Article:

Nisar, S and Pandit, H orcid.org/0000-0001-7392-8561 (2021) Outcome measures in trauma and orthopaedics: a guide to evaluating your practice. Orthopaedics and Trauma, 35 (5). pp. 260-266. ISSN 1877-1327

https://doi.org/10.1016/j.mporth.2021.07.001

Outcome measures in T&O – a guide to evaluating your practice

Mr Sohail Nisar (MBBS, MRes, PGCert, MRCS (Ed))[1,2],
Professor Hemant Pandit (D Phil)[1]

[1]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, UK.
[2]Academic Department of Trauma and Orthopaedics, LGI, University of Leeds, Leeds, UK.

Corresponding Author:
Sohail Nisar
sohailnisar@doctors.org.uk
07868739036

No conflicts of interest declared for this manuscript

- Improvements in Health-Related Quality of Life (HRQoL) are key in orthopaedic surgery, however appropriate, timely and meaningful measurement of HRQoL can be challenging
- This paper provides an overview of the different types of commonly used outcome measures along with their characteristics and key properties used in Trauma and Orthopaedics
- Orthopaedic surgeons are provided a basis to select the most appropriate measure to evaluate their clinical practice

Keywords: outcome, orthopaedics, satisfaction, psychometrics

Introduction

An outcome measure has been defined as 'a measure chosen to assess the impact of an intervention' [72]. Outcome measures are essential to ensure an intervention is benefiting a patient. Outcome measures may also be used to predict which patients may benefit from a particular intervention. As Orthopaedic surgeons it is important to evaluate our practice however, selection of the appropriate outcome measure can be challenging. This review paper aims to give Orthopaedic surgeons an overview of types and characteristics of outcome measures and a step-by-step guide to equip them with the knowledge to effectively evaluate their practice.

The 'Outcome Measures Hierarchy' is a conceptual model of outcome measurements introduced by Michael E. Porter in 2010 [46]. The model consists of three tiers 1) Health status achieved or retained, 2) Process of recovery, and 3) Sustainability of health (**Table 1**). Each tier consists of two further dimensions, Tier 1 consists of survival and of degree of recovery/health which for Orthopaedics would include functional ability and level of pain. Tier 2 consists of time to recovery or return to normal activities and disutility of care or treatment process which would include length of stay and complications such as infection of deep vein thrombosis. Tier 3 firstly consists of sustainability of recovery/health over time which would include revision rates and secondly long-term consequences of therapy which would include regional pain syndrome or stiffness due to limited rehabilitation. The principle of this model is that multiple outcomes that are most relevant to the patient over the short- and long-term should be measured. These outcomes can be assessed by the clinician or be patient reported, we will discuss further the merits of each.

One of the first considerations in selecting an appropriate outcome measure should be the acceptability and feasibility of the test or questionnaire. For example, a questionnaire that takes too long to complete may not be acceptable to patients. If the outcome requires a rater, the availability of a member of staff to administer the test could affect the feasibility.

For the purposes of research, a primary outcome measure is the most important outcome measure in a study. The primary outcome measure will form the basis of the sample size for the study. Any other outcome measure used in a trial will be either a secondary or tertiary outcome measure. These can provide useful information, but the study may not be powered for these to be statistically conclusive.

1. Is the outcome measure reliable?

A reliable outcome measure will give the same result in a situation where no change has occurred, and as such patients can be distinguished from each other. Certain outcome measures may only be reliable for use in a specific population. Errors in reliability may stem from the test itself, the rater or the patient. An example of an unreliable outcome measure would be a patient reported outcome in which the questionnaire is difficult to understand and therefore interpreted differently on different occasions. A reliable rater administered test would be consistent when administered by the same rater on multiple occasions, this is the intra-rater reliability. The test would also be consistent when administered by separate raters, this is the intra-rater reliability. The extent to which the various items of the outcome measure are correlated is the internal consistency[47].

2. Has the outcome measure been validated?

A test may be reliable however it still may not measure what we would like to measure. Validity is the degree to which a test measures what a clinician intends to measure. Validity is not an inherent characteristic of the test, it is the tests ability to answer a particular question in a particular population [50]. There are three main methods of assessing the validity of a test:

- Content validity is the ability of a test to assess all aspects of a condition and as such be appropriate for use in all patients with that condition[47]. It is important to understand the populating under investigation including usual activities, any limitations secondary to their condition to be able to assess if all the salient issues are being addressed. The trade-off for being applicable to wider group of patients is that the reliability of the test tends to decrease.

- Criterion validity assesses the tests validity relative to the gold standard test administered either at the same time or at a later date[47]. This does however rely on the availability of a 'gold standard'.

- Construct validity is used when a gold standard is not available[47]. This method may use the known groups method. This involves administering the test to two different groups for which the results should differ. If the results do indeed differ it is an indication the test does measure what is intended.

Figure 1 demonstrates the relationship between validity and reliability.

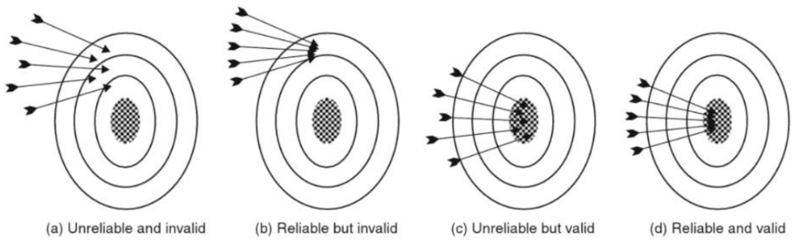

Figure 1: This diagram represents the relationship between reliability and validity. The bullseye is the outcome of interest and the arrow represent one instance of administering the test. Reproduced from [47].

3. Is the outcome measure responsive to change?

Responsiveness is the ability of a test to identify a clinically meaningful change [40] and is also referred to as 'sensitivity to change'. This would not be applicable if the outcome was, for example, mortality. When a test is administered before and after an intervention the ability of a test to detect this change is termed internal responsiveness. External responsiveness is the ability of a test to detect changes related to a change in another health status. A tests responsiveness very much relies on the reliability of that test. To be responsive a test must allow for scoring of included items to demonstrate a change. The responsiveness is of course specific to a particular question in a particular population.

The responsiveness of a test must be considered alongside the minimal clinically important difference (MCID). We will discuss MCID further in the next paragraph but for now we must understand that if the MCID is small then the test would need to be highly responsive to detect this change. If a test was not responsive to change it would lead to a false negative finding in a clinical study [39].

4. What is the MCID?

The MCID is '…the smallest difference in score in the domain of interest which patients perceive as beneficial…'[69,24]. The MCID is important in study design to determine sample size to ensure the study is adequately powered. A difference in outcome may be statistically significant, however this change is not necessarily clinically significant. The challenge is defining an MCID, there are many methods with no consensus on a particular one [9,55]. A brief overview of two of these methods:

- Anchor based methods compare the reported outcomes to an external measurement such as a laboratory bloods test[56]. The external test must be relevant to the condition and related to the outcome measure.
- Distribution based methods are statistical methods and express the change in outcome using metrics such as standard error of measurement or effects size[56].

As such there may be a number of values stated for an MCID of a particular test which may also vary due to the patient population in which they were calculated. Therefore, interpretation of an MCID requires careful consideration.

5. Does the outcome measure have ceiling or floor effects?

Ceiling and floor effects occur when a large proportion of the test group attain either the best or worst outcome[32,19]. In these situations, it is not possible to discern a meaningful difference between a group of high functioning patients or alternatively a group with a poor outcome. For example, if a test of functional ability is too easy and the patients all attain the highest score, it will not be possible to discern any meaningful difference between these patients. In Orthopaedic Surgery a score is deemed to display ceiling or floor effects if 15% or more of the respondents achieve the best or worst score [32,19].

6. Is a generic outcome measure or a disease specific outcome measure more appropriate?

Health Related Quality of Life (HRQoL) is an individual's or a group's perceived physical, psychological and social health over time. Improving HRQol is the most important goal of orthopaedic surgery [25]. HRQoL instruments are generic outcome measures designed to be used in a broad range of conditions. The use of a generic outcome measure allows for comparisons of treatments for different conditions which can be very useful in planning healthcare allocation. The condition of interest can also be compared to the general population. Another advantage is the potential ability to identify effects not directly related to the condition or treatment under investigation and as such, provide a more holistic view of the patient [5]. However, the limitation is that a generic outcome measure may not be responsive to specific changes related to the condition under investigation. Commonly used generic outcome measures are SF-36 (or shorter version SF-12) and EQ-5D.

SF-36 [63] is a widely used generic quality of life questionnaire. It has eight domains, which can be grouped into Physical Health and Mental Health. The shorter SF-12 questionnaire has only 12 questions in the two domains of Physical and Mental Health. The SF-36 scale does not have ratio properties so can be difficult to compare across a heterogenous group of clinical conditions. To enable comparison across different clinical conditions a utility measure such as the EQ-5D [48] is required. The EQ-5D has five domains and allows for the estimation of a health summary score on a scale 0 to 1 in which '0' is death and '1' is full health. The 5 dimensions of this score are mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The limitation of the EQ-5D is that it can be insensitive to small but clinically important change[25].

A condition specific outcome is also necessary in most cases of T&O; for example, following an arthroplasty (joint replacement) procedure. A test that is responsive to the change that occurs as a result of the arthroplasty is obviously very desirable. However, these tests may not be suitable for use in the general population of patients without a joint replacement and so comparative data with normal subjects may not be available. A condition specific test will also not pick-up other health benefits of the intervention for example improvements in mental health following an arthroplasty procedure. Ideally both generic and condition specific measures should be used in conjunction ad they provide complimentary information.

7. Is the outcome measure clinician assessed or patient reported?

Outcome scores may be completed independently by patients or administered by a clinician. Traditionally outcome has been assessed by a specific clinical parameter such as range of movement, radiographic findings or rate of complications.

An example of a clinician assessed outcome score is the Harris Hip Score [20]. The score covers pain, function, absence of deformity and range of motion. When using clinician assessed outcome measures surgeons must be aware of a discrepancy between what patients report independently and what is recorded when the test is completed by a rater. There is also a risk of variability in administering the score between raters.

Studies have shown when clinicians complete the same test or questionnaire patients report less pain and better function [6,15,28]. This finding may be due to patients not wanting to disappoint their surgeon and reporting improved outcomes. Or rather it may be what constitutes a successful outcome to the surgeon differs to the patient. Therefore, patient reported outcomes which minimise the potential of bias from the operating surgeon were developed. A Patient Reported Outcome Measure (PROM) is defined as 'a measurement based on a report that comes directly from the patient about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else. A PROM can be measured by self-report or by interview provided that the interviewer records only the patient's response'[73]. PROMs can be disease specific or generic questionnaires. The downsides of patient reported outcome scores is that some groups of patients may have difficulty completing them independently. Guidance on interpretation of incompletely filled questionnaires as well as validation of PROMs developed in different languages (to enable use in different patient populations) is important. For example the Oxford Knee Score (OKS) has been translated and validated in several languages[36,33].

8. Does the outcome measure assess patient satisfaction?

Assessing a patient's judgment about the success of an intervention rather than specific health status questions can provide a different perspective to an outcome. Matching patient expectation and satisfaction is vital in improving outcomes following surgical intervention[18,71,64]. In total knee arthroplasty the strongest predictor of dissatisfaction post-surgery is unmet expectations [4]. Satisfaction questionnaires also tend to be short and so acceptable to patients. However many available satisfaction questionnaires tend not to be validated [71]. Some commonly used orthopaedic outcome measures may contain questions about satisfaction but have not been validated to assess patient satisfaction[16].

Patient satisfaction may be with the outcome of the surgery or the process of care [16]. Surgeons tend to be more interested in satisfaction with outcome rather than process, however satisfaction with the process of care has a substantial effects on overall satisfaction [58].

Common Orthopaedic Outcome measures and their psychometric properties

Hip:

The Harris Hip Score (HHS)[20] was initially published in 1969 to evaluate the post-operative outcome. It is clinician administered which may be the Orthopaedic Surgeon, research nurse or physiotherapist. It takes 5 minutes to complete. The score has excellent internal consistency for reliability[42]. Content and construct validity of the score has been tested[60,61]. Responsiveness of the score is excellent for short term follow-up but weaker for long-term follow-up[57]. The use of the score is limited by large ceiling effects of between 20% and 32%[68]. The MCID for the HHS post-hip arthroplasty is 15.9–18 points[59]. The maximum score is 100, <70 is considered a poor outcome[42].

The Hip Disability and Osteoarthritis Outcome Score (HOOS)[30] is a PROM developed to assess patients with hip problems, it takes 10-15 minutes to complete[42]. The internal consistency for reliability ranges from good to excellent[42]. Content and construct validity of the score has been tested[30,43]. The responsiveness of the score is high following THR[45]. Ceiling effects of 19% have been reported in the pain subscale[43]. Floor effects of up to 17.8% have also been reported in the Sport/Recreation subscale[42]. The MCID post THR has not been reported for this score. The total scores for each subscale are: Pain 40 points, Symptoms 20 points, Activity limitations daily living 68 points, Function in sport and recreation 16 points, and Hip related quality of life 16 points. To enhance the interpretation, HOOS is transformed into a 0–100 worst to best scale, with 100 being the best and 0 the worst[43].

The Oxford Hip Score (OHS)[12] is an outcome measure specific to the hip. It is a patient reported outcome measure assessing pain and function. The score takes between 2 and 15 minutes to complete[38]. The internal consistency for reliability ranges from good to excellent [12,13]. Content validity has been measured for the OHS[11]. The score highly correlates with the HHS in patient post THR[26]. Responsiveness of the score has been compared with SF-36, EQ-5D and WOMAC with results in favour of OHS[42]. Ceiling effects of up to 13.5% have been reported but low floor effects[42]. The MCID for the OHS is 8 points[2]. 48 is the maximum score attainable in the OHS.

Knee:

The Knee Injury and Osteoarthritis Outcome Score (KOOS)[51] is a PROM developed for use in traumatic knee injuries which may lead to post-traumatic OA. The score takes 10 minutes to complete[52]. For severe OA, ceiling effects have been reported for pain (15%) and QoL (17%), floor effects of between 16-73.3% have been reported for the sport/recreation subscale[8]. In knee OA, the internal consistency for reliability between subscales varies from good to excellent, apart from for the symptoms and QoL subscales which score lower[8]. Content and construct validity of the score has been tested [52,54]. The scoring system is responsive to change post ACL reconstruction and TKR[8]. MCID for the KOOS is 16.7 for Pain, 10.7 for Symptoms, 18.4 for Activities of Daily Living, 12.5 for Sport/Recreation and 15.6 for Quality of Life. The five subscales of the KOOS (Pain, Symptoms, ADL Function, Sport and Recreation Function, and Quality of Life) are scored 0 (no problems) – 4 (extreme problems). Scores are transformed to a 0–100 scale, with zero representing extreme knee problems and 100 representing no problems[54].

The **Oxford Knee Score (OKS)** is a PROM developed for patients undergoing TKR[12]. The score takes 1-5 minutes to complete. The score has adequate internal consistency for reliability[8]. Content and construct validity of the score has been tested[12]. In patients following TKR ceiling effects of 27% have been reported with no floor effects[21]. The score has been shown to be demonstrate good responsiveness to change following TKR[8]. The MCID for the OKS in one study was 7 points[2] and in another study the MCID was between 4 and 5[7]. 48 is the maximum score attainable in the OKS.

The **Knee Society Score (KSS)** was initially developed in 1989 followed by an updated version in 2012[44] to assess expectations, satisfaction, and function of patients undergoing TKR. The score has clinician administered and PROMs. Content and construct validity of the score has been tested[44,10]. The score takes 15 minutes to complete[49]. The score has shown good responsiveness post TKR[35]. The score does not seem to demonstrate any floor or ceiling effects[35]. The KSS consists of four subscales: Objective Knee Score (100 points), Satisfaction Score (40 points), Expectation Score (15 points), and Functional Activity Score (100 points).

The **Forgotten Joint Score (FJS)** [66] is a PROM to assess patient awareness of the arthroplasty to determine satisfaction and functional level. Content and construct validity has been tested[66,1]. Internal consistency for reliability is excellent[1]. The score has shown good responsiveness post TKR between follow up appointments[17]. Floor and ceiling effects have been reported to be low (<15%)[1]. The MCID for the FJS has been reported at 14 points[22]. The FJS consists of 12 questions. Scores are transformed to a 0 to 100 scale. Higher scores reflect the ability of the patient to forget about the prosthesis[17].

Upper Limb:

QuickDASH score[3] is a short PROM developed from the Disabilities of the Arm, Shoulder and Hand (DASH) score. Internal consistency for reliability is excellent[3,27]. QuickDASH measures functional ability in patients with disorders of the upper limb. Content and construct validity has been tested[3]. A weakness of the score is poor responsiveness[27]. The MCID for the QuickDASH in patients with upper limb disorders is 15.91 points[14] and 14 in another study[62]. In patients with an elbow dislocation the QuickDash demonstrated ceiling effects of 29% from 3 months, increasing with time[23], however no ceiling effects were demonstrated for patients with shoulder pain[53]. Scores range from 0 (no disability) to 100 (most severe disability)[3].

Foot and Ankle:

The **American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale** [29] is one of the most commonly used outcome scores for patients with ankle or hindfoot injury. It combines a clinical administered and PROM. Construct validity has been measured for this score[29,37,34]. The scoring system is responsive in end stage ankle arthritis[34]. In patients with ankle arthritis the score demonstrates floor effects but no ceiling effects[67]. There is no available data on the MCID for the AOFAS Ankle-Hindfoot Scale. The score consists of three subscales: Pain with a maximal score of 40 points, Function with a maximal score of 50 points, Alignment with a maximal score of 10 points. The maximum score is 100 points, indicating no symptoms or impairments[31].

Activity Scales

As outcomes have improved following Orthopaedic procedures particularly joint replacement so have patient expectations with regard to participation in sporting activity. Activity rating scales to assess participation in sporting activity are becoming increasingly popular, two such scales are the UCLA scale[70] and Tegner score[65]. The UCLA scale ranges from 1 (no physical activity, dependant on other) to 10 (regular participation in impact sports). The Tegner score is also a maximum of 10 for participation in national and international elite competitive sport. Between the two, the UCLA scale demonstrates superior psychometric properties[41].

Conclusions

Overall, this paper provides an overview of the types of outcome measures as well as the characteristics of outcome measures. This will equip Orthopaedic Surgeons with the foundation to be able to select the most appropriate outcome measure to evaluate their practice.

References:

1. Adriani M, Malahias MA, Gu A, Kahlenberg CA, Ast MP, Sculco PK. Determining the Validity, Reliability, and Utility of the Forgotten Joint Score: A Systematic Review. J Arthroplasty. 2020 Apr 1;35(4):1137–44.
2. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol. 2015 Jan 1;68(1):73–9.

3. Beaton DE, Wright JG, Katz JN, Amadio P, Bombardier C, Cole D, et al. Development of the QuickDASH: COmparison of three item-reduction approaches. J Bone Jt Surg - Ser A. 2005 May;87(5):1038–46.
4. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ. Patient satisfaction after total knee arthroplasty: Who is satisfied and who is not? In: Clin Orthop. Springer New York; 2010. p. 57–63.
5. Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. Rheumatology. 1999 Sep 1;38(9):870–7.
6. Bullens PHJ, Van Loon CJM, De Waal Malefijt MC, Laan RFJM, Veth RPH. Patient satisfaction after total knee arthroplasty: A comparison between subjective and objective outcome assessments. J Arthroplasty. 2001;16(6):740–7.
7. Clement ND, MacDonald D, Simpson AHRW. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. Knee Surgery, Sport Traumatol Arthrosc. 2014;22(8):1933–9.
8. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster. Arthritis Care Res. 2011 Nov;63(SUPPL. 11):S208.
9. Cook CE. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. J Man Manip Ther. 2008 Oct;16(4):82E-83E.
10. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Validity and internal consistency of the new knee society knee scoring system. Clin Orthop Relat Res. 2018 Jan 1;476(1):77–84.
11. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Jt Surg - Ser B. 1996 Mar 1;78(2):185–90.
12. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Jt Surg - Ser B. 1998 Jan;80(1):63–9.
13. Fitzpatrick R, Morris R, Hajat S, Reeves B, Murray DW, Hannen D, et al. The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. Qual Saf Heal Care. 2000;9(3):146–50.
14. Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (quickDASH). J Orthop Sports Phys Ther. 2014;44(1):30–9.
15. Gioe TJ, Pomeroy D, Suthers K, Singh JA. Can patients help with long-term total knee arthroplasty surveillance? Comparison of the American Knee Society Score self-report and surgeon assessment. Rheumatology. 2009;48(2):160–4.
16. Graham B, Green A, James M, Katz J, Swiontkowski M. Measuring Patient Satisfaction in Orthopaedic Surgery. J Bone Jt Surg. 2015 Jan 7;97(1):80–4.
17. Hamilton DF, Loth FL, Giesinger JM, Giesinger K, MacDonald DJ, Patton JT, et al. Validation of the English language Forgotten Joint Score-12 as an outcome measure for total hip and knee arthroplasty in a British population. Bone Joint J. 2017 Feb 1;99-B(2):218–24.
18. Hamilton DF, Lane J V., Gaston P, Patton JT, MacDonald D, Simpson AHRW, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. BMJ Open. 2013;3(4).
19. Harris K, Lim CR, Dawson J, Fitzpatrick R, Beard DJ, Price AJ. The Oxford knee score and its subscales do not exhibit a ceiling or a floor effect in knee arthroplasty patients: an analysis of the National Health Service PROMs data set. Knee Surgery, Sport Traumatol Arthrosc. 2017 Sep 1;25(9):2736–42.
20. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am. 1969 Jun 1;51(4):737–55.
21. Impellizzeri FM, Mannion AF, Leunig M, Bizzini M, Naal FD. Comparison of the Reliability, Responsiveness, and Construct Validity of 4 Different Questionnaires for Evaluating Outcomes after Total Knee Arthroplasty. J Arthroplasty. 2011 Sep;26(6):861–9.
22. Ingelsrud LH, Roos EM, Terluin B, Gromov K, Husted H, Troelsen A. Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. Acta Orthop. 2018 Sep 3;89(5):541–7.
23. Iordens GIT, Den Hartog D, Tuinebreijer WE, Eygendaal D, Schep NWL, Verhofstad MHJ, et al. Minimal important change and other measurement properties of the Oxford Elbow Score and the Quick

Disabilities of the Arm, Shoulder, and Hand in patients with a simple elbow dislocation; validation study alongside the multicenter FuncSiE trial. PLoS One. 2017 Sep 1;12(9).

24. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407–15.
25. Jansson KÅ, Granath F. Health-related quality of life (EQ-5D) before and after orthopedic surgery. Acta Orthop. 2011 Feb;82(1):82–9.
26. Kalairajah Y, Azurza K, Hulme C, Molloy S, Drabu KJ. Health outcome measures in the evaluation of total hip arthroplasties - A comparison between the harris hip score and the Oxford hip score. J Arthroplasty. 2005 Dec;20(8):1037–41.
27. Kennedy CA, Beaton DE, Smith P, Van Eerd D, Tang K, Inrig T, et al. Measurement properties of the QuickDASH (Disabilities of the Arm, Shoulder and Hand) outcome measure and crosscultural adaptations of the QuickDASH: A systematic review. Vol. 22, Quality of Life Research. Qual Life Res; 2013. p. 2509–47.
28. Khanna G, Singh JA, Pomeroy DL, Gioe TJ. Comparison of Patient-Reported and Clinician-Assessed Outcomes Following Total Knee Arthroplasty. J Bone Jt Surg. 2011 Oct 19;93(20):e117.
29. Kitaoka HB, Alexander IJ, Adelaar RS, Nunley JA, Myerson MS, Sanders M. Clinical Rating Systems for the Ankle-Hindfoot, Midfoot, Hallux, and Lesser Toes. Foot Ankle Int. 1994;15(7):349–53.
30. Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score: An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. Scand J Rheumatol. 2003;32(1):46–51.
31. Van Lieshout EMM, De Boer AS, Meuffels DE, Den Hoed PT, Van Der Vlies CH, Tuinebreijer WE, et al. American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Score: A study protocol for the translation and validation of the Dutch language version. BMJ Open. 2017 Feb 1;7(2):e012884.
32. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: An analysis of the NHS PROMs data set. BMJ Open. 2015 Jul 1;5(7):e007765.
33. Lin K, Bao L, Wang J, Fujita K, Makimoto K, Liao X. Validation of the Chinese (Mandarin) Version of the Oxford Knee Score in Patients with Knee Osteoarthritis. Clin Orthop Relat Res. 2017 Dec 1;475(12):2992–3004.
34. Madeley NJ, Wing KJ, Topliss C, Penner MJ, Glazebrook MA, Younger ASE. Responsiveness and validity of the SF-36, Ankle osteoarthritis scale, AOFAS ankle hindfoot score, and foot function index in end stage ankle arthritis. Foot Ankle Int. 2012 Jan;33(1):57–63.
35. Maniar RN, Maniar PR, Chanda D, Gajbhare D, Chouhan T. What is the Responsiveness and Respondent Burden of the New Knee Society Score? Clin Orthop Relat Res. 2017 Sep 1;475(9):2218–27.
36. Martín-Fernández J, García-Maroto R, Sánchez-Jiménez FJ, Bau-González A, Valencia-García H, Gutiérrez-Teira B, et al. Validation of the Spanish version of the Oxford knee score and assessment of its utility to characterize quality of life of patients suffering from knee osteoarthritis: A multicentric study. Health Qual Life Outcomes. 2017 Sep 29;15(1).
37. Martin RL, Irrgang JJ. A survey of self-reported outcome instruments for the foot and ankle. Vol. 37, Journal of Orthopaedic and Sports Physical Therapy. Movement Science Media; 2007. p. 72–84.
38. McMurray R, Heaton J, Sloper P, Nettleton S. Measurement of patient perceptions of pain and disability in relation to total hip replacement: The place of the Oxford hip score in mixed methods. Qual Heal Care. 1999;8(4):228–33.
39. McPhail SM, Bagraith KS, Schippers M, Wells PJ, Hatton A. Use of Condition-Specific Patient-Reported Outcome Measures in Clinical Trials among Patients with Wrist Osteoarthritis: A Systematic Review. Adv Orthop. 2012;2012:1–10.
40. Moretti C, Viola S, Pistorio A, Magni-Manzoni S, Ruperto N, Martini A, et al. Relative responsiveness of condition specific and generic health status measures in juvenile idiopathic arthritis. Ann Rheum Dis. 2005 Feb 1;64(2):257–61.
41. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin Orthop Relat Res. 2009;467(4):958–65.
42. Nilsdotter A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (A. Arthritis Care Res (Hoboken). 2011 Nov 1;63(S11):S200–7.
43. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) - Validity and responsiveness in total hip replacement. BMC Musculoskelet Disord. 2003 May 30;4:1–8.

44. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, et al. Development of a new knee society scoring system. In: Clinical Orthopaedics and Related Research. Springer Science and Business Media, LLC; 2012. p. 20–32.
45. Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. Osteoarthr Cartil. 2010 Apr;18(4):522–9.
46. Porter ME. What Is Value in Health Care? N Engl J Med. 2010 Dec 23;363(26):2477–81.
47. Pynsent P, Fairbank J, Carr A. Outcome Measures in Orthopaedics and Orthopaedic Trauma. Second. Outcome Measures in Orthopaedics and Orthopaedic Trauma, 2Ed. CRC Press; 2004.
48. Rabin R, De Charro F. EQ-5D: A measure of health status from the EuroQol Group. In: Annals of Medicine. Royal Society of Medicine Press Ltd; 2001. p. 337–43.
49. Reynaud V, Verdilos A, Pereira B, Boisgard S, Costes F, Coudeyre E. Clinical Medicine Core Outcome Measurement Instruments for Clinical Trials of Total Knee Arthroplasty: A Systematic Review.
50. Roach KE. Measurement of health Outcomes: Reliability, validity and responsiveness. J Prosthetics Orthot. 2006 Jan 1;18(6 PROCEEDINGS):P8–12.
51. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): From joint injury to osteoarthritis. Vol. 1, Health and Quality of Life Outcomes. BioMed Central; 2003. p. 64.
52. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - Validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes. 2003 May 25;1:17.
53. Rysstad T, Grotle M, Klokk LP, Tveter AT. Responsiveness and minimal important change of the QuickDASH and PSFS when used among patients with shoulder pain. BMC Musculoskelet Disord. 2020 May 27;21(1):328.
54. Salavati M, Akhbari B, Mohammadi F, Mazaheri M, Khorrami M. Knee injury and Osteoarthritis Outcome Score (KOOS); reliability and validity in competitive athletes after anterior cruciate ligament reconstruction. Osteoarthr Cartil. 2011 Apr;19(4):406–10.
55. Sedaghat AR. Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. Otolaryngol - Head Neck Surg (United States). 2019 Oct 1;161(4):551–60.
56. Shah A, Bailey CR. Outcomes following surgery: are we measuring what really matters? Anaesthesia. 2019 Jun 8;74(6):696–9.
57. Shi HY, Mau LW, Chang JK, Wang JW, Chiu HC. Responsiveness of the Harris Hip Score and the SF-36: Five years after total hip arthroplasty. Qual Life Res. 2009 Oct;18(8):1053–60.
58. Shirley ED, Sanders DO. The orthopaedic forum :Patient satisfaction: Implications and predictors of success. Vol. 95, Journal of Bone and Joint Surgery - Series A. Journal of Bone and Joint Surgery Inc.; 2013.
59. Singh JA, Schelck C, Harmsen S, Lewallen DG. Clinically Important Improvement Thresholds for Harris Hip Score and Its Ability To Predict Revision Risk after Primary Total Hip Arthroplasty. Ann Rheum Dis. 2016 Jun 1;75(Suppl 2):1164.1-1164.
60. Söderman P, Malchau H, Herberts P. Outcome of total hip replacement: A comparison of different measurement methods. Clin Orthop Relat Res. 2001;(390):163–72.
61. Söderman P, Malchau H. Is the Harris hip score system useful to study the outcome of total hip replacement? Clin Orthop Relat Res. 2001;(384):189–97.
62. Sorensen AA, Howard D, Tan WH, Ketchersid J, Calfee RP. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg Am. 2013 Apr;38(4):641–9.
63. Stewart M. The medical Outcomes Study 36-item short-form health survey (SF-36). Vol. 53, Australian Journal of Physiotherapy. Australian Physiotherapy Association; 2007. p. 208.
64. Swarup I, Henn CM, Gulotta L V., Henn RF. Patient expectations and satisfaction in orthopaedic surgery: A review of the literature. Vol. 10, Journal of Clinical Orthopaedics and Trauma. Elsevier B.V.; 2019. p. 755–60.
65. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries - PubMed. Clin Orthop Relat Res. 1985;198:43–9.
66. Thomsen MG, Latifi R, Kallemose T, Barfod KW, Husted H, Troelsen A. Good validity and reliability of the forgotten joint score in evaluating the outcome of total knee arthroplasty: A retrospective cross-sectional survey-based study. Acta Orthop. 2016 May 3;87(3):280–5.
67. Veltman ES, Hofstad CJ, Witteveen AGH. Are current foot- and ankle outcome measures appropriate for the evaluation of treatment for osteoarthritis of the ankle?: Evaluation of ceiling effects in foot- and ankle outcome measures. Foot Ankle Surg. 2017 Sep 1;23(3):168–72.
68. Wamper KE, Sierevelt IN, Poolman RW, Bhandari M, Haverkamp D. The Harris hip score: Do ceiling

effects limit its usefulness in orthopedics? A systematic review. Acta Orthop. 2010 Dec;81(6):703–7.
69. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID).
70. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. J Arthroplasty. 1998;13(8):890–5.
71. Zywiel MG, Mahomed A, Gandhi R, Perruccio A V., Mahomed NN. Measuring expectations in orthopaedic surgery: A systematic review. In: Clinical Orthopaedics and Related Research. Springer New York LLC; 2013. p. 3446–56.
72. Outcome measures and case definition. In: Field Trials of Health Interventions. Oxford University Press; 2015. p. 198–215.
73. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims | FDA [Internet]. [cited 2021 Apr 6].

<u>Tables</u>

Outcome measures Hierarchy	Examples of measurement
Health Status achieved or retained	
- Survival	Peri-operative mortality
- Degree of recovery/health	Pain, function, return to ADLs and other recreational activities
Process of recovery	
- Time to recovery or return to normal activities	Time to fracture union, return to ADLs/work or to be symptom free
- Disutility of care/treatment process	Length of hospital stay, requirement of analgesia, post-operative infection
Sustainability of health	
- Sustainability of recovery or health over time	Revision or re-operation rates, maintenance of quality of life
- Long-term consequences of therapy	Aseptic loosening, infection, stiffness, implant survival

Table 1: Porters Outcome Measures Hierararchy