in measuring cervical spine and shoulder range of motion in healthy subjects.** *Rheumatology* 2000, **39:**382-388.

27. MacDermid JC, Chesworth BM, Patterson S, Roth JH: **Intratester and intertester reliability of goniometric measurement of passive lateral shoulder rotation.** *J Hand Ther* 1999, **12:**187-192.

28. Mayerson NH, Rose MA, Milano A: **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, **6:**655-657.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2474/5/18/prepub