JSES International 7 (2023) 153–157

Contents lists available at ScienceDirect

JSES International

journal homepage: www.jsesinternational.org

Substantial clinical benefit values demonstrate a high degree of variability when stratified by time and geographic region

Ian J. Wellington, MD[a,*], Annabelle P. Davey, MD[a], Mark P. Cote, DPT[b], Benjamin C. Hawthorne, BS[a], Caitlin G. Dorsey, BS[a], Patrick M. Garvin, DO[a], James C. Messina, MD[a], Cory R. Hewitt, MD[a], Augustus D. Mazzocca, MS, MD[b]

[a]*Department of Orthopaedic Surgery, University of Connecticut School of Medicine, Farmington, CT, USA*
[b]*Massachusetts General Brigham Sports Medicine, Harvard Medical School, Boston, MA, USA*

ARTICLE INFO

Keywords:
Substantial clinical benefit
Arthroplasty
Outcomes
Statistics
Methodology
Shoulder
Reverse

Level of evidence: Basic Science Study; Validation of Outcome Instruments

Background: A Substantial Clinical Benefit (SCB) value is the amount of change in a patient-reported outcome measure required for a patient to feel they significantly improved from an intervention. Previously published SCB values are often cited by researchers when publishing outcomes data. Where these SCB values are set can have a large impact on the conclusions drawn from a study citing them. As such, the goal of this study was to determine the generalizability of SCB values for a procedure when stratified by time from surgery and geographic region.

Methods: A nationwide outcomes database was utilized to obtain preoperative, one-year, and two-year postoperative outcome measurements for patients who underwent anatomic total shoulder arthroplasty (TSA) or reverse TSA. The data were divided into three geographic regions: the South, the Midwest, and the West. An East region was not included due to its limited number of patients. SCB values were calculated for four outcomes measures: Single Assessment Numeric Evaluation score, American Shoulder Elbow Surgeons score, Visual Analog Scale, and Western Ontario Osteoarthritis of the Shoulder score. SCB values were calculated for each region, for each procedure, and at both one and two years postoperatively. To determine the variability of potential SCBs within each region, simulated datasets were created to determine a distribution of possible calculated SCBs.

Results: A total of 380 anatomic TSA patients and 543 reverse TSA patients were included for analysis. There was a high degree of variability of SCB values when stratified by procedure, time, and region. While some simulated datasets did produce homogenous SCB distributions among regions, some outcome measures demonstrated a large heterogeneity in distribution among regions, with concomitant large distributions of values within individual regions.

Conclusions: There is notable heterogeneity of SCB values when stratified by region or time. The current method of citing previously published SCB values for determining the efficacy of an intervention may be inappropriate. It is likely that this variability holds true in other areas of orthopedics.

Patient-reported outcome measures (PROMs) have become a mainstay of orthopedic research, providing quantitative measurements of patients' function and perceptions of their health. However, the wide variety of PROMs with different scoring systems can be difficult to interpret, and statistically significant results do not always correlate with clinically meaningful differences.[9] The minimal clinically important difference (MCID), which defines the change in PROM score that represents a clinically meaningful difference, was proposed to address this problem.[10] The MCID has since become a common value to reference when publishing on the efficacy of a procedure. The substantial clinical benefit (SCB) was subsequently developed to provide a more thorough framework for interpreting PROMs, with the MCID acting as a threshold for change that patients can perceive and the SCB acting as a target for defining clinical success.[8] These values have become increasingly prevalent in orthopedic literature in recent years, especially with regards to shoulder conditions.[18]

An SCB value is the amount of change after an intervention for a particular outcome measure that is needed to predict a patient reporting that they substantially benefited from the intervention.

Approval for this study was obtained from the University of Connecticut's Institutional Review Board. Approval #16-042-1.

*Corresponding author: Ian J. Wellington, MD, Department of Orthopaedic Surgery University of Connecticut School of Medicine, Farmington, CT, USA.
E-mail address: iwellington@uchc.edu (I.J. Wellington).

https://doi.org/10.1016/j.jseint.2022.10.003

For example, if an SCB for an outcome measure, such as the Single Assessment Numeric Evaluation (SANE), after reverse shoulder arthroplasty is determined to be 50, a patient who had an increase in that outcome measure of 50 or greater is likely to report that they substantially improved from the arthroplasty. Subsequently, if a patient had an improvement of less than 50, they are likely to report that they did not improve substantially.

The SCB value is determined for one PROM for a specific pathology and/or procedure via a concomitantly administered anchor question.[8,10] This anchor question seeks to provide a patient reported, clear response regarding the efficacy of the intervention. Researchers then use these values to determine the efficacy of an intervention. Often, investigators will determine the percentage of patients who crossed these previously published SCB thresholds and report this as a measurement of surgical success.[2,5,15] Utilization of SCBs in this manner relies on an assumption of high degrees of homogeneity between patient populations, and any variation in the SCB value can have a drastic effect on the outcomes of the study.

The SCB is likely influenced by factors beyond symptomatic state alone, as indicated by the high variability between SCB values for different procedures.[18] Despite this, the SCB is frequently determined in a population receiving care at one institution, often from one surgeon.[18] SCBs are calculated using both the change in outcome measure and, more importantly, their expectations for symptomatic improvement. As such, any factor impacting patients' expectations, whether they be cultural/societal influences or the expectations set by a surgeon preoperatively, would significantly affect an SCB value. There may be significant heterogeneity between populations presenting to different institutions, which limits the utility of broadly applying the SCB determined in one patient population as a threshold for clinical success. For example, an SCB value published by an institution in the West may not apply to a patient from the South.

Understanding the consistency of these values is critical, as the current state of outcomes research relies on the assumption that they are highly homogenous. If there were significant differences in SCB values based on time or location, it would call into question the current methods by which SCB values are employed in outcomes research. The purpose of the present study was to determine the SCB values for four outcome measures for patients undergoing either anatomic total shoulder arthroplasty (aTSA) or reverse total shoulder arthroplasty (rTSA) stratified by both time from surgery and region within the United States, and to see if there is significant heterogeneity between these values. We hypothesized that a high degree of variability will exist for SCBs when separated out by region and time.

Methods

This study uses data collected under institutional review board number 16-042-1 Total Shoulder Multi-Center Registry provided by an industry sponsor (Arthrex Total Shoulder Registry). Patients were consented before surgery to be followed for 10 years. A database of outcome measures collected for aTSA and rTSA comprised of patients from 12 states across the United States was utilized for this study. Four PROMs were used: SANE score, American Shoulder Elbow Surgeons (ASES) score, Visual Analog Scale (VAS), and Western Ontario Osteoarthritis of the Shoulder (WOOS) score. To be included in the SCB calculation for each of these scores, patients required preoperative scores, both one- and two-year postoperative scores, and an answer to the anchor question: "How well did the treatment meet your expectations with regard to reducing your pain level?". Patients who responded to this question that they either "exceeded" or "met" surgical expectations were classified as having an SCB and those who responded "did not meet" were classified as having not substantially benefitted. This methodology was based on prior studies calculating SCBs.[4,8,13,14] While some studies use Likert-based anchor questions and exclude those that report a "slight improvement" of their symptoms,[17] we elected to include those who "met" surgical expectations for pain relief in the substantial improvement group as patients likely expect a significant improvement in their symptomatic pain after a procedure. Thus, those who "met expectations" for surgical pain relief likely experienced a significant improvement in their symptoms.

The states in the dataset were divided into three regions for analysis: the Southeast (Kentucky, North Carolina, South Carolina, Texas, and Virginia), the Midwest (Ohio, Illinois, Michigan, and Missouri), and the West (Arizona, Colorado, and Oregon). Analyzing based on region allowed for larger sample sizes to be used for calculations, as well as allowed for the incorporation of states which had too few patients to conduct an SCB analysis on their own. Initially, an East region was to be included; however, there were too few patients in this region to perform complete statistical analyses. SCBs for each of the outcome measures were then calculated at 1- and 2-years from the index procedure.

Statistical analysis

Patients were grouped according to their response to the anchor question. Patients indicating that treatment met or exceed expectations were considered to have experienced an SCB and those who responded that the treatment did not meet expectations were considered to have failed to experience substantial benefit. For each of the four PROMs evaluated, the difference in score was calculated by subtracting the postoperative score at each time interval from the preoperative score. The receiver operating characteristics (ROC) type analysis was conducted to determine the difference in score that most accurately separated the two groups. The optimal SCB value was defined as difference in score that maximized Youden index, which is the sum of sensitivity and specificity. For example, in a group of 30 hypothetical patients of which half noted significant improvement after surgery and half did not, ROC-analysis plots each subject's change in an outcome measure of choice and aims to find a cutoff that most accurately separates out those who did or did not significantly improve. This value is deemed the SCB.

ROC-analysis does not address potential variability in SCB values determined from a single patient population. To determine variation in potential SCB scores calculated from an identical dataset, data were simulated using the sample size, mean, and standard deviation of the difference scores for each procedure, time point, and region. From each simulated dataset the optimal SCB values were determined as described above. This process was repeated 1000 times resulting in 1000 SCB values for each of the groupings. The mean and standard deviation of these simulated values were calculated as a measure of variation in SCB values. Analyses were performed using the cutpointr package for the R programming language.

Results

A total of 380 patients were included for analysis in the aTSA group, while 543 were included in the rTSA group. For aTSA there were 51 patients in the South group, 70 patients in the Midwest group, and 259 patients in the West group. For rTSA there were 92 in the South group, 141 in the Midwest group, and 310 in the West group. Table I demonstrates the SCB values calculated for SANE, ASES, VAS, and raw WOOS scores at one and two years postoperatively for each region and all data combined (Table I).

Table I
SCB values at 1 y and 2 y postoperatively stratified by region.

	Postoperative time	aTSA			
		South (n = 51)	Midwest (n = 70)	West (n = 259)	Combined (n = 380)
SANE	1 y	30	46	40	40
SANE	2 y	31	40	50	33
ASES	1 y	33	30	38	38
ASES	2 y	30	40	33	33
VAS	1 y	−4.00	−4.90	−3.69	−3.63
VAS	2 y	−1.76	−4.00	−3.36	−3.50
WOOS	1 y	−813	−6.73	−671	−674
WOOS	2 y	−977	−532	−1035	−1002
		rTSA			
		South (n = 92)	Midwest (n = 141)	West (n = 310)	Combined (n = 543)
SANE	1 y	1	26	33	38
SANE	2 y	59	51	31	31
ASES	1 y	32	21	28	31
ASES	2 y	24	24	27	27
VAS	1 y	−3.12	−3.02	−3.88	−3.68
VAS	2 y	−2.08	−3.73	−3.40	−3.73
WOOS	1 y	−541	−627	−763	−706
WOOS	2 y	−528	−797	−753	−814

SCB, Substantial Clinical Benefit; *aTSA*, anatomic total shoulder arthroplasty; *SANE*, Single Assessment Numeric Evaluation; *ASES*, American Shoulder Elbow Surgeons; *VAS*, Visual Analog Scale; *WOOS*, Western Ontario Osteoarthritis of the Shoulder; *rTSA*, reverse total shoulder arthroplasty

The mean SCBs were then calculated for 1000 simulated datasets from each region to determine the potential variability of SCB values within region. The distribution of these values was then plotted (Fig. 1). These values were grouped by time from surgery, region, procedure, and PROM. While some of these distributions were consistent between regions, other outcome measures had a high degree of variability. Some regional distributions of SCBs did not overlap at all, such as the VAS and SCB distributions at 1 year for rTSA between the Midwest and West groups. Additionally, there was a significant amount of heterogeneity of distribution sizes among regions. For example, the WOOS SCB distribution at 1 year for rTSA had a large distribution in the Midwest group, while there was a small distribution seen in the South group.

Discussion

Our study found notable differences in the observed SCB values for both rTSA and aTSAs based on time from index procedure as well as geographical region. When these differences were explored further with data simulation, there was considerable variation in the SCB suggesting that a wide range of potential values are possible for a given outcome measure and patient population. These findings illustrate the inherent heterogeneity of SCB values for each outcome measure and demonstrate the lack of generalizability when using a single SCB value for outcome measures across varying patient populations.

These results are important as the use of a previously reported SCB values as a surrogate for identifying patients who have experienced substantial benefit has become more prevalent in shoulder arthroplasty research,[2,5,7,16] and the orthopedic literature in general.[1,3,6,11] While this well-intended approach allows the results to be viewed from the perspective of clinical significance, it is reliant upon the stability of the SCB as it relates to the outcome measure and patient population of interest. If the SCB being utilized is considerably different than the true SCB of the patient group under study, the interpretation of these results could be misleading. For instance, an SCB that's too large would result in an intervention appearing less efficacious, while an SCB that's too small would make it appear more effective than it truly is. Interestingly, a systematic review by Kolin et al found similar heterogeneity previously reported MCID values for shoulder arthroplasty outcome measurements.[12]

Table I demonstrates the large disparities in SCB values between regions. While certain values were consistent, such as those for rTSA ASES scores at 1 year (range: 24-38), other values like aTSA WOOS scores at 2 years (range: −373 to −1035) were highly variable. Figure 1 shows that some regions had no overlap in potential SCB values. For example, the Midwest and West regions had no overlap in their potential SCB values for VAS and ASES pain scores for rTSA at 1 year. This suggests that if a researcher in the Midwest published an SCB value for ASES for rTSA at 1 year, it would have no applicability to a researcher in the West using this value. While some potential SCB values were similar among regions, such as the WOOS for rTSA at 2 years, this was not consistently the case for other outcome measures at other time points.

Figure 1 demonstrated that some SCB ranges had large standard deviations, such as rTSA WOOS at 1 year in the Midwest, which suggests that even when drawn from the same patient population there is a high degree of variability in SCB values. The simulated dataset for rTSA in the Midwest at 1 year produced a mean WOOS SCB value of −498 with a standard deviation of 238. This suggests that a researcher looking to establish an SCB for WOOS in this region at 2 years might report a value between −736 and −260. This almost 476-point disparity between potential reported values would have a large impact on the results of future outcomes research quoting this value, even if the future research was conducted with the same patient pool.

It is important to note that the use of regions as a comparator in this study was arbitrary. The dataset could have just as well been divided into East coast vs. West coast, North vs. South, or even states with an "A" in their name vs. those without. The goal of the study was not to establish regional or temporal SCBs for rTSA and TSA, but rather to show the high degree of lability that is associated with these values.

The SCB values for each region in this study were not consistently homogenous. There are many potential explanations for this heterogeneity, such as differences in culture, socioeconomic status, or even simply the preoperative expectations set by the treating surgeons. Additionally, our study showed that SCB values fluctuated with time. While reported SCB values may specify the

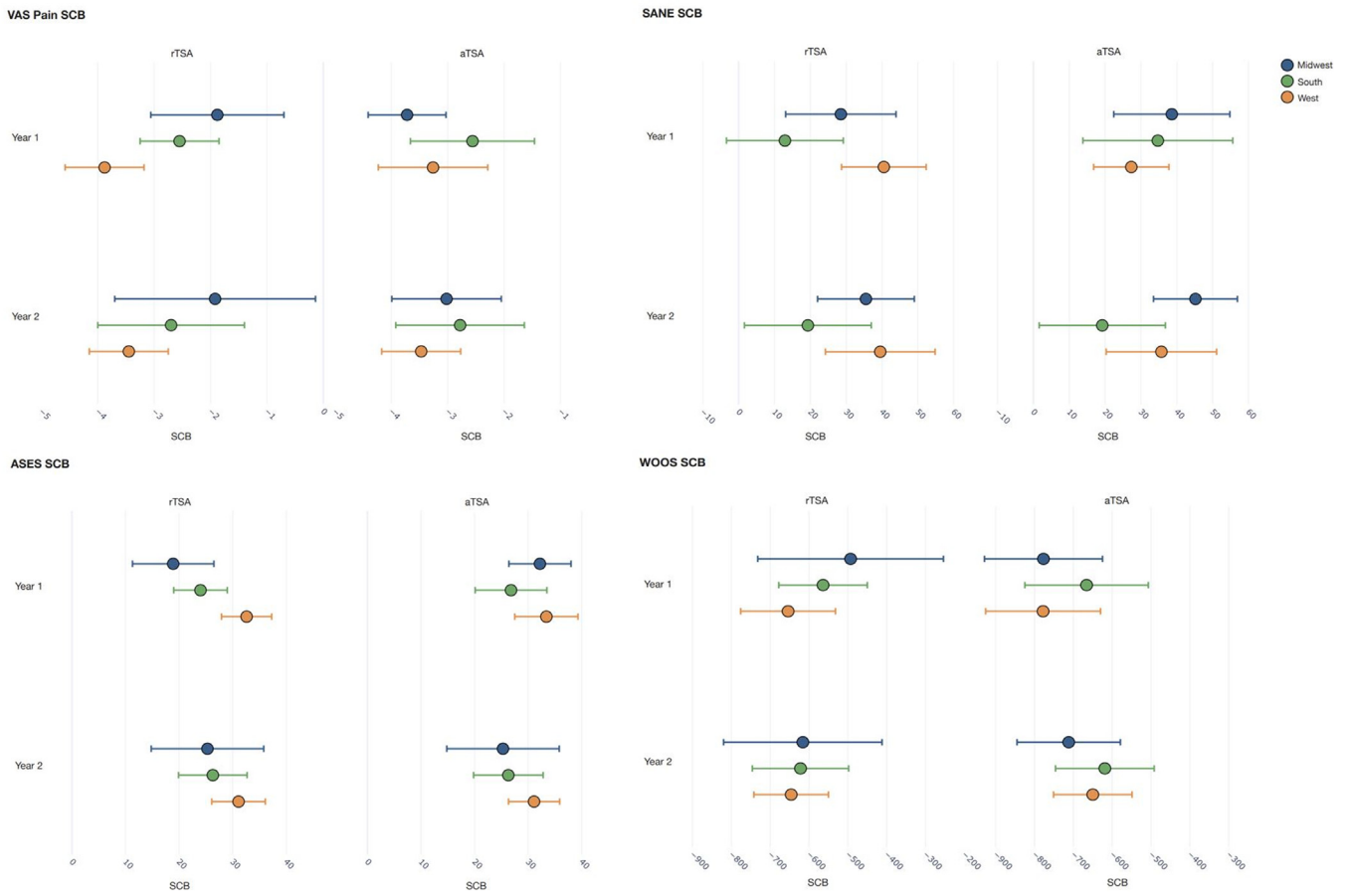

Figure 1 Mean SCB values for VAS Pain, SANE, ASES, and WOOS scores calculated at one y and two y postoperatively from 1000 simulated datasets. The error lines indicate the range of potential SCB values calculated from these datasets. *SCB*, Substantial Clinical Benefit; *VAS*, Visual Analog Scale; *SANE*, Single Assessment Numeric Evaluation; *ASES*, American Shoulder Elbow Surgeons; *WOOS*, Western Ontario Osteoarthritis of the Shoulder.

follow-up time at which they were calculated, this is not always the case.[4] Even if an SCB value is reported at a particular follow-up interval, researchers using previously established SCB values may apply a value calculated at 2 years postoperatively for a study evaluating clinical efficacy at 1 year. Our results suggest that this usage of SCBs in this manner may be suboptimal given the variations seen at one year and two years postoperatively.

While the results of this study call into question a foundational methodology of modern clinical outcomes research, there are simple ways for future outcomes research projects to avoid this issue of heterogeneity among SCB values. One option would be to calculate SCB values on an individual study basis. This can be accomplished by incorporating an anchor question into patient questionnaires from which an SCB can be calculated. Even more simply, SCBs can be done away with entirely. Instead, studies can report the percentage of patients who respond with a significant improvement to an anchor question. This would allow researchers to report the percentage of patients who substantially benefitted from an intervention calculating an SCB in the first place (ie 94% of patients reported significant improvement of their symptoms postoperatively).

While this study was successful in its aims, it is not without limitations. Firstly, this study did not investigate the level of heterogeneity of other outcomes thresholds such as the MCID or the Patient Acceptable Symptomatic State values. These were not addressed in this study as they require different anchor questions for their calculation which were not available in the utilized dataset. The MCID values have been shown to be highly variable between previous publications.[12] It is likely that Patient Acceptable Symptomatic State values are variable as well; however, further study is warranted to make this conclusion. This study included those who “met” surgical expectations for pain relief in the substantially improved group with the assumption that patients’ surgical expectations are likely a significant improvement in their preoperative pain. While this would not accurately characterize patients that expected only mild symptomatic improvements from surgery, we believe this would be a minority of patients. Additionally, this study was only conducted using data from aTSA and rTSA. It was limited to these procedures due to the availability of data suitable for this kind of analysis; however, future research is needed to validate if this trend is seen in other orthopedic procedures. Finally, as these are PROMs, there is always an influence of human error such as patients misunderstanding of outcome measure questions and responses or errors in recording of these responses.

Conclusions

SCB values for aTSA and rTSA were inconsistent when stratified by time and region. This suggests that the practice of applying previously established SCB values to new datasets may be suboptimal.

Disclaimers:

Funding: The authors report no funding for this project.
Conflicts of interest: This study utilized a database maintained by Arthrex inc. but Arthrex was not involved with the study design, data analysis, or manuscript preparation. ADM reports research grants from Arthrex, Inc. and is a consultant for Arthrex, Inc. The other authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. Beletsky A, Gowd AK, Liu JN, Manderle BJ, Yanke AB, Forsythe B, et al. Time to achievement of clinically significant outcomes after isolated arthroscopic partial meniscectomy: a multivariate analysis. Arthrosc Sport Med Rehabil 2020;2:e723-33. https://doi.org/10.1016/J.ASMR.2020.06.002.
2. Berthold DP, Muench LN, Kia C, Ziegler CG, Laurencin SJ, Witmer D, et al. Surgeon and patient upper extremity dominance does not influence clinical outcomes after total shoulder arthroplasty. Orthop J Sports Med 2020;8:2325967120932106. https://doi.org/10.1177/2325967120932106.
3. Campbell A, Emara AK, Klika A, Piuzzi NS. Does implant selection affect patient-reported outcome measures after primary total hip arthroplasty? J Bone Joint Surg Am 2021;103:2306-17. https://doi.org/10.2106/JBJS.20.01931.
4. Cohn MR, Kunze KN, Polce EM, Nemsick M, Garrigues GE, Forsythe B, et al. Establishing clinically significant outcome thresholds for the Single Assessment Numeric Evaluation 2 years following total shoulder arthroplasty. J Shoulder Elbow Surg 2021;30:e137-46. https://doi.org/10.1016/J.JSE.2020.07.011.
5. Cusano A, Kanski G, Uyeki C, Adams K, Cote MP, Muench LN, et al. Outcomes of reverse shoulder arthroplasty following failed superior capsular reconstruction. J Shoulder Elbow Surg 2022;31:1426-35. https://doi.org/10.1016/j.jse.2021.12.044.
6. Djurasovic M, Glassman SD, Dimar JR, Howard JM, Bratcher KR, Carreon LY. Does fusion status correlate with patient outcomes in lumbar spinal fusion? Spine (Phila Pa 1976) 2011;36:404-9. https://doi.org/10.1097/BRS.0B013E3181FDE2C4.
7. Drager J, Polce EM, Fu M, Nemsick M, Huddleston HP, Forsythe B, et al. Patients undergoing anatomic total shoulder arthroplasty achieve clinically significant outcomes faster than those undergoing reverse shoulder arthroplasty. J Shoulder Elbow Surg 2021;30:2523-32. https://doi.org/10.1016/J.JSE.2021.02.015.
8. Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg 2008;90:1839-47. https://doi.org/10.2106/JBJS.G.01095.
9. Harris JD, Brand JC, Cote MP, Faucett SC, Dhawan A. Research pearls: the significance of statistics and perils of pooling. Part 1: clinical versus statistical significance. Arthroscopy 2017;33:1102-12. https://doi.org/10.1016/j.arthro.2017.01.053.
10. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
11. Jung JM, Chung CK, Kim CH, Yang SH, Won YI, Choi Y. Effects of total psoas area index on surgical outcomes of single-level lateral lumbar interbody fusion. World Neurosurg 2021;154:e838-45. https://doi.org/10.1016/J.WNEU.2021.08.031.
12. Kolin DA, Moverman MA, Pagani NR, Puzzitiello RN, Dubin J, Menendez ME, et al. Substantial inconsistency and variability exists among minimum clinically important differences for shoulder arthroplasty outcomes: a systematic review. Clin Orthop Relat Res 2022;480:1371-83. https://doi.org/10.1097/CORR.0000000000002164.
13. Lyman S, Lee Y-Y, 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;476:2432. https://doi.org/10.1097/CORR.0000000000000456.
14. Martin RL, Kivlan BR, Christoforetti JJ, Wolff AB, Nho SJ, Salvo JP Jr, et al. Minimal clinically important difference and substantial clinical benefit values for the 12-item international hip outcome tool. Arthroscopy 2019;35:411-6. https://doi.org/10.1016/j.arthro.2018.09.028.
15. Muench LN, Kia C, Jerliu A, Williams AA, Berthold DP, Cote MP, et al. Clinical outcomes following biologically enhanced patch augmentation repair as a salvage procedure for revision massive rotator cuff tears. Arthroscopy 2020;36:1542-51. https://doi.org/10.1016/j.arthro.2020.02.006.
16. Routman HD, Simovitch RW, Wright TW, Flurin PH, Zuckerman JD, Roche CP. Acromial and scapular fractures after reverse total shoulder arthroplasty with a medialized glenoid and lateralized humeral implant: an analysis of outcomes and risk factors. J Bone Joint Surg Am 2020;102:1724-33. https://doi.org/10.2106/JBJS.19.00724.
17. Simovitch R, Flurin PH, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the substantial clinical benefit. J Shoulder Elbow Surg 2018;27:903-11. https://doi.org/10.1016/j.jse.2017.12.014.
18. Su F, Allahabadi S, Bongbong DN, Feeley BT, Lansdown DA. Minimal clinically important difference, substantial clinical benefit, and patient acceptable symptom state of outcome measures relating to shoulder pathology and surgery: a systematic review. Curr Rev Musculoskelet Med 2021;14:27-46. https://doi.org/10.1007/s12178-020-09684-2.