RESEARCH ARTICLE

Can Patients Accurately Recall their Preoperative Pain and Functional Scores Following Rotator Cuff Repair and Total Shoulder Arthroplasty?

Michael J. Gutman, MD[1]; Manan S. Patel, MD[1]; Matthew T. Weintraub, BS[2]; Arjun Singh, BS[1]; Eric M. Padegimas, MD[1]; Joseph A. Abboud, MD[1]; Surena Namdari, MD[1]

Research performed at the Rothman Orthopaedic Institute, Philadelphia, PA, USA

Received: 02 November 2021 *Accepted: 19 December 2021*

Abstract

Objectives: Accurate analysis of preoperative shoulder pain and function is important for understanding treatment efficacy and producing high-quality research. Oftentimes, preoperative patient-reported outcomes (PROs) are missing. Therefore the accuracy of recalled preoperative PROs may be important. We investigate the ability of patients who underwent rotator cuff repair (RCR) or shoulder arthroplasty (TSA) to recall their preoperative PROs.

Methods: We identified 145 patients who underwent either RCR or TSA and had preoperative PROs. All patients completed the ASES, SANE, SST, and VAS surveys within 3 months prior to surgery. Patients were contacted between one and four years after surgery and asked to recall their baseline pain and shoulder function prior to surgery. The mean difference was calculated by determining the difference between the mean recalled score and the mean actual score. Intraobserver reliability analysis was performed, comparing recall and actual score for each using the 2-way mixed-effects intraclass correlation coefficient (ICC) model. The ICC values > 0.75 were considered excellent, values between 0.4 and 0.75 were considered moderate, and values of < 0.4 demonstrated a weak agreement.

Results: For patients who underwent RCR, the mean differences between actual and recalled ASES, SANE, SST and VAS pain were 6.3 (*P=0.004*), 2.0 (*P=0.155*), -0.04 (*P=0.625*) and - 1.0 (*P<0.001*), respectively. In patients who underwent TSA, the mean differences between actual and recalled ASES, SANE, SST and VAS pain were 4.5 (*P =0.038*), -3.9 (*P=0.262*), -1.2 (*P=0.001*) and -1.5 (*P<0.001*), respectively. ASES, SST, and VAS show moderate reliability, and SANE reliability was weak in both RCR and TSA populations. Patients had a tendency to recall higher pain scores than actual preoperative pain scores.

Conclusion: In patients who underwent RCR or TSA, there was too much variability between individual patient's ability to accurately recall preoperative pain and function to reliably use recall data for research purposes.

Level of evidence: IV

Keywords: ASES, Recall, Rotator Cuff Repair, SANE, SST, Total shoulder arthroplasty VAS

Introduction

In the clinical and research setting, recall of pain and function prior to intervention are useful to assess the impact of treatment. [1-3] In orthopedic research, the most common way of determining patient outcomes is by comparing outcomes instruments given to patients prior to treatment the same instruments collected postoperatively. If preoperative patient-reported outcomes are not available for a population,

Corresponding Author: Surena Namdari, Department of Orthopaedic Surgery, Rothman Orthopaedic Institute at Thomas Jefferson University Hospital, Philadelphia, PA, USA
Email: surena.namdari@rothmanortho.com

THE ONLINE VERSION OF THIS ARTICLE ABJS.MUMS.AC.IR

retrospective studies may rely on a patient's ability to recall pretreatment pain and functional status to assess the intervention's effectiveness. Though recall patient-reported questionnaires are convenient, reliance on patients' memory can confound an intervention's perceived success due to the introduction of recall bias. Previous studies in the knee, hip, and hand populations have studied patients' ability to postoperatively recall preoperative pain and functional status. [4-8]

The purpose of this study was to evaluate if patients who underwent rotator cuff repair (RCR) or total shoulder arthroplasty (TSA) could postoperatively recall their preoperative function and pain accurately. We hypothesized that patients would be able to accurately remember their baseline preoperative pain and shoulder function; however, the ability to recall preoperative scores would diminish with increasing time from surgery.

Materials and Methods

Study Design

Following institutional review board (IRB) approval, patients who underwent RCR or TSA from January 2015 to June 2018 were queried in June 2019. Patients included in the study had ASES, SANE, SST, and VAS preoperative scores completed within 3 months of surgery and were over 18 years old preoperatively. Patients were excluded if the initial surgery was non-elective, a revision surgery, or if the patient had subsequent shoulder surgery, trauma, or infection. Patients who underwent revision surgery or subsequent surgery were excluded due to concern for confusion between recall of pain and function prior to index procedure versus revision procedure. Cases with trauma and infection were excluded since previous studies have demonstrated that recall of pain is not shown to be reliable in the setting of acute trauma or infection. [3,9] All cognitively impaired patients lacked English fluency or were unable to provide verbal consent were excluded.

Between June 2019 and November 2019, all eligible patients were contacted using a standardized phone script at a minimum 1-year follow-up. Recalled preoperative shoulder pain and functional scores via the Visual Analog Scale (VAS) pain, American Shoulder and Elbow Surgeons (ASES), the Single Assessment Numerical Evaluation (SANE), and Simple Shoulder Test (SST) were collected. Previous studies have demonstrated ASES, SANE, SST, and VAS pain are reliable and validated questionnaires to evaluate shoulder pain and function. [10-15]

When calling patients, the research staff members were blinded to the individual preoperative scores. The time from the preoperative survey to the time of the phone encounter determined the time of the recall period. Additionally, demographic data such as age and gender were recorded, as well as the past medical history of depression, anxiety, stroke, anxiety, transient ischemic attack (TIA), and smoking history.

Statistical Analysis

Analysis of recall scores and actual scores for all functional and pain instruments was carried out utilizing a nonparametric Wilcoxon test. The mean difference was calculated by determining the difference between the mean recalled score and the mean actual score. Additionally, the difference between the ASES, SANE, SST, and VAS scores for each individual was calculated. Descriptive statistics were performed on the individual differences in recall scores. Patients' recall scores were compared based on demographic factors and past medical history of smoking, anxiety, depression, TIA, stroke, and depression. Intraobserver reliability analysis was performed, comparing recall and actual score for each using the 2-way mixed-effects intraclass correlation coefficient (ICC) model. ICC values > 0.75 were considered excellent, values between 0.4 and 0.75 were considered moderate, and values of < 0.4 demonstrated a weak agreement. [14]

Correlational analysis was carried out via Spearman's Rho to evaluate the effect of time on patients' ability to recall. All statistical analysis was carried out on Statistical Package for the Social Science (SPSS) version 26 (IBM Corp., Armonk, NY). Statistical significance was defined at the *P-value* of <0.05 for all outputs.

Results

Demographics

After applying exclusion and inclusion criteria, 145 patients (68 RCR patients and 77 TSA patients) were included in the analysis. The mean age was 64.7 (range 25 to 88), 71 (49.0%) were males, and the mean time to recall was 2.7±0.9 years (1.0-4.4).

Patient Recall

In the RCR population, the mean differences between actual and recalled ASES, SANE, SST and VAS pain were 6.3 (*P=0.004*), 2.0 (*P=0.155*), -0.04 (*P=0.625*) and -1.0 (*P<0.001*), respectively [Table 1]. When each person's individual recall score was compared to their actual reported pre-operative score, the average delta in PRO score for ASES was 15.9±11.6 (0.1-55.0), SANE was 20.7±20.0 (0.0-90.0), SST was 2.4±2.1 (0-9) yes responses, and VAS pain was 1.8±1.6 (0-7) [Figure 1A]. ICC analysis demonstrated that patients had a moderate reliability in recalling ASES (ICC= 0.420) , SST (ICC=0.388), and VAS (ICC= 0.435), and weak reliability in recalling SANE(ICC=0.242) scores.

In the TSA population, the mean differences between actual and recalled ASES, SANE, SST and VAS pain were 4.5 (*P=0.038*), -3.9 (*P=0.262*), -1.2 (*P=0.001*) and -1.5 (*P<0.001*), respectively [Table 1]. When each person's individual recall score was compared to their actual reported pre-operative score, the average delta in PRO score for ASES was 14.4±12.1 (0.2-51.1), SANE was 21.0±16.6 (0.25-70.9), SST was 2.3±1.7 (0-8) yes responses, and VAS pain was 2.2±1.8 (0-8) [Figure 1B]. ICC analysis demonstrated that patients had a moderate reliability in recalling ASES (ICC=0.456), SST (ICC=0.451), and VAS (ICC= 0.389), and weak reliability in recalling SANE (ICC=0.161) scores.

Additionally, patients tended to overestimate the amount of pain they had prior to surgery in both RCR and TSA (*P<0.001*).

Table 1. Recall score and actual scores

	Survey	Actual	Re-call	Mean Difference	*P-value*	ICC
RCR	ASES	41.9±19.3	35.6±16.1	6.3±18.7	0.004	0.42
	SANE	36.6±23.6	34.6±23.2	2.0±28.8	0.155	0.242
	SST	4.2±3.1	4.3±2.6	-0.04±3.2	0.625	0.388
	VAS	6.0±2.3	7.0±2.0	-1.0±2.2	<0.001	0.435
TSA	ASES	36.9±17.2	32.4±18.3	4.5±18.3	0.038	0.456
	SANE	28.4±18.2	32.4±22.7	-3.9±26.6	0.262	0.161
	SST	2.8±2.4	4.0±2.7	-1.2±2.6	0.001	0.451
	VAS	5.9±2.6	7.4±2.0	-1.5±2.4	<0.001	0.389

ASES, American shoulder and elbow surgeons score; SANE, Single assessment numeric evaluation score; SST, Simple shoulder test score; VAS, Visual analog score for pain

RCR Patients

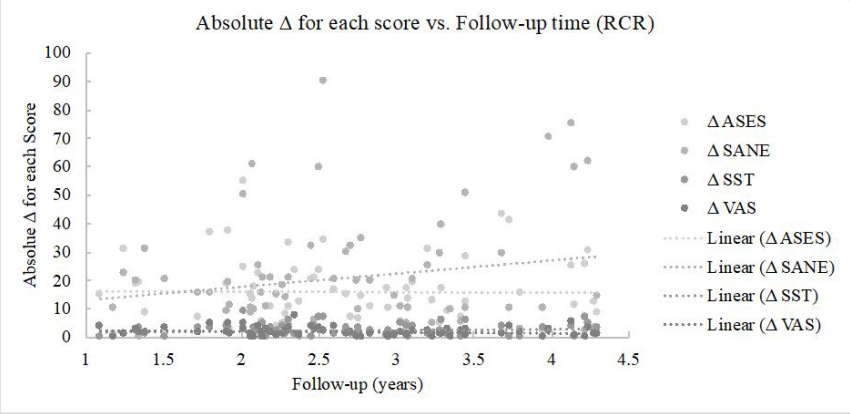

Figure 1A. Correlation between Actual Baseline Scores and Recall Scores over time of RCR Patients.

TSA Patients

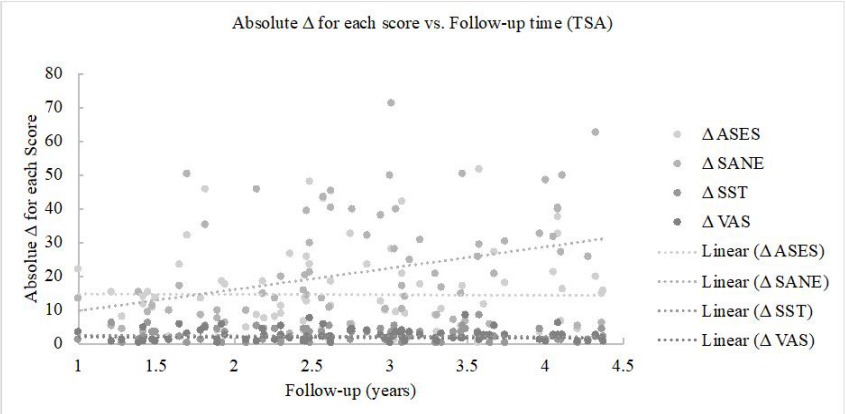

Figure 1B. Correlation between Actual Baseline Scores and Recall Scores over time in TSA Patients.

Follow-up time for reliable recall

The average follow-up for RCR was 2.6 ± 0.8 years (1.1-4.3 years) and for TSA was 2.8 ± 0.9 years (1.0-4.4 years) (*P=0.487*). Correlational analysis was performed to evaluate if absolute differences between actual and recall scores were affected by the length of time since the surgery in question. None of the scores were found to be correlated with the length of time for RCR and TSA patients (*P >0.05*). Additionally, analysis was performed by grouping absolute differences in scores into five categories based on time of follow-up (i.e. 1-2 years, 2-3 years, 3-4 years, and greater than 4 years). None of the scores were found to be correlated by length of time for RCR and TSA patients (*P >0.05*). Only SANE for TSA was found to be significantly different between 1-2 years and 3-4 years (*P=0.005*).

Risk factors for poor recall

Non-parametric analysis was undertaken to evaluate for risk factors that would predict poor patient recall. For RCR and TSA, age, gender, smoking, mental health, and medical comorbidities were not associated with poor recall (*P >0.05*).

Discussion

In the present study, we attempted to determine the ability of patients to recall their pain and functional limitation before RCR and TSA. This study demonstrates that RCR and TSA patients are unable to accurately recall their ASES, SANE, SST, and VAS scores more than 1 year after surgery. In both the RCR and TSA populations, on the individual patient level, recalled scores often differed largely from actual scores and are subject to recall bias. This was demonstrated in the intraobserver reliability analysis showing that patients after RCR and TSA could only moderately recall their ASES, SST, and VAS scores, and could only weakly recall their SANE scores. The highest ICC value in our study was 0.456, indicating that at best, the recall was at the low end of moderate. Furthermore, the correlation analysis demonstrated that patient recall after 1 year of surgery was consistently inaccurate. This study is the first to determine that recall should not be used for any of the most commonly used PROM (ASES, SANE, and SST) after RCR and TSA.

Previous studies in the knee, hip, spine, and hand populations compared recall after treatment and demonstrated conflicting results on patients' ability to recall their prior function. [4-7,16] Marsh et al demonstrated that at 3 weeks after hip arthroplasty, patients had excellent ability to recall preoperative functional status when asked to complete disease-specific questionnaires. [6] Lingard et al analyzed patients' ability to recall preoperative function three months after total knee arthroplasty, with results showing only moderate ability to recall the preoperative function. [4] In contrast to our study, past recall studies have focused on recall at a single point in time rather than over a large time interval. [4,6,16-18] Additionally, many of these studies analyzed recall bias within a few months after the intervention, with only a few studies looking at distant outcomes after 2 years or more. [8,19] In studies looking at recall more than 1 year after surgery, patient's ability to accurately recall their function appears mixed.

Similar to previous studies, our study found that ASES can not be recalled after RCR and TSA. 19,20 Recently, Gotlin et al. found that after a minimum of 1 year postoperatively, patients were unable to recall their ASES scores after RCR. Additionally, Lowe et al. found that patients overestimated their recalled preoperative pain levels after 6 weeks postoperative from TSA. [20] They noted that since VAS pain comprises 50% of the ASES scores, recalled ASES scores should be avoided after 6 weeks postoperatively. Similarly, in our study, RCR and TSA patients overestimated their recalled preoperative pain levels. As a result, ASES scores were significantly different from the actual preoperative ASES scores for both RCR (*P=0.001*) and TSA (*P=0.037*). Interestingly, in our study, the mean difference between the recalled and actual preoperative ASES scores was very small in both the RCR and TSA populations. However, when the absolute difference between a recalled score and an actual score was calculated for each patient, the average delta was significantly greater. Furthermore, the intraobserver reliability analysis demonstrated that patients could only moderately recall their ASES after both RCR and TSA, and that recall was not reliable in these populations. As a result of our findings, our authors recommend that intraobserver reliability analysis be performed when trying to assess recall ability rather than mean differences. [8,20] Additionally, in our study, after a minimum of 12 months postoperatively, the mean recalled ASES score was 35.6 after RCR and 32.4 after TSA. Similarly, Lowe, et al. found that at 1 year postoperatively from TSA, the mean recalled ASES was 25.3, and Gotlin et al. found that after a minimum of 12 months postoperatively after RCR the mean recalled ASES score was 30.7. [19-20] Our authors suspect that patients tend to recall ASES scores between 25-35 after both RCR and TSA after 1 year postoperatively, regardless of the patients' actual preoperative function. The specific nature of many of the questions in the ASES survey (such as: "is it difficult for you to throw a ball overhand?") may contribute to patients' decreased ability to recall their shoulder function. Additionally, our study's increased length of time likely contributed to decreased ability to recall their preoperative pain and function.

In our study, patients were unable to accurately recall SST or SANE functional scores. The intraobserver reliability analysis demonstrated that patients had the worst recall of SANE after both RCR and TSA, compared to the other shoulder PROM (i.e: SST and ASES). Since the SANE survey is a single question regarding perceived shoulder function at a specific moment in time, the SANE score is likely to be variable on a day-to-day basis. As a result, this likely contributed to patients in both the RCR and TSA populations having a weaker ability to recall their SANE score compared to their ASES and SST scores.

Just as patient-perceived shoulder function (SANE score) fluctuates at any given moment, patient-perceived pain at any moment also fluctuates greatly. As a result, patient pain levels are very difficult to recall. Previous studies have shown that patients have experienced a better ability to recall function rather than pain. [6,21]

Furthermore, patients tend to exaggerate preoperative pain. [4,21,22] Similarly, this was seen in our study in patients after both RCR and TSA where patients' recalled pain was significantly elevated from their actual preoperative pain levels. While literature shows conflicting results regarding the effect of time on recall of pain, our study's longer duration until time to recall may have contributed to the inaccurate recall scores of preoperative pain. [7,8,21,22]

This study had several limitations. While recall scores for ASES, SANE, and SST were investigated in the shoulder patients who underwent operative treatment, our results may not apply to patients who did not undergo surgical intervention. We did not assess recall at earlier time points; so, shorter durations of recall may lead to more accurate results.

In patients who underwent RCR or TSA, there is too much variability between individual patients' ability to accurately recall preoperative pain and function to reliably use recall data for research purposes.

Michael J. Gutman MD[1]
Manan S. Patel MD[1]
Matthew T. Weintraub BS[2]
Arjun Singh BS[1]
Eric M. Padegimas MD[1]
Joseph A. Abboud MD[1]
Surena Namdari MD[1]
1 Department of Orthopaedic Surgery, Rothman Orthopaedic Institute at Thomas Jefferson University Hospital, Philadelphia, PA, USA
2 Rutgers Robert Wood Johnson School of Medicine, New Brunswick, NJ, USA

References

1. Broderick JE, Stone AA, Calvanese P, Schwartz JE, Turk DC. Recalled pain ratings: A complex and poorly defined task. J Pain. 2006; 7(2):142-149
2. Gavaruzzi T, Carnaghi A, Lotto L, Rumiati R, Meggiato T, Polato F, et al. Recalling pain experienced during a colonoscopy: Pain expectation and variability. Br J Health Psychol. 2010; 15:253-264
3. Singer AJ, Kowalska A, Thode J. Ability of patients to accurately recall the severity of acute painful events. Acad Emerg Med. 2001; 8(3):292-295
4. Lingard EA, Wright EA, Sledge CB. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. J Bone Joint Surg AM. 2001;83(8):1149–1156
5. Bryant D, Norman G, Stratford P, Marx RG, Walter SD, Guyatt G. Patients undergoing knee surgery provided accurate ratings of preoperative quality of life and function 2 weeks after surgery. J Clin Epidemiol. 2006; 59(9):984-993
6. Marsh J, Bryant D, MacDonald SJ. Older patients can accurately recall their preoperative health status six weeks following total hip arthroplasty. J Bone Joint Surg AM. 2009;91(12):2827–2837
7. Mancuso CA, Charlson ME. Does recollection error threaten the validity of cross-sectional studies of effectiveness? Med Care. 1995; 33(4 Suppl):AS77-88
8. Stepan JG, London DA, Boyer MI, Calfee RP. Accuracy of patient recall of hand and elbow disability on the quickdash questionnaire over a two-year period. J Bone Joint Surg AM. 2013; 95(22):e17
9. Easton RM, Bendinelli C, Sisak K, Enninghorst N, Regan D, Evans J, et al. Recalled pain scores are not reliable after acute trauma. Injury. 2012; 43(7):1029-1032
10. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg AM. 2005; 87(9):2006-2011
11. Roy JS, Macdermid JC, Woodhouse LJ. Measuring shoulder function: A systematic review of four questionnaires. Arthritis Care Res. 2009; 61(5):623-623
12. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: Reliability, validity, and responsiveness. J Shoulder Elbow Surg. 2002;11(6):587-594
13. Godfrey J, Hamman R, Lowenstein S, Briggs K, Kocher M. Reliability, validity, and responsiveness of the simple shoulder test: Psychometric properties by age and injury type. J Shoulder Elbow Surg. 2007;16(3):260-267
14. Williams GN, Gangel TJ, Arciero RA, Uhorchak JM, Taylor DC. Comparison of the single assessment numeric evaluation method and two shoulder rating scales. Outcomes measures after shoulder surgery. Am J Sports Med. 1999; 27(2):214-221.
15. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. World J Orthop. 2014; 5(5):623-633
16. Lindsay GM, Niven KA, Brodie EE, Gaw A, Belcher PR. Accuracy of patient recall of preoperative symptom severity (angina and breathlessness) at one year following aorta-coronary artery bypass grafting. J Clin Nurs. 2009; 18(3):418-425
17. Wilson J, Baker P, Rangan A. Is retrospective application of the Oxford Shoulder Score valid? J Shoulder Elbow Surg. 2009;18(4):577-580
18. Howell J, Xu M, Duncan CP, Masri BA, Garbuz DS. A Comparison Between Patient Recall and Concurrent Measurement of Preoperative Quality of Life

Outcome in Total Hip Arthroplasty. J Arthroplasty. 2008;23(6):843–849.
19. Gotlin MJ, Kingery MT, Baron SL, McCafferty J, Jazrawi LM, Meislin RJ. Recall Bias in Retrospective Assessment of Preoperative Patient-Reported American Shoulder and Elbow Surgeons Scores in Arthroscopic Rotator Cuff Repair Surgery. Am J Sports Med. 2020; 48(6):1471-1475
20. Lowe JT, Li X, Fasulo SM, Testa EJ, Jawa A. Patients recall worse preoperative pain after shoulder arthroplasty than originally reported: a study of recall accuracy using the American Shoulder and Elbow Surgeons score. J Shoulder Elbow Surg. 2017;26(3):506-511
21. Pellisé F, Vidal X, Hernández A, Cedraschi C, Bagó J, Villanueva C. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. Spine (Phila Pa 1976). 2005; 30(3):365-8.
22. Erskine A, Morley S, Pearce S. Memory for pain: a review. Pain. 1990;41(3):255-265