ORIGINAL RESEARCH

Deconstructing the Minimum Clinically Important Difference (MCID)

Janine Molino[1,2], Joseph Harrington[2], Jennifer Racine-Avila[2], Roy Aaron [2]

[1]Lifespan Biostatistics Core, Rhode Island Hospital, Providence, RI, USA; [2]Department of Orthopedic Surgery, Warren Alpert Medical School of Brown University, Providence, RI, USA

Correspondence: Roy Aaron, Email Roy_Aaron@Brown.edu

Purpose: The minimal clinically important difference (MCID) is a way of dichotomizing data for assessment of success or failure based on clinically meaningful changes. The magnitude of the MCID is often misunderstood to be a singular quantity applicable across studies. However, substantial differences have been reported among MCIDs for the same outcome measures usually based upon differences extrinsic to the calculation. This study explores the effects of variabilities intrinsic to the calculation of the MCID.

Methods: The MCIDs for two knee replacement patient-reported outcomes measures of pain and function were calculated at 1 year postoperative with an integrative anchor and distribution-based method using external anchor questions and receiver operator characteristic (ROC) curves. The effects upon the magnitude and precision of the MCIDs of varying the anchor questions, the thresholds for success/failure, and the sample sizes were examined.

Results: Wide variabilities were observed in both the magnitudes and precision of the MCIDs. The threshold for success had the largest effect on magnitude of pain scores, while the sample size had the largest effect on precision. For function scores, the sample size had the largest effect on magnitude, and the anchor question had the largest effect on precision.

Conclusion: Comparisons among MCIDs are difficult to interpret if elements of the calculations are different and influence the results. While factors extrinsic to the calculations, e.g., study population, trial design, methods of calculation, etc., are known to produce differences in the magnitude of MCIDs, this study shows that more subtle and less obvious factors intrinsic to the calculations have profound effects on both the magnitude and precision of MCIDs. Comparisons among MCIDs should be made with caution and call for greater transparency in reporting intrinsic methods. It is probably advisable for individual studies to calculate their own MCIDs and not rely on published values.

Keywords: outcome assessment, categorical measure, clinical improvement

Introduction

For a variety of clinical, quality improvement, and research applications, it is often advantageous to express outcomes in binary or categorical terms, reflecting the success or failure of a therapeutic intervention. To be most useful, the outcomes should represent clinically meaningful changes, that is, changes that are acknowledged by the patient to be of sufficient magnitude to represent a successful or unsuccessful result. Patient-reported outcome measures (PROMs) express clinically meaningful outcomes.[1] However, they are usually expressed as continuous scales with no criteria for success/failure. One way of transforming PROMs to categorical scales is with the minimum clinically important difference (MCID) that reflects important health status changes on the patient level and can represent success or failure of a therapeutic intervention. "MCIDs are patient derived scores that reflect changes in a clinical intervention that are meaningful for the patient".[2]

The concept of MCIDs in outcome assessment was introduced in 1989 and quickly became an important outcome instrument, supported by the FDA and NIH.[3] It was adopted by the AAOS to assess clinical significance and publications appeared purporting to identify the MCID for a number of patient-reported outcome instruments.[4] While early reports suggested that the MCID was a well-defined and singular quantity, subsequent studies demonstrated differences among reported values for the same outcome measures.[5–8] MCIDs have been shown not to be singular values and not necessarily

35

transferable among studies, and comparisons of MCIDs among studies are fraught with difficulties. Part of the dilemma is that (1) there are several methods of estimating the MCID; (2) the MCID depends upon the clinical population, disease entity and severity; and (3) the calculations themselves depend upon a variety of methodological techniques.[5,6,9] Most of the inconsistencies have been ascribed to factors extrinsic to the actual calculation of the MCID such as patient population characteristics, including socioeconomic status, mental health, and social support, disease type and severity, and methods used to calculate the MCID.[6,8,10–14]

The hypothesis of this study is that both the precision and magnitude of the MCID can be influenced also by methods intrinsic to the calculations and, together with extrinsic factors, influence how MCIDs can be interpreted and compared. While much has been written about external features influencing MCIDs, this report demonstrates effects of features internal to the calculation and draws conclusions about comparing MCIDs using different criteria for the calculation. This study uses the Knee Injury and Osteoarthritis Outcomes Score (KOOS) Pain subscale and the Veterans Rand 12-Item Health Survey Physical Component Summary (VR-12 PCS) scale in the setting of total knee replacement (TKR) to examine the dependency of the MCID calculations upon the intrinsic elements of the methods used and, thereby, explain the wide variations that have been reported in MCID calculations. The importance of the study derives from the observation that 10–30% of patients report suboptimal pain and function status after TKR and that clinically relevant, reliable, binary descriptors of clinical success and failure are needed.[15] To ensure that the MCID is interpreted and compared accurately, sources of uncertainty need to be identified.

Methods

This study required, and was granted, approval by the Institutional Review Board of the Lifespan Academic Medical Center. Deidentified clinical data from patients undergoing TKR were obtained from the Functional Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) data registry of The Miriam Hospital Total Joint Center. As part of the registry, the validated PROMs, KOOS pain subscale, VR-12 PCS, and Patient-Reported Outcomes Measurement Information System (PROMIS), were prospectively collected preoperatively and at 3 and 12 months postoperatively. The KOOS pain and the PCS represent two independent domains of pain and function, respectively. This study used the preoperative and 12-month postoperative data from 101 consecutive patients for its calculations. Inclusion criteria included primary TKR for osteoarthritis and completed preoperative and 1-year post-operative KOOS pain and PCS scores, and the PROMIS questionnaire.

The MCIDs for the KOOS pain and PCS scales at 12 months postoperative were calculated using an integrative anchor and distribution-based method. In this method, MCIDs are calculated by using anchor questions to categorize patients based on clinical improvement and applying receiver operator characteristic (ROC) curves to identify the value on the health status instrument under study (ie, KOOS pain or PCS) that characterizes patient outcomes most precisely.[14,16] The anchor questions were external to the PROMs and responses to them were collected concurrently with those of the PROMs.

The study hypothesis that both the precision and magnitude of the MCID can be influenced by methods intrinsic to the calculation was tested by examining the effects on the magnitude and precision of the MCIDs for clinical improvement of (1) varying the anchor questions; (2) varying the threshold for success; and (3) varying the sample size. The patient population was the same for all three tests.

Varying the Anchor Question

In an anchor-based MCID method, outcome scores are compared with an external, relevant "anchor" question with which patients report their degree of improvement.[13] Anchors can be transition questions, Patient Global Impression of Change (PGIC) or Patient Global Assessment (PGA) of treatment effectiveness.[13,17] We used PGA questions as our external anchors. The effects on the MCIDs of two external anchor questions were examined for each domain of pain and function (Table 1). One pain anchor question was obtained from the VR-12 and the comparison anchor question for pain was obtained from the PROMIS. One function anchor question was obtained from the KOOS ADL subscale and the comparison was obtained from the PROMIS.

Table 1 Anchor Questions

Outcome Domain	Measurement Source	Anchor Question
KOOS Pain	VR12	"During the past 4 weeks, how much did **pain** interfere with your normal work (including both work outside the home and housework)?"
	PROMIS	"How would you rate your **pain** on average?"
VR12 PCS	KOOS ADL	"The following question concerns your **physical function**. By this we mean your ability to move around and to look after yourself. For the following activity please indicate the degree of difficulty you have experienced in the last week due to your surgical knee. If you are not able to do the activity listed, tell us how difficult it would be if you attempted to do the activity. 'Rising from sitting'"
	PROMIS	"To what extent are you able to carry out your everyday **physical activities** such as walking, climbing stairs, carrying groceries, or moving a chair?"

Note: Bold text indications assessment domain.

Varying the Threshold for Success

Because of its greater precision, the PROMIS anchor questions for pain and function were used to examine the effects on the MCID of varying the criteria for success. Responses to the anchor questions were obtained on 0–10 Likert scale for KOOS pain and a 1–5 Likert scale for PCS function. Two sets of clinically realistic thresholds for success/failure were compared for their effects on the MCID. For KOOS pain, one group of patients with a response of 0–3 were considered as clinical success while patients with a response of 4–10 were considered as failures. They were compared to another group of patients with a response of 0–6, considered as clinical success, and a response of 7–10, considered as clinical failures. For PCS function, one group of patients with a response of 1–2 were considered as clinical failures while patients with a response of 3–5 were considered as clinical successes. They were compared to a group with a response of 1–3 considered as clinical failures and 4–5 considered as clinical successes.

Varying the Sample Size

The study of the effect of sample size on the MCID compared samples of 50 with 101 patients. The PROMIS anchor questions for the KOOS pain scale and PCS function score were used to examine the effects on the MCID of varying the sample size. A response of 0–3 was considered as the threshold for clinical success for KOOS pain and a response of 1–3 was considered as the threshold for clinical success for PCS function.

Statistical Analysis

Data was imported in SAS version 9.4 (SAS Institute Inc., Cary, NC) for data management and analysis. The MCID was calculated separately for the KOOS pain and PCS scales at 12-months postoperatively. For each scale, the patient cohort was divided into two groups, successfully and unsuccessfully treated patients, according to the responses to the anchor question. The magnitude and precision of the MCID for clinical success were calculated for each scale by using the ROC threshold method. In the context of calculating the MCID, the health status instrument (ie, the KOOS Pain Scale or the VR12 PCS) was considered the diagnostic test while the quantification of clinical success was defined by the responses to the anchor question. The ROC threshold estimated the magnitude of the MCID and was calculated by finding the value of the health status instrument that was maximal by Youden's J statistic, which was calculated as follows.[18]

J = sensitivity + specificity - 1

Precision was described by the concordance index (C-statistic) of the ROC curve and was compared within scenarios using Z-tests. Results are reported as mean ± SD. A $p<0.05$ was used to determine statistical significance.

Results

The patient population was representative of patients undergoing TKR. The mean age was 67; 70% were female; the mean BMI was 32. The results indicate that the magnitudes and precisions of the MCIDs of the two PROM domains examined after TKR were affected by factors intrinsic to their calculations. Varying the anchor questions, thresholds for
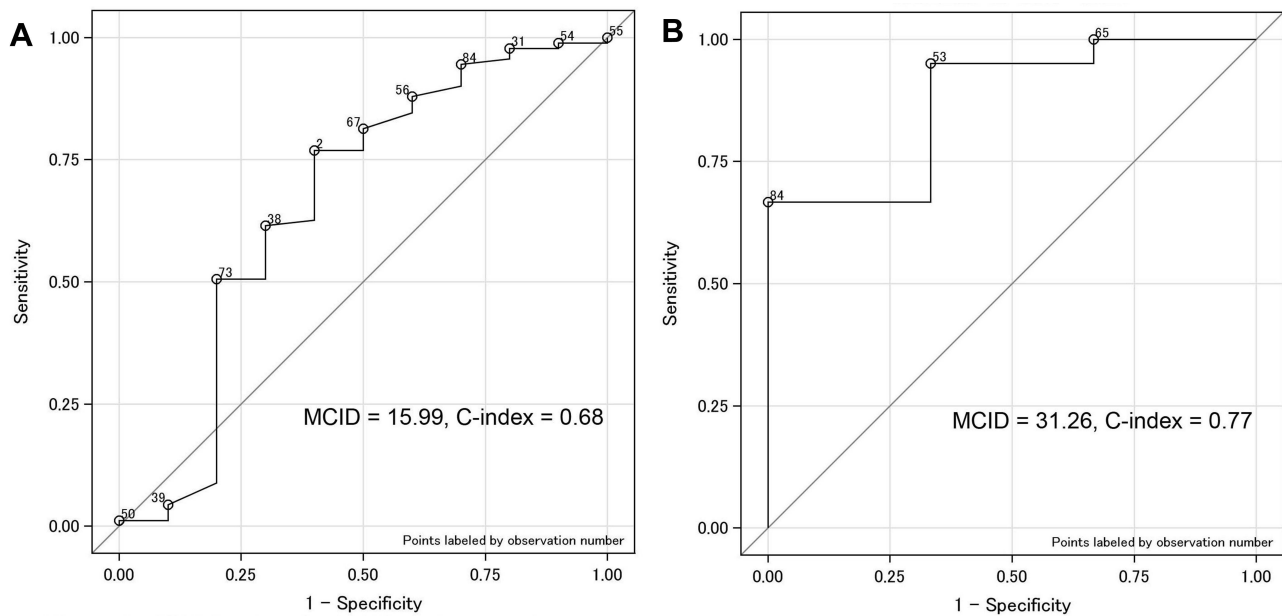
Figure 1 ROC curves for KOOS pain subscale demonstrating the effects of the anchor question upon both magnitude and precision of the MCID. (**A**) Anchor question derived from the VR-12. (**B**) Anchor question derived from PROMIS. The statistical significance (p value) applies to the measure of precision (C-index).

success/failure, and the sample size each exerted substantial effects on the magnitude and precision of the MCIDs of the KOOS pain subscale and the PCS function scale. Examples of the ROC curves for KOOS pain with varying anchor questions are shown in Figure 1A and B. Compared with the VR-12 question, the PROMIS question doubled the MCID from 15.99 to 31.26 and significantly increased the precision from a C-index of 0.68 to 0.77 (p < 0.001). An example of changing the threshold for success/failure on the ROC curves for PCS function scores is shown in Figure 2A and B. Compared to a threshold of success of 3–5, using a threshold of success of 4–5 doubled the MCID from 1.03 to 2.46; however, the precision significantly decreased from a C-index of 0.75 to 0.58 (p < 0.001).
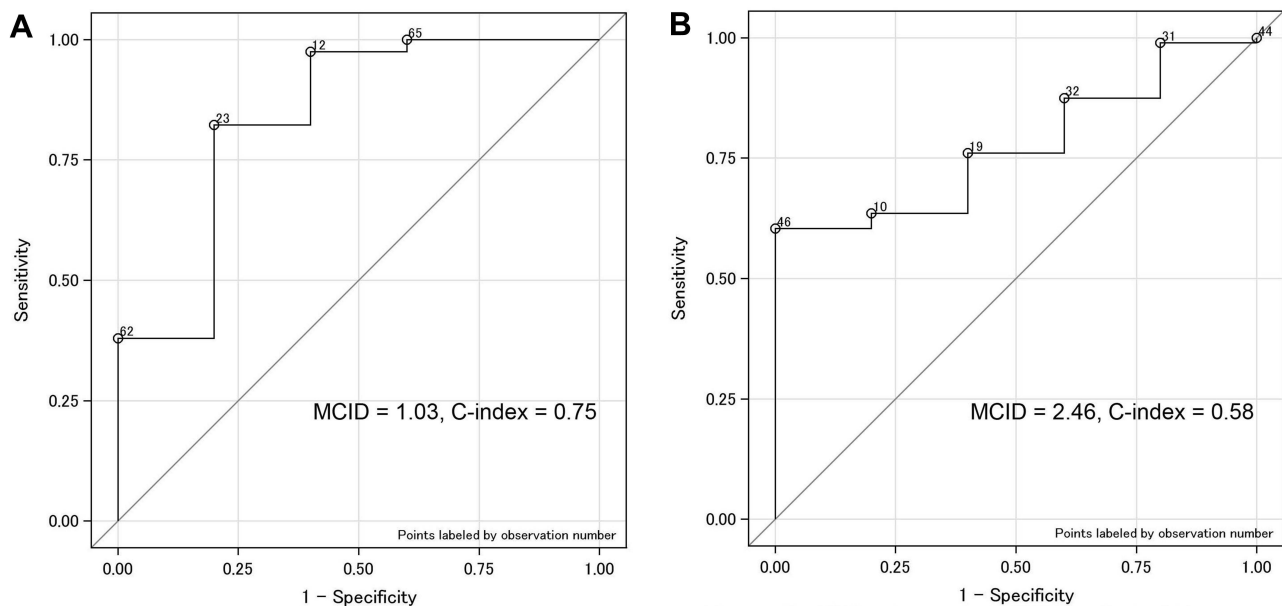

Figure 2 ROC curves for PCS function domain demonstrating the effects of the threshold criteria from a Likert scale upon both magnitude and precision of the MCID. (**A**) Success criteria 3–5. (**B**) Success criteria 4–5. The statistical significance (p value) applies to the measure of precision (C-index).

Table 2 Estimated MCID for KOOS Pain

	MCID	C-Index	P-value
Varying Anchor Questions			<0.0001
VR-12 Anchor	15.99	0.68	
PROMIS Anchor	31.26	0.77	
Varying Criteria for Success/Failure			<0.0001
Criteria 1 (0–3 Clinical Success)	31.26	0.77	
Criteria 2 (0–6 Clinical Success)	6.26	0.56	
Varying Sample Size			<0.0001
Sample Size of 50	25.06	0.53	
Sample Size of 100	31.26	0.77	

The mean preoperative KOOS pain score was 47.4 ± 19.3 and at 1-year follow-up, increased to 81.2 ± 20.4. The magnitude of the MCID of the KOOS pain subscale ranged from 6.26 to 31.26 and the precision from 0.53 (poor) to 0.77 (excellent) ($p < 0.001$) (Table 2). Both the magnitude and precision of the KOOS pain MCID were sensitive to all 3 changing scenarios. The threshold for success had the largest effect on magnitude while the sample size had the largest effect on precision. The sample size exerted the smallest effect on magnitude.

The mean preoperative PCS score was 36.2 ± 9.8 and at 1-year follow-up, increased to 45.8 ± 9.5. The magnitude of the MCID of the PCS ranged from 1.03 to 12.19 and the precision from 0.50 (poor) to 0.77 (excellent) ($p < 0.001$) (Table 3). The magnitude and precision of the PCS MCID were also sensitive to all 3 changing scenarios. The sample size had the largest effect on magnitude and the anchor question had the largest effect on precision.

Discussion

The goal of an MCID is to express, with as much precision as possible, the clinical significance of an increment of change in a patient's medical status. Therefore, it has to express both a clinical change that is meaningful to a cohort of patients and it has to be statistically rigorous enough to deflect bias. The integrative anchor-based ROC MCID method best reflects clinical change, offers a degree of precision, and dichotomizes continuous data. However, comparisons among MCIDs are fraught with error if elements of the calculations are different and influence the results.

Several studies have shown that factors extrinsic to the calculations themselves can produce variability in the MCID. These factors include characteristics of the study populations including sociodemographics, trial design including various clinical measurement scales, and the methods used to calculate the MCID.[7,10,13] Assessing outcome after TKR with quality of life (SF-36), disease-specific (WOMAC), and knee specific (KOOS) instruments will yield different MCIDs. Different MCIDs have been calculated depending upon the condition being assessed and the outcome assessment instruments used.[19,20] MCIDs of TKR, THR, and rehabilitation differ from one another.[7,8] Preoperative baseline PROM threshold scores have been shown to affect the KOOS pain scores 1 year after TKR.[6,12] Length of follow-up can

Table 3 Estimated MCID for PCS

	MCID	C-index	P-value
Varying Anchor Questions			<0.0001
VR-12 Anchor	6.67	0.50	
PROMIS Anchor	1.03	0.77	
Varying Criteria for Success/Failure			<0.0001
Criteria 1 (3–5 Clinical Success)	1.03	0.75	
Criteria 2 (4–5 Clinical Success)	2.46	0.58	
Varying Sample Size			<0.0001
Sample Size of 50	12.19	0.50	
Sample Size of 100	2.46	0.53	

also affect the MCID. A study of the long-term variability of the MCID of TKR patients demonstrated that the magnitude of the MCIDs fluctuated between 1 and 7 years postoperative so that the time of calculation of the MCID is an important extrinsic factor.[21] One-third of patients exhibited changes in MCID within 1–2 years post therapy. A meta-analysis demonstrated the dependency of MCID upon the external factors of time of assessment, study population, diagnosis, baseline status and patient demographics but did not include MCID analytics for KOOS or PCS after TKR.[11]

While extrinsic factors have been well reported to affect the MCID, less attention has been paid to factors intrinsic to the calculations. Our data indicate that, in addition to factors extrinsic to the calculation of the MCID, elements intrinsic to the calculations themselves can produce differences in both the magnitude and the precision of the MCID. With two commonly used PROMs in the context of TKR, KOOS pain subscale and VR-12 PCS, our data have shown that both the magnitude and the precision of the MCID calculation can be affected by the anchor question, the threshold for success, and the sample size. With the KOOS pain score MCID, we observed up to a 25-point (5-fold) range in magnitude and a 0.24 range in C-index (from poor to excellent), depending upon the variables used in the calculations. For the MCID of the PCS function score, we observed an 11.16 range (over 10 fold) in magnitude and a 0.27 range in C-index (from poor to excellent). These results are examples of the susceptibility of the MCID calculations to intrinsic factors.

There are some limitations to this study. Perhaps the most important one is the uncertainty of the generalizability of its observations since they were done with a particular set of methods and in a particular population of patients. As the general theme of this report points out, MCIDs may not be easily transferable from one study to another. Additionally, while the results are striking, there may well be other factors not studied which also influence the MCID calculations. The reasons that certain factors influence the calculation of the MCID to a greater degree than do others is uncertain at this time and requires further study. The anchor-based method contains a degree of subjectivity in the choice of the anchor question representing the assessment domain and the selection of the threshold levels of success. What is successful for one individual may not be for another. The addition of the ROC curve allows the most precise assessment to be made of the aggregate responses and adds an important quantitative factor to the anchor method. The study also had strengths one of which was the concurrent, contemporaneous collection of the anchor questions and PROMs that reduced bias. The advantage of using the integrative approach is combining clinical relevance with quantitative rigor. Anchor questions provide clinical significance of the MCID, but their use alone may not take into account measurement variability; distribution-based methods account for measurement variability but can lack clinical relevance. By combining the two approaches, using the ROC curves to provide quantitative support to the clinical anchor questions, the integrative approach addresses the clinical and statistical aspects of the MCID calculation.[14,16]

Conclusion

There is not yet a single approach for establishing the MCID nor are there consensus values in MCIDs in the TKR population.[1,2,13] For example, the MCIDs of KOOS pain subscales have been reported to vary from 10 to 38, indicating that no one value can be applied universally.[22] Both intrinsic and extrinsic factors can have marked influences upon the precision and magnitude of the MCID that make comparisons among MCIDs difficult. Our data as well as the literature reviewed indicate the need for caution when interpreting the MCID and especially when it is used to compare studies or populations. Determining whether a treatment is clinically efficacious is of importance on both person- and policy-levels. On a person-level, quantifying outcomes is important for the identification of risk profiles and informs patient selection, perioperative risk mitigation efforts, and informed consent. On a policy-level, regulatory decisions, reimbursement, and best practice guidelines depend upon accurate quantitative outcome data. To enhance the use of the MCID as a comparator of outcomes in more than one study, the details of how the MCID was computed should be specified. Transparency of methods including characteristics of patient populations, methodological specificity including anchor questions, thresholds and sample sizes, time course of evaluation, and precisions of calculations should be provided. Nonetheless, because other factors may influence the MCID, comparisons among MCIDs from different studies should be made with caution. It is probably advisable for individual studies to calculate their own MCIDs and not rely on published values.

Funding

Disclosure

References

1. Hung M, Bounsanga J, Voss MW, Saltzman CL. Establishing minimum clinically important difference values for the patient-reported outcomes measurement information system physical function, hip disability and osteoarthritis outcome score for joint reconstruction, and knee injury and osteoarthritis outcome score for joint reconstruction in orthopaedics. *World J Orthop*. 2018;9(3):41–49. doi:10.5312/wjo.v9.i3.41

2. Cook CE. Clinimetrics corner: the Minimal Clinically Important Change Score (MCID): a necessary pretense. *J Man Manip Ther*. 2008;16(4):E82–3. doi:10.1179/jmt.2008.16.4.82E

3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–415. doi:10.1016/0197-2456(89)90005-6

4. Jevsevar D, Shea K, Cummins D, Murray J, Sanders J. Recent changes in the AAOS evidence-based clinical practice guidelines process. *J Bone Joint Surg Am*. 2014;96(20):1740–1741. doi:10.2106/jbjs.N.00658

5. Monticone M, Ferrante S, Salvaderi S, Motta L, Cerri C. Responsiveness and minimal important changes for the knee injury and osteoarthritis outcome score in subjects undergoing rehabilitation after total knee arthroplasty. *Am J Phys Med Rehabil*. 2013;92(10):864–870. doi:10.1097/PHM.0b013e31829f19d8

6. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? *Clin Orthop Relat Res*. 2017;475(1):149–157. doi:10.1007/s11999-016-4770-y

7. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. *Acta Orthop*. 2014;85(1):39–48. doi:10.3109/17453674.2013.867782

8. Lyman S, Lee YY, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res*. 2018;476(12):2432–2441. doi:10.1097/corr.0000000000000456

9. Kuo AC, Giori NJ, Bowe TR, et al. Comparing methods to determine the minimal clinically important differences in patient-reported outcome measures for veterans undergoing elective total hip or knee arthroplasty in veterans health administration hospitals. *JAMA Surg*. 2020;155(5):404–411. doi:10.1001/jamasurg.2020.0024

10. Hossain FS, Konan S, Patel S, Rodriguez-Merchan EC, Haddad FS. The assessment of outcome after total knee arthroplasty: are we there yet? *Bone Joint J*. 2015;97-B(1):3–9. doi:10.1302/0301-620X.97B1.34434

11. Çelik D, Çoban Ö, Kılıçoğlu Ö. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. *J Clin Epidemiol*. 2019;113:44–57. doi:10.1016/j.jclinepi.2019.04.017

12. Escobar A, García Pérez L, Herrera-Espiñeira C, et al. Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage*. 2013;21(12):2006–2012. doi:10.1016/j.joca.2013.09.009

13. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *J Orthop Surg Res*. 2015;10(1):24. doi:10.1186/s13018-014-0144-x

14. Hu G, Huang Q, Huang Z, Sun Z. [Methods to determine minimal clinically important difference]. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*. 2009;34(11):1058–1062. Chinese.

15. Haydel A, Guilbeau S, Roubion R, Leonardi C, Bronstone A, Dasa V. Achieving validated thresholds for clinically meaningful change on the knee injury and osteoarthritis outcome score after total knee arthroplasty: findings from a university-based orthopaedic tertiary care safety net practice. *J Am Acad Orthop Surg Glob Res Rev*. 2019;3(11):e00142. doi:10.5435/JAAOSGlobal-D-19-00142

16. de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res*. 2007;16(1):131–142. doi:10.1007/s11136-006-9109-9

17. Farrar JT, Young JP Jr., LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94(2):149–158. doi:10.1016/s0304-3959(01)00349-9

18. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–35. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3

19. Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthr Cartil*. 2007;15(3):273–280. doi:10.1016/j.joca.2006.09.001

20. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, medical outcomes study questionnaire short form 36, and pain scales. *Spine J*. 2008;8(6):968–974. doi:10.1016/j.spinee.2007.11.006

21. Wylde V, Penfold C, Rose A, Blom AW. Variability in long-term pain and function trajectories after total knee replacement: a cohort study. *Orthop Traumatol Surg Res*. 2019;105(7):1345–1350. doi:10.1016/j.otsr.2019.08.014

22. White DK, Master H. Patient-reported measures of physical function in knee osteoarthritis. *Rheum Dis Clin North Am*. 2016;42(2):239–252. doi:10.1016/j.rdc.2016.01.005