**Title**

Minimal Clinically Important Difference of Liverpool Elbow Score in Elbow Arthroplasty

**Running Title**

Minimal Clinically Important Difference: Liverpool Elbow Score

**Authors:**

1**. Karthik Vishwanathan.** MBBS, MS (Orth), DNB (Orth), MRCS, MSc (Orth), MCh (Orth) - Shri Krishna Hospital and Pramukhswami Medical College, Karamsad, India

2**. Dr Omid Alizadehkhaiyat.** Associate Professor in Health Sciences (MD, PhD) - School of Health Sciences, Liverpool Hope University, Liverpool, UK

3. **Professor Graham Kemp**  (MA, DM, FRCPath, FHEA, CSci, FRSB) - Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK

4. **Professor Simon Frostick**. Professor of Orthopaedics (MA, DM, FRCSEng, FRCSEd (ad hom), FFSTEd) - Musculoskeletal Science Research Group, Institute of Translational Medicine, University of Liverpool, Liverpool, UK

**Corresponding Author:**

Professor Simon P. Frostick

Musculoskeletal Science Research Group,

Institute of Translational Medicine,

University of Liverpool

Liverpool UK, L69 3GA

Phone: +44 151 706 4120 Fax: +44 151 706 5815

Email: [s.p.frostick@liverpool.ac.uk](http://uk.mc240.mail.yahoo.com/mc/compose?to=s.p.frostick@liverpool.ac.uk)

**Disclaimer:** No external funding received for this study. Authors report no conflict of interest.

**Ethical Approval:** Sefton Research Ethics Committee - REC Number: 08/H1001/109

**Trust Study Number**: 3735

**ABSTRACT**

**Background**

The minimal clinically important difference (MCID) that allows the interpretation of small but meaningful changes following intervention has not been reported for the Liverpool Elbow Score (LES). This study aimed to determine the MCID for the LES in patients undergoing total elbow replacement (TER).

**Methods**

This observational study is based on preoperative and 1-year postoperative clinical outcome of TER (Discovery Elbow System) in 71 patients using the LES. A four-point Likert-like transition scale was used to evaluate patient satisfaction following TER. A combination of distribution-based (standard deviation of change in the LES; standard error of mean [SEM], smallest detectable change [SDC]) and anchor-based methods (receiver operating curve [ROC]; difference of mean of change in LES) was used to determine range of MCID values.

**Results**

The mean change in the LES value was 2.4 (SD: 2.1). The estimated SDC value with upper limit of 90% confidence interval was 1.5. The mean change in LES of “satisfied” and “somewhat satisfied” patient groups was 2.4 (SD = 2.1) and 1.1 (SD = 1.4) respectively and the difference between the both means (MCID based on difference of mean in two sub-groups) was 1.3. According to ROC analysis, the value of MCID was 1.6.

**Conclusion**

The MCID value for the LES was estimated to range between 0.7 and 1.8. The estimated SDC value was 1.5. We propose that the “true” MCID value of LES would be between 1.6 and 1.8 to ensure that the value is higher than the measurement error of LES.

**Keywords:** Elbow; Total Arthroplasty; Discovery Elbow; Clinical Outcome; Joint Replacement, Elbow Prostheses; Minimal Clinically Important Changes.

Level of Evidence: Therapeutic II

**INTRODUCTION**

The Liverpool Elbow Score (LES) is a region-specific outcome score that is completed by both the clinician and the patient. Validation study by the developers of the score demonstrated that LES is valid, reliable and responsive to change in clinical condition of the patient in different elbow conditions.25 Furthermore, a recent study reported satisfactory responsiveness of LES in elbow arthroplasty.32

Establishing minimally clinically important difference (MCID) for clinical outcomes scores is an important component of outcomes research to understand treatment effectiveness, particularly from the patient's perspective. The MCID is defined as the smallest change in the value of an outcome instrument that patients perceive as important, beneficial, or harmful.18 In other words, MCID value differentiates patients who improve from those who do not improve after a therapeutic intervention.16 The concept of MCID assists in differentiating statistical significance from clinical significance. A statistical test might reveal a significant difference between preoperative and postoperative scores of an outcome instrument; however if the difference is lower than the MCID value of outcome instrument then this statistically significant difference is not deemed to be clinically significant. The MCID value is also helpful in evaluating cost-effectiveness, estimating appropriate sample size for randomised control trials, and evaluating power of a non-randomised study.18

To our knowledge, the MCID value has not been determined for the LES. Hence, the present study aimed to critically evaluate MCID of LES in large cohort of patients who underwent total elbow arthroplasty (TEA) due to various underlying pathologies.

**METHODS**

A prospective database of patients who had undergone TEA using the Discovery Elbow System (Biomet Inc., Warsaw IN, USA) was reviewed in order to identify patients with completed preoperative LES and 1-year postoperative LES and satisfaction questionnaire. Identified patients had undergone TEA for degenerative arthritis (osteoarthritis, posttraumatic arthritis), inflammatory arthritis (rheumatoid arthritis, haemophiliac arthropathy, and psoriatic arthritis), comminuted distal humerus fracture, and loosening of previous elbow prostheses using the Discovery prosthesis between April 2003 and March 2013. Patient demographics are presented in the results section. All identified cases (n=71) were operated and followed-up in a single Upper Limb centre by the same surgeon.

The Discovery system was initially developed with the intention to decrease polyethylene-bushing wear, anatomic stem design, restore elbow joint biomechanics and produce a hinge that could be easily revised, if needed. This elbow replacement system is one of the latest generations of linked prostheses, whose hinge linkage allows its use in elbows with severe bony and ligamentous deficiency. This single centre study received approval from the local ethics committee and the local hospital audit committee. All participants gave written consent for participating in the study and for operative intervention.

**Outcome assessment**

Clinical and functional outcome after TEA was assessed using LES. Prior to the operation and at 1-year follow-up patients first completed the Patient-Answered Questionnaire of the LES (PAQ-LES) followed by completion of the Clinical Assessment part of the score (CAS–LES) by independent research fellows. PAQ-LES includes 9 questions comprising domains of pain (1 question), functional ability to do activities of daily living (7 questions) and functional ability to participate in sporting and recreational activities (1 question). These questions are answered on a 5-point adjectival scale from 0 (maximum disability) to 4 (no functional disability). The CAS–LES includes assessment of range of motion (4 items), muscle strength (1 item) and ulnar nerve function (1 item). The points from PAQ-LES and CAS–LES are then entered individually in a mathematical formula to determine the total LES. In this scoring system, 0 and 10 points indicate the worst and best outcome respectively.25

As there is no “gold standard” external criterion to assess change and improvements in the clinical condition of the patient, a four-point Likert-like transition scale was used to evaluate patient satisfaction following TEA. The options on this scale were “very satisfied”, “satisfied”, “somewhat satisfied” and “unsatisfied”. Patients answered this question at postoperative follow-ups.

**Estimation of Minimal Clinically Important Difference (MCID)**

There is no “gold standard” method to measure MCID, which can be estimated using either anchor-based or distribution-based methods. It has been recommended that studies use both anchor-based and distribution-based methods to give range of values for MCID and finally triangulate to converge on possible MCID value.5,35 It is also suggested that anchor-based methods be given greater weight than distribution-based methods for converging on a single value or to narrow the range of possible MCID values. Distribution-based methods are solely used only when suitable external anchors have not been used or not available for use.22

The present study determined the MCID using both anchor-based and distribution-based methods. Patient satisfaction was used as a global transition external anchor. This is in accordance with a similar approach by previous studies to estimate clinically meaningful change in various studies.4,26 Adjusting for the change in unsatisfied patients, the MCID can be calculated as the mean change score for satisfied patients minus the mean change score for somewhat satisfied patients.4,12,26,29 Receiver operating curve (ROC) analysis is then used to evaluate the point that is closest to the upper left hand corner of the curve representing MCID.29,33 A diagonal is drawn from the upper left corner of the ROC to the lower right corner. The point at which this diagonal intersects the curve is considered to be the point closest to the upper left corner and hence, the value of change in LES at this site of intersection represents the MCID.29 For ROC analysis, patients who were unsatisfied and somewhat satisfied were grouped into “not improved” group and those who were “satisfied” and “very satisfied” were grouped into “improved” group. The entire cohort was included in the ROC analysis rather than just the values adjacent to the point of dichotomy, as this has been shown to increase precision of MCID estimation.30

Sensitivity is the proportion of patients who are definitely satisfied and whose change in LES is above the threshold MCID value. Specificity is the proportion of patients who are not definitely satisfied and whose change in LES is below the threshold MCID value. For distribution-based approach, we first estimated the Standard Error of Mean (SEM) and Smallest Detectable Change (SDC). It has been reported that estimates based on measurement precision of outcome measurement (SEM) are better than estimates based on sample variation (effect size) or those based on statistical significance (paired t-test).7 SEM is an indicator of random error during single use of an outcome instrument and is believed generally to be stable across different population and different studies.7,8 SDC or minimum detectable change (MDC) refers to the smallest change in the value of an outcome instrument that is greater than random measurement error associated with use of the instrument.8 Both SEM and SDC are determined in a stable sub-group of patients in the study cohort. These patients have either perceived no change in clinical condition after an intervention or have experienced negligible or minimal change in their clinical condition.

Repeated application of an outcome instrument in the same patient should give a similar value if the condition has remained stable with no change. However; this is infrequently seen, and more often repeated application gives rise to slight changes in the value of the outcome instrument. This minimum change in value is likely to occur due to the measurement error of the outcome instrument. SDC or MDC represents the threshold value beyond which any increase in the score of outcome instrument is likely to indicate “true” change in clinical condition instead of error due to repeated administration of the outcome tool. A change in the value of an outcome instrument lower than the value of SDC might not indicate “true” change in the clinical condition, as this is likely to be due to the measurement error of the outcome instrument.

In the present study, SEM was calculated as SEM = [SD of baseline preoperative LES] x [square root of 1 – alpha], wherein alpha represents reliability coefficient of test-retest value of outcome instrument in stable group of patients.3,8,15,17 Alpha can be represented as either Cronbach’s alpha or the intra-class correlation coefficient. Commonly, 90% confidence limit is chosen for MDC and is calculated as MDC90 = (1.65) x (square root of 2) x (SEM).3,8,15,17 Cronbach’s alpha based on standardised items was used to measure reliability coefficient of test-retest in stable group of patients.17 Based on patients’ response to the 4-point Likert-like satisfaction scale, those patients who felt “somewhat satisfied” after the TEA were considered to be stable patients, as they probably did not have significant change in their clinical condition. Various threshold values have been reported for the estimation of clinically meaningful change based on SEM including 1 SEM,1.96 SEM and 2.77 SEM.20,34 Norman et al21observed that a value of one half the standard deviation of the change in score of the outcome instrument was equal to MCID in a variety of studies though it is believed that this is a conservative estimate of MCID. SPSS Version 18 was used to do the statistical analysis.

**RESULTS**

**Study Participants**

Study included 71 patients who had the required preoperative and postoperative data on LES and patient satisfaction (there was no missing data). The mean age of patients was 64 years (range – 22 to 93 years), 49 cases were female (69 %) and 22 cases were male (31%). Fifty-two (73%) and 19 (27%) were primary and revision TEA cases, respectively. Forty-five cases (63%) presented with unilateral elbow and 26 cases (37%) with bilateral elbow involvement. Underlying pathology was rheumatoid arthritis in 22 cases (31%), loosening of elbow prosthesis in 19 cases (27%), posttraumatic arthritis in 11 cases (16%), primary osteoarthritis in 11 cases (16%), fracture of distal humerus in 5 cases (7%), haemophiliac arthropathy in two cases (3%), and arthritis due to synovial chondromatosis in one case (1%).

**LES and Patient Satisfaction**

The mean preoperative LES improved from 3.7 (SD: 1.8; range: 0.2 to 8.3) to 6.1 (SD: 1.71; range: 2.5 to 9.2). Forty patients (56%) were very satisfied, 18 patients (25%) satisfied, ten patients (14%) somewhat satisfied, and three patients (4%) unsatisfied after TEA.

**Distribution-based Approach**

1. MCID: Based on SD of change in LES, the mean change in the value of LES was 2.4 (range: -1.7 to 8.7). The value of SD for change in LES was 2.13. Using criteria of one half times SD the estimated MCID was 1.1.

2. SDC and SEM: Ten patients chose “somewhat satisfied” response at 1-year follow-up. These patients were considered to have a stable clinical condition as they did not have significant satisfaction. In these patients, the mean preoperative LES was 3.8 (SD: 2.1; range: 0.2 to 6.8) and the mean postoperative LES was 4.9 (SD: 1.6; range: 2.7 to 7.0). Cronbach’s alpha based on standardised items was estimated as 0.87. The SD of the baseline preoperative LES of the entire cohort of 71 patients was 1.8. The SEM was estimated to be 0.66 and using criteria of 1SEM, 1.96 SEM and 2.77 SEM resulted in the values of 0.7, 1.3, and 1.8 respectively. Based on the mathematical formula already described in the methods section, the SDC with upper limit of 90% confidence interval was estimated as 1.5.

**Anchor-based approach**

1. MCID based on the difference of mean of change in LES: The mean change in LES of the satisfied group was 2.4 (SD: 2.1; range: -1.5 to 6.2) while the mean change in LES of the somewhat satisfied group was 1.1 (SD: 1.4; range: -0.7 to 3.5) and the difference between both the means (MCID based on difference of mean in two sub-groups) was 1.3.

2. MCID based on ROC analysis: ROC analysis revealed the value of change of LES closest to the upper left hand corner of the graph (MCID) as 1.6(sensitivity value of 0.69 and specificity value of 0.69) (Figure 1). The area under curve (AUC) was 0.74 (p = 0.007; 95% confidence interval: 0.61 to 0.88).

**Triangulation approach: Combined distribution- and anchor-based approaches**

Thus using various methods it was estimated that the true value of MCID for the LES ranged from minimum value of 0.7 to maximum value of 1.8. The SDC value was estimated to be 1.5 in the present study. It is proposed that the “true” MCID value of LES would be between 1.6 and 1.8 in order to ensure that the value is higher than the measurement error of LES.

**DISCUSSION**

The present study has estimated the range of possible values of MCID using combination of both anchor-based and distribution-based methods. Some of the values were lower than the SDC values whereas some were higher than the SDC value of 1.5. Most studies have used ROC analysis to estimate MCID and in our study the MCID using ROC analysis was 1.6. We propose using the value of 1.6 for calculating sample size for research trials, as it is the value just higher than the SDC value.

Patient satisfaction has been previously used as a global transition external criterion to estimate clinically meaningful change.4,26 It is recommended that correlation of change in value of the outcome score with external anchor must be at least 0.30 for it to be useful for evaluation of minimal clinically important difference (MCID).23 A previously published study has shown positive correlation between change in value of LES and patient satisfaction (correlation coefficient 0.35). This justifies our choice of using patient satisfaction as an external anchor.32

Some authors have used the 90% upper confidence limit to estimate the SDC, calculated as the product of SEM with square root values of 2 and 1.645.3,15 These authors have recommended using 90% upper confidence limit of the SDC instead of the 95% upper confidence limit because of its higher precision.3 By contrast, some authors have used the 95% upper confidence limit to estimate SDC, calculated as the product of SEM with square root values of 2 and 1.96.14,31 SEM in the stable group of patients has been determined by the formula SD divided by square root of 2 wherein SD was the change in value of the outcome instrument in stable group of patients. Minimum detectable change (MDC) was determined by obtaining the product of SEM with square root values of 2 and 1.65. MDC calculated by this method represented the upper 90% confidence interval of the MDC. Calculating MDC using the above technique tended to overestimate the value of SDC.6 There is still no consensus regarding whether to use 90% or 95% upper confidence limit for calculating SDC.

The results of the present study concur with the findings of De Boer et al11 who used patient satisfaction after elbow surgery as an external anchor. In that study a larger difference in the LES value was associated with higher satisfaction while a smaller difference was observed in patients with lower satisfaction. The present study used a 4-point Likert-like scale while De Boer et al11 used a 10 cm visual analogue scale to assess patient satisfaction, converted into ordinal data by arbitrarily categorizing patients scoring 0 - 2.5 points as “dissatisfied”; 2.5 - 5 points as “somewhat satisfied”; 5 - 7.5 points as “moderately satisfied”; and those scoring 7.5 - 10 points as “very satisfied”. This classification may could be criticized as patients rating 2.5 points on the VAS could theoretically end in either “dissatisfied” or “somewhat satisfied” group and similarly, patients rating 5 points on the VAS could end up in either “somewhat satisfied” or “moderately satisfied” group. Moreover, the authors did not report the results of satisfaction level including the proportion of patients in various groups of satisfaction level. It would also have facilitated the data interpretation if the results had been published in form of mean, SD, and range for various satisfaction level groups.

De Boer et al11 used patient satisfaction as one of the external anchors and calculated the MCID by estimating the mean change in value of patients classified to be “somewhat satisfied”. Second external anchor used in that study was patient’s global perceived effect of intervention on a 5-point Likert scale (much improved/slightly improved/ no change/slightly worsened/much worsened). For ROC analysis, much improved response was categorized as “improved” group and those who experienced slight improvement, no change and slightly worsened were categorized as “no change” group. One case out of 25 patients who selected “much worsened” option was excluded from the analysis. It is recommended that for ROC analysis, all groups be included and no group should be excluded from the analysis.28 The study by De Boer et al11 relied solely on anchor-based method for MCID calculation while SDC and SEM values were not calculated. It is recommended that author should use both anchor-based and distribution-based methods to estimate MCID.28 Moreover, assessment of MCID was not at a uniform interval after the intervention: though mean follow-up period was seven months, the assessment was done at varying time intervals ranging from two months postoperative to 15 months after the index procedure.11

Dawson et al10 calculated the MCID of Oxford Elbow Score in 74 patients that underwent different elbow operations. It is unclear as to how many of these patients had TEA. Though SDC of OES-pain subscale was smaller than MCID, the SDC of OES-function and OES-social-psychological subscales was higher than the MCID values of corresponding subscales and hence the results of OES must be interpreted with caution. The authors had calculated SDC using 90% confidence upper limit of SEM. As the main intended benefits of TEA in elbow conditions are both pain relief and improved function, it is crucial that both pain and functional subscales show lower values of SDC as compared to MCID. Postoperative assessment was done at a similar interval of six months after the intervention and the authors used both anchor-based and distribution-based methods to estimate the MCID. Test retest reliability was estimated in a sample of 52 subjects with a delay of 48 hours between the readings.9 In the present study test retest reliability was estimated in a smaller sample of 10 subjects with longer time duration of 12 months between the readings. This could be a possible explanation for lower value of reliability coefficient observed in our study. Higher value of reliability coefficient is likely to lead to lowering of the value of SDC as it depends on the value of the reliability coefficient. It is also possible that a larger sample size would show lower variability in the LES values.

In the study by Dawson et al10, the authors did not calculate the value of reliability coefficient for the DASH and chose to impute the reliability coefficient from an earlier study by Beaton et al2. This approach of using previously published values of reliability coefficient in order to estimate MCID was also used by Angst et al1. The original validation study on LES showed that the value of reliability coefficient was 0.93.25,26 If we had used this value of reliability coefficient then the value of SDC would be 1.1. However, we chose to calculate the reliability coefficient in the present study as patient population and intervention administered are different than in the initial validation study.

A recent review24 on various outcome instruments for the TEA compared Mayo Elbow Performance Score (MEPS), Hospital for Special Surgery scoring system (HSS), Hospital for Special Surgery Total Elbow scoring system (HSS 2), Elbow Functional Assessment (EFA), and LES found MEPS the most commonly reported elbow score. The authors concluded that comparison of psychometric properties of the elbow scores failed to show any major difference between the outcome instruments. We agree with the view of Riedel and Beaton24 that a single total value of outcome instrument should be presented instead of several values for various sub-domains like pain, function and psychosocial components. Another review observed that measurement error of LES has not been reported. The present study addresses the issue of measurement error by describing SEM and SDC.28

Oneof the strengths of the present study is to show satisfactory interpretability of psychometric property of the LES. Interpretability is defined as the ability to assign and infer qualitative information from quantitative data.19 There is no consensus regarding the ideal way to check interpretability of an outcome measure. Terwee et al27 gave positive rating if the outcome score was presented as mean and SD in at least four sub-groups of participants and if values of MCID or minimal important change were presented. The present study has described MCID values for LES. Eechaute et al13 considered that interpretability of an outcome measure could be inferred if any two of the following four criteria were achieved: scores presented as mean and SD, comparison of data in various sub-groups, evaluating correlation between change in score and patient’s global assessment of change, and correlation between change in score and other commonly used outcome measures. Another study described LES in four subgroups of patients based on patient satisfaction after the total elbow replacement.32 The mean change in LES and the SD of the change score was estimated for unsatisfied, somewhat satisfied, satisfied and very satisfied patients. The mean change in LES increased as patient satisfaction improved.

The present study has its limitations. The value of reliability coefficient was 0.85, which was lower than the reliability coefficient (0.93) reported for the initial validation study.25 Possible explanations for this could be the difference in the outcome evaluation time. It is likely that evaluation at 1-year after the index procedure was long. Test-retest validity in the original validation study was performed between one and three days after the initial evaluation. Evaluation at one to weeks is more likely to give a more accurate result.12 It is recommended that the value of test-retest reliability coefficient should be more than 0.70 in order to be considered satisfactory.13 Our value is still acceptable as it is higher than this recommended value.

**CONCLUSION**

The MCID value for LES is expected to be between 0.7 and 1.9. As the SDC value was estimated as 1.5, we recommend using the MCID value of 1.6, which is greater than the value of SDC, and hence more likely to represent a true change in clinical condition of the patient.

**References**

1. Angst F, Goldhahn J, Drerup S, Kolling C, Aeschlimann A, Simmen BR, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. Arthritis Care Res 2012;64: 1749-1755. http://dx.doi: 10.1002/acr.21744

2. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. J Hand Ther 2001;14: 128-146.

3. Beninato M, Portney LG. Applying concepts of responsiveness to patient management in neurologic physical therapy. J Neurol Phys Ther 2011;35: 75-81. http://dx.doi:10.1097/NPT.0b013e318219308c

4. Bessette L, Sangha O, Kuntz KM, Keller RB, Lew RA, Fossel AH, et al. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. Med Care 1998;36: 491–502

5. Cella D, Eton DT, Lai JS, Peterman AH, Merkel DE. Combining anchor and distribution based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy anemia and fatigue scales. J Pain Symptom Manage 2002;24: 547-61. DOI: http://dx.doi.org/10.1016/S0885-3924(02)00529-8

6. Cleland JA, Whitman JM, Houser JL, Wainner RS, Childs JD. Psychometric properties of selected tests in patients with lumbar spinal stenosis. Spine 2012;12: 921-931. http://dx.doi:10.1016/j.spinee.2012.05.004

7. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56: 395-407. http://dx.doi:10.1016/S0895-4356(03)00044-1

8. Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, Jenkinson C. J Clin Epidemiol 2014;67: 697-705. http://dx.doi: 10.1016/j.jclinepi.2014.01.003

9. Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, et al. The development and validation of a patient reported questionnaire to assess outcomes of elbow surgery. J Bone Joint Surg [Br] 2008;90-B: 466-473. http://dx.doi:10.1302/0301-620X.90B4

10. Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, et al. Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. Qual Life Res 2008;17: 1257-67. http://dx.doi:10.1007/s11136-008-9409-3

11. De Boer YA, Hazes JMW, Winia PCA, Brand R, Rozing PM. Comparative responsiveness of four elbow scoring instruments in patients with rheumatoid arthritis, J Rheumatol 2001;28: 2616-23.

12. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Controlled Clin Trial 1991;12: 142–158.

13. Eechaute C, Vaes P, Van Aerschot L, Asman S, Duquet W. The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: a systematic review. BMC Musculoskelet Disord 2007;18: 1-11. http://dx.doi:10.1186/1471-2474-8-6

14. Gärtner FR, Nieuwenhuijsen K, van Dijk FJ, Sluiter JK. Interpretability of change in the Nurses Work Functioning Questionnaire: minimal important change and smallest detectable change. J Clin Epidemiol. 2012;65:1337-47. http://dx.doi:10.1016/j.jclinepi.2012.06.013

15. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Phys Ther. 2006;86: 735-743.

16. Irrgang JJ, Lubowitz JH. Measuring arthroscopic outcome. Arthroscopy 2008;24: 718-722. http://dx.doi:10.1016/j.arthro.2007.10.007

17. Kim JK, Park ES. Comparative responsiveness and minimal clinically important differences for idiopathic ulnar impaction syndrome. Clin Orthop Relat Res 2013;471: 1406-1411. http://dx.doi:10.1007/s11999-013-2843-8.

18. Kirkley A, Griffin S. Development of disease-specific quality of life measurement tools. Arthroscopy 2003;19:1121-8. http://dx.doi:10.1016/j.arthro.2003.10.028

19. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality of life and health status instruments: Development of scientific review criteria. Clin Ther 1996;18: 979-992. http://dx.doi:10.1016/S01492918(96)80054-3

20. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate. Qual Life Res 1995;4: 293–307.

21. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation, Med Care 2003;41: 582–592. http://dx.doi:10.1097/01.MLR.0000062554.74615.4C

22. Revicki DA, Erickson PA, Sloan JA, Dueck A, Guess H, Santanello NC. Interpreting and Reporting Results Based on Patient-Reported Outcomes. Value Health 2007;10 Suppl 2:S116-24. http://dx.doi:10.1111/j.1524-4733.2007.00274.x

23. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61:102 – 109. http://dx.doi:10.1016/j.jclinepi.2007.03.012

24. Riedel K, Beaton DE. Update on the state of outcome measurement in total elbow arthroplasty research: identifying a need for consensus. J Bone Joint Surg Am. 2013;95: e97 (1-8). http://dx.doi:10.2106/JBJS.K.01420

25. Sathyamoorthy P, Kemp GJ, Rawal A, Rayner V, Frostick SP. Development and validation of an elbow score. Rheumatology 2004;43: 1434-40. http://dx.doi:10.1093/ rheumatology/keh367

26. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition specific and health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol 1995;48: 1369–78. http://dx.doi:10.1016/0895-4356(95)00054-2

27. Terwee CB, Bot SDM, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Clin Epidemiol 2007;60: 34-42. http://dx.doi:10.1016/j.jclinepi.2006.03.012

28. The B, Reininga IH, El Moumni M, Eygendaal D. Elbow-specific clinical rating systems: extent of established validity, reliability, and responsiveness. J Shoulder Elbow Surg 2013;22:1380-94. http://dx.doi:10.1016/j.jse.2013.04.013.

29. Turner D, Schunemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et al. The minimal detectable change cannot reliably replace the minimal important difference. . J Clin Epidemiol 2010; 63:28-36. http://dx.doi:10.1016/j.jclinepi.2009.01.024.

30. Turner D, Schunemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol 2009;62: 374-9. http://dx.doi:10.1016/j.jclinepi.2008.07.009

31. Van Kampen DA, Willems WJ, Van Beers LW, Castelein RM, Scholtes VA, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res 2013;8: 40. http://dx.doi: 10.1186/1749-799X-8-40.

32. Vishwanathan K, Alizadehkhaiyat O, Kemp GJ, Frostick SP. Responsiveness of the Liverpool Elbow Score in elbow arthroplasty. J Shoulder Elbow Surg. 2013;22:312-7. http://dx.doi:10.1016/j.jse.2012.09.003.

33. Ward MM, Marx AS, Barry NN. Identification of clinically important changes in health status using receiver operating characteristic curves. J Clin Epidemiol 2000;53:279-84. http://dx.doi:10.1016/S0895-4356(99)00140-7

34. Wolinsky F,Wan G, TierneyW. Changes in the SF-36 in 12 months in a sample of disadvantaged older adults. Med Care 1998;36: 1589–98.

35. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. Eval Health Prof 2005;28:172-91. http://dx.doi:10.1177/0163278705275340

**Figure Legends**

**Figure1**. Figure shows estimation of MCID using ROC analysis. A diagonal is drawn from the upper left corner to the lower right corner. The point at which this diagonal intersects the curve represents the point closest to the upper left corner and hence, this represents the MCID value of 1.6. Vertical and horizontal reference lines are drawn to X-axis and Y-axis respectively. The vertical dotted reference line to X-axis is through a point traversing 1-specificity of 0.308. The horizontal dotted reference line to Y-axis is through a point traversing sensitivity of 0.690.