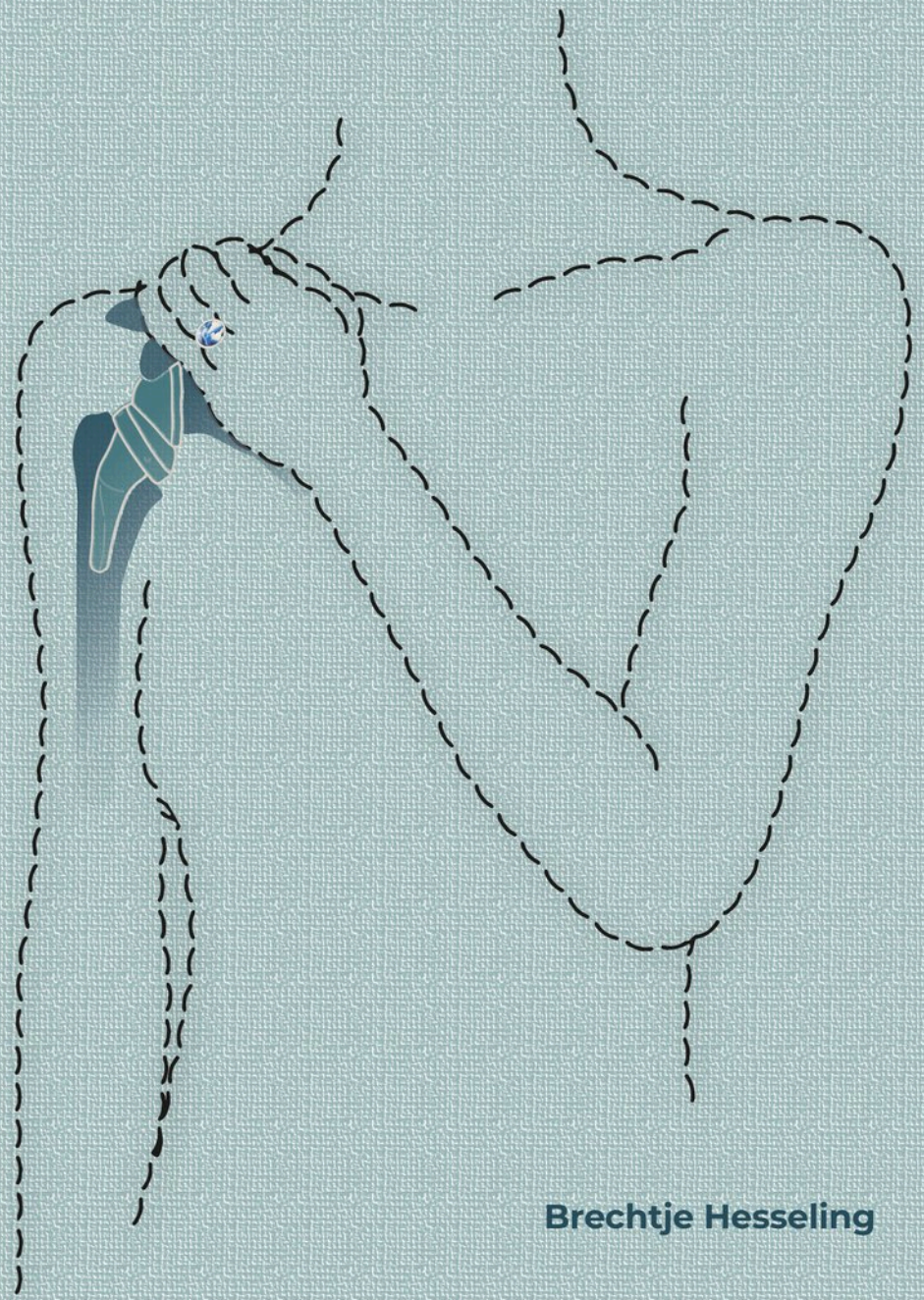
IMPROVING THE PATIENT'S EXPERIENCE IN SHOULDER ARTHROPLASTY

TOWARDS MORE PATIENT-SPECIFIC COUNSELING AND EXPECTATION MANAGEMENT



Brechtje Hesseling

Improving the Patient's Experience in Shoulder Arthroplasty

Towards more patient-specific counseling and expectation management

Improving the Patient's Experience in Shoulder Arthroplasty
Towards more patient-specific counseling and expectation management

Het verbeteren van de patiëntervaring bij schouderarthroplastiek
Richting meer patiëntspecifieke voorlichting en verwachtingsmanagement

The research described in this thesis was performed at the Department of Orthopedics, Reinier de Graaf, Delft, The Netherlands and at Reinier Haga Orthopaedic Centre, Zoetermeer, the Netherlands.

Parts of the research were supported by grants from the Scientific Research Council / the Board of Directors of Reinier de Graaf Hospital, Delft, the Netherlands.

Financial support for the printing of this thesis was generously provided by (in no particular order) Reinier Haga Orthopaedic Centre, Reinier de Graaf hospital, department of Orthopedic Surgery & Sports Medicine at Erasmus MC, Erasmus University Rotterdam, Nederlandse Orthopaedische Vereniging (NOV) and ChipSoft.

Cover & layout design: Anastasia Gangaev | www.ridderprint.nl
Printed by: Ridderprint | www.ridderprint.nl
ISBN: 978-94-6568-146-7

Copyright 2026 © Brechtje Hesseling
All rights reserved. No part of this thesis may be reproduced or transmitted in any way or by any means, without the prior permission of the author.

Improving the Patient's Experience in Shoulder Arthroplasty
Towards more patient-specific counseling and expectation management

Het verbeteren van de patiëntervaring bij schouderarthroplastiek
Richting meer patiëntspecifieke voorlichting en verwachtingsmanagement

Thesis

to obtain the degree of Doctor from
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr.ir. A.J. Schuit

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Wednesday 16 December 2026 at 10.30 hrs
by

Brechtje Hesseling
born in Groningen, the Netherlands.

Erasmus University Rotterdam

Doctoral Committee:

Promoter:	Prof.dr. D. Eygendaal
Co-promoters:	Dr. N.M.C. Mathijssen Dr. B.A.M. Snoeker
Assessment committee:	Prof.dr. S.M.A. Bierma-Zeinstra Prof.dr. T.P.M. Vliet Vlieland Dr. A. van Noort

Table of Contents

	Preface	7
Chapter 1	General introduction	9
PART I	**What can we tell patients about their outcomes after shoulder arthroplasty?**	**19**
Chapter 2	Patient-related prognostic factors for function and pain after shoulder arthroplasty: a systematic review	21
Chapter 3	Trajectories of pain recovery during the first 8 weeks after shoulder arthroplasty: results from the shoulder diary study using latent growth curve modeling	63
PART II	**Patients' expectations regarding shoulder arthroplasty**	**99**
Chapter 4	Complete fulfillment of expectations is associated with greater satisfaction after shoulder arthroplasty: results from a prospective multicenter cohort study	101
PART III	**Measuring change in shoulder function of shoulder arthroplasty patients**	**131**
Chapter 5	The Dutch patient-based Constant-Murley score cannot be used interchangeably with the clinician-based Constant-Murley score in shoulder arthroplasty patients	133
Chapter 6	General discussion	161
Appendices	English summary	178
	Nederlandse samenvatting	181
	List of publications	184
	Portfolio	188
	Curriculum Vitae	191
	Dankwoord	192

Preface

When I started the research for this thesis, I had nearly a decade of experience as a physical therapist, specializing in shoulder conditions. I frequently saw patients in the first postoperative week following shoulder arthroplasty, and what struck me was how many patients expressed uncertainty about what they were feeling in their shoulder, what to expect for the coming weeks, and whether their recovery was normal compared to others. Witnessing this uncertainty repeatedly fueled my drive to improve patients' experiences of shoulder arthroplasty through this research. During the research, many participating patients also acknowledged the studies' relevance.

This validation, along with the unwavering support from my family, friends, supervisors and coworkers, is what kept me going through the challenges of writing this thesis.

1

General introduction

The shoulder joint and the impact of shoulder disorders

The shoulder is a crucial joint for many tasks of daily living; not just for overhead activities, but also for tasks in front of the body or to the side or back. The shoulder girdle consists of multiple joints: the acromioclavicular (AC) joint, the sternoclavicular (SC) joint, the scapulothoracic (ST) joint and the glenohumeral (GH) joint (figure 1).

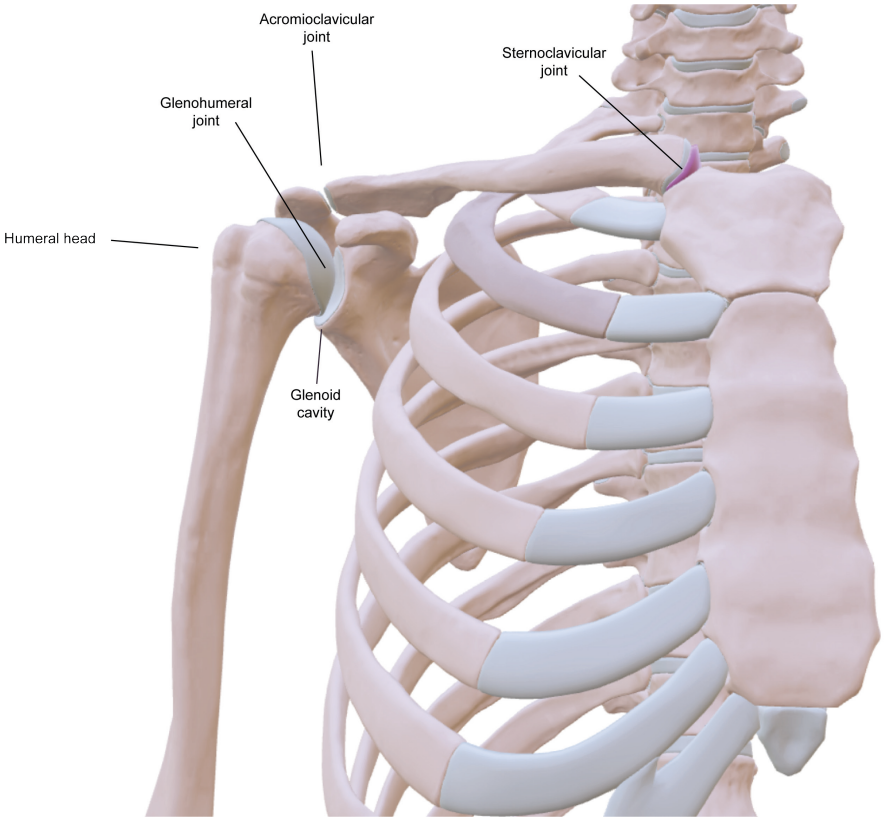


Fig. 1 3D anatomical model of the upper limb, shoulder girdle and thorax.
From Open3DModel – Upper limb – English labels, AnatomyTOOL.org by the Open3D project, licensed under Creative Commons Attribution-ShareAlike (CC BY-SA). Available at: https://anatomytool.org/content/open3dmodel-upper-limb-english-labels (accessed 22 Feb 2026).

Of these, the GH joint is the largest. It consists of the large convex head of the humerus and the much smaller concave glenoid cavity of the scapula and is surrounded by capsule, ligaments and the rotator cuff – a group of

muscles and tendons that stabilize the shoulder. However, several disorders (acute, traumatic, chronic, or degenerative) can disrupt proper shoulder function. Depending on source populations and exact definitions, the prevalence of shoulder pain in the general population has been estimated between 0.6% to 55.2% [1].

Shoulder disorders often greatly affect patients' lives in multiple domains. Pain and impaired function of the shoulder joint often lead to limitations in daily activities[2-5]. For example, self-care activities such as washing, dressing and combing hair become painful or severely limited due to decreased range of motion. Work and recreational activities often also suffer. Moreover, many patients experience disruptions of their sleep. In turn, all these may lead to emotional distress, such as frustration and anxiety and, for some, even depressed feelings[4,5].

Two of the most common shoulder disorders in the aging population are osteoarthrosis (OA, also known as osteoarthritis in the literature) and cuff tear arthropathy (CTA) (figure 2).

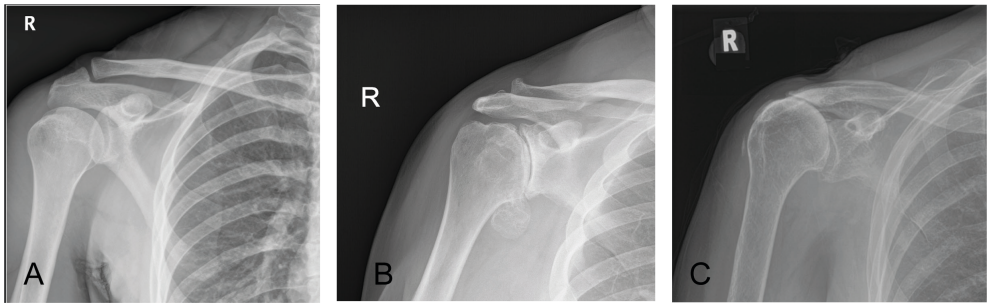


Fig. 2 X-ray images of a A) healthy shoulder, B) shoulder with OA and C) shoulder with CTA. Cases courtesy of Radiopaedia.org, by A) dr. A. Murphy - rID 72387, B) dr. B. Botz - rID 79230 and C) dr. E. Collins - rID 189406.

OA is a complex process that affects the entire joint, although the most notable effects are the degradation of articular cartilage, joint space narrowing and formation of osteophytes[6,7]. Although the articular cartilage is relatively insensitive, other tissues like subchondral bone and the synovium are a likely source of pain. Pain, together with inflammation and stiffness, may subsequently limit range of motion. In CTA, the stabilizing function of the rotator cuff is decreased or absent because of

fatty infiltration, muscle atrophy or tendon tears. This leads to superior and anterior migration of the humeral head and subsequent (painful) wear/arthritis of the head and glenoid[8]. In addition, the superior migration of the humeral head also severely limits the ability to raise the arm.

Shoulder arthroplasty

First-line treatment for degenerative shoulder disorders is usually non-surgical, consisting of analgesia (often NSAIDs) and physical therapy[9,10]. In selected cases, intra-articular injections provide additional pain relief by reducing inflammation[9,10].

When non-surgical treatment fails, shoulder arthroplasty (SA) is a viable option in selected patients[10]. The type of SA used depends on the clinical indication. In hemi-arthroplasty (HA, figure 3a), only the humeral head is replaced, leaving the glenoid intact. In anatomic total shoulder arthroplasty (aTSA, figure 3b), both the glenoid and humeral head are replaced to match natural anatomy. In reverse total shoulder arthroplasty (rTSA, figure 3c), both are again replaced, but the positions are reversed: the socket is placed in the humerus and the ball in the scapula.

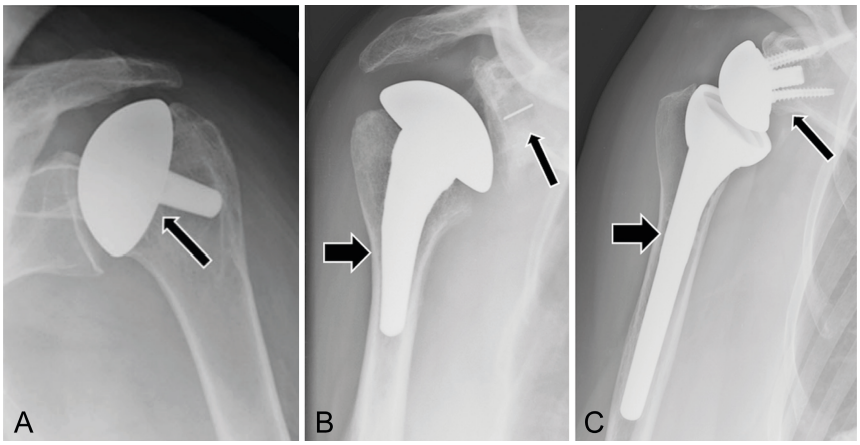


Fig. 3 X-ray images of A) hemi prosthesis (arrow depicts humeral head), B) total shoulder prosthesis (thick arrow depicts humeral prosthetic stem, thin arrow depicts glenoid implant) and C) reverse total shoulder prosthesis (thick arrow depicts humeral prosthetic stem, thin arrow depicts the glenosphere with humeral component).
Adapted from Combes D, Lancigu R, Desbordes de Cepoy P, et al. Imaging of shoulder arthroplasties and their complications: a pictorial review. Insights Imaging. 2019;10:90. doi:10.1186/s13244-019-0756-z

Although SA is the third most performed arthroplasty in the Netherlands, it is performed approximately ten times less frequently than total knee arthroplasty (TKA) or total hip arthroplasty (THA). In 2024, 3,508 reverse, 629 total, and 143 hemi shoulder arthroplasties were registered in the Dutch Arthroplasty Register (Landelijke Registratie Orthopedische Interventies, LROI), with an additional 492 revision arthroplasties[11]. By comparison, more than 37,000 THAs and 28,000 TKAs were performed in the same year[11].

SA is generally successful in relieving pain and improving shoulder function. In both clinical practice and biomedical research, these improvements are commonly measured and quantified using patient-reported outcome measures (PROMs) – questionnaires completed by patients to assess their health status or health-related quality of life. Common shoulder-related PROMs include the Oxford Shoulder Score[12,13] (OSS), the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment[14] (ASES), the Simple Shoulder Test[15,16] (SST) and the Disabilities of the Arm, Shoulder and Hand questionnaire[17,18] (DASH). However, PROMs are by definition subjective. When objective evaluation is desired, the Constant-Murley Score[19,20] (CMS) is often used. In addition to subjective items on pain and function, the CMS includes an objective assessment of ROM and strength performed by the clinician. However, it is time-consuming: Kohn & Geyer[21] found that even an experienced examiner required an average of almost seven minutes to complete it. This is a significant portion of the allotted time for a consultation in the Netherlands, as a first consultation is often scheduled for 20 minutes, with follow-up appointments as short as 10 minutes.

Tailored patient education in arthroplasty

A total joint arthroplasty (TJA) is a major surgical procedure and, for some patients, even a life-altering event[22]. The prospect of undergoing such a procedure can cause patients to experience uncertainty and perioperative anxiety, especially if they do not know quite what to expect[23,24]. These feelings may be exacerbated by the fact that SA is much less often performed than THA and TKA. Whereas patients often know someone in their environment who has had a new hip or knee, they may have not yet encountered someone who has had SA. This preoperative uncertainty and anxiety may also be related to poorer outcomes[25].

Qualitative studies have shown that to combat this uncertainty, patients desire informational support from their surgeon, not only on longer-term outcomes but also on short-term recovery trajectories and pain management[23,24,26]. Not adequately receiving this support may lead to lower satisfaction with the TJA process.

In addition, postoperative satisfaction is linked to patients' preoperative expectations. More specifically, when expectations are not met postoperatively, satisfaction scores are lower[27-29]. This relationship is supported by studies that intervene on preoperative expectations[30,31]. Preoperative educational programs that aimed to set more realistic preoperative expectations led to higher expectation fulfilment and higher postoperative satisfaction. This relationship appears to be stronger for TKA than for THA[32,33].

From the above, it can be concluded that orthopedic surgeons can improve their care by better preparing the patient for their arthroplasty journey. More accurate information about what the patient can expect can reduce fear and anxiety and help the patient set more realistic expectations. To be able to do this, however, the surgeon must have access to high-quality evidence-based knowledge; knowledge that helps to tailor the counseling to the patient's situation, needs and values. This patient-centered approach also aligns well with healthcare system shifts towards shared decision-making and value-based healthcare[34-36].

Although research on THA and TKA has been extensive, research on shoulder OA and SA is lagging behind[37], likely because of the relatively small numbers involved. Literature is especially sparse on early postoperative recovery and expectations in SA patients. While some authors have studied preoperative expectations, little is known about the fulfilment of expectations after SA and its possible relationship with postoperative satisfaction[38].

Results from one type of TJA cannot necessarily be extrapolated to other types, as illustrated by the finding that interventions to set more realistic expectations have a clearer effect in TKA than in THA. To provide adequate, tailored patient education for SA patients, it is essential to deepen our understanding of recovery and the role of expectations.

General aim and outline of this thesis

The general aim of this thesis is to improve the patient's experience of the arthroplasty journey, specifically the preoperative and early postoperative period.

Part I therefore focuses on the substantive part of tailored patient education for SA patients. In **Chapter 2,** we systematically review what is known to date in the existing literature on the prognostic effect of patient characteristics on pain and functional outcomes after SA. More specifically, we focus on the individual prognostic effect of each factor by only including studies that used multivariable modelling. **Chapter 3** studies the recovery during the first eight weeks after SA in detail. We investigate whether trajectories in pain recovery during these weeks differ between patients and, if so, how many different subgroups can be distinguished.

In **Part II,** we deepen our understanding of patient expectations in SA. In **Chapter 4,** two types of expectations are studied: probabilistic expectations (what does the patient realistically expect) and value-based expectations (what is important to the patient). For both types, we analyze what patients expect before surgery and to what extent these expectations are fulfilled at 6 and 12 months. Moreover, we examine whether greater expectation fulfilment is associated with greater postoperative satisfaction.

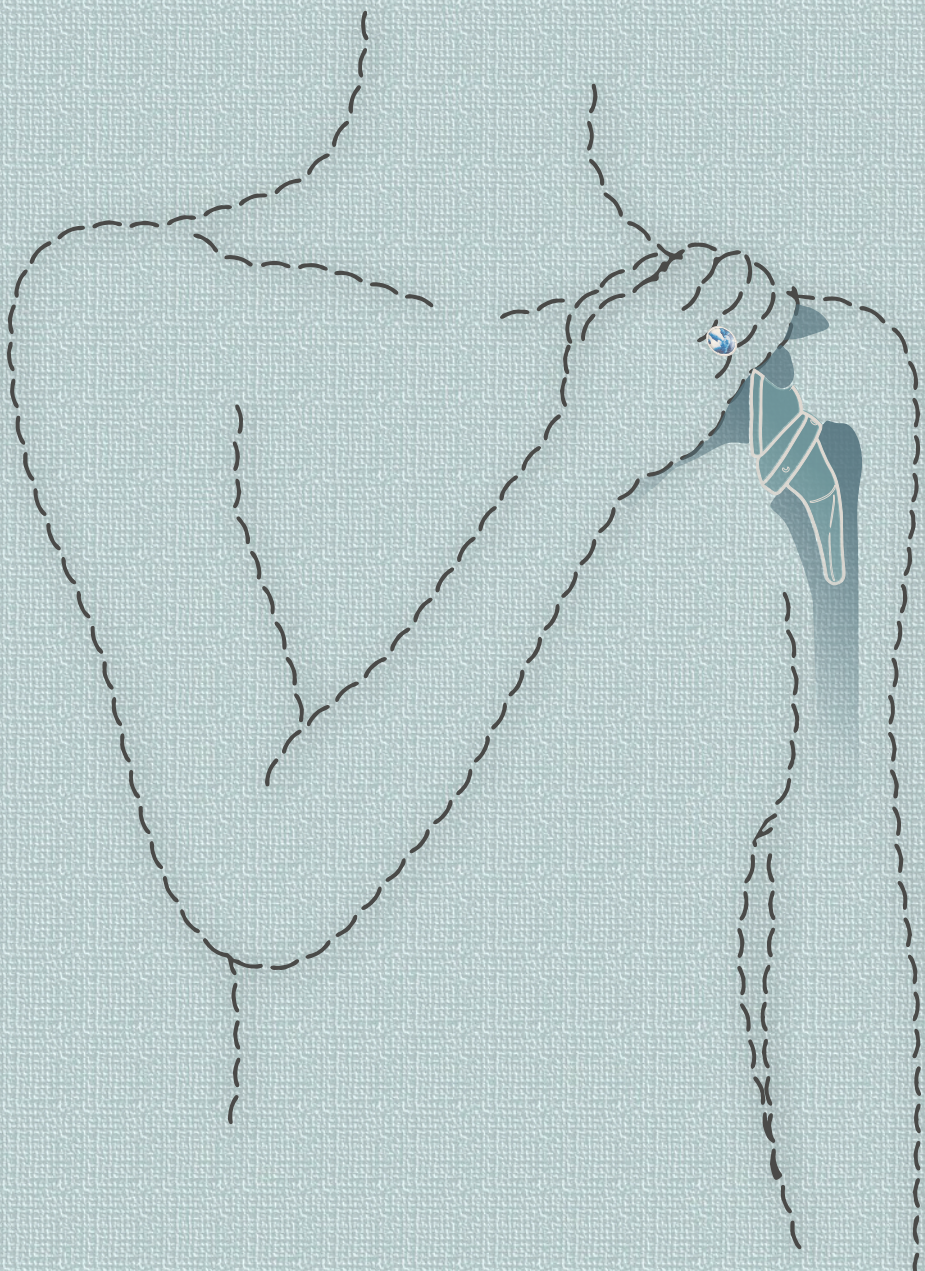
In **Part III,** we focus on the CMS. Shifting the CMS to be completed prior to the consultation leaves more time to align patient education and care activities with the patient's needs and preferences. Therefore, **Chapter 5** investigates whether a Dutch patient-based CMS is a viable alternative to the original clinician-based CMS. We analyze and describe the interchangeability of the two methods, as well as the test-retest reproducibility of each method.

In **Chapter 6**, the general discussion, the main findings of this thesis are discussed. The results are placed in a larger perspective, and implications for future research are addressed.

References

1. Lucas J, van Doorn P, Hegedus E, Lewis J, van der Windt D. A systematic review of the global prevalence and incidence of shoulder pain. *BMC Musculoskelet Disord* 2022; **23**(1): 1073.
2. Largacha M, Parsons IMt, Campbell B, Titelman RM, Smith KL, Matsen F, 3rd. Deficits in shoulder function and general health associated with sixteen common shoulder diagnoses: a study of 2674 patients. *J Shoulder Elbow Surg* 2006; **15**(1): 30-9.
3. Page MJ, Huang H, Verhagen AP, Buchbinder R, Gagnier JJ. Identifying a core set of outcome domains to measure in clinical trials for shoulder disorders: a modified Delphi study. *RMD Open* 2016; **2**(2): e000380.
4. Maxwell C, Robinson K, McCreesh K. Understanding Shoulder Pain: A Qualitative Evidence Synthesis Exploring the Patient Experience. *Phys Ther* 2021; **101**(3).
5. Page MJ, O'Connor DA, Malek M, et al. Patients' experience of shoulder disorders: a systematic review of qualitative studies for the OMERACT Shoulder Core Domain Set. *Rheumatology (Oxford)* 2019.
6. Abramoff B, Caldera FE. Osteoarthritis: Pathology, Diagnosis, and Treatment Options. *Med Clin North Am* 2020; **104**(2): 293-311.
7. Ibounig T, Simons T, Launonen A, Paavola M. Glenohumeral osteoarthritis: an overview of etiology and diagnostics. *Scand J Surg* 2021; **110**(3): 441-51.
8. Rugg CM, Gallo RA, Craig EV, Feeley BT. The pathogenesis and management of cuff tear arthropathy. *J Shoulder Elbow Surg* 2018; **27**(12): 2271-83.
9. NHG-werkgroep Schouderklachten, Winters J, Van der Windt D, Spinnewijn W, De Jongh A, Van der Heijden G. NHG-Standaard Schouderklachten. 2019. https://richtlijnen.nhg.org/standaarden/schouderklachten (accessed January 20 2026).
10. Federatie Medisch Specialisten. Richtlijn Schouderprothese. 2023. https://richtlijnendatabase.nl/richtlijn/schouderprothese/startpagina_-_schouderprothese.html (accessed January 20 2026).
11. Landelijke Registratie Orthopedische Interventies (LROI). Annual LROI Report 2025. 2025. https://www.lroi.nl/jaarrapportage.
12. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. *Arch Orthop Trauma Surg* 2009; **129**(1): 119-23.
13. Berendes T, Pilot P, Willems J, Verburg H, te Slaa R. Validation of the Dutch version of the Oxford Shoulder Score. *J Shoulder Elbow Surg* 2010; **19**(6): 829-36.
14. Richards RR, An KN, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg* 1994; **3**(6): 347-52.
15. Lippit S, Matsen F, Fu F, Hawkins R. A practical tool for evaluating function: the Simple Shoulder Test. *The shoulder: a balance of mobility and stability Rosemont (IL): American Academy of Orthopaedic Surgeons* 1993.
16. van Kampen DA, van Beers LW, Scholtes VA, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. *J Shoulder Elbow Surg* 2012; **21**(6): 808-14.
17. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med* 1996; **29**(6): 602-8.
18. Veehof MM, Sleegers EJ, van Veldhoven NH, Schuurman AH, van Meeteren NL. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *J Hand Ther* 2002; **15**(4): 347-54.
19. Constant C, Murley A. A clinical method of functional assessment of the shoulder. *Clinical orthopaedics and related research* 1987; (214): 160-4.

20. Constant CR, Gerber C, Emery RJ, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. *Journal of shoulder and elbow surgery* 2008; **17**(2): 355-61.
21. Kohn D, Geyer M. The subjective shoulder rating system. *Arch Orthop Trauma Surg* 1997; **116**(6-7): 324-8.
22. Gustafsson B, Ekman SL, Ponzer S, Heikkilä K. The hip and knee replacement operation: an extensive life event. *Scand J Caring Sci* 2010; **24**(4): 663-70.
23. Buus A, Hejlsen OK, Dorisdatter Bjørnes C, Laugesen B. Experiences of pre- and postoperative information among patients undergoing knee arthroplasty: a systematic review and narrative synthesis. *Disabil Rehabil* 2021; **43**(2): 150-62.
24. Cronström A, Jönsson T, Limbäck G, Ljung M, Ståhl C, Östlind E. Stairway to heaven via the highway to hell: a qualitative study on patients' experience of knee joint replacement surgery. *J Orthop Surg Res* 2025; **20**(1): 570.
25. Kagan I, Bar-Tal Y. The effect of preoperative uncertainty and anxiety on short-term recovery after elective arthroplasty. *J Clin Nurs* 2008; **17**(5): 576-83.
26. Goldsmith LJ, Suryaprakash N, Randall E, et al. The importance of informational, clinical and personal support in patient experience with total knee replacement: a qualitative investigation. *BMC Musculoskelet Disord* 2017; **18**(1): 127.
27. Deakin AH, Smith MA, Wallace DT, Smith EJ, Sarungi M. Fulfilment of preoperative expectations and postoperative patient satisfaction after total knee replacement. A prospective analysis of 200 patients. *Knee* 2019; **26**(6): 1403-12.
28. Jain D, Nguyen LL, Bendich I, et al. Higher Patient Expectations Predict Higher Patient-Reported Outcomes, But Not Satisfaction, in Total Knee Arthroplasty Patients: A Prospective Multicenter Study. *J Arthroplasty* 2017; **32**(9s): S166-s70.
29. Tilbury C, Haanstra TM, Leichtenberg CS, et al. Unfulfilled Expectations After Total Hip and Knee Arthroplasty Surgery: There Is a Need for Better Preoperative Patient Information and Education. *J Arthroplasty* 2016; **31**(10): 2139-45.
30. Nam HS, Yoo HJ, Ho JPY, Kim YB, Lee YS. Preoperative education on realistic expectations improves the satisfaction of patients with central sensitization after total knee arthroplasty: a randomized-controlled trial. *Knee Surg Sports Traumatol Arthrosc* 2023; **31**(11): 4705-15.
31. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. *Bone Joint J* 2021; **103-b**(4): 619-26.
32. Scott CE, Bugler KE, Clement ND, MacDonald D, Howie CR, Biant LC. Patient expectations of arthroplasty of the hip and knee. *J Bone Joint Surg Br* 2012; **94**(7): 974-81.
33. Tolk JJ. Expectations and Outcome in Knee and Hip Arthroplasty. 2020.
34. Makkar D, Sauhta R, Jaganmohan S, Priyadarshini T, Devanand V, Raja A. Insights of shared decision-making and personalized treatment in orthopedics - A qualitative systematic review. *J Clin Orthop Trauma* 2025; **65**: 102989.
35. Ribbons K, Walker FR, Pollack M, Nilsson M. Assessing the Ability of an Online Education Program to Modify Patient Expectations of Total Knee Arthroplasty Outcomes: Protocol for a Randomized Controlled Trial. *JMIR Res Protoc* 2025; **14**: e71898.
36. Wilson CD, Probe RA. Shared Decision-making in Orthopaedic Surgery. *J Am Acad Orthop Surg* 2020; **28**(23): e1032-e41.
37. Ackerman IN, Buchbinder R. Let's talk about shoulder osteoarthritis. *Rheumatology (Oxford)* 2022; **61**(9): 3507-8.
38. Claes A, Mesel A, Verborgt O, Struyf F. Effects of pre-operative patient expectations on the outcome after total shoulder arthroplasty: A systematic review. *Shoulder Elbow* 2025; **17**(4): 388-400.



PART I

What can we tell patients about their outcomes after shoulder arthroplasty?

2

Patient-related prognostic factors for function and pain after shoulder arthroplasty: a systematic review

Hesseling B, Prinsze N, Jamaludin F, Perry SI, Eygendaal D, Mathijssen NM, Snoeker BA.

Systematic Reviews 2024; 13(1): 286

Abstract

Background

While shared decision making is a cornerstone of orthopedic care, orthopedic surgeons face challenges in tailoring their advice and expectation management to individual shoulder arthroplasty patients due to the lack of systematically summarized evidence-based knowledge. This systematic review aims to provide an overview of current knowledge on independent predictive effects of patient-related factors on functional and pain-related outcomes after shoulder arthroplasty.

Methods

We included longitudinal cohort studies including patients receiving total or reverse shoulder arthroplasty or hemiarthroplasty for primary osteoarthritis or cuff tear arthropathy. Studies with only univariable analyses were excluded. MEDLINE, Embase and CINAHL databases were last searched on June 27th 2023. Risk of bias was evaluated using the QUIPS tool. For the analyses, we divided outcomes into three domains (Functional Recovery, Pain, and Functional Recovery & Pain) and four time points (short term, medium-short term, medium-long term and long term). When appropriate, meta-analyses were conducted to pool regression coefficients or odds ratios. Otherwise, results were summarized in a qualitative analysis. We used the GRADE approach to rate the certainty of the evidence.

Results

Thirty-three studies analyzing over 6900 patients were included; these studied 16 PROMs and 52 prognostic factors. We could perform meta-analyses for six combinations of prognostic factor, domain and time point. Only the meta-analysis for medium-long term poor ASES scores indicated worse outcomes for previous shoulder surgery (OR (95%CI) of 2.10 (1.33 – 3.33)). The majority of reported factors showed unclear or neutral independent effects on functional outcomes.

Conclusions

Methodological heterogeneity and selective/incomplete reporting prevented us from pooling most results, culminating in a largely qualitative analysis. Depression, preoperative opioid use, preoperative ASES and SST scores, surgery on the dominant side, previous surgery, male gender, no. of patient-reported allergies, back pain, living alone, CTA vs OA, diabetes and greater preoperative external ROM predicted neutral to worse or worse outcomes. In contrast, higher electrical pain threshold on the operative side, OA/RCA vs other diagnosis and private insurance vs Medicaid/Medicare predicted neutral to better or better outcomes.

These results can help orthopedic surgeons tailor their advice and better manage expectations.

Systematic review registration

PROSPERO CRD42021284822

Keywords: shoulder arthroplasty; prognostic factor; functional recovery; pain; PROMs

Introduction

Currently, it is a challenge for surgeons to provide evidence-based advice to their shoulder arthroplasty (SA) patients, since an overview of patient-related prognostic factors on functional and pain-related outcomes after shoulder arthroplasty does not exist. Thousands of patients undergo SA yearly, and the incidence continues to rise. For the United States, the incidence has been predicted to increase to at least 174.810 SAs for 2025[1]. To each of these patients, orthopedic surgeons would ideally tailor their advice and expectation management regarding functional recovery and pain. After all, explaining evidence-based benefits and risks of treatment is an essential step in shared-decision making[2].

However, such an overview of the independent predictive effect of patient-related factors on outcomes after SA does not exist. Although Vajapey[3] described the influence of only psychological factors on outcomes after SA in a systematic review, the included studies mostly used univariable analysis or correlations and did not contain the full scope of relevant prognostic factors.

Since patient-related prognostic factors do not present in isolation as the outcome of surgery is determined by multiple factors, univariable analyses are insufficient to predict outcomes. Adjustment for other prognostic factors can affect not only the magnitude of the predictive effect but even the direction of the effect. To enable surgeons to tailor their advice to each individual patient's circumstances, the first step is a thorough overview of evidence from studies investigating prognostic factors in multivariable analyses. To the best of our knowledge, no such overview is currently available in the literature.

Therefore, this systematic review aims to create an overview of the current body of evidence on the independent predictive effects of patient-related factors on functional and pain-related outcomes after SA.

Materials and methods

Eligibility criteria

We included longitudinal cohort studies with multivariable analysis to study patient-related prognostic factors on patient-reported pain and functional recovery after total shoulder arthroplasty (TSA), reverse shoulder arthroplasty (RSA), and hemiarthroplasty (HA). We only included studies with patients older than 18 years and in which at least 75% of included patients received SA for primary osteoarthritis or cuff tear arthropathy. If data could be extracted for SA alone, we also included studies comparing SA to another treatment. No restrictions were placed on the timing of outcome measurement.

Studies that were cross-sectional or case reports, only reported univariable analysis or correlations, included intra- or postoperative variables in the multivariable analysis, or used concomitant treatment were excluded.

Prognostic factors under study

Any patient-related prognostic factor was of interest for this review, including but not limited to socio-demographic, physical, psychological, and work-related factors.

Outcomes under study

For this review, only patient-reported outcome measures (PROMs) on pain and functional status were of interest. Measures with a component performed by a health professional (e.g. Constant-Murley score) were excluded. No limitation was placed on the timing of outcome measurement.

To aid in interpretation of the results, we grouped PROMS under three separate constructs:

- Functional recovery and pain - combined: American Shoulder and Elbow Surgeons score (ASES), Shoulder Pain and Disability Index (SPADI), Western Ontario Osteoarthritis of the Shoulder Index (WOOS) and Oxford Shoulder Score (OSS)
- Functional recovery - isolated: Quick Disabilities of the Arm, Shoulder

and Hand score (QuickDASH), Disabilities of the Arm, Shoulder and Hand score (DASH), Simple Shoulder Test (SST), Patient-Reported Outcomes Measurement Information System (PROMIS) Upper Extremity (UE) subscale, PROMIS Physical Functioning (PF) subscale, PROMIS Pain Interference (PI) subscale, Single Assessment Numeric Evaluation (SANE), Shoulder Activity Scale (SAS), Subjective Shoulder Value (SSV), and ASES function score

- Pain - isolated: Visual Analogue Scale (VAS) pain and ASES pain score

Timing of the measurements are grouped under 'short-term' (1 day to 3 months), 'medium-short term' (3 months up to and including 1 year), 'medium-long term' (after 1 year, up to and including 5 years) and 'long term' (after 5 years).

Search methods

The following databases were used up to July 5th, 2021: MEDLINE (through Pubmed), EMBASE (through Ovid), and CINAHL . A search update was performed on June 27th, 2023, using the same strategy. A medical librarian (FJ) designed and performed a sensitive search strategy guided by the PICOTS format[4, 5] (Table 1). She deduplicated the results of both searches. For the full PubMed search strategy, see Table A1 in Appendix 1.

Table 1 – PICOTS format

Patients	Patients who received HA, TSA or RSA for primary OA or CTA
Index prognostic factors	Patient-related prognostic factors measured before the intervention
Comparator factor	-
Outcome	PROMs assessing functional recovery and pain
Time	Any range of prediction interval
Setting	Hospital care

HA = hemi-arthroplasty, TSA = total shoulder arthroplasty, RSA = reverse shoulder arthroplasty, OA = osteoarthritis, CTA = cuff tear arthropathy, PROMs = patient-reported outcome measures

Data collection, data extraction and risk of bias

All search results were imported into Rayyan, a web and mobile app for systematic reviews[6]. In Rayyan, two reviewers (NP, BH) independently screened titles and abstracts for eligibility and, subsequently, the full-text articles of studies that were deemed eligible. Next, for the included studies, the two reviewers independently extracted data on study characteristics, prognostic factors and outcomes and assessed the risk of bias (RoB) using the Quality in Prognostic Studies (QUIPS) tool[7]. When studies used more than one multivariable model to analyze their data, we extracted only the most complex model (i.e. adjusting for the highest number of factors).

For all stages, disagreements were solved by discussion and, if needed, a third reviewer (BS).

The QUIPS tool is divided into six domains: 'study participation', 'study attrition', 'prognostic factor measurement', 'outcome measurement', 'adjustment for other prognostic variables' and 'statistical analysis and reporting'.

For the domain 'adjustment for other prognostic variables', we decided to only use moderate or high RoB, given the absence of a predefined evidence-based core set of variables to include in a multivariable analysis.

Studies received an overall 'Low RoB' rating when all six domains had low RoB, an overall 'High RoB' rating when at least one domain had high RoB, and an overall 'Moderate RoB' rating when at least one domain had moderate RoB and none had high RoB. For all stages, disagreements were solved by discussion and, if needed, a third reviewer (BS).

Extracted data and RoB results were recorded on a detailed Excel worksheet by each of the two reviewers, and a final version was created after discussion and reaching an agreement. Since only studies that used multivariable analysis were eligible for this review, we extracted for each combination of prognostic factor and outcome either the β-coefficient or odds ratio (OR), its 95% confidence interval (CI) and its p-value. When data was unclear or missing, authors were contacted up to two times to request clarification by email. If multiple reports for a single study were

encountered, both reports were only used if they reported on different prognostic factor/outcome combinations.

Quantitative data synthesis

For performing meta-analysis, a minimum of two studies per prognostic factor and outcome combination was necessary. First, we assessed whether clinical and methodological heterogeneity, including similar time points of outcome measurement, were sufficiently low to enable meta-analysis. If so, statistical heterogeneity needed to be acceptable (I^2 < 75%) to pool results. We used random-effects models with Restricted Maximum Likelihood (REML) estimation[8, 9] to conduct the analyses, using R version 4.3.1[10] in RStudio version 2023.6.1.524[11] and the metafor package[12].

When multiple reports with the same source population were eligible for meta-analysis, we included the report with the lowest RoB or, in case of similar RoB, the largest sample size. We then performed sensitivity analyses by repeating the meta-analysis using the other study/studies instead, to examine whether this would substantively affect the results.

Qualitative data synthesis

When meta-analysis was not possible, we performed a qualitative analysis. For the qualitative analysis, the following data from all studies were tabulated for each prognostic factor: the studied PROM, its method (e.g. raw score, change score) and timing of measurement, effect estimate with 95% CI and p-value, the number of other factors for which each study adjusted and overall RoB.

Because we did not find guidelines for defining or summarizing the direction of effect in a qualitative data synthesis of prognostic factor studies, we created and adhered to the following guidelines using the minimal clinically important difference (MCID) to determine the direction and size of effect:

- Better or worse outcomes: when all effect estimates and CI limits of involved studies exceeded either the MCID of that outcome (in case of an established MCID in the literature) or the neutral value (0 for linear

regression coefficients and 1 for OR, when no established MCID was found) in a positive or negative direction, respectively

- Neutral to better/neutral to worse outcomes: when one end of the CI of at least one study exceeded the MCID or when no established MCID was found, the effect estimate of at least one study exceeded the neutral value after rounding
- Neutral direction of effect: when both CI limits of involved studies fell short of reaching the MCID or when no established MCID was found, the effect estimates of involved studies equalled the neutral value after rounding
- Conflicting direction of effect: when effect estimates or CIs of involved studies exceeded the MCID on both the positive and negative sides of effect
- Unclear direction of effect: when no effect estimates or only effect estimates without CI could be extracted from all involved studies (except when the effect estimate of at least one study exceeded the MCID, then 'neutral to better/neutral to worse outcomes' was defined)

Table A2 in Appendix 1 lists the MCIDs we have utilized.

Certainty of the evidence

We used the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach to evaluate the certainty of the evidence[13], following the recommendations for the assessment of evidence about prognostic factors[14]. The certainty of the evidence can be rated down on the domains 'risk of bias', 'inconsistency', 'indirectness', 'imprecision' and 'publication bias'. The certainty can be rated up when there is a clear dose-response gradient, a large effect, or when plausible confounding is present, adjustment for that confounding would increase the confidence in the found effect being the true effect. However, Foroutan et al.[14] report that they have not yet found examples of systematic reviews that mandate rating up for prognostic factors.

Since the interpretations of some aspects of GRADE are mainly described for quantitatively pooled estimates, we clarify our interpretation of the 'imprecision' and 'inconsistency' domains for use in our qualitative analysis as follows:

- Imprecision: Foroutan et al.[14] recommend evaluating imprecision by relating the upper and lower CI limits of the pooled estimate to a clinical decision threshold (in our case, the MCID). Because no pooled estimate is available in a qualitative analysis, we evaluated whether our conclusion on the overall direction of effect would change if the CIs of the individual studies were more precise (e.g. thereby changing from a CI with one end surpassing the MCID to a CI which falls short of reaching the MCID on both ends).
- Inconsistency: Foroutan et al.[14] state that to decide whether to rate down the quality based on inconsistency, 'one should consider the contribution of the divergent study to the pooled estimate', for example, by looking at the study's weight in the meta-analysis. Instead, we evaluated the variability in point estimates and the extent of overlap in CIs of all involved studies.

The GRADE approach was applied by two reviewers to each combination of prognostic factor and outcome construct when data from at least two studies could be extracted, and resulted in one of the following levels of certainty (disagreements were solved by discussion):

- High certainty: further research is very unlikely to change our confidence in the association between the prognostic factor and the outcome construct
- Moderate certainty: further research is likely to impact our confidence in the association between the prognostic factor and the outcome construct
- Low certainty: further research is very likely to impact our confidence in the association between the prognostic factor and the outcome construct
- Very low certainty: any association between the prognostic factor and the outcome construct is very uncertain

Finally, for each construct, the overall judgments on effect and certainty of evidence for the prognostic factors were incorporated in Summary of Findings (SoF) figures, ordered per timing of measurements (medium-short term in the main text, other time points in the Appendix). Since the GRADE approach is meant to rate certainty of aggregated evidence, prognostic factors for which the effect could only be based on a single study were incorporated separately into the SoF figures without a certainty

rating. The same was done for prognostic factors for which the direction of effect was deemed unclear.

Results

After removing duplicates (n=1,289) the search resulted in 1,840 articles. After screening titles and abstracts, 180 articles were retrieved for full-text screening. Five articles remained unavailable and were excluded[15-19]. In total, 33 studies were included in this systematic review[20-52]. For a complete overview, see the PRISMA flow diagram in Figure 1.

Included studies

Most of the included studies were (partly) conducted in the USA (n=28). Of the 16 different PROMs studied in total, studies most frequently reported ASES (n=25), VAS pain (n=10), and SST (n=7).

A total of 52 different prognostic factors were studied (in many different exact definitions), of which gender (n=28), age (n=25), and preoperative ASES score (n=11) were the most frequently encountered. We did not find any results for short-term functional recovery and pain, short-term pain, or long-term pain. Table 2 provides a detailed overview of all included studies.

Several authors appeared to have used the same source population for their studies; these have been indicated in Table 2. These related study reports were not analyzed together in any meta-analysis. Taking these related source populations into account, over 6900 patients were analyzed in the included studies.

Of the 33 studies, 9 scored moderate RoB, and 24 scored high RoB. No study scored low RoB. Moderate and high RoB were primarily due to the domains 'study attrition' (e.g. excluding patients that were lost to follow-up (LtFU) where LtFU could be related to the prognostic factor or outcome), 'adjustment for other factors' (e.g. unclear for which factors the final analysis adjusted) and 'statistical analysis and reporting' (e.g. reporting only effect estimates without CI or p-value, only p-value, or not reporting any data at all). For the RoB overview, see Figure 2.

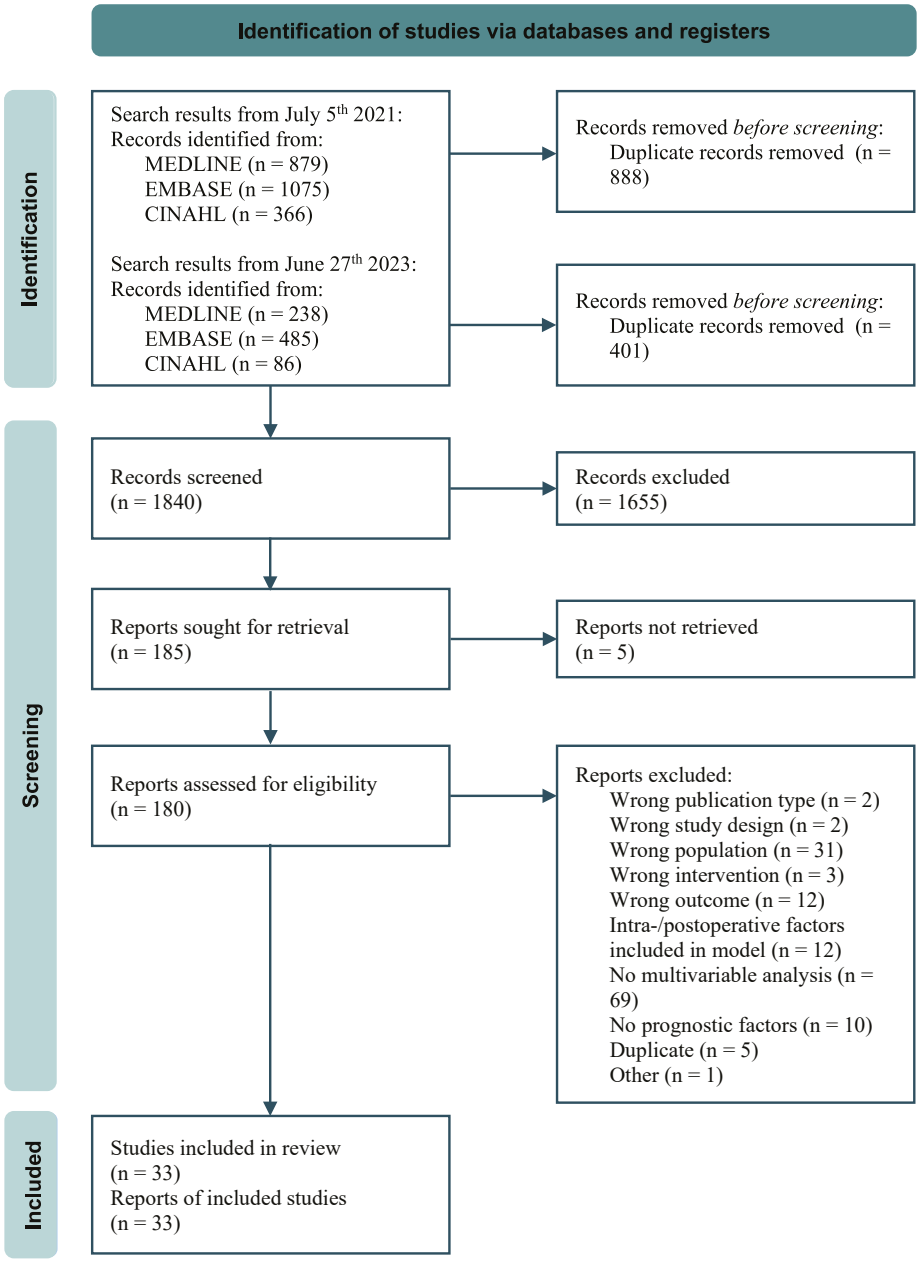


Fig. 1 PRISMA flow diagram

Risk of Bias



RoB = Risk of Bias, SP = Study Participation, SA = Study Attrition, PF = Prognostic Factor measurement, OM = Outcomes Measurement, AF = Adjustment for other Factors, SR = Statistical analysis and Reporting

Fig. 2 Risk of Bias overview

Table 2 – Characteristics of included studies

Study ID	Study design	Country, setting	Analyzed participants
Carducci 2019*	Prospective cohort	USA, New England Baptist Hospital, single surgeon	137 RSA patients with DJD, RCA or RCT, 2013 – 2016
Chang 2022	Retrospective cohort	USA + Canada, orthopedic departments, 25 surgeons in 23 centers	227 stemless TSA patients with GH arthritis (KL ≥ 3), recruitment period unknown
Chawla 2022	Retrospective cohort	USA, University of Washington Medical Center, 3 surgeons	33 RSA patients and 39 CTA-H patients diagnosed with CTA, 2010 – 2013
Cho 2017	Prospective cohort	USA, Massachusetts General Hospital, single surgeon	46 TSA patients with GH arthritis, 2013 – 2015
Dekker 2022	Retrospective cohort	USA, Steadman Philippon Research Institute, single surgeon	168 TSA patients with mostly primary OA, 2006 – 2016
Edwards 2020	Retrospective cohort	UK, Southmead Hospital, single surgeon	32 TSA patients with GH arthritis, 2006 – 2011
Fehringer 2002	Retrospective cohort	USA, University of Washington, single surgeon	102 TSA patients with GH arthritis, 1993 – 1997

Time point	Prognostic factors/factors adjusted for	Outcomes	Type of multivariable regression
Mean 29 months (± 8)	· RCA/RCT/DJD diagnosis · Preoperative ASES score · Opioid use · Previous shoulder surgery	· Poor ASES score · (bottom 30th percentile) · Poor ASES improvement · (bottom 30th percentile)	Logistic
Min. 2 years	· Gender · Age · Preoperative ASES score	· Improved ASES score · (no further definition)	Logistic
Mean 2.2 years (± 0.3)	· Medicare/WC/Commercial insurance · Preoperative SST score (for CTA-H group only)	· SST raw score	Linear
1 year	· Gender · Age · HADS Anxiety · HADS Depression · Duration of symptoms · Involved side	· ASES raw score · SSV raw score · VAS pain raw score$^{0-10}$	Linear
Mean 4.6 years (2 – 12 years)	· Gender · Age · Glenoid retroversion · Glenoid inclination · Walch classification · Preoperative ASES score · Preoperative SANE score · Preoperative QuickDASH score · Subluxation	· ASES raw score · SANE raw score · QuickDASH raw score	Linear
Mean 67 months (49 – 94 months)	· Fatty infiltration · Atrophy · Tendinopathy grade · Age · Gender	· OSS raw score · QuickDASH raw score	Linear
30 – 60 months	· Preoperative SST score · Age · Gender	· SST raw score · SST change score · Change in no. of functions that could be performed expressed as percentage of no. that could not be performed before surgery · Change in ability to perform each of SST functions	General linear model analysis of variance

Study ID	Study design	Country, setting	Analyzed participants
Forlizzi 2022*	Retrospective cohort	USA, New England Baptist Hospital, single surgeon	162 RSA patients with mostly OA or CTA, 2015 – 2018
Franovic 2020	Retrospective cohort	USA, Henry Ford Health System, single surgeon	73 RSA patients with CTA, recruitment period unknown
Friedman 2018	Retrospective cohort	USA, University of Florida and NYU Langone Orthopedic Hospital, 13 surgeons	660 RSA patients with CTA or OA with RC tear, 2007 – 2014
Green 2020	Retrospective cohort	USA, Brown University, single surgeon	176 TSA patients with advanced OA, recruitment period unknown
Huber 2020	Retrospective cohort	International, 5 clinics specialized in shoulder surgery (3 in Germany, 2 in France, 1 in Switzerland)	168 RSA patients with cuff arthropathy, recruitment period unknown
Kadum 2018**	Prospective cohort	Sweden, Sundsvall Teaching Hospital, 2 surgeons	63 TSA patients with OA, 2014 – 2016

Time point	Prognostic factors/factors adjusted for	Outcomes	Type of multivariable regression
Mean 28.3 months (± 8.1)	· OA/CTA/other diagnosis · Private/Medicaid/Medicare insurance · No. of patient-reported allergies · Depression · Previous surgery · Preoperative ASES score · Preoperative opioid use	· ASES top vs bottom quartile	Logistic
Mean 9.6 months (± 5.0)	· Preoperative PROMIS-UE score · Preoperative PROMIS-PI score · Preoperative PROMIS-D score · Age · Gender · BMI · ASA	· MCID PROMIS-UE · MCID PROMIS-PI · MCID PROMIS-D	Logistic
Variable time points (range <3 – 96 months)	· Gender · Age	· ASES change score · SST change score · SPADI change score	Linear mixed model
2 years	· Mean preoperative expectation score · Gender · Age · Dominant side · College education · Work status · Comorbidities · Depression/anxiety · Preoperative DASH score · Preoperative SST score · Preoperative VAS pain	· DASH raw score · DASH change score · SST raw score · SST change score · VAS pain score · VAS pain change	Linear
2 years	· Hamada grade · Age · Gender · Dominant side · ASA score	· ASES TE	Linear
1 year	· Gender · Age · Preoperative electrical pain threshold (operative side and contralateral side) · Preoperative pain at rest · Preoperative pain on exertion · Preoperative QuickDASH score	· QuickDASH raw score	Linear

Study ID	Study design	Country, setting	Analyzed participants
Kohan 2020	Retrospective cohort	USA, Washington University, 4 surgeons	111 TSA patients with OA, 2015 – 2017
Lansdown 2021*****	Prospective cohort	USA, University of California, 4 surgeons	491 TSA/RSA/HA patients with mostly OA/CTA, 2012 – 2017
Lapner 2015	Retrospective cohort	Canada, multicenter	62 TSA patients with OA, 2006 – 2009
Matsen 2019	Retrospective cohort	International, multicenter (11 centers)	691 TSA patients with OA, 2000 – 2016
McFarland 2021	Retrospective cohort	USA, Johns Hopkins University, single surgeon	44 RSA patients, 2009 – 2016

Time point	Prognostic factors/factors adjusted for	Outcomes	Type of multivariable regression
Mean 20.5 months (12 – 29 months)	· Glenoid concentricity · Anxiety · Depression · Gender · Age · CCI · Smoking status · Dominant side	· ASES raw score · WOOS raw score · Pain on 0 – 10 scale · PROMIS-UE raw score · PROMIS- PF score · PROMIS-PI score	Linear
Min. 2 years	· Age · BMI · Gender · CCI · Private/Medicare/Medicaid insurance · Zip code-based income	· ASES raw score	Linear
1 year	· Preoperative SSP FI · Preoperative SSP occupation area · Preoperative SSP atrophy grade · Preoperative ISP FI · Preoperative ISP atrophy grade · Age · Gender · Hand dominance · Preoperative SSP strength · RA vs OA	· ASES raw score · WOOS raw score	Linear
1 year & 2 years	· Gender · Preoperative glenoid version (continuous) · Preoperative glenoid anteversion/retroversion/ · neutral version · Previous surgery · Gender · Age · Walch classification · Preoperative ASES score · Preoperative SST100 score	· ASES raw score · ASES %MPI · SST100 raw score · SST100 %MPI	Linear
Mean 34 months (24 – 62 months (SSDI group)), mean 36 months (24 – 96 months (WC group))	· 2 or more previous surgeries · Age · Gender · Dominant side · College education · Marital status · Smoking · BMI · Comorbidities · Level of activity at work	· Poor ASES score (cut-point unknown)	Logistic

Study ID	Study design	Country, setting	Analyzed participants
Moverman 2021*	Retrospective cohort	USA, New England Baptist Hospital, single surgeon	480 TSA/RSA patients with DJD/ RCA, 2015 – 2018
Okoroha 2019****	Retrospective cohort	International, multicenter (3 in USA, 1 in France)	2364 TSA/RSA patients with mostly OA/RCT/RCA, 2007 – 2015
Patel 2019****	Retrospective cohort	International, multicenter (3 in USA, 1 in France)	1135 TSA patients with OA, 2006 – 2015
Pettit 2022*	Retrospective cohort	USA, New England Baptist Hospital, single surgeon	197 RSA patients with OA, 2015 – 2018
Polce 2021	Retrospective cohort	USA, Rush University, 5 surgeons	105 RSA patients with OA/CTA/ mRCT, 2016 – 2018
Rauck 2018***	Retrospective cohort	USA, Hospital for Special Surgery, multiple surgeons	137 RSA patients with mostly OA/ CTA, 2008 – 2015

	Time point	Prognostic factors/factors adjusted for	Outcomes	Type of multivariable regression
	2 years	· Presence of functional somatic disorder · No. of patient-reported allergies · Gender · Anxiety · Chronic preoperative opioid use · Diagnosis · Preoperative ASES score · Preoperative SANE score · Preoperative VAS score	· ASES raw score · SANE raw score · VAS pain raw score	Linear
	Mean 45.9 months (± 23.7 (women)), 46.4 months (± 23.6 (men))	· Gender · Age · BMI · History of surgery · Diagnosis	· ASES raw score · SST raw score	Linear
	Mean 4.2 years (± 2.3 (age<55)), 4.3 years (± 2.4 (age>55))	· Age · Gender · BMI · Previous surgery · Diagnosis	· ASES raw score · ASES change score · SST raw score · SST change score · SPADI raw score · SPADI change score · VAS pain raw score · VAS pain change score	Linear
	Mean 28 months (± 7.6)	· Age · Gender · Previous shoulder surgery · Preoperative ASES score · Walch classification	· ASES raw score	Linear
	Min. 2 years	· Gender · WC/other insurance · Diabetes mellitus · Dominant side	· ASES %MOI · SANE %MOI	Logistic
	2 years	· Total no. of preoperative expectations · Expectation 'Relieving nighttime pain' · Expectation 'Improving nonoverhead sports' · Age · Gender · SF-36 subscales · Preoperative VAS pain score · Preoperative VAS fatigue score · Preoperative VAS general health score	· ASES raw score · SAS raw score · VAS pain raw score · VAS pain change score	Linear

Study ID	Study design	Country, setting	Analyzed participants
Saini 2022*	Retrospective cohort	USA, New England Baptist Hospital, single surgeon	311 RSA patients with OA/CTA, 2015 – 2018
Sayed-Noor 2018**	Prospective cohort	Sweden, Sundsvall Teaching Hospital	63 TSA patients with OA, 2014 – 2016
Shields 2017	Retrospective cohort	USA, Beaumont Health, single surgeon	272 RSA patients with CTA/OARCT/ mRCT, 2007 – 2014
Somerson 2016	Retrospective cohort	USA, University of Texas Health Science Center San Antonio, 2 surgeons	42 HA patients with CTA, 1991 – 2007
Strotman 2020	Retrospective cohort	USA, Loyola University Medical Center	91 TSA/RSA patients with mostly OA/CTA, 2013 – 2016
Swarup 2017***	Retrospective cohort	USA, Hospital for Special Surgery, multiple surgeons	67 TSA patients with OA, 2007 – 2008
Werner 2016***	Retrospective cohort	USA, Hospital for Special Surgery	490 TSA/RSA patients with GH arthritis/CTA, 2007 – 2013

Time point	Prognostic factors/factors adjusted for	Outcomes	Type of multivariable regression
Mean 28.1 months (± 7.6 (OA)), 27.6 months (± 7.3 (CTA))	· Age · Gender · Previous surgery · Preoperative ASES score · Diagnosis	· ASES raw score	Linear
3 months 1 year	· Gender · Age · Preoperative SSP FI · Preoperative SSP atrophy · Preoperative ISP FI · Preoperative ISP atrophy · Preoperative QuickDASH score	· QuickDASH raw score	Linear
Mean 25 months (± 13 (Previous RCR)) Mean 26 months (± 13 (Control))	· Previous RCR · Age · Gender · BMI	· ASES raw score · ASES change score · ASES ADL raw score · ASES ADL change score · SSV raw score · SSV change score · VAS pain raw score · VAS pain change score	Linear
Mean 48 months (24 – 132 months)	· Preoperative active external rotation · Preoperative SST score · Preoperative VAS pain score · Tear limited to SSP and ISP	· 30% of MPI in SST score	Logistic
Min. 1 year	· Private/Medicare/WC/ Medicaid insurance · Preoperative ASES score	· ASES raw score · VAS pain raw score	Linear mixed model
2 years	· No. of 'Very important expectations' · Age · Gender · Preoperative VAS pain score · Preoperative VAS general health score · SF-36 subscales	· ASES raw score · ASES change score · VAS pain raw score · VAS pain change score	Linear
2 years	· Preoperative ASES score · Living alone · Comorbidity back pain · Diagnosis	· MCID ASES score · SCB ASES score	Logistic

Study ID	Study design	Country, setting	Analyzed participants
Werner 2017***	Retrospective cohort	USA, Hospital for Special Surgery	264 TSA patients with GH arthritis, 2007 – 2013
Wong 2017*****	Prospective cohort	USA, University of California, 3 surgeons	117 RSA patients with RCA/OARCT, 2009 – 2015

For factors in Italic print, it is unknown which factors were retained in the final model
* Forlizzi 2022, Moverman 2021, Pettit 2022 and Saini 2022 appear to have used the same source population (Carducci 2019 partly)
** Kadum 2018 and Sayed-Noor 2018 appear to have used the same source population
*** Werner 2016 and Werner 2017 appear to have used the same source population and possibly overlap with Rauck 2018 and Swarup 2017
**** Okoroha 2019 and Patel 2019 appear to have (partly) used the same source population
***** Lansdown 2021 and Wong 2017 appear to have partly used the same source population
(m)RCT = (massive) rotator cuff tear, (Quick)DASH = (Quick) Disabilities of Arm, Shoulder and Hand, ASA = American Society of Anesthesiologists, ASES = American Shoulder and Elbow Surgeons, BMI = body mass index, CCI = Charlson Comorbidity Index, CTA = cuff tear arthropathy, CTA-H = cuff tear arthropathy – hemiarthroplasty, DJD = degenerative joint disease, FI = fatty infiltration, GH = glenohumeral, HA = hemiarthroplasty, HADS = Hospital Anxiety and Depression Scale, ISP = infraspinatus, KL = Kellgren & Lawrence, MCID = minimal clinically important difference, MOI = maximal outcome improvement, MPI = maximal possible improvement, OA = osteoarthritis, OARCT = osteoarthritis with rotator cuff tear, OSS = Oxford Shoulder Score, PROMIS-D = Patient-Reported Outcomes Measurement Information System – Depression, PROMIS-PF = Patient-Reported Outcomes Measurement Information System – Physical Functioning, PROMIS-PI = Patient-Reported Outcomes Measurement Information System – Pain Interference, PROMIS-UE = Patient-Reported Outcomes Measurement Information System – Upper Extremity, RA = rheumatoid arthritis, RC = rotator cuff, RCA = rotator cuff arthropathy, RCR = rotator cuff repair, RSA = reverse shoulder arthroplasty, SANE = Single Assessment Numeric Evaluation, SCB = substantial clinical benefit, SF-36 = 36-Item Short Form Health Survey, SPADI = Shoulder Pain and Disability Index, SSDI = social security disability insurance, SSP = supraspinatus, SST = Simple Shoulder Test, SSV = Subjective Shoulder Value, SSV = Subjective Shoulder Value, TE = treatment effects, TSA = total shoulder arthroplasty, USA = United States of America, UK = United Kingdom, VAS = Visual Analogue Scale, WC = workers' compensation, WOOS = Western Ontario Osteoarthritis of the Shoulder

	Time point	Prognostic factors/factors adjusted for	Outcomes	Type of multivariable regression
	2 years	· Depression · Age · Gender · ASA · Previous surgery · Subscapularis management · RC status · Total no. of comorbidities · BMI · Preoperative ASES score	· ASES change score	Linear
	Min. 1 year	· Gender · BMI · Age · CCI · Dominant side · Smoking	· ASES Function raw score · ASES pain raw score · ASES raw score · VAS pain raw score	Linear

Meta-analyses

There was a high degree of methodological heterogeneity: outcomes were measured using different methods even when the same PROM was used. For example, male gender as prognostic factor for ASES scores was studied with raw scores, change scores, treatment effects, percentage of maximal possible improvement and dichotomized scores. Therefore, we could only pool results for six combinations of prognostic factor and outcome measurements for the domain 'Functional recovery & pain': one for medium-short term (age on raw ASES scores[23, 36] and five for medium-long term (male gender[36, 41], age[36, 44], Walch A2 vs A1[36, 41] and previous surgery[36, 44] on raw ASES scores, and previous surgery on poor ASES scores[27, 37]).

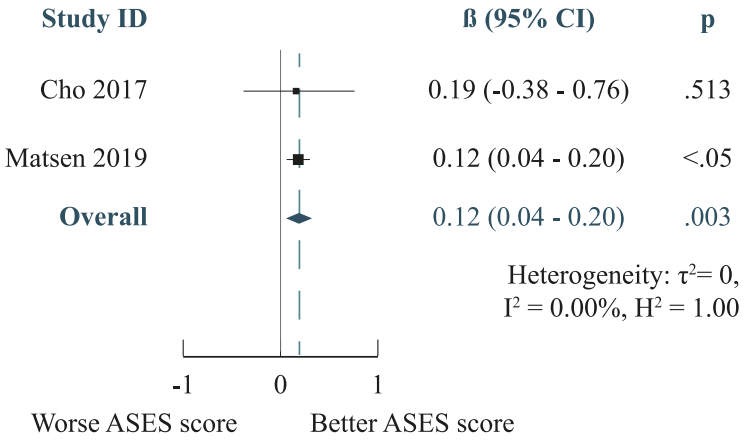


Fig. 3 Meta-analysis results for prognostic factor "age" on raw ASES scores, medium-short term

Only the meta-analysis for medium-long term poor ASES scores indicated worse outcomes for patients who had undergone previous shoulder surgery (OR (95%CI) of 2.10 (1.33 – 3.33)). All other meta-analyses (for raw ASES scores) led to ORs and 95% CIs below the MCID, indicating neutral effects. The exact results are displayed in Figures 3 – 8. The planned sensitivity analyses did not alter the meaning of the pooled effect estimates and confidence intervals (Figures A1 – A2 in Appendix 2).

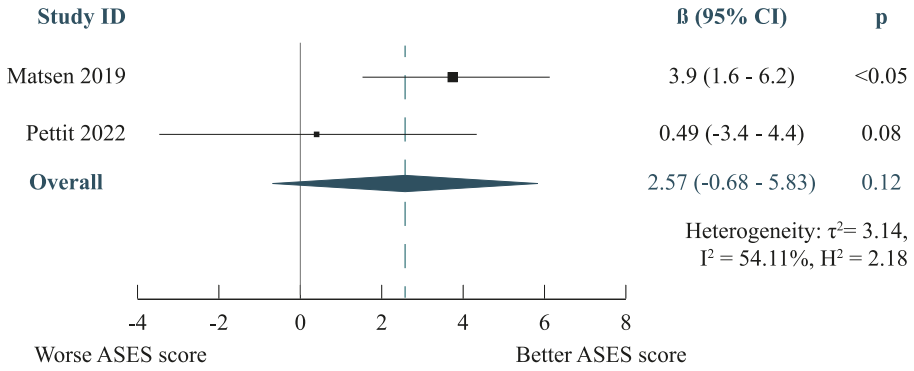


Fig. 4 Meta-analysis results for prognostic factor "gender" on raw ASES scores, medium-long term

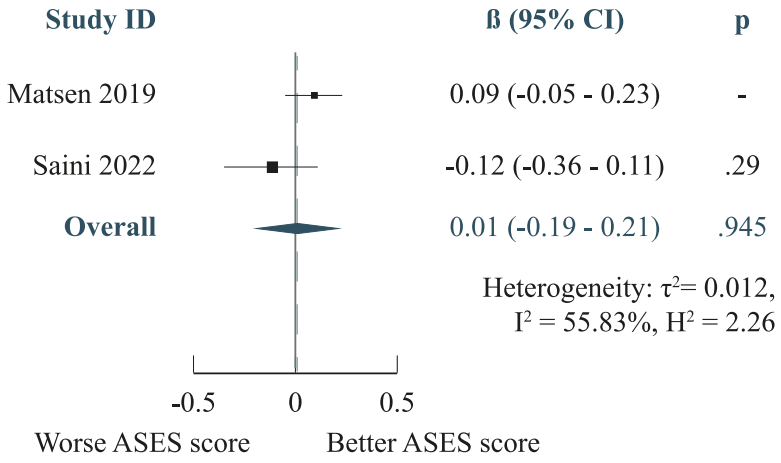


Fig. 5 Meta-analysis results for prognostic factor "age" on raw ASES scores, medium-long term

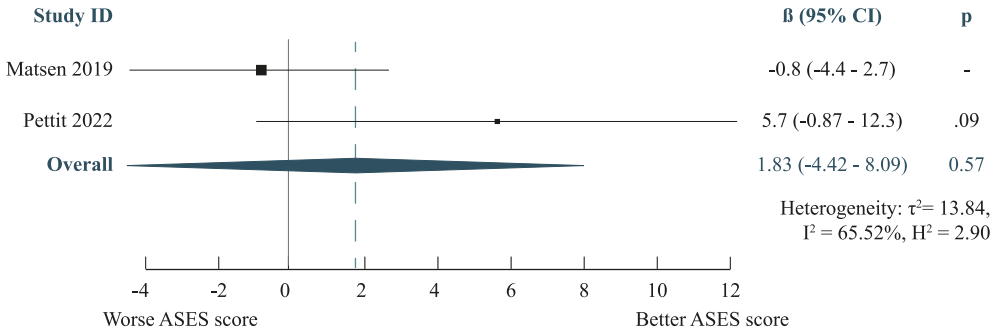


Fig. 6 Meta-analysis results for prognostic factor "Walch A2 vs A1" on raw ASES scores, medium-long term

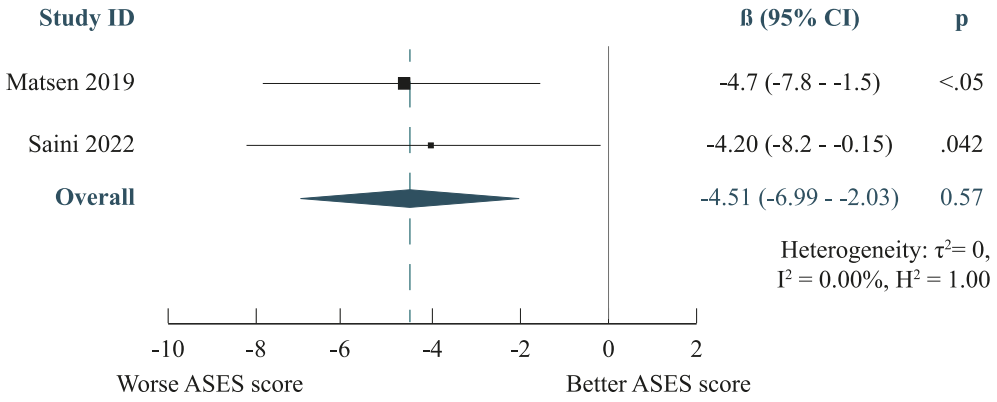


Fig. 7 Meta-analysis results for prognostic factor "Previous surgery" on raw ASES scores, medium-long term

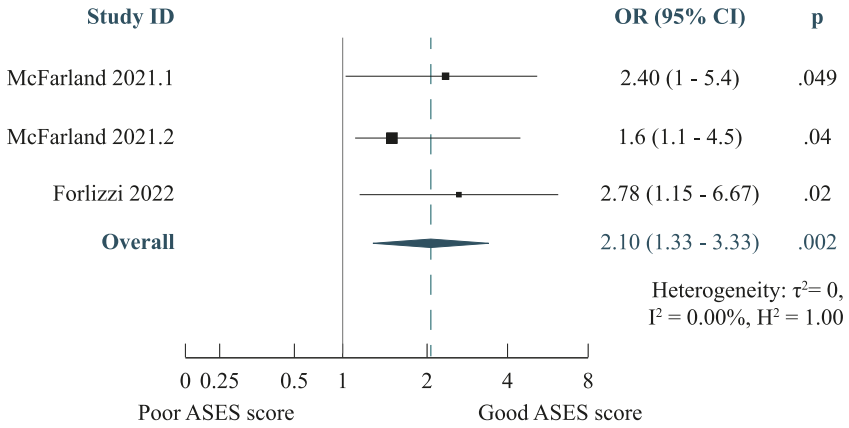


Fig. 8 Meta-analysis results for prognostic factor "Previous surgery" on the odds of achieving poor ASES scores, medium-long term

Qualitative analysis and GRADE

For each construct per time point, SoF Figures 9 - 11 summarize which prognostic factors were studied for medium-short-term outcomes, the direction of their effect and the certainty of the evidence according to GRADE. Figures A3 – A8 (Appendix 2) depict the short-term, medium-long and long-term outcomes. For the outcomes for which meta-analysis was possible, the direction and certainty were determined by combining that pooled result with the results from the additional studies that could not be meta-analyzed.

As the SoF figures show, no factor investigated in at least two studies showed distinct better or worse outcomes, but at most neutral to better or neutral to worse outcomes with varying degrees of certainty. In total, only nine factors showed better or worse outcomes, all examined in single studies with thus unknown certainty. The majority of factors were found to be either neutral or unclear in direction of effect based on the MCID.

For more details, tables A3 – A17 in Appendix 1 show all results from the individual studies, ordered per prognostic factor. It is clear that the effect estimates were not always reported for all variables included in the analyses. Table A18 in Appendix 1 provides additional details on which domains the evidence was rated down or up.

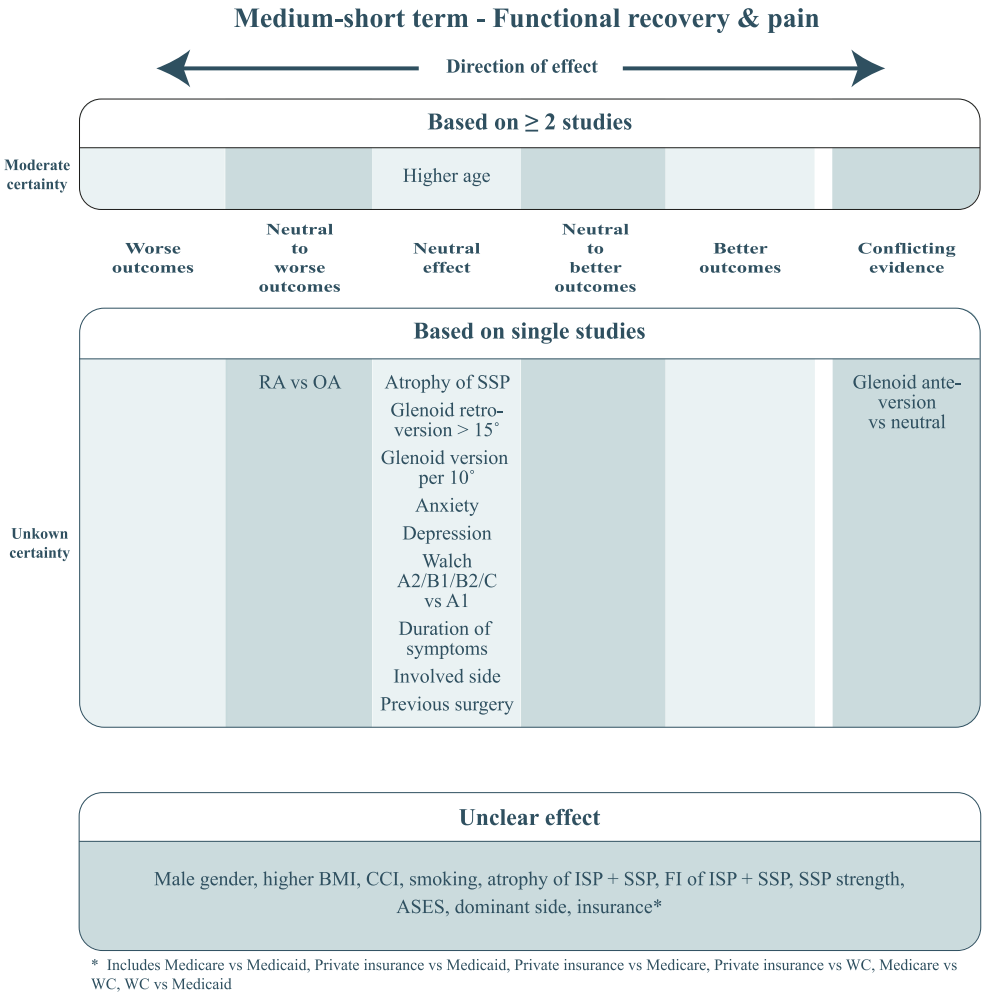


Fig. 9 Summary of Findings for the domain "Functional Recovery & Pain", medium-short term

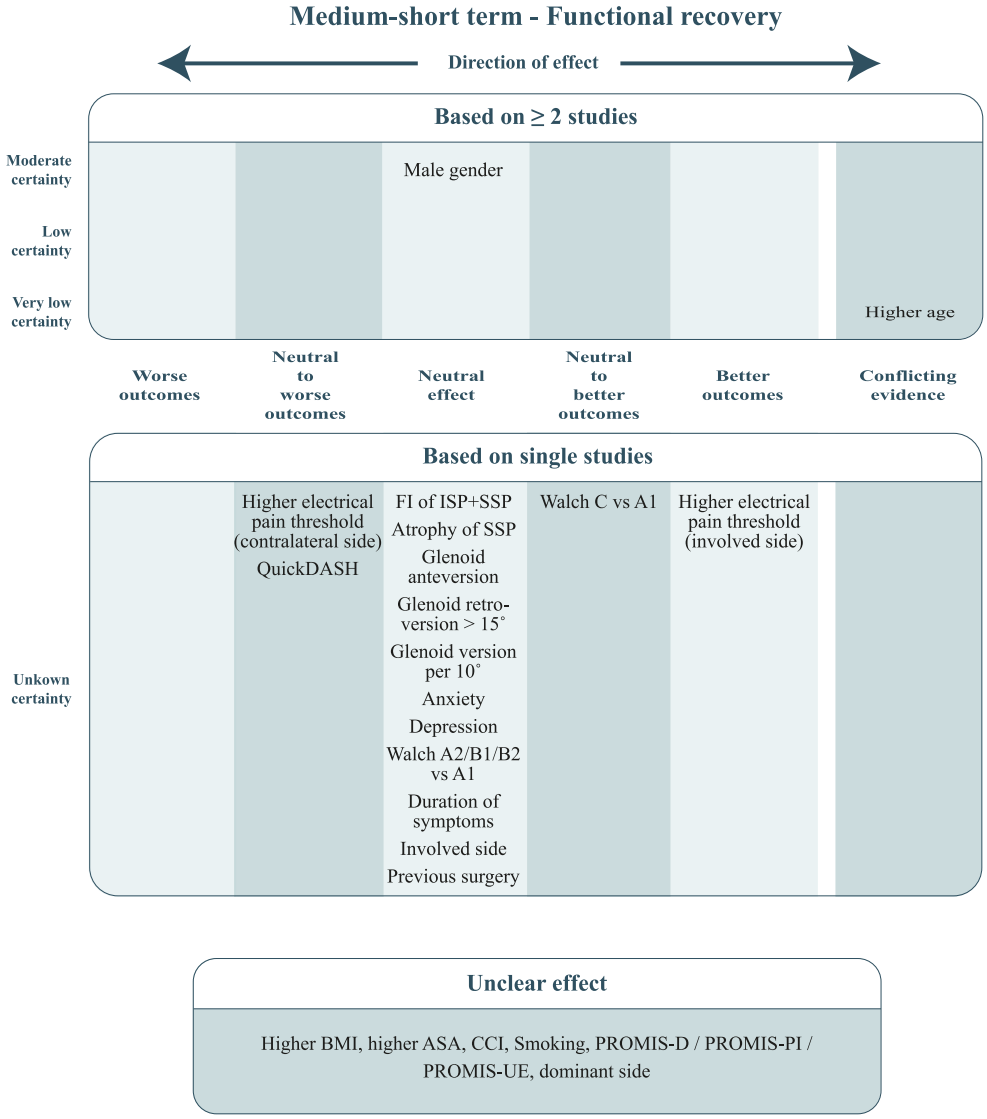


Fig. 10 Summary of Findings for the domain "Functional Recovery", medium-short term

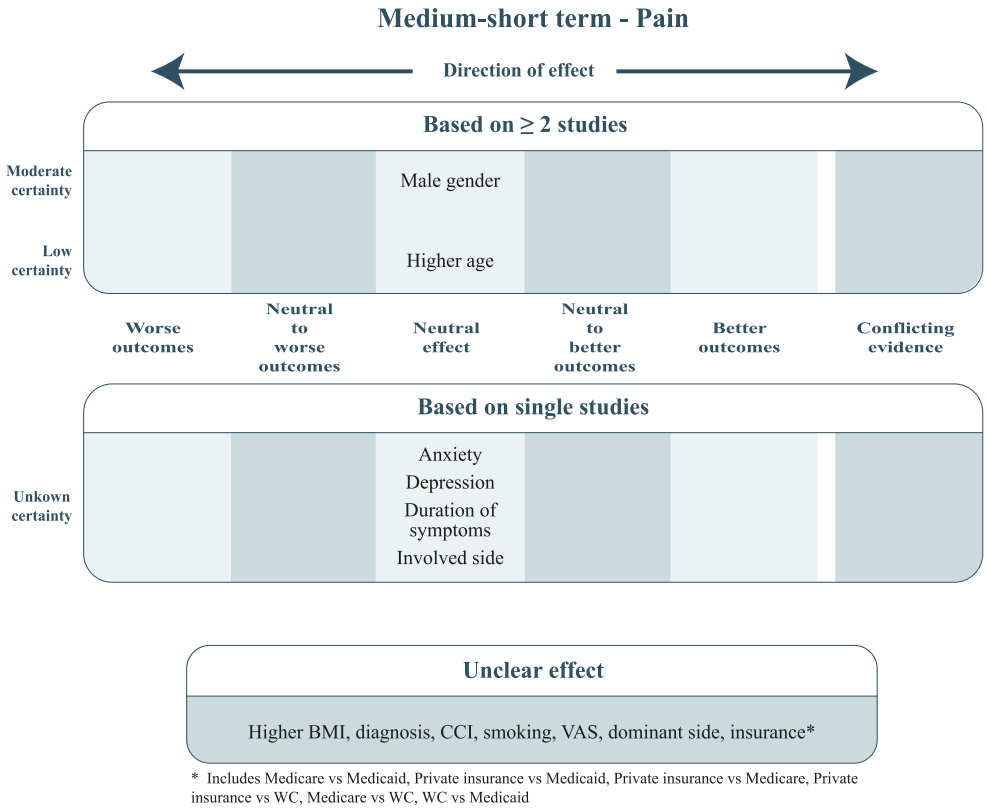


Fig. 11 Summary of Findings for the domain "Pain", medium-short term

Discussion

The results of our review suggest that the strongest independent prognostic factors for better functional and pain-related outcomes after SA are 'OA vs other diagnosis', 'private insurance' and 'higher electrical pain threshold on the operative side', and for worse outcomes 'number of patient-reported allergies', 'back pain', 'living alone', 'CTA vs OA', 'diabetes' and 'greater preoperative range of motion in external rotation'. The evidence for these strongest predictors is of unknown certainty. Since they were only examined in single studies, using the GRADE system was not possible.

Prognostic factors that were examined in at least two studies and predicted neutral to better outcomes are 'male gender' (low certainty). Neutral to worse outcomes were predicted by 'depression' and 'preoperative opioid use' (moderate certainty), 'preoperative ASES scores', 'surgery on the dominant side' and 'previous surgery' (low certainty) and 'male gender' (very low certainty).

Our results support those of recent systematic reviews that focused on a single predictor and used the results of univariable analyses. For example, both Mirghaderi et al.[53] and Ardebol et al.[54] also concluded that SA patients with previous shoulder surgery had worse postoperative outcomes than those without previous surgery. Where Ardebol et al.[54] only studied patients with massive rotator cuff tears (RCT) without glenohumeral arthritis, our included studies mainly focused on patients with osteoarthritis[36, 44, 46], thereby strengthening the evidence for this prognostic factor. In addition, Al-Mohrej et al.[55] found similar results to ours for preoperative opioid use: patients with preoperative opioid use had worse outcomes after shoulder surgery (including but not limited to SA) than patients without. Interestingly, in the included studies, the preoperative scores were worse for patients with preoperative opioid use as well, suggesting that maybe worse postoperative outcomes were a result of already having worse preoperative scores. However, the studies included in our current review[20, 27] used multivariable analyses that adjusted for preoperative scores and still found neutral to worse effects.

Interestingly, for the majority of reported factors, the direction of effect was unclear or neutral. This does not necessarily mean that these factors predict no effect at all; for the 'unclear' factors it was simply impossible to extract sufficient data to determine the most likely direction. For the 'neutral' factors, some studies did report results that were statistically significantly different from predicting no effect, but the effect estimate and 95% CI failed to reach the MCID.

Only a few factors predicted distinct better or worse outcomes on their own when adjusted for other factors. These were reported only in single studies, and most originated from studies using dichotomized endpoints with high contrasts. For example, Forlizzi et al.[27] compared patients scoring within the top 25th percentile of ASES scores with those scoring within the bottom 25th percentile and Werner et al.[50] used the cut-off of having

reached the substantial clinical benefit (SCB, in most instances larger than the MCID). Using smaller contrasts could still be relevant to patients and could lead to different and/or smaller effects.

Strengths and limitations

An obvious limitation of our study is that we could not provide pooled estimates for many combinations of prognostic factors and outcomes despite the large number of studies we included. Two mechanisms were responsible for this. Firstly, studies used many different ways to analyze the same PROMs, resulting in important methodological heterogeneity. Secondly, the sheer amount of incomplete reporting and selective reporting in prognostic factor studies was a major problem, as has been noted before by other authors[56, 57]. The grouping of individual study results in Appendix 1 shows that for prognostic factors 'higher age', 'male gender', 'anxiety' and 'dominant side', additional meta-analyses would have been possible if only all results from the multivariable analysis were reported.

We can also highlight several strengths of this study.

Firstly, we did not limit ourselves to a predefined set of prognostic factors to include in the review, allowing us to compile all evidence from multivariable analyses currently available on this topic.

Secondly, although we could only perform a few meta-analyses, we have achieved a rigorous qualitative synthesis by using clear, transparent definitions to determine the direction of effect. Most systematic reviews without meta-analysis simply list the individual study results, while our approach enabled us to summarize the results. Thirdly, we incorporated all combinations of prognostic factors and outcomes of the individual studies into our tables, whether the effect estimates were reported or not. This ensures transparency for which combinations information is missing, especially if only 'p > 0.05' was reported. Lastly, we applied the GRADE system to our qualitative analysis. While GRADE is designed foremost to grade quantitatively pooled estimates, it is no less important to state the overall certainty of available evidence in qualitative data synthesis. This also enabled us to take incomplete and selective reporting into account by downgrading the certainty of the evidence on the domain 'publication bias' when present.

Conclusion

Our systematic review has provided the first overview of current knowledge on the independent predictive effect of patient-related prognostic factors on functional recovery and pain-related outcomes after SA.

Based on the current body of evidence, a limited number of factors have a clinically relevant independent predictive effect, with an unknown certainty of evidence for predicting better or worse outcomes and mostly (very) low certainty of evidence for predicting neutral to better or neutral to worse outcomes. Promising factors for use in daily practice are a diagnosis of OA, multiple types of comorbidities, gender, insurance, living alone, multiple measures of preoperative status (e.g. ASES score, depression, opioid use, previous surgery, ROM) and surgery on the dominant side. However, further research is certainly necessary to confirm or nuance these results.

We strongly encourage future prognostic factor studies to report their complete findings and use more homogeneous outcomes. This will enable future research to refine the results of our first overview with more pooled results and clearer levels of certainty.

Supplementary materials

The supplementary materials can be viewed online at https://doi.org/10.1186/s13643-024-02694-y

Appendix 1 - Supplementary tables

Appendix 2 - Supplementary figures

References

1. Wagner ER, Farley KX, Higgins I, Wilson JM, Daly CA, Gottschalk MB. The incidence of shoulder arthroplasty: rise and future projections compared with hip and knee arthroplasty. J Shoulder Elbow Surg. 2020;**29**(12):2601-9.
2. Wilson CD, Probe RA. Shared Decision-making in Orthopaedic Surgery. J Am Acad Orthop Surg. 2020;**28**(23):e1032-e41.
3. Vajapey SP, Cvetanovich GL, Bishop JY, Neviaser AS. Psychosocial factors affecting outcomes after shoulder arthroplasty: a systematic review. J Shoulder Elbow Surg. 2020;**29**(5):e175-e84.
4. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. Bmj. 2017;356:i6460.
5. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;**11**(10):e1001744.
6. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;**5**(1):210.
7. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;**158**(4):280-6.
8. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Res Synth Methods. 2019;**10**(1):83-98.
9. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods. 2016;**7**(1):55-79.
10. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
11. Posit team. RStudio: Integrated Development Environment for R. Boston, MA: Posit software, PBC; 2023.
12. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of statistical software. 2010;**36**(3):1-48.
13. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Bmj. 2008;**336**(7650):924-6.
14. Foroutan F, Guyatt G, Zuk V, Vandvik PO, Alba AC, Mustafa R, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. J Clin Epidemiol. 2020;121:62-70.
15. Livesey M, Horneff JG, 3rd, Sholder D, Lazarus M, Williams G, Namdari S. Functional Outcomes and Predictors of Failure After Rotator Cuff Repair During Total Shoulder Arthroplasty. Orthopedics. 2018;**41**(3):e334-e9.
16. Matsoukis J, Tabib W, Guiffault P, Walch G. [Shoulder arthroplasty for osteoarthritis after prior surgery for anterior instability: a report of 27 cases]. Rev Chir Orthop Reparatrice Appar Mot. 2003;**89**(7):580-92.
17. Matsoukis J, Tabib W, Mandelbaum A, Walch G. [Shoulder arthroplasty for non-operated anterior shoulder instability with secondary osteoarthritis]. Rev Chir Orthop Reparatrice Appar Mot. 2003;**89**(1):7-18.

18. Moore AR, Brolin TJ, Smith RA, Azar FM, Throckmorton TW. Outcomes after primary reverse total shoulder arthroplasty in patients with comorbid anxiety and depressive disorders: a retrospective cohort study. Current Orthopaedic Practice. 2021;**32**(1):37-42.
19. Ode GE, Ling D, Finocchiaro A, Lai EY, Taylor SA, Dines J, et al. Clinical characteristics and patient-reported outcomes of total shoulder arthroplasty after anterior stabilization: a retrospective matched control study. J Shoulder Elbow Surg. 2020;**29**(7s):S59-s66.
20. Carducci MP, Zimmer ZR, Jawa A. Predictors of unsatisfactory patient outcomes in primary reverse total shoulder arthroplasty. J Shoulder Elbow Surg. 2019;**28**(11):2113-20.
21. Chang NB, Bicknell R, Krupp R, Wiater JM, Levy J, Athwal GS. Sex-related differences in stemless total shoulder arthroplasty. JSES Int. 2022;**6**(1):26-31.
22. Chawla SS, Schiffman CJ, Whitson AJ, Matsen FA, 3rd, Hsu JE. Drivers of inpatient hospitalization costs, joint-specific patient-reported outcomes, and health-related quality of life in shoulder arthroplasty for cuff tear arthropathy. J Shoulder Elbow Surg. 2022;**31**(12):e586-e92.
23. Cho CH, Song KS, Hwang I, Coats-Thomas MS, Warner JJP. Changes in Psychological Status and Health-Related Quality of Life Following Total Shoulder Arthroplasty. J Bone Joint Surg Am. 2017;**99**(12):1030-5.
24. Dekker TJ, Grantham WJ, Lacheta L, Goldenberg BT, Dey Hazra RO, Rakowski DR, et al. Glenoid retroversion does not impact clinical outcomes or implant survivorship after total shoulder arthroplasty with minimal, noncorrective reaming. JSES Int. 2022;**6**(4):596-603.
25. Edwards GA, McCann PA, Whitehouse MR, Wakeley CJ, Sarangi PP. The influence of fatty infiltration and muscle atrophy of the rotator cuff muscles on midterm functional outcomes in total shoulder resurfacing at six years' follow-up. Shoulder Elbow. 2020;**12**(2):91-8.
26. Fehringer EV, Kopjar B, Boorman RS, Churchill RS, Smith KL, Matsen FA, 3rd. Characterizing the functional improvement after total shoulder arthroplasty for osteoarthritis. J Bone Joint Surg Am. 2002;**84**(8):1349-53.
27. Forlizzi JM, Puzzitiello RN, Hart PA, Churchill R, Jawa A, Kirsch JM. Predictors of poor and excellent outcomes after reverse total shoulder arthroplasty. J Shoulder Elbow Surg. 2022;**31**(2):294-301.
28. Franovic S, Kuhlmann N, Schlosser C, Pietroski A, Buchta AG, Muh SJ, editors. Role of preoperative PROMIS scores in predicting postoperative outcomes and likelihood of achieving MCID following reverse shoulder arthroplasty. Seminars in Arthroplasty: JSES; 2020: Elsevier.
29. Friedman RJ, Cheung EV, Flurin PH, Wright T, Simovitch RW, Bolch C, et al. Are Age and Patient Gender Associated With Different Rates and Magnitudes of Clinical Improvement After Reverse Shoulder Arthroplasty? Clin Orthop Relat Res. 2018;**476**(6):1264-73.
30. Green A, Neradilek MB, Thompson KM, Mayer A, editors. Expectations affect outcome and satisfaction after anatomic total shoulder arthroplasty. Seminars in Arthroplasty: JSES; 2020: Elsevier.
31. Huber J, Irlenbusch U, Kääb MJ, Reuther F, Kohut G, Judge A. Treatment effects of reverse total shoulder arthroplasty - a simple method to measure outcomes at 6, 12, 24 and 60 months for each patient. BMC Musculoskelet Disord. 2020;**21**(1):397.
32. Kadum B, Inngul C, Ihrman R, Sjödén GO, Sayed-Noor AS. Higher preoperative sensitivity to pain and pain at rest are associated with worse functional outcome after stemless total shoulder arthroplasty: a prospective cohort study. Bone Joint J. 2018;100-b(4):480-4.

33. Kohan EM, Aleem AW, Chamberlain AM, Keener JD, editors. The influence of mental health on outcomes following total shoulder arthroplasty. Seminars in Arthroplasty: JSES; 2020: Elsevier.
34. Lansdown DA, Ma GC, Aung MS, Gomez A, Zhang AL, Feeley BT, Ma CB. Do patient outcomes and follow-up completion rates after shoulder arthroplasty differ based on insurance payor? J Shoulder Elbow Surg. 2021;**30**(1):65-71.
35. Lapner PL, Jiang L, Zhang T, Athwal GS. Rotator cuff fatty infiltration and atrophy are associated with functional outcomes in anatomic shoulder arthroplasty. Clin Orthop Relat Res. 2015;**473**(2):674-82.
36. Matsen FA, 3rd, Iannotti JP, Churchill RS, De Wilde L, Edwards TB, Evans MC, et al. One and two-year clinical outcomes for a polyethylene glenoid with a fluted peg: one thousand two hundred seventy individual patients from eleven centers. Int Orthop. 2019;**43**(2):367-78.
37. McFarland EG, Meshram P, Binkley M, Joseph J, Pushkin GW, Srikumaran U, editors. Clinical and patient-reported outcomes of reverse shoulder arthroplasty in patients receiving Social Security Disability Insurance or workers' compensation. Seminars in Arthroplasty: JSES; 2021: Elsevier.
38. Moverman MA, Puzzitiello RN, Pagani NR, Moon AS, Hart PA, Kirsch JM, et al. Functional somatic syndromes are associated with suboptimal outcomes and high cost after shoulder arthroplasty. J Shoulder Elbow Surg. 2022;**31**(1):48-55.
39. Okoroha KR, Muh S, Gabbard M, Evans T, Roche C, Flurin PH, et al. Early outcomes of shoulder arthroplasty according to sex. JSES Open Access. 2019;**3**(1):43-7.
40. Patel RB, Muh S, Okoroha KR, Wright TW, Flurin PH, Roche C, Zuckerman JD. Results of total shoulder arthroplasty in patients aged 55 years or younger versus those older than 55 years: an analysis of 1135 patients with over 2 years of follow-up. J Shoulder Elbow Surg. 2019;**28**(5):861-8.
41. Pettit RJ, Saini SB, Puzzitiello RN, Hart PJ, Ross G, Kirsch JM, Jawa A. Primary reverse total shoulder arthroplasty performed for glenohumeral arthritis: does glenoid morphology matter? J Shoulder Elbow Surg. 2022;**31**(5):923-31.
42. Polce EM, Cohn MR, Kunze KN, Fu MC, Forsythe B, Nicholson GP, et al., editors. Defining maximal outcome improvement thresholds for patient satisfaction after reverse total shoulder arthroplasty: a minimum two-year follow-up study. Seminars in Arthroplasty: JSES; 2021: Elsevier.
43. Rauck RC, Swarup I, Chang B, Dines DM, Warren RF, Gulotta LV, Henn RF, 3rd. Effect of preoperative patient expectations on outcomes after reverse total shoulder arthroplasty. J Shoulder Elbow Surg. 2018;**27**(11):e323-e9.
44. Saini SS, Pettit R, Puzzitiello RN, Hart P-A, Shah SS, Jawa A, Kirsch JM. Clinical outcomes after reverse total shoulder arthroplasty in patients with primary glenohumeral osteoarthritis compared with rotator cuff tear arthropathy: does preoperative diagnosis make a difference? Journal of the American Academy of Orthopaedic Surgeons. 2022;**30**(3):e415-e22.
45. Sayed-Noor AS, Pollock R, Elhassan BT, Kadum B. Fatty infiltration and muscle atrophy of the rotator cuff in stemless total shoulder arthroplasty: a prospective cohort study. J Shoulder Elbow Surg. 2018;**27**(6):976-82.
46. Shields EJW, Koueiter DM, Maerz T, Schwark A, Wiater JM. Previous Rotator Cuff Repair Is Associated With Inferior Clinical Outcomes After Reverse Total Shoulder Arthroplasty. Orthop J Sports Med. 2017;**5**(10):2325967117730311.
47. Somerson JS, Sander P, Bohsali K, Tibbetts R, Rockwood CA, Jr., Wirth MA. What Factors are Associated With Clinically Important Improvement After Shoulder Hemiarthroplasty for Cuff Tear Arthropathy? Clin Orthop Relat Res. 2016;**474**(12):2682-8.

48. Strotman P, Perry M, LeDuc R, Joyce C, Garbis N. Effect of Insurance Status on Clinical Outcomes After Shoulder Arthroplasty. Orthopedics. 2020;**43**(6):e523-e8.
49. Swarup I, Henn CM, Nguyen JT, Dines DM, Craig EV, Warren RF, et al. Effect of pre-operative expectations on the outcomes following total shoulder arthroplasty. Bone Joint J. 2017;99-b(9):1190-6.
50. Werner BC, Chang B, Nguyen JT, Dines DM, Gulotta LV. What Change in American Shoulder and Elbow Surgeons Score Represents a Clinically Important Change After Shoulder Arthroplasty? Clin Orthop Relat Res. 2016;**474**(12):2672-81.
51. Werner BC, Wong AC, Chang B, Craig EV, Dines DM, Warren RF, Gulotta LV. Depression and Patient-Reported Outcomes Following Total Shoulder Arthroplasty. J Bone Joint Surg Am. 2017;**99**(8):688-95.
52. Wong SE, Pitcher AA, Ding DY, Cashman N, Zhang AL, Ma CB, Feeley BT. The effect of patient gender on outcomes after reverse total shoulder arthroplasty. J Shoulder Elbow Surg. 2017;**26**(11):1889-96.
53. Mirghaderi P, Azarboo A, Ghaseminejad-Raeini A, Eshraghi N, Vahedi H, Namdari S. Shoulder Arthroplasty After Previous Nonarthroplasty Surgery: A Systematic Review and Meta-Analysis of Clinical Outcomes and Complications. JBJS Rev. 2024;**12**(3).
54. Ardebol J, Menendez ME, Narbona P, Horinek JL, Pasqualini I, Denard PJ. Reverse shoulder arthroplasty for massive rotator cuff tears without glenohumeral arthritis can improve clinical outcomes despite history of prior rotator cuff repair: A systematic review. J isakos. 2024;**9**(3):394-400.
55. Al-Mohrej OA, Prada C, Madden K, Shanthanna H, Leroux T, Khan M. The role of preoperative opioid use in shoulder surgery-A systematic review. Shoulder Elbow. 2023;**15**(3):250-73.
56. Riley RD, Moons KG, Snell KI, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. bmj. 2019;364.
57. Sauerbrei W, Haeussler T, Balmford J, Huebner M. Structured reporting to improve transparency of analyses in prognostic marker studies. BMC medicine. 2022;**20**(1):184.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during this systematic review are available from the corresponding author (BH) on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

BH wrote the first draft of the manuscript. BH and NP contributed to the review's conception and design, the acquisition, analysis and interpretation of the data and the article's drafting.

FJ contributed to the design and acquisition of the work. SP contributed to the acquisition, analysis and interpretation of the data. DE revised the manuscript. NM and BS contributed to the review's conception and design, the analysis and interpretation of the data, and the revision of the manuscript.

All authors have read and approved the submitted version of the manuscript and take responsibility for the accuracy and integrity of the review.

3

Trajectories of pain recovery during the first 8 weeks after shoulder arthroplasty: results from the Shoulder Diary study using latent growth curve modeling

Hesseling B, Snoeker BA, Ten Brinke B, Slot K, Zwitser EW, van Rosmalen J, Eygendaal D, Mathijssen NM

Journal of Shoulder and Elbow Surgery 2026; 35(2); 438

Abstract

Objective

Perioperative stress or discomfort in shoulder arthroplasty (SA) patients can be reduced using more individually tailored patient education and expectation management. Most published studies assess pain and function for the first time at six weeks or three months. Consequently, there is no thorough understanding of day-to-day recovery trajectories within the first postoperative weeks, hindering effective patient education and expectation management in the early postoperative phase.

In this study, we explored the distinct pain recovery trajectories that emerge for SA patients during their first eight postoperative weeks and examined how patients in the identified subgroups differ in terms of sociodemographic and psychological factors.

Methods

In our prospective multicenter cohort study, we included 230 SA patients who completed an eight-week postoperative diary containing daily Numeric Rating Scale of pain (NRS) scores and medication use, weekly function scores and twice-weekly quality of life scores. In addition, patients completed preoperative questionnaires regarding pain, function and demographic and psychological factors.

We used Latent Growth Curve Modeling (LGCM) to classify groups of patients based on their early pain recovery trajectories; models included smooth functions based on natural cubic splines to represent the different trajectories of pain scores over time in the latent classes.

Results

Our final model contained six different classes whose trajectories differed during the first two weeks. The model contained random intercepts (i.e., allowed for between-person variability around the initial pain score) and fixed slopes (i.e., did not allow for between-person variability in subsequent change in pain scores over time) within each class.

After the first two weeks, classes 1 through 4 (83.7%) were similarly stable with very low pain scores (the 'Faster group'). Classes 5 and 6 (16.3%) had a slower decline in pain scores (the 'Slower group'), but comparable scores to the Faster group at week eight.

The Slower group also more frequently had ASA ≥ 3, were less often employed, and had lower baseline OSS and EQ-5D VAS scores. Both groups had similar recovery rates in OSS and EQ-5D VAS scores, although the Slower group had lower scores than the Faster group.

Conclusion

In this study, we distinguished six early recovery trajectories after TSA.

Our results enable clinicians to reassure their patients before surgery, as five out of six patients likely have very low pain scores (Numeric Rating Scale ≤ 2) after only two weeks. Also, the sixth patient is at almost similar low pain scores at eight weeks post-surgery.

Level of Evidence: level I (High-quality prospective cohort study with >80% follow-up)

Keywords: shoulder arthroplasty, short-term pain recovery, longitudinal pain recovery trajectories, latent growth curve modeling, latent class growth analysis, growth mixture modeling

Introduction

In most published studies focusing on postoperative pain and clinical outcomes in shoulder arthroplasty (SA), the first reported follow-up is generally set at six weeks or even three months after surgery[8, 9, 21, 44]. Consequently, we still have no thorough understanding of early recovery trajectories within the first eight weeks after surgery.

While we know from daily practice that recovery trajectories may strongly differ between patients in the first weeks after surgery, our knowledge of these specific trajectories and the risk factors for delayed recovery or increased pain scores is limited in the current literature. A deeper understanding of SA recovery trajectories and their predictors will enable more personalized preoperative education, expectation management and postoperative support, potentially reducing anxious feelings or insecurity which may then reduce perioperative stress and discomfort. Such trajectories can be studied using Latent Growth Curve Modeling (LGCM), which can be used to discern previously unobserved subgroups of patients that have similar recovery patterns.

For longer-term recovery, multiple demographic and surgical factors have already been studied. Several studies have, for example, shown that male gender predicts better outcomes after SA[15, 26, 31, 46], as does osteoarthritis as indication for surgery compared to other indications[5, 27, 43]. Examples of predictors for worse outcomes were having a history of previous shoulder surgery[31, 46] and a higher degree of preoperative opioid use[5, 14].

While demographic and surgical factors have been extensively studied in relation to longer-term recovery (e.g. recovery measured at one year follow-up), the role of psychological factors in postoperative outcomes has been a relatively recent focus in joint replacement research[13, 28, 49]. Factors such as catastrophizing, optimism, expectations and anxiety could significantly influence postoperative recovery trajectories and warrant further investigation.

The majority of these studies have been done in total knee (TKA) and total hip (THA) arthroplasty. However, results from TKA and THA studies cannot easily be extrapolated to SA as major differences exist in patient groups

(with SA patients generally being older and more often having ASA class 3 or higher[11]), underlying pathology and postoperative rehabilitation protocols. Consequently, little is known about the relation between psychological factors and clinical outcomes after SA.

Therefore, the primary aim of this study is to assess how many and which different types of pain recovery patterns can be discerned during the first eight weeks after SA. The secondary objective is exploratory in nature and is to investigate differences in baseline demographic characteristics, patient-reported outcome measures (PROMs) and psychological factors between the different trajectories.

Materials and Methods

Study population

This multicenter, prospective cohort study was assessed by our regional Medical Ethics Committee, who decided that the study did not fall under the scope of the Medical Research Involving Human Subjects Act because of the minimal patient burden. (METC Leiden, Den Haag, Delft in the Netherlands, METC-no. 17-117). Patients were recruited between April 2018 and September 2021 from the orthopedic departments in four different hospitals in The Netherlands: Reinier de Graaf Hospital and Haga Hospital (merged in 2020 to become the Reinier Haga Orthopedic Center (RHOC)), Alrijne Hospital and NorthWest Clinics. All patients placed on the waiting list for SA were screened consecutively and, when meeting study criteria, contacted for participation.

Inclusion criteria were age ≥ 18 years, scheduled to undergo primary or revision shoulder arthroplasty (either hemi arthroplasty (HA), anatomic or reverse total shoulder arthroplasty (aTSA or rTSA)), good command of the Dutch language, and able and willing to participate. Exclusion criteria were cognitive impairment, arthroplasty for acute fractures and difficulty with the Dutch language. All participating patients provided written informed consent prior to the study.

Data collection

As part of a larger study[33], patients completed a set of questionnaires before surgery (six weeks to one day preoperatively) and after surgery (at six months and one year). These questionnaires consisted of general demographic questions and questions regarding pain (Numeric Rating Scale (NRS) for pain (at rest and during activity)), function (Oxford Shoulder Score (OSS))[1], quality of life (EQ-5D-5L)[39] and psychological factors (Hospital for Special Surgery (HSS) Expectation questionnaire[30], Sunnybrook expectation questionnaire[41], Pain Catastrophizing Scale (PCS)[47], Life Optimism Test – Revised (LOT-R)[22, 45], Fear of Pain Questionnaire – 9 items (FPQ-9)[32] and Central Sensitization Index (CSI)[24]). In addition, the Tilburg Frailty Indicator[16] was used to assess frailty among the participants. For the LOT-R, we specifically used the optimism subscale instead of the total score, as Ten Klooster et al.[22] found that the Dutch version does not reflect a unidimensional construct and they thus advocate to use subscales.

The day after surgery, patients received a diary, which they were requested to fill in every day for eight weeks. In this diary, NRS scores were prompted daily (both average NRS and worst NRS for each day), EQ-5D-5L scores twice per week and OSS scores at the end of each week.

The current article focuses on the preoperative and diary data from the more extensive study. Results for data measured with the questionnaires at six months and one year will be described in future articles.

All study data was collected onto electronic case report forms (eCRFs) in Castor electronic data capture[6]. Deviations of ± one day were allowed between the actual day after surgery and the reported date in the diary. Data for days that deviated more than one day were entered into the eCRF as missing.

Surgical procedure, pain medication and rehabilitation protocol

To ensure the generalizability of our results, we did not standardize the surgical procedure and rehabilitation protocol; each hospital could follow its own procedures and had its own postoperative rehabilitation protocol. Standard postoperative pain medication consisted of paracetamol

combined with either Naproxen or Celecoxib, using Oxycodone (either instant or extended-release) as rescue medication.

In each hospital, patients were instructed to use a sling for six weeks. In the first two weeks, only passive exercises were allowed, after which exercises could gradually increase to active assisted and, finally, active exercises.

Statistical analysis

Descriptive statistics were used to describe the characteristics of the entire sample, using number and frequency for categorical variables. For normally distributed continuous data, the mean and standard deviation (SD), while for non-normally distributed data, the median, interquartile range (IQR) and range were used.

For the primary objective, we used Latent Growth Curve Modeling (LGCM) to classify groups of patients based on their early pain recovery trajectories, using the NRS for average pain scores from the diaries. With LGCM, it is possible to detect latent (i.e., previously unobserved) subgroups within the data: separate groups in which patients are similar with respect to the trajectory in health status over time. More specifically, LGCM can be used to group people together into separate classes by minimizing differences between individuals *within* each class and maximizing the differences between individuals *across* the classes.

To perform the LCGM analysis, we used R software[38] and RStudio[36], with the hlme() function from the 'lcmm' package[37] and natural cubic splines to model the relationship between time and pain scores. In doing so, we allowed for nonlinear growth trajectories, i.e., the trajectories were free to take on any shape or form. Since the computational burden of modeling each day of the eight weeks was too high, we entered every other day of the first two and last two weeks and every third day of the remaining weeks into the models. For the splines, knots were placed on days 6, 14, 22, 28, 34, 43, and 49.

Three different types of models were specified: a Latent Class Growth Analysis (LCGA, with fixed intercept and fixed slope) and two types of Growth Mixture Models (GMM-1, with random intercept and fixed slope

and GMM-2, with random intercept and random slope). For more details, see the Appendix.

After running all models, we used a combination of fit statistics (Akaike information criterion (AIC), Bayesian information criterion (BIC) and -log likelihood), visual inspection of the estimated mean trajectories with observed individual line plots, entropy and clinical relevance of the models to choose our final model. R syntax for the LCGA and GMM models can be found in the Appendix.

Since the literature has no concrete advice on sample size for LGCM, the sample size of 230 subjects was chosen arbitrarily and based on the feasibility of including the number of subjects.

For our secondary objective, we planned a priori to compare the different classes on baseline characteristics to investigate whether the trajectories in pain scores were associated with patient characteristics in an exploratory analysis. To compare these characteristics between the latent classes, nominal data were tested with chi-square tests or Fisher's exact tests, while normally distributed data were tested with independent samples t-tests (when comparing two groups) or ANOVA (when comparing more than two groups). Non-normally distributed data were tested with Mann-Whitney U tests or Kruskal-Wallis tests.

In addition, we studied whether OSS and EQ-5D VAS scores during the eight weeks differed between the groups by specifying linear mixed models to account for the repeated measurements within subjects, using the lmer() function from the 'lme4' package in R. For model specifications, see the Appendix. For medication use, we calculated the percentage of patients indicating they had used medication on the last day of each week. This was calculated for the overall use of any type of medication and for paracetamol, NSAIDs and opioids separately. Statistically significant differences between the classes were tested with chi-square tests or Fisher's exact tests.

Not all diaries were complete: some participants had missing NRS data for several days (consecutively or nonconsecutively), for which we found no apparent reasons for missingness. Four participants started with the diary but found it too strenuous to complete every day and withdrew from the study. For each participant, all available NRS data were used in the LGCM analyses; missing data were not imputed. Given the exploratory nature of the secondary analysis, no correction for multiple testing was implemented.

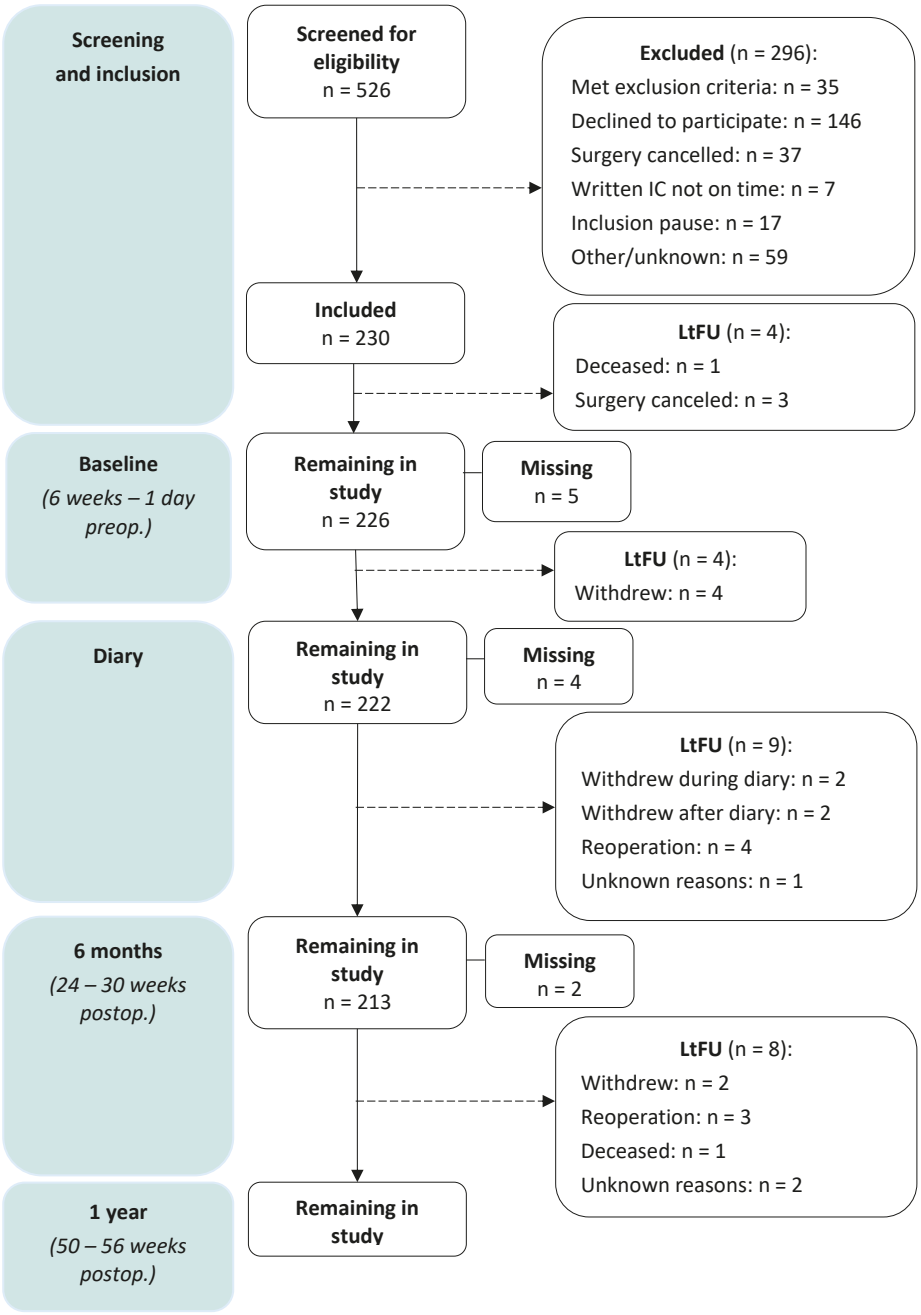
Results

Patient characteristics

A total of 491 patients were found eligible during the recruitment period, of whom 230 gave informed consent. Two hundred twenty-two patients remained in the study until after the diary, and 205 patients completed the entire study period of one year. For details, see the study flowchart in Figure 1.

The majority of patients were female (69.6%), and the mean age was 69.9 (SD 8.5) years. The most frequent indication for surgery was osteoarthritis (60.4%). rTSA was the most commonly used type of implant (66.1%). In all instances, surgeons used the deltopectoral approach and general anesthesia combined with an interscalene nerve block.

Table 1 displays all baseline characteristics for the entire sample.



*: The transition from Reinier de Graaf Hospital and Haga Hospital into Reinier Haga Orthopedic Center (RHOC) resulted in an inclusion pause of approximately two months for these centers

Fig. 1 Study flowchart

Table 1 Descriptive Statistics of Preoperative Patient Characteristics and Surgery Characteristics of the Entire Sample

Variable	Entire Sample (N = 230)
Demographic	
Age *(mean (SD))*	69.9 (8.5)
Sex *(no. (%))*	
Male	67 (29.1%)
Female	160 (69.6%)
ASA *(no. (%))*	
Class I	22 (9.6%)
Class II	138 (60.0%)
Class III or higher	65 (28.3%)
BMI *(no. (%))*	
Normal weight	61 (26.5%)
Overweight (BMI 25 – 30)	94 (40.9%)
Obese (BMI ≥ 30)	70 (30.4%)
Duration of complaints in years *(median [IQR])*	3.0 [1.5 – 6.0]
Indication *(no. (%))*	
OA	139 (60.4%)
mRCT/CTA	51 (22.2%)
Other	35 (15.2%)
Education *(no. (%))*	
Low	110 (51.2%)
Middle	63 (29.3%)
High	35 (16.3%)
Other	3 (1.3%)
Work status (paid/unpaid) *(no. (%))*	
No work / retired	167 (72.6%)
< 12 hrs/wk	6 (2.6%)
12-35 hrs/wk	26 (11.3%)
≥ 36 hrs/wk	19 (8.3%)
Other	3 (1.3%)
Cultural background *(no. (%))*	
Dutch	212 (92.2%)
Surinamese	2 (0.9%)
Other	7 (3.0%)
Religion *(no. (%))*	
Christian	121 (56.3%)
Catholic	4 (1.9%)
Jewish	1 (0.5%)
Other	5 (2.3%)
Not religious	66 (30.7%)
I'd rather not say	7 (3.3%)
TFI *(no. (%))*	
Frail	148 (64.4%)
Not frail	64 (27.8%)

Variable	Entire Sample (N = 230)
Primary/revision *(no. (%))*	
Primary	220 (95.7%)
Revision	5 (2.2%)
Type of prosthesis *(no. (%))*	
aTSA	64 (27.8%)
rTSA	152 (66.1%)
HA	7 (3.0%)
Dominant side *(no. (%))*	
Yes	100 (43.5%)
No	110 (47.8%)
Mean pain *(mean (SD))*	5.4 (2.1)
Worst pain *(mean (SD))*	6.7 (2.2%)
OSS *(mean (SD))*	20.2 (8.0)
EQ-5D item 'Mobility' *(no. (%))*	
No problems in walking about	130 (56.5%)
Slight problems in walking about	34 (14.8%)
Moderate problems in walking about	36 (15.7%)
Severe problems in walking about	14 (6.1%)
Unable to walk about	1 (0.4%)
EQ-5D item 'Self-Care' *(no. (%))*	
No problems washing or dressing	52 (22.6%)
Slight problems washing or dressing	87 (37.8%)
Moderate problems washing or dressing	55 (23.9%)
Severe problems washing or dressing	18 (7.8%)
Unable to wash or dress	3 (1.3%)
EQ-5D item 'Usual Activities' *(no. (%))*	
No problems doing usual activities	17 (7.4%)
Slight problems doing usual activities	64 (27.8%)
Moderate problems doing usual activities	94 (40.9%)
Severe problems doing usual activities	34 (14.8%)
Unable to do usual activities	6 (2.6%)
EQ-5D item 'Pain / Discomfort' *(no. (%))*	
No pain or discomfort	3 (1.3%)
Slight pain or discomfort	36 (15.7%)
Moderate pain or discomfort	105 (46.7%)
Severe pain or discomfort	66 (28.7%)
Extreme pain or discomfort	5 (2.2%)
EQ-5D item 'Anxiety / Depression' *(no. (%))*	
Not anxious or depressed	136 (59.1%)
Slightly anxious or depressed	48 (20.9%)
Moderately anxious or depressed	27 (11.7%)
Severely anxious or depressed	4 (1.7%)
Extremely anxious or depressed	-
EQ-5D VAS *(median [IQR])*	75.0 [64.0 – 82.0]
PCS *(median [IQR])*	17.0 [9.0 – 26.0]
LOTR optimism subscale *(mean (SD))*	8.1 (1.9)

Variable	Entire Sample (N = 230)
FPQ-9 *(median [IQR])*	
Total	15.0 [12.0 – 18.0]
Fear of severe pain	7.0 [5.0 – 9.0]
Fear of minor pain	4.0 [3.0 – 5.0]
Fear of medical/dental pain	4.0 [3.0 – 5.0]
CSI *(mean (SD))*	30.5 (12.1)
HSS expectations *(mean (SD))*	5.7 (4.0)
Sunnybrook expectation 'Pain relief' *(no. (%))*	
N/A	4 (1.7%)
No	-
Yes, but just a little	1 (0.4%)
Yes, somewhat	47 (20.4%)
Yes, a lot	163 (70.9%)
Sunnybrook expectation 'Pain-free range of motion' *(no. (%))*	
N/A	3 (1.3%)
No	5 (2.2%)
Yes, but just a little	10 (4.4%)
Yes, somewhat	80 (34.8%)
Yes, a lot	117 (50.9%)
Sunnybrook expectation 'Ability to carry out normal activities of daily living' *(no. (%))*	
N/A	5 (2.2%)
No	4 (1.7%)
Yes, but just a little	15 (6.5%)
Yes, somewhat	78 (33.9%)
Yes, a lot	113 (49.1%)
Sunnybrook expectation 'Ability to care for others' *(no. (%))*	
N/A	59 (25.7%)
No	7 (3.0%)
Yes, but just a little	24 (10.4%)
Yes, somewhat	66 (28.7%)
Yes, a lot	59 (25.7%)
Sunnybrook expectation 'Participate in leisure, sports or recreational activities like you did before' *(no. (%))*	
N/A	49 (21.3%)
No	11 (4.8%)
Yes, but not as much as before	105 (45.7%)
Yes, as much as before	49 (21.3%)
Sunnybrook expectation 'Shoulder back to the way it was before having problems' *(no. (%))*	
No	18 (7.8%)
No, but a little improved	3 (1.3%)
No, but somewhat improved	115 (50.0%)
Yes, completely	78 (33.9%)

ASA = American Society of Anesthesiologists, BMI = Body Mass Index, OA = osteoarthritis,

mRCT = massive rotator cuff tear, CTA = cuff tear arthropathy, TFI = Tilburg Frailty Indicator, aTSA = anatomic total shoulder arthroplasty, rTSA = reverse total shoulder arthroplasty, HA = hemiarthroplasty, OSS = Oxford Shoulder Score, PCS = Pain Catastrophizing Scale, FPQ-9 = Fear of Pain Questionnaire – 9 items, CSI = Central Sensitization Inventory, HSS = Hospital for Special Surgery, N/A = not applicable

Latent Growth Curve Modeling

The model that best fit our data was a GMM-1 model with six classes, due to the combination of good clinical interpretability, good fit statistics and lower heterogeneity within classes than other models. As described in the methods section, patients were allocated to the class for which they had the highest posterior probability.

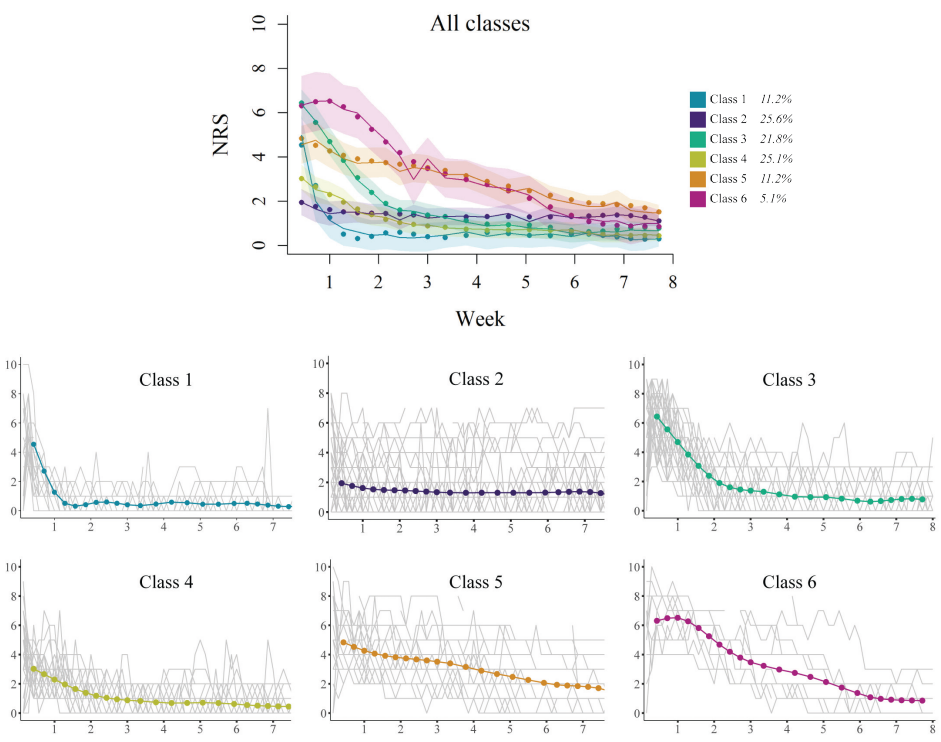


Fig. 2 Mean predicted and observed individual trajectories per class of the GMM-1 6-class model.
Shades around the mean predicted trajectories represent the 95% confidence interval.
GMM = Growth Mixture Model; NRS = Numeric Rating Scale

Figure 2 shows the estimated mean trajectories of the entire six-class model, as well as the estimated mean trajectory and observed individual trajectories per class separately. Classes 2 and 4 start with relatively low pain scores and maintain low pain scores or decrease even further. Classes 1 and 5 start with moderate pain scores; pain scores decline very quickly in class 1 but much slower in class 5. Classes 3 and 6 start with the highest scores. Pain scores in class 3 decline faster and more dramatically than in class 5. For exact intercepts, standard errors and class sizes, see Table 2.

The Appendix contains graphical plots for all the models that were run (seven models each for the LCGA (Figure A-1), GMM-1 (Figure A-2) and GMM-2 models (Figure A-3)), as well as their fit statistics, class size and entropy (Table A-I).

Table 2 Model Parameters for the six-class GMM-1 model

Class	Intercept	SE	Patients per class
Class 1	4.5	0.19	24 (11.2%)
Class 2	1.9	0.23	55 (25.6%)
Class 3	6.4	0.20	47 (21.8%)
Class 4	3.0	0.19	54 (25.1%)
Class 5	4.8	0.38	24 (11.2%)
Class 6	6.3	0.44	11 (5.1%)

GMM = Growth Mixture Model, SE = Standard Error

Comparing groups

While from a statistical standpoint the six-class model fit the observed data best, on visual inspection classes 1 – 4 evidently recover more quickly (NRS < 3 within two weeks) than classes 5 – 6 (NRS < 3 after five weeks). Hence, from a clinical standpoint, it is very interesting to compare these two groups: knowing how these patients differ from each other could help clinicians tailor their patient education and expectation management. We therefore made the post-hoc decision to compare classes 1 – 4 (collapsed into the 'Faster' group) to classes 5 – 6 (collapsed into the 'Slower' group) for our secondary objective instead of the six classes separately (Figure 3).

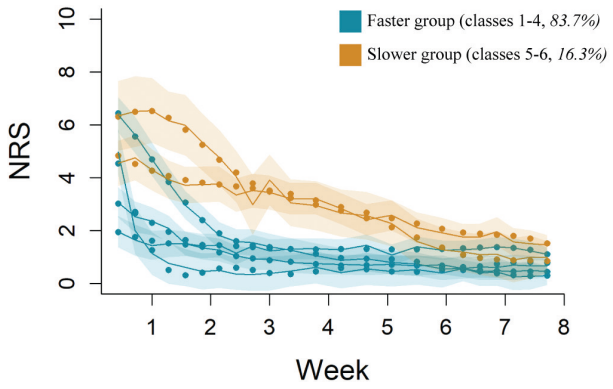


Fig. 3 Faster group vs. Slower group. NRS = Numeric Rating Scale

Compared to the Faster group, patients in the Slower group more often had ASA III scores (instead of ASA I or II, p = 0.002) and less often worked part-time or full-time (p = 0.012). They also had slightly lower baseline OSS scores (p = 0.044) and lower baseline EQ-5D VAS scores (p = 0.024), although the former did not exceed the smallest detectable change of 6.6 points[29]. For exact values of the univariable comparisons on all baseline variables, see Table 3. Multivariable analyses were not feasible due to the limited number of patients in the Slower group. The a priori planned comparisons between all six classes can be found in Table A-II in the Appendix.

Table 3 Preoperative Patient Characteristics and Surgery Characteristics of the Faster group versus the Slower group

Variable	Faster group (N = 180)	Slower group (N = 35)	p-value
Demographic			
Age *(mean (SD) [95% CI])*	69.9 (8.1) [68.7 – 71.1]	70.9 (9.0) [68.0 – 73.9]	.482
Sex *(no. (%))*			.52
Male	49 (27.2%)	12 (34.3%)	
Female	131 (72.8%)	23 (65.7%)	
ASA *(no. (%))*			.002
Class I	21 (11.7%)	1 (2.9%)	
Class II	117 (65.0%)	17 (48.6%)	
Class III or higher	42 (23.3%)	17 (48.6%)	

Variable	Faster group (N = 180)	Slower group (N = 35)	p-value
BMI *(no. (%))*			.792
Normal weight	50 (27.8%)	7 (20.0%)	
Overweight (BMI 25 – 30)	72 (40.0%)	18 (51.4%)	
Obese (BMI ≥ 30)	58 (32.2%)	10 (28.6%)	
Duration of complaints in years *(median [IQR])*	3.0 [1.5 – 6.0]	3.0 [1.25 – 6.5]	.932
Indication *(no. (%))*			.171
OA	117 (65.0%)	17 (48.6%)	
mRCT/CTA	38 (21.1%)	10 (28.5%)	
Other	25 (13.9%)	8 (22.9%)	
Education *(no. (%))*			.946
Low	92 (51.1%)	18 (51.4%)	
Middle	53 (29.4%)	10 (28.6%)	
High	30 (16.7%)	5 (14.3%)	
Other	3 (1.7%)	-	
Work *(no. (%))*			.012
No work	133 (73.9%)	26 (74.3%)	
<12 hours per week	6 (3.3%)	-	
12 – 35 hours per week	21 (11.7%)	2 (5.7%)	
≥36 hours per week	17 (9.4%)	2 (5.7%)	
Other	-	3 (8.6%)	
Cultural background *(no. (%))*			1.00
Dutch	171 (95.0%)	32 (91.4%)	
Surinamese	1 (0.6%)	-	
Other	5 (2.8%)	1 (2.9%)	
Religion *(no. (%))*			.702
Christian	102 (56.7%)	19 (54.3%)	
Catholic	2 (1.1%)	2 (5.7%)	
Jewish	1 (0.6%)	-	
Other	4 (2.2%)	1 (2.9%)	
Not religious	56 (31.1%)	10 (28.6%)	
I'd rather not say	7 (3.9%)	-	
TFI *(no. (%))*			.741
Frail	116 (64.4%)	23 (65.7%)	
Not frail	53 (29.4%)	9 (25.7%)	
Primary or revision *(no. (%))*			.592
Primary	176 (97.8%)	34 (97.1%)	
Revision	4 (2.2%)	1 (2.9%)	
Type of prosthesis *(no. (%))*			.517
aTSA	54 (30.0%)	8 (22.9%)	
rTSA	120 (66.7%)	27 (77.1%)	
HA	6 (3.3%)	-	
Dominant side *(no. (%))*			.110
Yes	89 (49.4%)	11 (31.4%)	
No	88 (48.9%)	22 (62.9%)	

Variable	Faster group (N = 180)	Slower group (N = 35)	p-value
Mean pain *(mean (SD) [95% CI])*	5.3 (2.1) [5.0 – 5.6]	5.7 (2.1) [5.0 – 6.4]	.272
Worst pain *(mean (SD) [95% CI])*	6.6 (2.2) [6.3 – 3.9]	7.1 (2.2) [6.4 – 7.8]	.234
OSS *(mean (SD) [95% CI])*	20.8 (7.7) [19.7 – 21.9]	17.7 (8.5) [14.9 – 20.6]	.044
EQ-5D item 'Mobility' *(no. (%))*			.273
No problems in walking about	104 (57.8%)	20 (57.1%)	
Slight problems in walking about	30 (16.7%)	4 (11.4%)	
Moderate problems in walking about	28 (15.6%)	4 (11.4%)	
Severe problems in walking about	10 (5.6%)	3 (8.6%)	
Unable to walk about	-	1 (2.9%)	
EQ-5D item 'Self-Care' *(no. (%))*			.400
No problems washing or dressing	44 (24.4%)	5 (14.3%)	
Slight problems washing or dressing	68 (37.8%)	15 (42.9%)	
Moderate problems washing or dressing	45 (25.0%)	7 (20.0%)	
Severe problems washing or dressing	13 (7.2%)	4 (11.4%)	
Unable to wash or dress	2 (1.1%)	1 (2.9%)	
EQ-5D item 'Usual Activities' *(no. (%))*			.050
No problems doing usual activities	16 (8.9%)	1 (2.9%)	
Slight problems doing usual activities	48 (26.7%)	11 (31.4%)	
Moderate problems doing usual activities	81 (45.0%)	9 (25.7%)	
Severe problems doing usual activities	23 (12.8%)	10 (28.6%)	
Unable to do usual activities	4 (2.2%)	1 (2.9%)	
EQ-5D item 'Pain / Discomfort' *(no. (%))*			.183
No pain or discomfort	3 (1.7%)	-	
Slight pain or discomfort	30 (16.7%)	4 (11.4%)	
Moderate pain or discomfort	86 (47.8%)	14 (40.0%)	
Severe pain or discomfort	52 (28.9%)	12 (34.3%)	
Extreme pain or discomfort	1 (0.6%)	2 (5.7%)	
EQ-5D item 'Anxiety / Depression' *(no. (%))*			.154
Not anxious or depressed	116 (64.4%)	17 (48.6%)	
Slightly anxious or depressed	32 (17.8%)	12 (34.3%)	
Moderately anxious or depressed	21 (11.7%)	3 (8.6%)	
Severely anxious or depressed	3 (1.7%)	-	
Extremely anxious or depressed	-	-	
EQ-5D VAS *(median [IQR])*	75.0 [71.3 – 76.0]	65.0 [56.2 – 71.7]	.024
PCS *(median [IQR])*	17.6 (10.8) [16.1 – 19.2]	21.2 (12.5) [17.0 – 25.3]	.099
LOTR optimism subscale *(mean (SD) [95% CI])*	8.2 (1.94) [7.9 – 8.5]	7.9 (1.93) [7.2 – 8.5]	.430

Variable	Faster group (N = 180)	Slower group (N = 35)	p-value
FPQ-9 *(median [IQR])*			
Total	15.5 [12.0 – 19.0]	14.5 [12.0 – 16.3]	.223
Fear of severe pain	7.0 [5.0 – 9.0]	8.0 [5.0 – 8.0]	.828
Fear of minor pain	4.0 [3.0 – 5.0]	3.0 [3.0 – 4.0]	.004
Fear of medical/dental pain	4.0 [3.0 – 5.0]	4.0 [3.0 – 5.0]	.422
CSI *(mean (SD) [95% CI])*	29.9 (11.9) [28.2 – 31.7]	34.2 (13) [29.9 – 38.5]	.067
HSS *(mean (SD) [95% CI])*	5.8 (4.0) [5.2 – 6.4]	6.1 (4.0) [4.8 – 7.4]	.714
Sunnybrook expectation 'Pain relief' *(no. (%))*			.617
N/A	4 (2.2%)	-	
No	-	-	
Yes, but just a little	1 (0.6%)	-	
Yes, somewhat	35 (19.4%)	9 (25.7%)	
Yes, a lot	132 (73.3%)	23 (65.7%)	
Sunnybrook expectation 'Pain-free range of motion' *(no. (%))*			.657
N/A	3 (1.7%)	-	
No	5 (2.8%)	-	
Yes, but just a little	10 (5.6%)	-	
Yes, somewhat	59 (32.8%)	13 (37.1%)	
Yes, a lot	95 (52.8%)	19 (54.3%)	
Sunnybrook expectation 'Ability to carry out normal activities of daily living' *(no. (%))*			.282
N/A	3 (1.7%)	1 (2.9%)	
No	2 (1.1%)	2 (5.7%)	
Yes, but just a little	13 (7.2%)	1 (2.9%)	
Yes, somewhat	62 (34.4%)	11 (31.4%)	
Yes, a lot	92 (51.1%)	17 (48.6%)	
Sunnybrook expectation 'Ability to care for others' *(no. (%))*			.966
N/A	49 (27.2%)	8 (22.9%)	
No	5 (2.8%)	1 (2.9%)	
Yes, but just a little	18 (10.0%)	4 (11.4%)	
Yes, somewhat	52 (28.9%)	9 (25.7%)	
Yes, a lot	48 (26.7%)	10 (28.6%)	
Sunnybrook expectation 'Participate in leisure, sports or recreational activities like you did before' *(no. (%))*			.107
N/A	36 (20.0%)	10 (28.6%)	
No	10 (5.6%)	1 (2.9%)	
Yes, but not as much as before	88 (48.9%)	10 (28.6%)	
Yes, as much as before	37 (20.6%)	11 (31.4%)	

Variable	Faster group (N = 180)	Slower group (N = 35)	p-value
Sunnybrook expectation 'Shoulder back to the way it was before having problems' *(no. (%))*			.687
No	12 (6.7%)	4 (11.4%)	
No, but a little improved	3 (1.7%)	-	
No, but somewhat improved	93 (51.7%)	16 (45.7%)	
Yes, completely	63 (35.0%)	12 (34.3%)	

ASA = American Society of Anesthesiologists, BMI = Body Mass Index, OA = osteoarthritis, mRCT = massive rotator cuff tear, CTA = cuff tear arthropathy, TFI = Tilburg Frailty Indicator, aTSA = anatomic total shoulder arthroplasty, rTSA = reverse total shoulder arthroplasty, HA = hemiarthroplasty, OSS = Oxford Shoulder Score, PCS = Pain Catastrophizing Scale, FPQ-9 = Fear of Pain Questionnaire – 9 items, CSI = Central Sensitization Inventory, HSS = Hospital for Special Surgery, N/A = not applicable

Function and quality of life

The mean OSS and EQ-5D VAS scores for both groups during the eight weeks are shown in Figure 4.

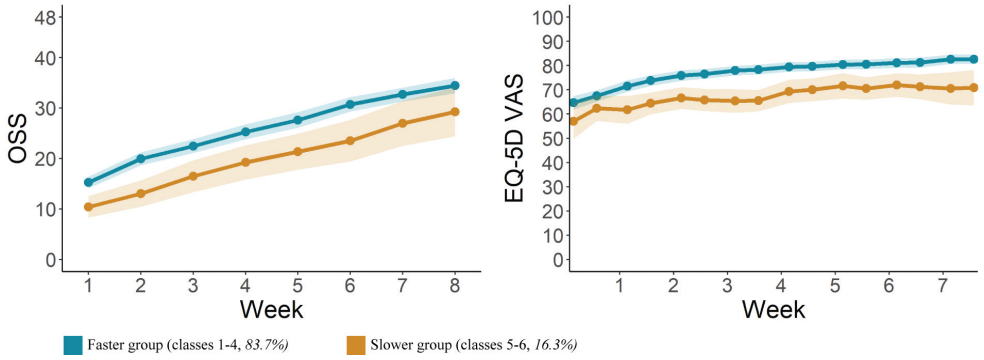


Fig. 4 Longitudinal change in Oxford Shoulder Score (OSS) and EQ-5D visual analog scale for Faster group vs. Slower group

For the OSS, the final linear mixed model contained group membership (Slower group vs. Faster group) and time as fixed effects, a random intercept and a random slope. Both groups progressed in OSS scores at a similar rate of 2.7 points per week. However, from the start, patients in the Slower group scored on average six points lower on the OSS than the Faster group. This was true for all time points since adding an interaction

term (p = 0.513) between group and time did not significantly improve the model fit.

For the EQ-5D VAS, the final linear mixed model was similar to the OSS model: group membership and time were fixed effects, and a random intercept and slope were included, but no interaction term (p-value for interaction term p = 0.120). Both groups improved on the EQ-5D VAS with 0.29 points per day on average, with the Slower group scoring a little over ten points lower than the Faster group. Table 4 shows the relevant model parameters for both final models.

Table 4 Model parameters for LMMs for OSS and EQ-5D VAS scores

LMM for OSS			
Fixed effects	**Estimates**	**95% CI**	**p-value**
Intercept	14.05	12.79 – 15.30	<.001
Slower group membership	-5.93	-8.86 - -3.00	<.001
Time (week)	2.66	2.48 – 2.84	<.001
Random effects	**Variance**		
Subject	65.14		
Time	1.36		
Residual	13.56		
LMM for EQ-5D VAS			
Fixed effects	**Estimates**	**95% CI**	**p-value**
Intercept	69.41	67.37 – 71.45	<.001
Slower group membership	-10.39	-14.87 - -5.92	<.001
Time (day)	0.29	0.25 – 0.32	<.001
Random effects	**Variance**		
Subject	190.20		
Time	5.96		
Residual	46.56		

LMM = Linear Mixed Model, OSS = Oxford Shoulder Score, CI = Confidence Interval, VAS = Visual Analog Scale

Medication use

The Slower group consistently used more medication, including opioids, than the Faster group during the first eight weeks. This difference was statistically significant at weeks 3 through 5 for any type of medication, weeks 3 through 8 for paracetamol and weeks 3 and 4 for opioids. For exact percentages of patients using any type of medication and paracetamol, NSAIDs and opioids specifically, see Figure 5 in the main text and Table A-III in the Appendix. Table A-IV also displays exact percentages of medication use during the first six days in detail.

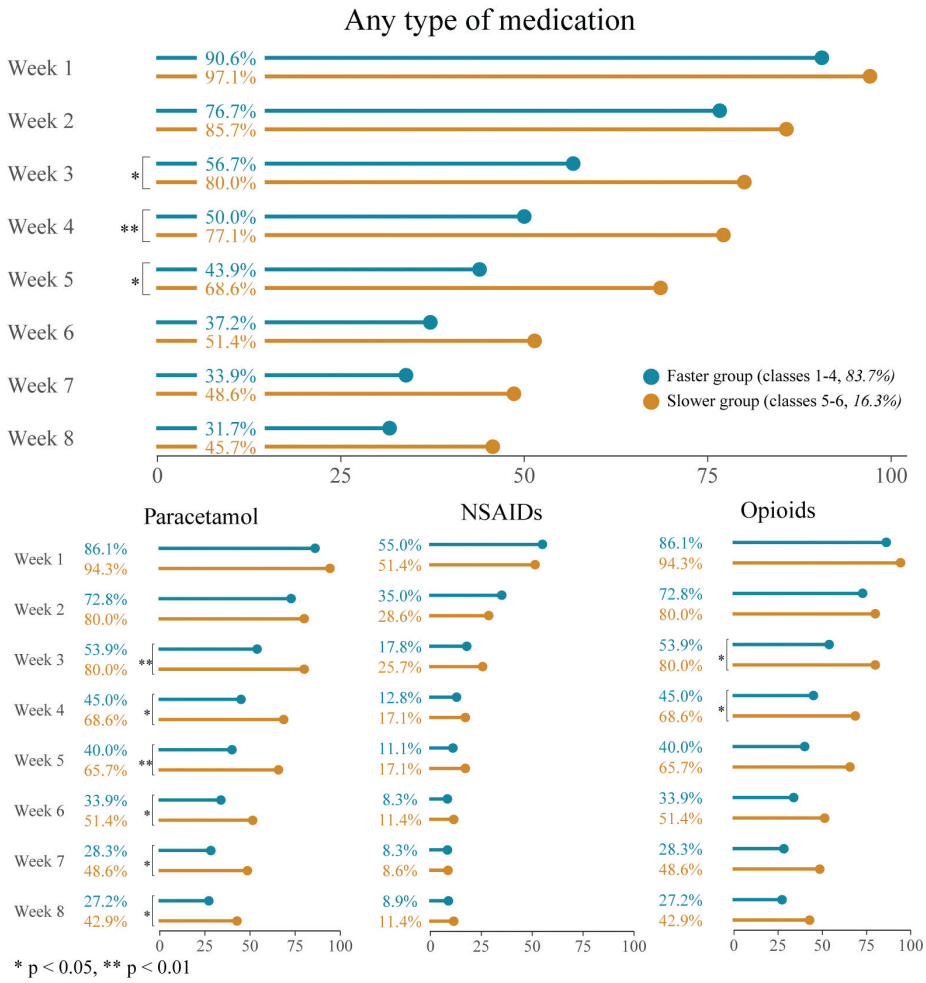


Fig. 5 Medication use for Faster group vs. Slower group. NSAIDs = nonsteroidal anti-inflammatory drug

Discussion

According to our study, patients who underwent shoulder arthroplasty can be grouped into six different classes based on their starting point and subsequent pain recovery trajectory during the first eight weeks after surgery. These six classes can furthermore be categorized into two clinically relevant subgroups: approximately 84% of patients with a fast decline in pain scores during the first two weeks and a smaller group of approximately 16% of patients in whom pain scores decline more slowly. Compared with the Faster group, patients in the Slower group more often had ASA scores of 3, were less often part-time or full-time employed and had lower baseline OSS and EQ-5D VAS scores. The pace of recovery in OSS and EQ-5D VAS were similar across the two groups, although the Slower group consistently had lower scores on these measures than the Faster group. This difference was statistically significant but not clinically relevant, as it did not pass the minimal clinically important difference of 6.9 that Liu et al.[29] found in SA patients.

Our results enable clinicians to reassure their patients before surgery, as five out of six patients likely have very low pain scores after only two weeks. Also, the sixth patient has almost similar low pain scores at eight weeks post-surgery.

Comparison with previous literature

To the best of our knowledge, no other study has yet intensively studied the pain trajectories during the first few weeks after SA. However, we wish to highlight two recent studies that did model growth trajectories after SA.

Rubinstein et al.[42] have used LCGA to investigate recovery trajectories after aTSA and rTSA using data from baseline and from six weeks, six months, one year and two years postoperatively; a much longer timespan than our study and without intensive early post-operative data. They found three groups for the cohort as a whole (aTSA and rTSA combined): Resistant Responders, Steady Progressors and High Performers. They also analysed aTSA and rTSA in separate analyses. In the separate aTSA and rTSA analyses, the Resistant Responders were replaced with Delayed Responders and Late Regressors, respectively.

Rubinstein et al.[42] did not find clear answers on which patient characteristics impact class membership. They also stated that their findings suggest other unmeasured variables, amongst which psychological elements, may be responsible for outcomes; in their study they only investigated age, sex, BMI, preoperative diagnosis and type of arthroplasty.

In our study, however, we did not find clear indications for such an association despite measuring multiple psychological factors preoperatively: pain catastrophizing, fear of pain, optimism, expectations and the Central Sensitization Inventory. None of these factors differed substantively between the Faster and Slower groups. Only the FPQ-9 subscale 'Fear of minor pain' had a statistically significant difference of 1 point on the median score, but we strongly doubt this difference to be clinically relevant.

Another relevant study that is particularly interesting to highlight is the recent study by Broekman et al.[2]. These authors also assessed postoperative trajectories after SA, but in a single-surgeon registry database study. They found that mental health was related to greater pain intensity at baseline but not to different rates of recovery. Their study differs from ours in some key aspects. Whereas we used LGCM to let previously undetected subgroups emerge from the data, they used growth models to study a priori selected subgroups stratified by quartiles of mental health measured with the mental component summary score (MCS) of the Veterans RAND 12.

Furthermore, whereas we allowed the trajectories to take any shape or form for each class, Broekman et al.[2] fit quadratic growth models, thus imposing similar shapes onto the subgroups.

Finally, both Rubenstein et al.[42] and Broekman et al.[2] were limited to data that had already been collected at regular care intervals and were more interested in longer-term trajectories than in the first eight weeks as we were in this study.

When comparing our study to the broader literature on outcomes after SA (not only studies that modeled growth trajectories), several previous studies did find possible associations between, for example, depression,

anxiety, pain catastrophizing, or expectations and outcomes after SA specifically[14, 17, 18, 23, 40, 48], or after other types of joint arthroplasty[4, 18, 35].

Several mechanisms may be responsible for this discrepancy. Firstly, although we have a large sample size for such an intensive prospective longitudinal study in this patient category, our Slower group is still small in absolute terms (35 patients). We therefore may lack power to detect statistically significant differences. Secondly, different methodological choices will yield different results. For example, when comparing subgroups of responders versus non-responders, different cut-off values can be chosen to define non-responders. This will inevitably alter which patients are assigned to each subgroup, thereby affecting the resulting associations. Finally, the choice of outcome measure and predictors may also affect the associations. For example, Swarup et al.[48] found a positive association between higher preoperative expectations and improvement in the American Shoulder and Elbow Surgeons (ASES) score, which measures both function and pain. Rauck et al.[40] found better ASES and VAS pain scores for patients with higher expectations for the surgery relieving nighttime pain, but found no association between the overall number of 'very important' expectations and 2-year outcome or improvement scores. Hence, our choice for pain as our outcome measure and the mean number of 'very important' expectations as predictor (instead of the separate expectations) could explain why we could not corroborate their results.

Strengths and limitations

A prominent strength is that we succeeded in including 230 subjects in our prospective multicenter cohort study. Many previous studies on SA have resorted to retrospective designs to gather study samples of over 100 subjects[7, 12, 15, 17, 19, 23, 34, 43]; shoulder arthroplasties only make up a small percentage of all arthroplasties. For example, a total of 3,581 SA were performed in The Netherlands in 2021, only 4.7% of all registered arthroplasties[25]. By using a prospective design, we had control over which data to collect at which time points while simultaneously actively minimizing missing data.

This leads us to another strength of our study: the intensive longitudinal data we collected with the daily diaries. To our knowledge, no other study to date has collected daily data for eight consecutive weeks starting the

day after SA. This enabled us to shed light on a previously unanswered research question.

Finally, we were able to recruit a sample that is representative for the overall Dutch SA population regarding general demographics (i.e. age, sex, BMI)[25].

However, we need to address some limitations as well.

Firstly, although we managed a very respectable sample size for a prospective study on SA patients, the Slower group still consisted of a small number of subjects. This precluded us from performing multivariable analyses and, thereby, from defining the independent predictive effect of preoperative patient characteristics on group membership. In addition, given the exploratory nature of the secondary analysis, we did not correct for multiple testing. Future studies may be needed to confirm or refute the findings of which factors are associated with Slower and Faster group membership.

Secondly, another possible limitation is that we included (both primary and revision) aTSA, rTSA and HA patients and analyzed our sample as a whole. Stratifying the analysis according to implant type could have led to different results. For example, Jones et al.[20] reported that rTSA patients required fewer opioids postoperatively than aTSA patients. However, when comparing the Faster and Slower groups, we found no statistically significant different proportions for implant type, although the Slower group did contain slightly more rTSA patients. Since the Slower group also had higher proportions of patients using opioids, our results do not match those of Jones et al.[20]. Our results do match with evidence for the HA patients: Craig et al.[10] and Bryant et al.[3] found in their systematic reviews slightly better pain and functional outcomes in favor of aTSA compared to HA, but these differences were not necessarily clinically relevant. Even though the included number of HA patients was minimal (n = 6), they were all classified into the Faster group, indicating that they did not fare worse than aTSA patients in a clinically relevant manner. In addition, the percentages primary and revision arthroplasty were similar between the Faster and Slower groups, suggesting that including both does not seem to influence our findings.

Thirdly, the reader should bear in mind that we cannot state that altering the factors on which the Faster and Slower groups in our sample differed will also alter the probability of becoming part of the Slower group, as our analysis is only of an exploratory nature and univariable. Nevertheless, our results can still make a significant contribution, as clinicians can now offer more accurate expectation management by explaining which factors (i.e. ASA scores of 3, not being employed and having lower baseline OSS and EQ-5D VAS scores) may predict a slower pain recovery in the early postoperative phase.

Lastly, the postoperative pain and rehabilitation protocols were not standardized among the different participating centers. Different protocols could influence the patient's postoperative pain trajectory and, thereby, our models. On the other hand, not standardizing the protocols increases the generalizability of our results to the broader population of SA patients treated in different hospitals and clinics; differences in protocols for rehabilitation and pain management will always exist in the real world. For example, in the USA first line pain management after SA is often opioid medication, whereas in the Netherlands paracetamol and NSAIDs are preferred. However, since a substantial proportion of our sample also used opioids as rescue medication in the first weeks postoperatively, we believe that our results are also generalizable to other countries such as the USA.

Conclusions

In this study, we distinguished six early recovery trajectories after SA. While recovery after shoulder arthroplasty clearly does not end at eight weeks, this early recovery period is extremely impactful in patients' lives. Being able to better manage expectations and reassure patients is a major advantage during preoperative consultations. This study has taken an essential first step in elucidating how patients experience their pain in the first weeks postoperatively. Surgeons can now show the figures within this article and explain how many patients generally experience a fast recovery or a slower recovery in pain. It enables clinicians to reassure their patients prior to surgery that within two weeks after surgery, over 80% of the patients have very low pain scores.

For future studies, it would be relevant to know whether our results can be replicated in different samples and if, using our models, the patient-

specific trajectory can be predicted based on daily pain data from the first two weeks. Also, we advise to explore if improved patient education focused on early recovery increases satisfaction in the first weeks after surgery, preferably with randomized controlled trials comparing patient education with and without detailed information regarding the first eight weeks.

Supplementary materials

The supplementary materials can be viewed online at https://doi.org/10.1016/j.jse.2025.06.016

Appendix 1

References

1. Berendes T, Pilot P, Willems J, Verburg H, te Slaa R. Validation of the Dutch version of the Oxford Shoulder Score. J Shoulder Elbow Surg 2010;**19**(6):829-36. doi:10.1016/j.jse.2010.01.017
2. Broekman MM, Brinkman N, Swanson D, Ring D, van den Bekerom M, Jawa A. Variations in 1-year Trajectories of Levels of Pain and Capability After Shoulder Arthroplasty Are Associated With Baseline Mental Health. Clin Orthop Relat Res 2024;**482**(3):514-22. doi:10.1097/corr.0000000000002821
3. Bryant D, Litchfield R, Sandow M, Gartsman GM, Guyatt G, Kirkley A. A comparison of pain, strength, range of motion, and functional outcomes after hemiarthroplasty and total shoulder arthroplasty in patients with osteoarthritis of the shoulder. A systematic review and meta-analysis. J Bone Joint Surg Am 2005;**87**(9):1947-56. doi:10.2106/jbjs.D.02854
4. Burns LC, Ritvo SE, Ferguson MK, Clarke H, Seltzer Z, Katz J. Pain catastrophizing as a risk factor for chronic pain after total knee arthroplasty: a systematic review. J Pain Res 2015;8:21-32. doi:10.2147/jpr.S64730
5. Carducci MP, Zimmer ZR, Jawa A. Predictors of unsatisfactory patient outcomes in primary reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2019;**28**(11):2113-20. doi:10.1016/j.jse.2019.04.009
6. Castor EDC. Castor Electronic Data Capture; 2019. Available from: https://castoredc.com.
7. Chang NB, Bicknell R, Krupp R, Wiater JM, Levy J, Athwal GS. Sex-related differences in stemless total shoulder arthroplasty. JSES Int 2022;**6**(1):26-31. doi:10.1016/j.jseint.2021.09.008
8. Cho CH, Song KS, Hwang I, Coats-Thomas MS, Warner JJP. Changes in Psychological Status and Health-Related Quality of Life Following Total Shoulder Arthroplasty. J Bone Joint Surg Am 2017;**99**(12):1030-5. doi:10.2106/jbjs.16.00954
9. Claes A, De Mesel A, Struyf T, Verborgt O, Struyf F. Factors Influencing Outcome After Shoulder Arthroplasty (FINOSA Study): Protocol of a Prospective Longitudinal Study With Randomized Group Allocation. JMIR Res Protoc 2024;13:e56522. doi:10.2196/56522
10. Craig RS, Goodier H, Singh JA, Hopewell S, Rees JL. Shoulder replacement surgery for osteoarthritis and rotator cuff tear arthropathy. Cochrane Database Syst Rev 2020;**4**(4):Cd012879. doi:10.1002/14651858.CD012879.pub2
11. Dean MC, Cherian NJ, Eberlin CT, Rudisill SS, LaPorte ZL, Kucharik MP, et al. Total shoulder vs. hip and knee arthroplasty: an analysis of perioperative outcomes. *Seminars in Arthroplasty: JSES*. 33. Elsevier; 2023:627-36.
12. Dekker TJ, Grantham WJ, Lacheta L, Goldenberg BT, Dey Hazra RO, Rakowski DR, et al. Glenoid retroversion does not impact clinical outcomes or implant survivorship after total shoulder arthroplasty with minimal, noncorrective reaming. JSES Int 2022;**6**(4):596-603. doi:10.1016/j.jseint.2022.02.011
13. Fernández-de-Las-Peñas C, Florencio LL, de-la-Llave-Rincón AI, Ortega-Santiago R, Cigarán-Méndez M, Fuensalida-Novo S, et al. Prognostic Factors for Postoperative Chronic Pain after Knee or Hip Replacement in Patients with Knee or Hip Osteoarthritis: An Umbrella Review. J Clin Med 2023;**12**(20). doi:10.3390/jcm12206624
14. Forlizzi JM, Puzzitiello RN, Hart PA, Churchill R, Jawa A, Kirsch JM. Predictors of poor and excellent outcomes after reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2022;**31**(2):294-301. doi:10.1016/j.jse.2021.07.009
15. Friedman RJ, Cheung EV, Flurin PH, Wright T, Simovitch RW, Bolch C, et al. Are Age and Patient Gender Associated With Different Rates and Magnitudes of Clinical Improvement After Reverse Shoulder Arthroplasty? Clin Orthop Relat Res 2018;**476**(6):1264-73. doi:10.1007/s11999.0000000000000270

16. Gobbens RJ, van Assen MA, Luijkx KG, Wijnen-Sponselee MT, Schols JM. Determinants of frailty. J Am Med Dir Assoc 2010;**11**(5):356-64. doi:10.1016/j.jamda.2009.11.008
17. Green A, Neradilek MB, Thompson KM, Mayer A. Expectations affect outcome and satisfaction after anatomic total shoulder arthroplasty. *Seminars in Arthroplasty: JSES*. 30. Elsevier; 2020:297-307.
18. Hardy A, Sandiford MH, Menigaux C, Bauer T, Klouche S, Hardy P. Pain catastrophizing and pre-operative psychological state are predictive of chronic pain after joint arthroplasty of the hip, knee or shoulder: results of a prospective, comparative study at one year follow-up. Int Orthop 2022;**46**(11):2461-9. doi:10.1007/s00264-022-05542-7
19. Huber J, Irlenbusch U, Kääb MJ, Reuther F, Kohut G, Judge A. Treatment effects of reverse total shoulder arthroplasty - a simple method to measure outcomes at 6, 12, 24 and 60 months for each patient. BMC Musculoskelet Disord 2020;**21**(1):397. doi:10.1186/s12891-020-03427-7
20. Jones CA, Chambers S, Renshaw AD, Throckmorton TW, Bernholt DL, Azar FM, Brolin TJ. Patients undergoing reverse total shoulder arthroplasty have less pain and require fewer opioid pain medications compared to anatomic total shoulder arthroplasty in the early postoperative period: a retrospective review. J Shoulder Elbow Surg 2024. doi:10.1016/j.jse.2024.04.024
21. Kadum B, Inngul C, Ihrman R, Sjödén GO, Sayed-Noor AS. Higher preoperative sensitivity to pain and pain at rest are associated with worse functional outcome after stemless total shoulder arthroplasty: a prospective cohort study. Bone Joint J 2018;100-b(4):480-4. doi:10.1302/0301-620x.100b4.Bjj-2017-1000.R1
22. Klooster Pt, Weekers A, Eggelmeijer F, Van Woerkom J, Drossaert C, Taal E, et al. Optimisme en/of pessimisme: factorstructuur van de Nederlandse Life Orientation Test-Revised. Psychologie en Gezondheid 2010;38:89-100. doi:10.1007/BF03089356
23. Kohan EM, Aleem AW, Chamberlain AM, Keener JD. The influence of mental health on outcomes following total shoulder arthroplasty. *Seminars in Arthroplasty: JSES*. 30. Elsevier; 2020:18-27.
24. Kregel J, Vuijk PJ, Descheemaeker F, Keizer D, van der Noord R, Nijs J, et al. The Dutch Central Sensitization Inventory (CSI): Factor Analysis, Discriminative Power, and Test-Retest Reliability. Clin J Pain 2016;**32**(7):624-30. doi:10.1097/ajp.0000000000000306
25. Landelijke Registratie Orthopedische Interventies (LROI). Annual LROI Report 2024; 2024. Available from: https://www.lroi.nl/jaarrapportage/shoulder/numbers/.
26. Lansdown DA, Ma GC, Aung MS, Gomez A, Zhang AL, Feeley BT, Ma CB. Do patient outcomes and follow-up completion rates after shoulder arthroplasty differ based on insurance payor? J Shoulder Elbow Surg 2021;**30**(1):65-71. doi:10.1016/j.jse.2020.04.028
27. Lapner PL, Jiang L, Zhang T, Athwal GS. Rotator cuff fatty infiltration and atrophy are associated with functional outcomes in anatomic shoulder arthroplasty. Clin Orthop Relat Res 2015;**473**(2):674-82. doi:10.1007/s11999-014-3963-5
28. Lewis GN, Rice DA, McNair PJ, Kluger M. Predictors of persistent pain after total knee arthroplasty: a systematic review and meta-analysis. Br J Anaesth 2015;**114**(4):551-61. doi:10.1093/bja/aeu441
29. Liu P, Afzal I, Asopa V, Clement ND, Patel V. Changes and thresholds in the Oxford Shoulder Score following shoulder arthroplasty: Minimal clinically important difference, minimal important and detectable changes, and patient-acceptable symptom state. Shoulder & Elbow 2023:17585732231176423. doi:10.1177/17585732231176423
30. Mancuso CA, Altchek DW, Craig EV, Jones EC, Robbins L, Warren RF, Williams-Russo P. Patients' expectations of shoulder surgery. J Shoulder Elbow Surg 2002;**11**(6):541-9. doi:10.1067/mse.2002.126764

31. Matsen FA, 3rd, Iannotti JP, Churchill RS, De Wilde L, Edwards TB, Evans MC, et al. One and two-year clinical outcomes for a polyethylene glenoid with a fluted peg: one thousand two hundred seventy individual patients from eleven centers. Int Orthop 2019;**43**(2):367-78. doi:10.1007/s00264-018-4213-3
32. McNeil DW, Kennedy SG, Randall CL, Addicks SH, Wright CD, Hursey KG, Vaglienti R. Fear of Pain Questionnaire-9: Brief assessment of pain-related fear and anxiety. Eur J Pain 2018;**22**(1):39-48. doi:10.1002/ejp.1074
33. Overview of Medical Research in the Netherlands (OMON). Trial The Shoulder Diary; 2017. Available from: https://onderzoekmetmensen.nl/en/trial/25734. [Accessed Sep 29 2024].
34. Pettit RJ, Saini SB, Puzzitiello RN, Hart PJ, Ross G, Kirsch JM, Jawa A. Primary reverse total shoulder arthroplasty performed for glenohumeral arthritis: does glenoid morphology matter? J Shoulder Elbow Surg 2022;**31**(5):923-31. doi:10.1016/j.jse.2021.10.022
35. Pinto PR, McIntyre T, Araújo-Soares V, Costa P, Ferrero R, Almeida A. A comparison of predictors and intensity of acute postsurgical pain in patients undergoing total hip and knee arthroplasty. J Pain Res 2017;10:1087-98. doi:10.2147/jpr.S126467
36. Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.
37. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. Journal of Statistical Software 2017;**78**(2):1-56. doi:10.18637/jss.v078.i02
38. R Development Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.
39. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;**33**(5):337-43. doi:10.3109/07853890109002087
40. Rauck RC, Swarup I, Chang B, Dines DM, Warren RF, Gulotta LV, Henn RF, 3rd. Effect of preoperative patient expectations on outcomes after reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2018;**27**(11):e323-e9. doi:10.1016/j.jse.2018.05.026
41. Razmjou H, Finkelstein JA, Yee A, Holtby R, Vidmar M, Ford M. Relationship between Preoperative Patient Characteristics and Expectations in Candidates for Total Knee Arthroplasty. Physiother Can 2009;**61**(1):38-45. doi:10.3138/physio.61.1.38
42. Rubenstein WJ, Warwick HSL, Aung MS, Zhang AL, Feeley BT, Ma CB, Lansdown DA. Defining recovery trajectories after shoulder arthroplasty: a latent class analysis of patient-reported outcomes. J Shoulder Elbow Surg 2021;**30**(10):2375-85. doi:10.1016/j.jse.2021.02.024
43. Saini SS, Pettit R, Puzzitiello RN, Hart P-A, Shah SS, Jawa A, Kirsch JM. Clinical outcomes after reverse total shoulder arthroplasty in patients with primary glenohumeral osteoarthritis compared with rotator cuff tear arthropathy: does preoperative diagnosis make a difference? Journal of the American Academy of Orthopaedic Surgeons 2022;**30**(3):e415-e22. doi:10.5435/JAAOS-D-21-00797
44. Sayed-Noor AS, Pollock R, Elhassan BT, Kadum B. Fatty infiltration and muscle atrophy of the rotator cuff in stemless total shoulder arthroplasty: a prospective cohort study. J Shoulder Elbow Surg 2018;**27**(6):976-82. doi:10.1016/j.jse.2017.12.021
45. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol 1994;**67**(6):1063-78. doi:10.1037//0022-3514.67.6.1063
46. Shields EJW, Koueiter DM, Maerz T, Schwark A, Wiater JM. Previous Rotator Cuff Repair Is Associated With Inferior Clinical Outcomes After Reverse Total Shoulder Arthroplasty. Orthop J Sports Med 2017;**5**(10):2325967117730311. doi:10.1177/2325967117730311

47. Sullivan MJ, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychological assessment 1995;**7**(4):524. doi:10.1037/1040-3590.7.4.524
48. Swarup I, Henn CM, Nguyen JT, Dines DM, Craig EV, Warren RF, et al. Effect of pre-operative expectations on the outcomes following total shoulder arthroplasty. Bone Joint J 2017;99-b(9):1190-6. doi:10.1302/0301-620x.99b9.Bjj-2016-1263.R1
49. Wood TJ, Gazendam AM, Kabali CB. Postoperative Outcomes Following Total Hip and Knee Arthroplasty in Patients with Pain Catastrophizing, Anxiety, or Depression. J Arthroplasty 2021;**36**(6):1908-14. doi:10.1016/j.arth.2021.02.018

Declarations

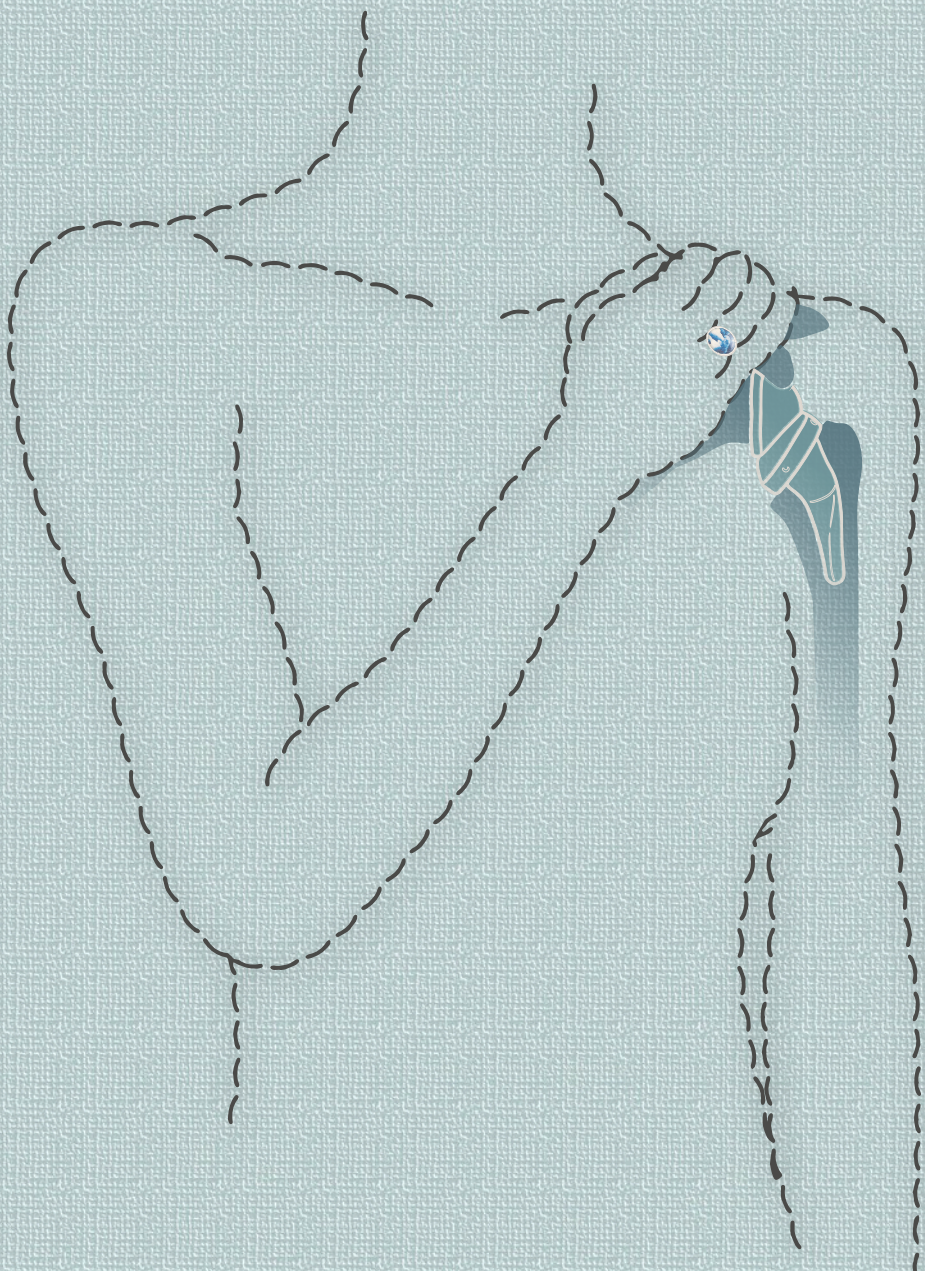
Acknowledgment

The authors would like to thank Roos Bazuin (RHOC), Elga Meurs and Joyce Benner (Northwest clinics), and Ilse Bouman (Haga Hospital) for their efforts in including patients and helping with data collection.

Disclaimers

Funding: The paper diaries were printed with financial support from the Scientific Research Council / the Board of Directors of the Reinier de Graaf Hospital, Delft, The Netherlands.

Conflicts of interest: The 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.



PART II

Patients' expectations regarding shoulder arthroplasty

4

Complete fulfillment of expectations is associated with greater satisfaction after shoulder arthroplasty: results from a prospective multicenter cohort study

Hesseling B, Berends Y, van der Linden S, Tolk JJ, van der Linde JA, Visser CP, Slot K, Eygendaal D, Snoeker BA, Mathijssen NM

JSES International 2026; 10(3)

Abstract

Background

With this study, we aim to explore preoperative probabilistic expectations (what does the patient realistically expect) and value-based expectations (what does a patient consider important), postoperative expectation fulfillment and the relationship with postoperative satisfaction in shoulder arthroplasty (SA) patients. We hypothesized that higher expectation fulfillment is related to higher postoperative satisfaction, corrected for baseline level of optimism.

Methods

In this prospective multicenter cohort study, 230 SA patients were included. These patients completed preoperative questionnaires measuring probabilistic (Sunnybrook questionnaire) and value-based expectations (Hospital for Special Surgery Shoulder Expectations questionnaire) as well as optimism (Life Orientation Test – Revised). Postoperatively, expectation fulfillment/improvement and postoperative satisfaction were measured. We used ordinal logistic regression to study whether expectation fulfillment is related to postoperative satisfaction, corrected for baseline level of optimism.

Results

For preoperative probabilistic expectations, patients expected improvement on a median of 5 domains (IQR 4 – 6) out of six possible domains. For preoperative value-based expectations, patients had a median of 5 (IQR 3 – 9) out of seventeen possible expectations that they valued as 'Very important'. At six months and twelve months, the median percentage of completely fulfilled probabilistic expectations per patient was 75% and 80%, respectively. The median percentage of completely improved 'very important' value-based expectations per patient was 7.7% and 14.3%. Patients with higher percentages of fulfilled or improved expectations had higher odds of being 'very satisfied' postoperatively.

Conclusions

This study shows that fulfillment of probabilistic expectations and improvement on value-based expectations is associated with postoperative satisfaction after shoulder arthroplasty.

Level of Evidence: level I (High-quality prospective cohort study with >80% follow-up)

Keywords: shoulder arthroplasty, preoperative expectations, postoperative expectation fulfillment, probabilistic expectations, value-based expectations, postoperative satisfaction

Introduction

In the past years, joint arthroplasty research has shifted its focus from mainly biomedical factors to more psychological factors, among which preoperative expectations and expectations fulfillment are included. In the literature, preoperative expectations can be measured with two different main paradigms: probabilistic (what a patient realistically expects) and value-based (what is important to a patient)[7]. Hafkamp et al.[4] found in their systematic review that higher expectation fulfillment was linked to higher postoperative satisfaction for total knee arthroplasty (TKA) and total hip arthroplasty (THA). For the level of preoperative expectations (as opposed to expectation fulfillment) they found conflicting evidence. The studies in this review used both paradigms to measure expectations and the review did not discriminate between these two.

Regarding the influence of expectations in shoulder arthroplasty (SA) patients the literature is scarcer. A recent systematic review on preoperative expectations by Claes et al.[1] could only include four studies. They concluded that there is moderate evidence for no association between pre-operative expectations and post-operative outcomes. Moreover, only one study has investigated expectation fulfillment in SA patients[3] and found weak correlations with satisfaction (r = 0.17 – 0.37) but did not quantify the magnitude of the relationship. Also, they only studied probabilistic and not value-based expectations.

To prepare patients well and to enable surgeons to align their expectations to those of their patients, we need more insight into the association between expectation fulfillment and postoperative satisfaction in this population. Furthermore, we need to understand if it matters with which paradigm (probabilistic or value-based) and corresponding instrument expectations are measured to provide patients with correct information. To date, we have not found any study that measured both paradigms at the same time in a single group of patients.

In this study we therefore set out to answer three research questions: (1) Preoperatively, what do SA patients expect (probabilistic expectations) and what is important to SA patients (value-based expectations)? (2) What percentage of preoperative expectations becomes fulfilled at six and

twelve months after shoulder arthroplasty? (3) Do patients with higher expectation fulfillment have higher postoperative satisfaction scores? We hypothesized that higher expectation fulfillment is related to higher postoperative satisfaction, corrected for baseline level of optimism.

Materials and Methods

Study design and setting

For this multicenter, prospective cohort study patients were recruited between April 2018 and September 2021 from four teaching and/or general hospitals in The Netherlands. All participating surgeons were fellowship-trained high volume shoulder surgeons. Patients were followed for one year and study data was collected from the electronic patient records, through a paper diary and through surveys (either through email or on paper).

Participants

We screened all patients indicated for SA consecutively for study eligibility. Patients were eligible for inclusion when they were 18 years or older, scheduled to undergo primary or revision SA (either hemi arthroplasty (HA), anatomic or reverse total shoulder arthroplasty (aTSA or rTSA)), able and willing to participate and had good understanding of the written and spoken Dutch language. Patients were excluded for cognitive impairment and arthroplasty for acute fractures. All participating patients provided written informed consent prior to study participation.

Description of treatment

Each surgeon typically used the deltopectoral approach and general anesthesia with interscalene block. Patients were kept overnight and were generally discharged the next day. They were instructed to use a sling for six weeks and to only perform passive exercises during the first two weeks. After two weeks, patients could increase their exercises to include active-assisted and subsequently active exercises.

Variables, outcome measures, data sources, and bias

Data collection

Patients completed questionnaires at baseline and at six and twelve months postoperatively; this was part of a larger study in which patients also completed a daily diary during the early recovery period. The questionnaires contained questions on general patient characteristics, pain, function, quality of life, psychological factors, expectations and (postoperatively only) satisfaction. For the research questions in this manuscript, only questionnaires for patient characteristics, optimism, expectations and satisfaction were used. Results on early recovery using the diary data were published previously in this journal[6].

Optimism

The validated Life Orientation Test – Revised (LOT-R)[15] was used to measure preoperative generalized optimism. It is a validated 10-item questionnaire answered on a 5-point Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree). The total score thus ranges from 0 (low optimism) to 24 (high optimism).

Expectations

Probabilistic expectations were measured pre- and postoperatively with the Sunnybrook expectation questionnaire[14]. This questionnaire contains six domains (i.e. pain, range of motion (ROM), ability to perform activities of daily living, ability to interact with and take care of others, ability to return to previous leisure/recreational/sports activities and potential to achieve full recovery following surgery; domains written out in figures in Results section) and asks whether patients preoperatively expect improvement on these domains and if so, how much improvement. For postoperative expectation fulfillment, the phrasing of the questions was changed from, for example, 'Do you expect pain relief from the surgery' to 'Did the surgery reduce your pain?' in a modified Sunnybrook questionnaire.

Value-based expectations were measured pre- and postoperatively with the Hospital for Special Surgery (HSS) Shoulder Expectations questionnaire[10]. This questionnaire contains seventeen items (similar to the domains of the Sunnybrook questionnaire but in more detail; domains written out

in figures in Results section) and asks patients to rate the importance of each expectation. In this value-based questionnaire patients reported how important each item was, but not if they actually expected improvement on this item after surgery. Therefore, fulfillment of these expectations could not be assessed; instead, we measured postoperative improvement. We reworded the instruction for the modified HSS questionnaire from 'How important are these expectations in the treatment for your shoulder?' to 'For these expectations, which answer fits your situation best?'. Answer categories were on an 8-point Likert scale ranging from 'complete improvement' to 'much deterioration'.

Both the preoperative and modified (postoperative) questionnaires can be found in the Supplementary file (Tables S1 – S4). Since no validated Dutch translations were available for both questionnaires at the start of the study, self-translated versions were used.

Satisfaction

We measured postoperative satisfaction with a single question ('How satisfied are you up to now with the results of your surgery?'). The answers were on a 5-point Likert scale ranging from 'very satisfied' to 'very dissatisfied'.

Ethical approval

METC Leiden, Den Haag, Delft (formerly METC Southwest Holland) in the Netherlands decided that the study did not fall under the scope of the Medical Research Involving Human Subjects Act because of the minimal patient burden (METC-no. 17-117). The study was conducted conform the Declaration of Helsinki and the European Medicines Agency Guidelines for Good Clinical Practice.

Statistical analysis, study size

We described characteristics of the entire sample with number and frequency for categorical variables, with mean and SD for normally distributed continuous data and with median and interquartile range (IQR) for non-normally distributed data.

All statistical analyses were performed in R version 4.4.1[12] with RStudio[11]. As described in our previous publication, the sample size of 230 subjects was based on the analysis of the early recovery results and on the feasibility of including the number of subjects. Missing data were not imputed. The statistical analyses per research question are described below.

Preoperative expectations (research question 1)

To describe the preoperative expectations, our analysis for both types of expectations was two-fold: on the level of each expectation and on the patient level.

First, for each separate expectation, we calculated the percentages of patients within each answer category.

Next, for each patient, the number of probabilistic (Sunnybrook) expectations that were actually present per patient was calculated (e.g. the number of expectations scored with any of the 'yes' categories). Similarly, the number of value-based (HSS) expectations that were rated as 'very important' per patient was calculated.

Expectation fulfillment and improvement (research question 2)

For probabilistic expectations, expectation fulfillment was categorized as 'completely fulfilled', 'somewhat fulfilled' or 'not at all fulfilled'.

For value-based expectations, we collapsed our 8-point Likert scale into similar categories 'completely improved', 'somewhat improved' or 'not at all improved'.

Similarly to question 1, the analysis was done on the expectation level and on the patient level.

First, on the expectation level, we calculated for each probabilistic and value-based expectation the percentages of patients rating that expectation as 'completely fulfilled' and as 'completely or somewhat fulfilled' (probabilistic) or 'completely improved' and 'completely or somewhat improved' (value-based), at both time points.

Next, on the patient level, we calculated for each patient the percentage of their preoperative expectations that were 'completely fulfilled/improved' and 'completely or somewhat fulfilled/improved', also at both time points.

Patients that did not complete the preoperative questionnaires were excluded from these analyses. For more details on this analysis, see Supplementary file.

Relationship of expectation fulfillment/improvement with postoperative satisfaction (research question 3)

To account for the ordinal nature of our satisfaction measure, we performed ordinal logistic regression analyses using the 'MASS' package[18]. In these models, the dependent variable was postoperative satisfaction and the focal predictor variable was the percentage probabilistic expectation fulfillment or percentage value-based expectation improvement. The percentages were first divided by ten so that the odds ratios represent a step of 10% increase in the predictor variable.

For ease of interpretation and due to small numbers of not satisfied/not dissatisfied and unsatisfied patients, the outcome categories 'not satisfied / not dissatisfied', 'dissatisfied' and 'very dissatisfied' were merged into a single category 'not satisfied'.

For each type of expectations (i.e. probabilistic and value-based), we ran four separate models:

- Focal predictor 'complete fulfillment/improvement' at six months
- Focal predictor 'complete fulfillment/improvement' at twelve months
- Focal predictor 'complete or some fulfillment/improvement' at six months
- Focal predictor 'complete or some fulfillment/improvement' at twelve months

Within each type of expectation we used the Bonferroni method to correct for multiple testing and set our p-value threshold at $p = 0.0125$. Patients that did not complete the preoperative questionnaires were excluded from these analyses.

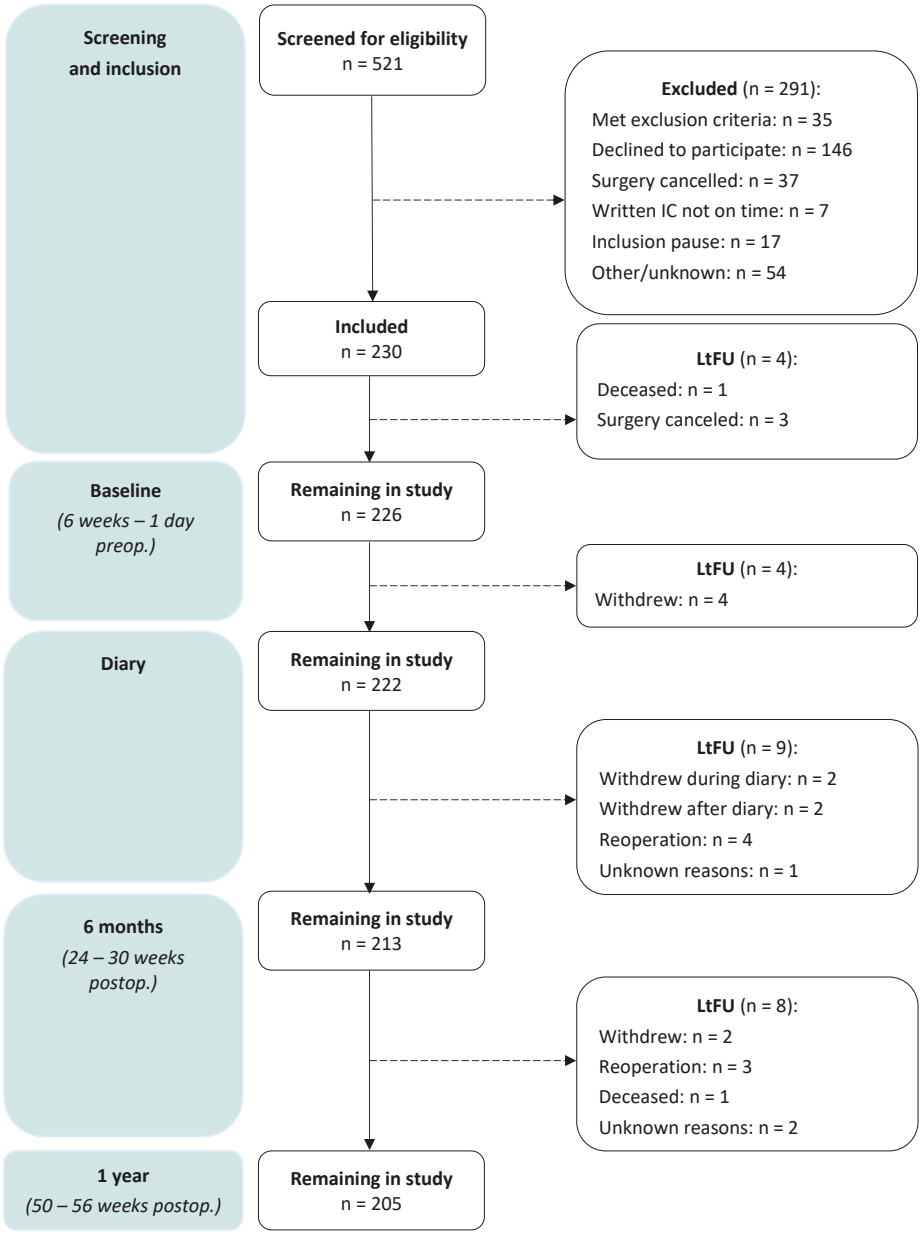
The degree of preoperative generalized optimism was accounted for by adjusting for baseline LOT-R scores in each model. Given the common rule of thumb to only add one variable per ten cases, adding more variables to the model was not considered prudent: at six and twelve months only 21 and 19 patients were 'not satisfied', respectively.

No multicollinearity was detected for any model. We used the Brant-Wald test to assess for each model whether the proportional hazard assumption held and the Lipsitz goodness of fit test to investigate whether the model fit our data well. Two models at twelve months proved to be problematic; the Supplementary file details how these problems were handled.

Finally, to aid in the interpretation of the models, we calculated and visualized the predicted probabilities for all different levels of the predictor variable.

Results

Figure 1 shows the flow of patients throughout the study. We included 230 patients with a mean age of 69.9 (SD 8.5) years, of which 160 (69.6%) were female. The most common diagnosis was osteoarthritis (60.4%) and the most frequently used type of implant was rTSA (66.1%). For complete patient characteristics, see Table 1. Of the 230 included patients, 205 patients completed the one year follow-up period; 11% of patients were lost to follow up.



*: The transition from Reinier de Graaf Hospital and Haga Hospital into Reinier Haga Orthopedic Center (RHOC) resulted in an inclusion pause of approximately two months for these centers

LtFU = lost to follow-up

Fig. 1 Study flow chart

Table 1. Descriptive statistics of the study sample

Variable	Entire Sample (N = 230)
Demographic	
Age *(mean (SD))*	69.9 (8.5)
Sex *(no. (%))*	
Male	67 (29.1%)
Female	160 (69.6%)
ASA *(no. (%))*	
Class I	22 (9.6%)
Class II	138 (60.0%)
Class III or higher	65 (28.3%)
BMI *(no. (%))*	
Normal weight	61 (26.5%)
Overweight (BMI 25 – 30)	94 (40.9%)
Obese (BMI ≥ 30)	70 (30.4%)
Duration of complaints in years *(median [IQR])*	3.0 [1.5 – 6.0]
Indication *(no. (%))*	
OA	139 (60.4%)
mRCT/CTA	51 (22.2%)
Other	35 (15.2%)
Cultural background *(no. (%))*	
Dutch	212 (92.2%)
Surinamese	2 (0.9%)
Other	7 (3.0%)
Religion *(no. (%))*	
Christian	121 (56.3%)
Catholic	4 (1.9%)
Jewish	1 (0.5%)
Other	5 (2.3%)
Not religious	66 (30.7%)
I'd rather not say	7 (3.3%)
Primary/revision *(no. (%))*	
Primary	220 (95.7%)
Revision	5 (2.2%)
Type of prosthesis *(no. (%))*	
aTSA	64 (27.8%)
rTSA	152 (66.1%)
HP	7 (3.0%)
Dominant side *(no. (%))*	
Yes	100 (43.5%)
No	110 (47.8%)
LOT-R optimism subscale *(mean (SD))*	8.1 (1.9)

ASA = American Society of Anesthesiologists, BMI = Body Mass Index, OA = osteoarthritis, mRCT = massive rotator cuff tear, CTA = cuff tear arthropathy, aTSA = anatomic total shoulder arthroplasty, rTSA = reverse total shoulder arthroplasty, HA = hemiarthroplasty, LOT-R = Life Orientation Test - Revised

Preoperatively, what do SA patients expect (probabilistic expectations) and what is important to SA patients (value-based expectations)?

For preoperative probabilistic expectations, patients expected improvement on 5 (median, IQR 4 – 6) out of the six possible domains. For preoperative value-based expectations, patients had 5 (median, IQR 3 – 9) out of the seventeen possible expectations that they valued as 'Very important'. Pain relief and improvement in range of motion were both most often expected (probabilistic) and most often rated as 'very important' (value-based). Figures 2A and 3A show the distribution of answers for the Sunnybrook questionnaire and the HSS, respectively. Exact numbers and percentages can be found in Table S5 and S6 in the Supplementary file.

What percentage of preoperative expectations becomes fulfilled at six and twelve months after shoulder arthroplasty?

A higher percentage of probabilistic expectations were fulfilled than value-based expectations were improved at both time points. At six months and twelve months, the median percentage of completely fulfilled probabilistic expectations per patient was 75% and 80%, respectively (IQR 33.3% - 100% for both time points). The median percentage of completely improved 'very important' value-based expectations per patient was 7.7% (IQR 0% - 50%) and 14.3% (IQR 0% - 57.1%), at six and twelve months respectively. Figures 2B-C and 3B-C show the percentages 'completely fulfilled/improved', 'somewhat fulfilled/improved' and 'not at all fulfilled/improved' at both time points for the Sunnybrook questionnaire and HSS, respectively. Tables S7 and S8 in the Supplementary file show exact numbers and percentages.

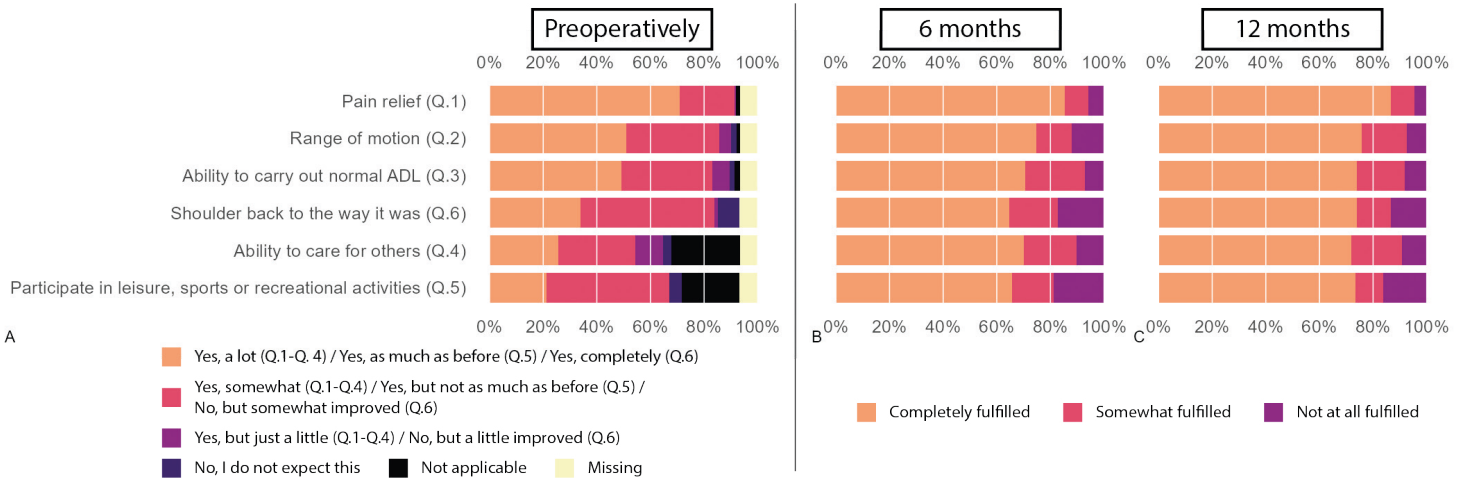


Fig. 2 (A-C) Barplots showing probabilistic expectations (Sunnybrook). (A) How many patients responded to each of the answer categories pre-operatively, and (for patients who had that expectation pre-operatively) the degree of fulfillment at (B) 6 months and (C) 12 months

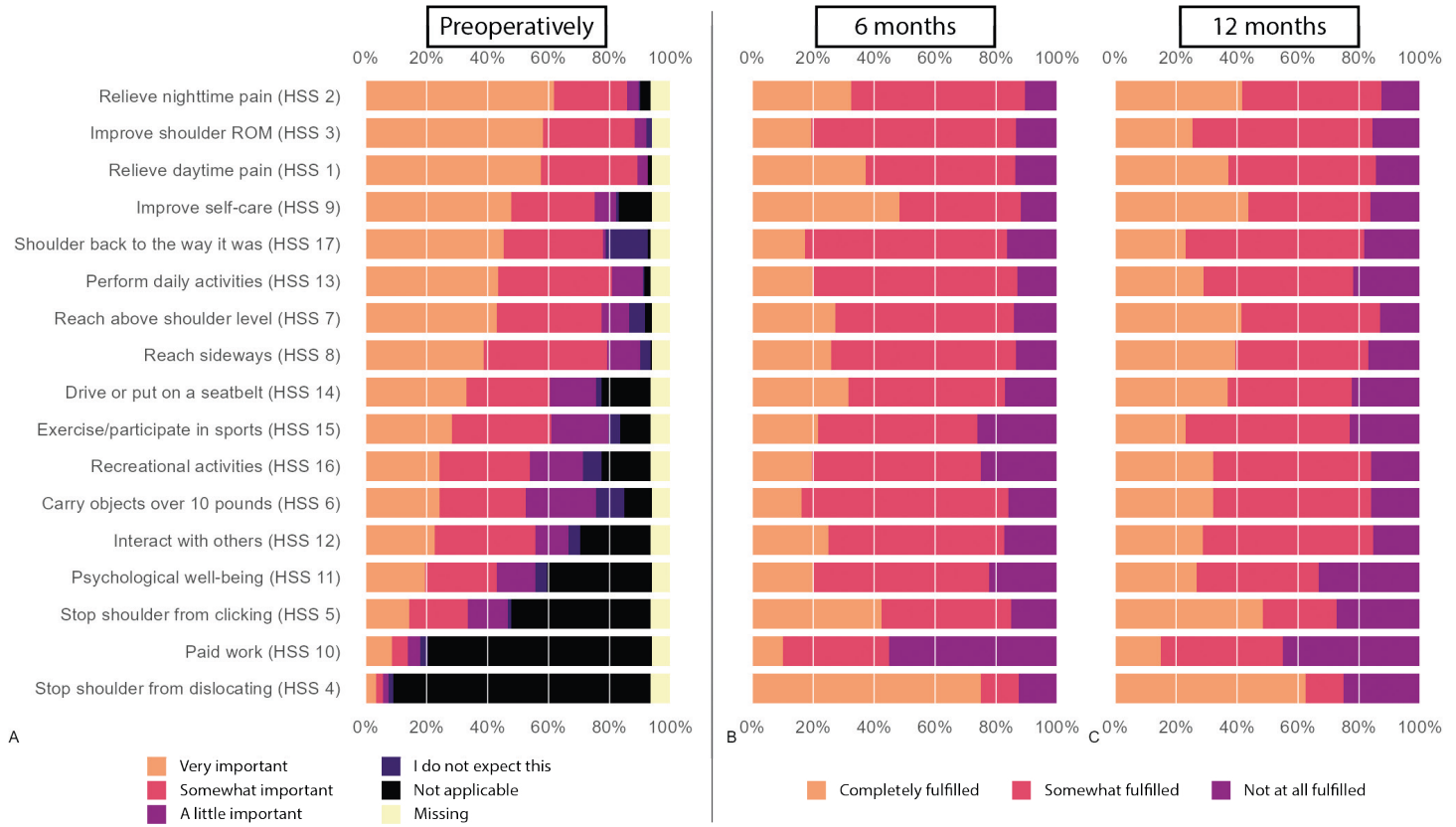


Fig. 3 (A-C) Barplots showing for value-based expectations (HSS) (A) how many patients responded to each of the answer categories pre-operatively, and (for patients who rated that expectation pre-operatively as 'very important') the degree of improvement at (B) 6 months and (C) 12 months

Do patients with higher expectation fulfillment have higher postoperative satisfaction scores?

Patients with higher percentages of fulfilled or improved expectations had higher odds of being 'very satisfied' postoperatively. At six and twelve months, 50.4% (116 of 230 patients) and 53.9% (124 of 230 patients) were 'very satisfied', 30.4% (70 patients) and 24.8% (57 patients) were 'satisfied', 9.1% (21 patients) and 8.3% (19 patients) were 'not satisfied'. 10% (23 patients) and 13% (30 patients) had missing satisfaction scores. Table 2 displays the odds ratios, CIs and corresponding *p*-values for all models, as well as model diagnostics. The model for 'completely fulfilled' probabilistic expectations at twelve months was not reliable and is therefore not reported in Table 2 (see Supplementary file for details).

Table 2. Model estimates and diagnostics for all ordinal logistic regression models

Model	OR* (95% CI)	p-value	p-value Lipsitz	p-value Brant-Wald
Probabilistic expectations (Sunnybrook)				
6 months				
Model 1			0.32	.19
Per 10% completely fulfilled	1.44 (1.30 – 1.59)	< .001		.07
Baseline optimism	0.98 (0.90 – 1.07)	.63		.77
Model 2			.53	.45
Per 10% completely/ somewhat fulfilled	1.49 (1.32 – 1.68)	< .001		.23
Baseline optimism	0.99 (0.91 – 1.07)	.78		.58
12 months				
Model 3			.02	.88
Per 10% completely fulfilled	-[1]			.97
Baseline optimism	-[1]			.62
Model 4			.52	.89
Per 10% completely/ somewhat fulfilled	1.25 (1.12 – 1.39)	< .001		.69
Baseline optimism	1.08 (0.98 – 1.19)	.10		.78

Model	OR* (95% CI)	p-value	p-value Lipsitz	p-value Brant-Wald
Value-based expectations (HSS)				
6 months				
Model 5			.51	.40
Per 10% completely improved	1.66 (1.42 – 1.93)	< .001		.18
Baseline optimism	.91 (.83 – 1.01)	.07		.84
Model 6			.27	.51
Per 10% completely/ somewhat improved	1.59 (1.37 – 1.84)	< .001		.39
Baseline optimism	1.01 (.92 – 1.10)	.89		.39
12 months				
Model 7			.19	.33
Per 10% completely improved	1.53 (1.32 – 1.78)	< .001		.18
Baseline optimism	1.04 (0.94 – 1.15)	0.48		.43
Model 8			.41	.19
Per 10% completely/ somewhat improved	1.24 (1.11 – 1.38)	< .001		.08
Baseline optimism	1.10 (1.00 – 1.21)	.05		.78

* OR = Odds Ratio for moving one step upward in satisfaction (i.e. 'Not satisfied' → 'Satisfied' or 'Satisfied' → 'Very satisfied')

¶: this model was not reliable

To aid in the interpretation of the different models, Figures 4 and 5 visualize the predicted probabilities of the reliable models. These plots show for each level of percentage expectation fulfillment/improvement what the probability is for a patient to be 'very satisfied', 'satisfied' and 'not satisfied'. For example, for patients who have 60% of their very important value-based expectations *completely* improved at six months, the predicted probability of being 'very satisfied' is 89% (Figure 5A). In contrast, when a patient has 60% of their expectations *completely or somewhat* improved, the predicted probability is only 23% (Figure 5B).

Regarding the qualitative examination of the model for probabilistic expectations at twelve months, Figure 4C shows that the median percentages 'completely fulfilled' probabilistic expectations at twelve months for 'very satisfied', 'satisfied' and 'not satisfied' correspond to the conclusion of the other models. In addition, Figures S1 and S2 in the Supplementary file show the distributions of all focal predictors at both time points split into the three satisfaction levels.

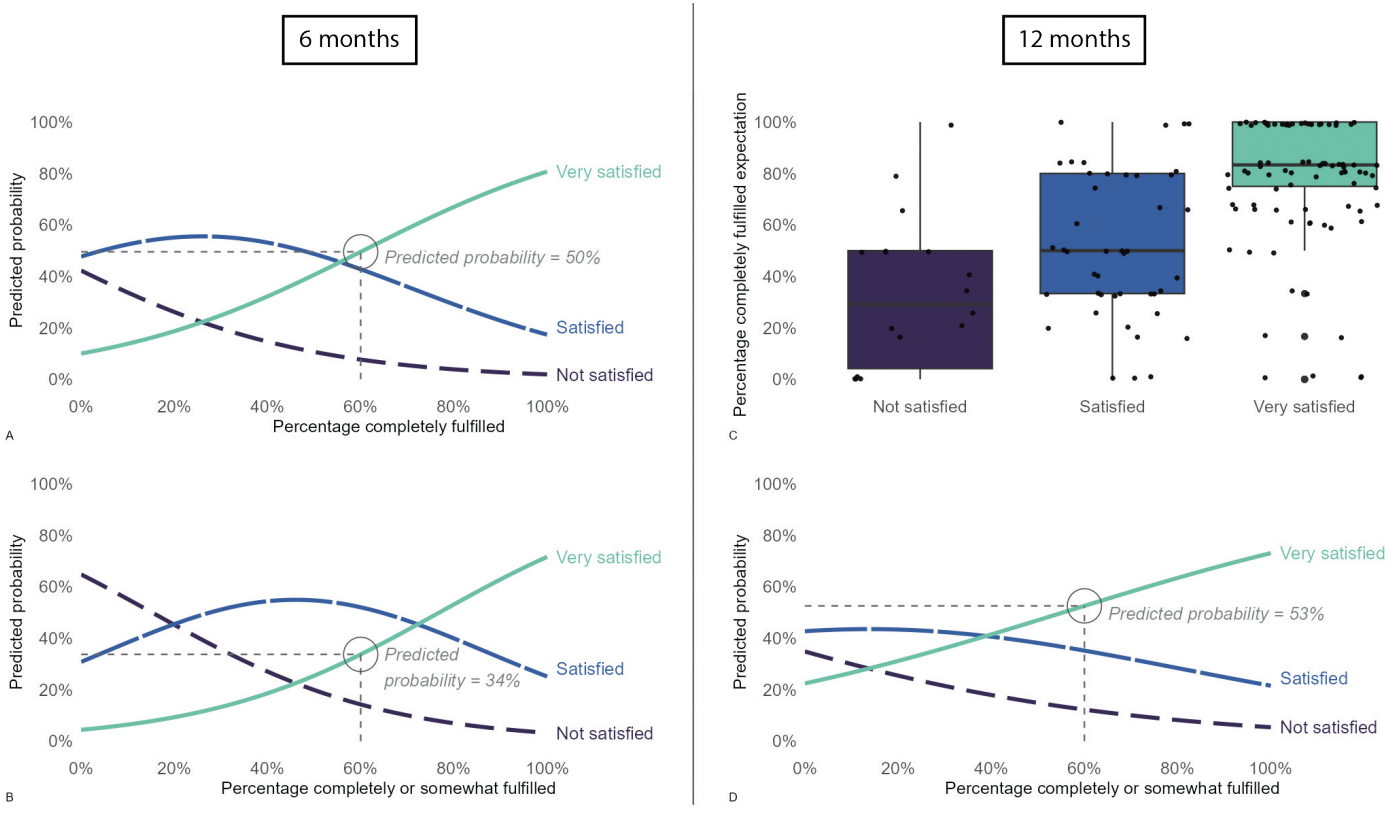


Fig. 4 (A-D) Plots showing the predicted probabilities of being 'Very satisfied,' 'Satisfied,' or 'Not satisfied' for probabilistic expectations (Sunnybrook), (A) for percentage 'completely fulfilled' expectations at 6 months, (B) for percentage 'completely or somewhat fulfilled' at 6 months, (C) median and IQR of percentage 'completely fulfilled' expectations per satisfaction level at 12 months (this model was unreliable), and (D) for percentage 'completely or somewhat fulfilled' at 12 months. The 60% fulfilled in the annotations was arbitrarily chosen to support interpretation of the plots

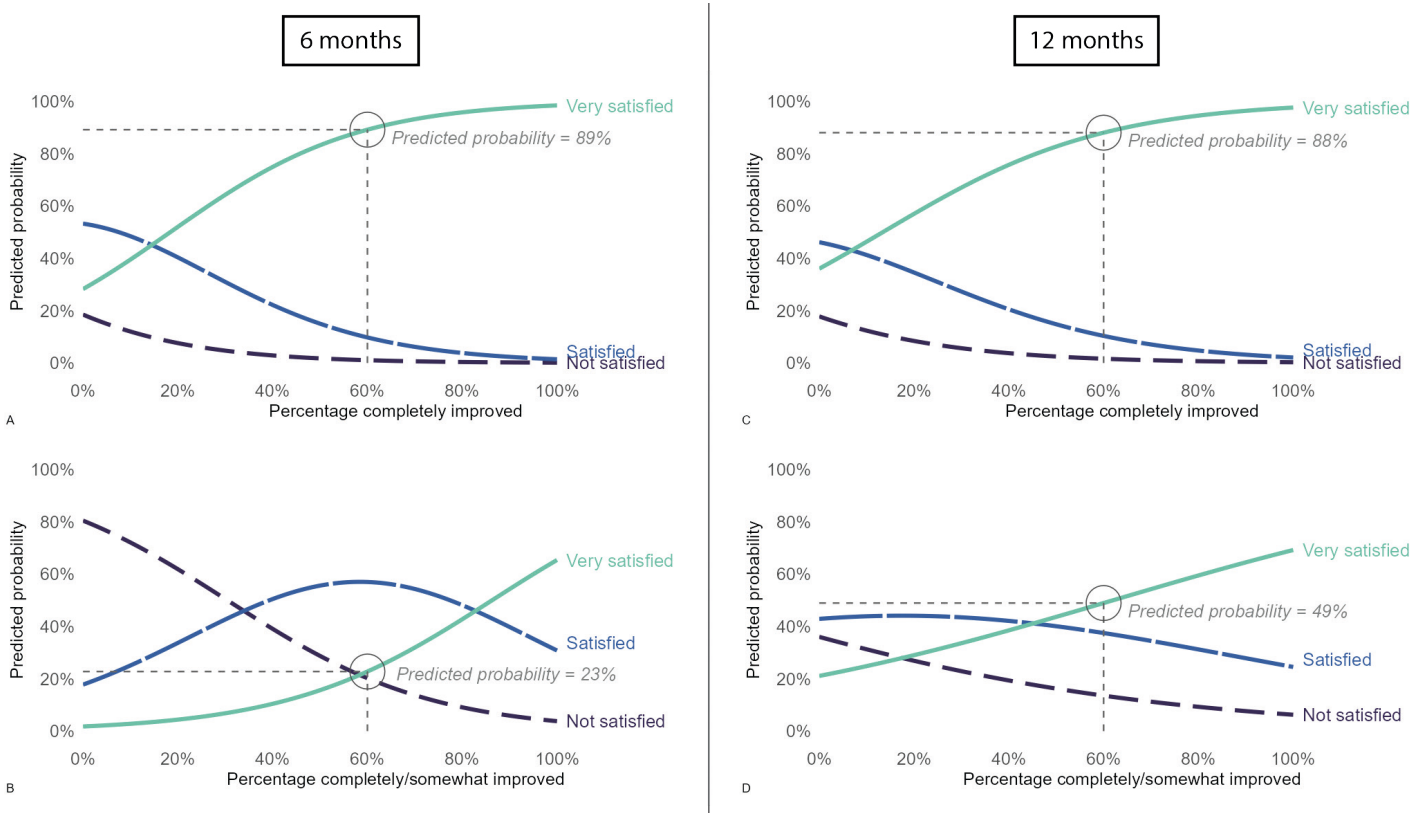


Fig. 5 A-D) Plots showing the predicted probabilities of being 'Very satisfied,' 'Satisfied,' or 'Not satisfied' for value-based expectations (HSS), (A) for percentage 'completely improved' expectations at 6 months, (B) for percentage 'completely or somewhat improved' at 6 months, (C) for percentage 'completely improved' expectations at 12 months, and (D) for percentage 'completely or somewhat improved' at 12 months. The 60% fulfilled in the annotations was arbitrarily chosen to support interpretation of the plots

Discussion

This is the first study to measure both probabilistic and value-based expectations within a single joint arthroplasty sample. We found that higher percentages of expectation fulfillment or expectation improvement were associated with higher levels of satisfaction. The predicted probabilities for being 'very satisfied' were higher for 'completely improved' value-based expectations than for 'completely or somewhat' improved value-based expectations.

In TKA and THA patients, expectation fulfillment has been consistently linked to better postoperative outcomes, but for SA patients the literature has been sparse. A better understanding of the relationship between expectation fulfillment and postoperative satisfaction in SA patients and of possible differences between the probabilistic and value-based paradigms will enable the orthopedic surgeon to adequately manage their patients' expectations.

Preoperatively, what do SA patients expect (probabilistic expectations) and what is important to SA patients (value-based expectations)?

In this study, SA patients preoperatively ranked pain relief most often as very important, which corresponds to results from other studies that focused on SA and used the value-based HSS questionnaire[5, 9, 13]. However, exact percentages varied between these studies: daytime pain relief was rated as very important by 86% of patients in the study by Henn et al.[5] and 85% in the study by Lizzio et al.[9], and by 48% in the study by Rauck et al.[13] and 57 % in our study. Improvement in ROM and improvement in self-care were also different between the studies, ranging from 35% to 89% for ROM and from 48% to 82% for self-care. The reason for these differences is not yet clear and could be multifactorial, for example due to type of arthroplasty or whether expectation fulfillment was determined through written or verbal questioning. Perhaps different methods of expectation management in different hospitals could affect how patients rate expectations, although it remains questionable if value-based expectations can be influenced by preoperative counseling to the same extent as probabilistic expectations[16, 17]. After all, knowing what is realistic may not change what is important to a patient.

What percentage of preoperative expectations becomes fulfilled at six and twelve months after shoulder arthroplasty?

Until this study, only Green et al.[3] investigated expectation fulfillment in aTSA patients, by asking to which degree preoperative probabilistic expectations (measured with the Musculoskeletal Outcomes Data Evaluation and Management Scale (MODEMS)) had been met at a follow-up of two years. They report that expectations were met for a high percentage of patients. The highest level of met expectations was for relief from symptoms and the lowest level for the ability to exercise and do recreational activities. These results show the same tendency as our results on fulfillment of probabilistic expectations at six and twelve months.

For value-based expectations, the 7.7% of 'completely improved' expectations in our results seems very low. However, we believe this low percentage can be explained by the definitions we employed in our methods. Firstly, the percentage was only calculated on expectations that patients rated preoperatively as 'very important'; therefore, improvement on expectations that were of some importance but not 'very important' was not included in this calculation. Secondly, 'complete improvement' is a high bar to set and perhaps not realistic for all expectations. Large (yet not large enough to rate as 'complete') improvements are again not included in the percentage of 'completely improved' expectations. Figures 3B and 3C show that for each expectation the large majority of patients reached either 'completely improved' or 'somewhat improved'. Fewer patients indicated 'not at all improved'.

To our knowledge, there are no studies that examined improvement or fulfillment of value-based expectations in SA patients. Our results therefore provide valuable input for surgeons and other healthcare professionals to use in their patient counseling; they may add another dimension to the conversation. For example, instead of overwhelming patients with information on all possible domains and generalized statements on effectiveness, information can be focussed on those specific domains that are very important to the patient and by showing the level of improvement of other patients who also valued those domains as very important.

At the same time, it is necessary for future studies to expand on our findings. For instance, by also measuring what patients actually expect on the 'very important' domains (measuring value-based and probabilistic expectations on the same questions). After all, a patient can find something very important but still anticipate little actual improvement on that domain. This could be reflected by the low percentages of improvement on the value-based expectations in our study.

Do patients with higher expectation fulfillment have higher postoperative satisfaction scores?

This is the first study to investigate the relationship between value-based expectation improvement and postoperative satisfaction after SA. Even for probabilistic expectations, we only found the aforementioned study by Green et al.[3]. They found statistically significant but weak to very weak correlations between met probabilistic expectations and satisfaction (r = 0.17 – 0.37), but they did not further quantify the magnitude of that relationship. Our study elaborates on their findings, not only by quantifying the relationship but also by correcting for generalized optimism. Future studies with larger sample sizes can correct for more psychological factors like anxiety, depression or pain catastrophizing, parallel to studies on THA and TKA.

A remarkable finding of our study is that the predicted probabilities for 'completely improved' expectations and 'completely or somewhat improved' expectations were very different at six months (see Figures 5A-B). Because a low percentage of 'completely or somewhat improved' expectations means there is little improvement at all, it would make sense that a patient would not be satisfied with that result. However, at twelve months the difference in predicted probabilities is much smaller. Deakin et al.[2] found similar results in TKA patients: fulfilled expectations were related to satisfaction, but a low percentage of fulfillment did not exclude satisfaction. Perhaps patients adjust their frame of reference to align more with the current situation, the longer ago the surgery has been – a psychological phenomenon called resilience. It would be interesting to gain more insight into such potential response shifts in future studies.

Limitations

Despite the study's strengths, we also need to address some clear limitations. The first limitation is that we did not measure the exact same items probabilistically as we did value-based. Assessing probabilistic expectations with the HSS would have enabled us to identify whether the 'very important' expectations were truly fulfilled, instead of only measuring how much patients improved on those domains. Nevertheless, since this is the first study that measured and evaluated both probabilistic and value-based expectations in the same sample, our results still deepen our understanding of how different kinds of expectations relate to satisfaction.

Second, we analyzed aTSA and rTSA patients together as a single group, but the extent to which certain expectations (e.g., range of motion (ROM)) can be fulfilled/improved can differ between aTSA and rTSA. We also included small numbers of patients undergoing HA or revision arthroplasty. However, we expect these differences to only affect the absolute levels of fulfillment/improvement and not the relationship between fulfillment/improvement and satisfaction. Moreover, Lawrence et al.[8] did not find important differences in cumulative expectation scores and single expectations between aTSA and rTSA patients.

Finally, the preoperative counseling (and thus preoperative expectation management), treatment and aftercare were not standardized between the participating hospitals and surgeons. It could therefore be that expectations were managed differently between patients. The upside of this limitation is that our results might be better applicable to other settings than only the participating hospitals, as these differences exist in the real world as well. We expect that our results are also generalizable to different healthcare system structures than in the Netherlands, since the expectations in our questionnaires specifically concern shoulder pain and function, rather than properties of the healthcare process itself.

Conclusions

This study shows that fulfillment of probabilistic expectations and improvement on value-based expectations is associated with postoperative satisfaction after shoulder arthroplasty. Surgeons and other healthcare professionals can use these results to further customize and optimize

patient education and expectation management: generalized statements about recovery can now be supplemented with statements on specific domains that are important to patients.

Future studies should elaborate on our results by measuring and comparing probabilistic and value-based expectations on the same instrument and items in one population, and by predicting the probability of complete expectation fulfillment/improvement on an individual level.

Supplementary materials

The supplementary materials can be viewed online at https://doi.org/10.1016/j.jseint.2026.101651

Appendix 1

References

1. Claes A, Mesel A, Verborgt O, Struyf F. Effects of pre-operative patient expectations on the outcome after total shoulder arthroplasty: A systematic review. Shoulder Elbow 2024;**17**(4):388-400. doi:10.1177/17585732241282021
2. Deakin AH, Smith MA, Wallace DT, Smith EJ, Sarungi M. Fulfilment of preoperative expectations and postoperative patient satisfaction after total knee replacement. A prospective analysis of 200 patients. Knee 2019;**26**(6):1403-12. doi:10.1016/j.knee.2019.07.018
3. Green A, Neradilek MB, Thompson KM, Mayer A. Expectations affect outcome and satisfaction after anatomic total shoulder arthroplasty. Seminars in Arthroplasty: JSES 2020;**30**(4):297-307. doi:10.1053/j.sart.2020.09.005
4. Hafkamp FJ, Gosens T, de Vries J, den Oudsten BL. Do dissatisfied patients have unrealistic expectations? A systematic review and best-evidence synthesis in knee and hip arthroplasty patients. EFORT Open Rev 2020;**5**(4):226-40. doi:10.1302/2058-5241.5.190015
5. Henn RF, 3rd, Ghomrawi H, Rutledge JR, Mazumdar M, Mancuso CA, Marx RG. Preoperative patient expectations of total shoulder arthroplasty. J Bone Joint Surg Am 2011;**93**(22):2110-5. doi:10.2106/jbjs.J.01114
6. Hesseling B, Snoeker BA, Ten Brinke B, Slot K, Zwitser EW, van Rosmalen J, et al. Trajectories of pain recovery during the first eight weeks after shoulder arthroplasty: results from the Shoulder Diary study using latent growth curve modeling. Journal of Shoulder and Elbow Surgery 2025. doi:10.1016/j.jse.2025.06.016
7. Kravitz RL. Patients' expectations for medical care: an expanded formulation based on review of the literature. Med Care Res Rev 1996;**53**(1):3-27. doi:10.1177/107755879605300101
8. Lawrence C, Lazarus M, Abboud J, Williams G, Namdari S. Prospective Comparative Study of Preoperative Expectations and Postoperative Outcomes in Anatomic and Reverse Shoulder Arthroplasty. Joints 2019;**7**(4):159-64. doi:10.1055/s-0041-1730982
9. Lizzio VA, James CL, Pietroski AD, Kuhlmann NA, Franovic S, Warren JR, Muh SJ. Characterizing preoperative expectations for patients undergoing reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2022;**31**(3):537-44. doi:10.1016/j.jse.2021.08.030
10. Mancuso CA, Altchek DW, Craig EV, Jones EC, Robbins L, Warren RF, Williams-Russo P. Patients' expectations of shoulder surgery. Journal of Shoulder and Elbow Surgery 2002;**11**(6):541-9. doi:10.1067/mse.2002.126764
11. Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.
12. R Development Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.
13. Rauck RC, Swarup I, Chang B, Ruzbarsky JJ, Dines DM, Warren RF, et al. Preoperative patient expectations of elective reverse shoulder arthroplasty. J Shoulder Elbow Surg 2019;**28**(7):1217-22. doi:10.1016/j.jse.2018.12.008
14. Razmjou H, Finkelstein JA, Yee A, Holtby R, Vidmar M, Ford M. Relationship between preoperative patient characteristics and expectations in candidates for total knee arthroplasty. Physiotherapy Canada 2009;**61**(1):38-45. doi:10.3138/physio.61.1.38
15. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. Journal of personality and social psychology 1994;**67**(6):1063. doi:10.1037//0022-3514.67.6.1063

16. Tolk JJ, Haanstra TM, Reijman M. Letter to the Editor on "What Do Scottish Patients Expect of Their Total Knee Arthroplasty?". J Arthroplasty 2016;**31**(10):2374. doi:10.1016/j.arth.2016.05.023
17. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. Bone Joint J 2021;103-b(4):619-26. doi:10.1302/0301-620x.103b4.Bjj-2020-0629.R3
18. Venables WN, Ripley BD. Statistics complements to modern applied statistics with S. Fourth ed. New York: Springer; 2002. (ISBN No. 0-387-95457-0)

Declarations

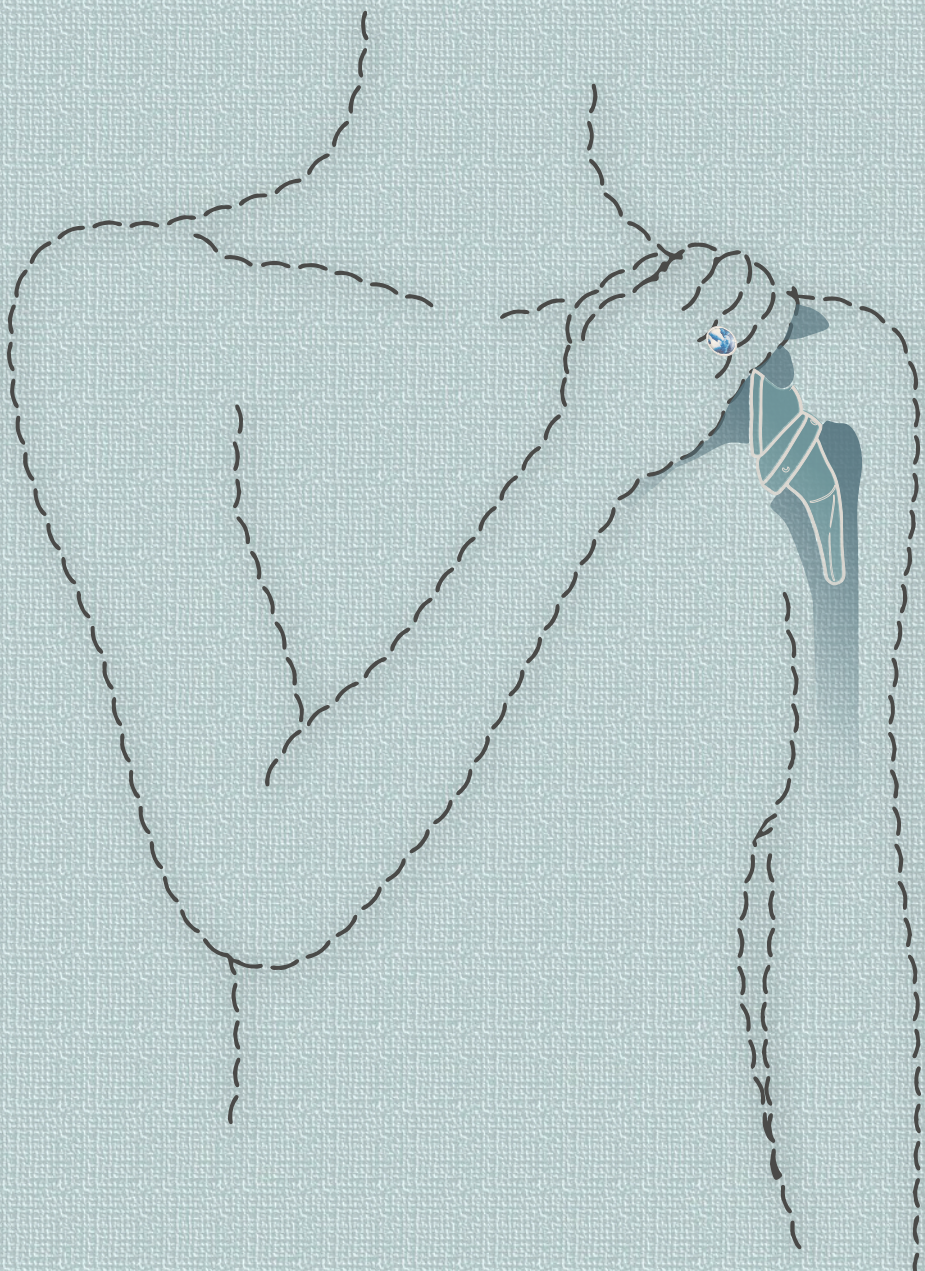
Acknowledgements

The authors would like to thank Roos Bazuin (RHOC), Elga Meurs and Joyce Benner (NorthWest clinics), and Ilse Bouman (Haga hospital) for their efforts in including patients and helping with data collection. In addition, the authors thank the Scientific Committee of the Reinier de Graaf Hospital, Delft, the Netherlands for their financial support (grant no. 621704).

Disclaimers

Funding: This study received financial support for printing costs from the Scientific Research Council / the Board of Directors of the Reinier de Graaf Hospital, Delft, the Netherlands (grant no. 621704).

Conflicts of interest: The 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.



PART III

Measuring change in shoulder function of shoulder arthroplasty patients

5

The Dutch patient-based Constant-Murley score cannot be used interchangeably with the clinician-based Constant-Murley score in shoulder arthroplasty patients

Hesseling B, Perry SIB, Hoelen MA, van Deurzen DFP, Eygendaal D, Mathijssen NM, Snoeker BA

Journal of Shoulder and Elbow Arthroplasty 2026; 10(3)

Abstract

Purpose

To find an alternative for the time-consuming Dutch clinician-based Constant-Murley Score (c-CMS), we assessed its interchangeability with the patient-based CMS (p-CMS) in patients undergoing shoulder arthroplasty. Also, we assessed test-retest reproducibility of both CMS versions.

Methods

The English p-CMS was translated into Dutch using forward and backward translation. We administered the c-CMS and p-CMS at two time points before surgery, and six months after surgery. Subsequently, we calculated absolute agreement parameters, both between p-CMS and c-CMS, to study interchangeability. For the repeated measurements of c-CMS and p-CMS agreement was calculated separately to study test-retest reproducibility.

Results

We recruited 139 patients from three different hospitals. Between total p-CMS and total c-CMS, the 95% limits of agreement (LoAs) were -20.4 to 22.6 with a mean bias of 1.1. For test-retest reproducibility of the c-CMS and the p-CMS, LoAs were -15.1 to 15.7 (mean bias 0.3) and -16.6 to 18.8 (mean bias 1.1), respectively. These LoAs were all wider than the SA-specific minimal clinically important differences (MCIDs) found in the literature.

Conclusions

The p-CMS cannot be used interchangeably with the c-CMS. In addition, the CMS has become obsolete for quantifying shoulder functionality; shoulder-specific PROMs with better reproducibility are currently available.

Level of Evidence: level II (Lower quality prospective cohort study with <80% follow-up)

Keywords: Constant Murley score, shoulder arthroplasty, agreement, test-retest reproducibility, clinimetric study

This study is registered in the Dutch trial register OMON (Overzicht van medisch-wetenschappelijk onderzoek in Nederland, NTR6088): https://www.onderzoekmetmensen.nl/nl/trial/28025

Introduction

To monitor shoulder function in patients with shoulder arthroplasty (SA) and to evaluate the effectiveness of shoulder interventions, the Constant Murley score (CMS)[7, 8] is widely used in clinical practice and research[22]. Compared to shoulder specific patient-reported outcome measures (PROMs), the CMS is characterized by the addition of an objective evaluation of shoulder function through measurements of range of motion and strength.

However, the CMS is time-consuming, taking time away from shared decision-making and patient counseling. In an effort to negate these time constraints, several authors have developed or translated a patient-based CMS (p-CMS) to enable CMS collection outside consultation visits; German[3], English[16], French[5] and Turkish[19] versions exist. However, the clinician-based CMS (c-CMS) can be replaced by a p-CMS only if they are interchangeable.

Interchangeability of existing p-CMS versions has been studied in various ways. Boehm et al.[3] found the German p-CMS to correlate well with the c-CMS ($\rho = 0.82$) and concluded that the p-CMS can be used for follow-up of patients after shoulder surgery. Levy et al.[16] created an English p-CMS and found substantial to almost perfect agreement. They concluded that the p-CMS can be used interchangeably with the c-CMS. Chelli et al.[5] designed a French version of the p-CMS and assessed correlation and mean differences. Although patients underestimated their score on average by 3 points compared to clinicians, the authors concluded that the French p-CMS is a good estimator of the c-CMS. Finally, Mısırlıoğlu et al.[19] translated Levy's p-CMS into Turkish and found lower weighted kappa values than Levy et al.[16]. Together with concerns about differences in test protocols and scoring methods, they therefore concluded that the p-CMS cannot be used interchangeably with the c-CMS.

To date, a Dutch patient-based version has not yet been created or translated. Given the mixed results described above, it is prudent to study whether a Dutch p-CMS could be used interchangeably with the c-CMS.

The primary aim of our study was therefore to translate the English p-CMS into Dutch and to assess absolute agreement between the p-CMS and c-CMS versions in patients undergoing shoulder arthroplasty. The

secondary aim was to analyze the test-retest reproducibility of both the Dutch p-CMS and c-CMS.

Materials and Methods

Study population

This multicenter, prospective cohort study was conducted in three hospitals in the Netherlands: Reinier de Graaf hospital and Haga hospital (merged in 2020 to the Reinier Haga Orthopedic Centre (RHOC)), and Onze Lieve Vrouwe Gasthuis (OLVG) hospital. From September 2016 to February 2021, patients of the orthopedic surgery departments were screened for eligibility according to the following inclusion criteria: age ≥ 18 years, scheduled to undergo primary or revision shoulder arthroplasty (either hemi arthroplasty (HA), anatomic or reversed total shoulder arthroplasty (aTSA or rTSA)), good command of the Dutch language, and able and willing to participate. Exclusion criteria were cognitive impairment, arthroplasty for acute fractures and difficulty with the Dutch language. All participating patients provided written informed consent before the study. The study was reviewed by the regional Medical Ethics Committee, which declared that the study did not fall within the scope of the Medical Research Involving Human Subjects Act (METC Leiden, Den Haag, Delft, in the Netherlands, METC-no. 16-084).

Constant Murley Score

The CMS measures shoulder function using ten items over four domains: pain (15 points), daily activities (20 points), range of motion (40 points), and strength (25 points). The total score ranges from 0 to 100, with higher scores indicating better shoulder function. For the p-CMS, we obtained permission to use and translate the version created by Levy et al.[16]. The translation into Dutch was performed using forward and backward translation by a translation service affiliated with the Dutch Association of Interpreters and Translators. We used the photographs and scoring guidelines from the original English version. The final Dutch version is available in the Appendix.

Data collection

We prospectively collected patient- and clinician-based CMS at three time points: twice preoperatively (ideally spaced two weeks apart, T0 and T1) and six months postoperatively (T2). During each visit, patients and clinicians were blinded to the clinician-based and patient-based scores, respectively.

Clinician-based scores were obtained by four orthopedic surgeons, three physician assistants or a researcher with a minimum of 10 years of experience as a physiotherapist. All raters were specialized in treating shoulder disorders. We deliberately did not standardize the CMS assessment protocol or provide study specific training across raters to incorporate existing practice variation.

Most patient-based scores were obtained via digital questionnaires. Paper questionnaires were used in case e-mail was not available to patients. Missing answers or multiple answers to single-answer items were clarified by phone. If a patient could not be reached, we marked the answers as missing (when completely missing in the questionnaire) or selected the worst option (when multiple answers on a single question were given).

In addition to CMS, we collected data on age, gender, preoperative diagnosis, type of prosthesis, primary or revision arthroplasty, length of hospital stay, and change in shoulder status at T1 and T2 with anchor questions. All data were stored on electronic case report forms (eCRF) using Castor EDC[4].

Statistical analysis

All statistical analyses were conducted in R 4.5.0[23].

We calculated descriptive statistics for our sample. Continuous data are presented as mean with standard deviation (SD) and 95% confidence interval (CI) or, in case not normally distributed, as median with interquartile range (IQR). Nominal and ordinal data are presented as numbers and percentages.

Similar to Levy et al.[16], we used the scores from both T0 and T2 for the comparison between the c-CMS and the p-CMS: scores at T0 and T2 from the same patient can be seen as unique measurements due to the surgery in between. This also ensures that we cover the largest possible range of CMS scores. We did not impute missing data.

Scoring differences between the c-CMS and p-CMS

As already stated by Mısırlıoğlu et al.[19], the scoring methods for the c-CMS and p-CMS differ on the items 'pain', 'work', 'recreational activities' and 'ability to position hand in space'. We have adapted their suggested modifications to the c-CMS scoring as follows.

For 'work' and 'recreational activities', the c-CMS is scored from 1 to 4 points, while the p-CMS is scored at 0, 2 or 4 points. C-CMS scoring was therefore modified by recoding 1 point to 0 points, and 3 points to 2 points.

For 'Ability to position the hand in space', the c-CMS allows for 0 points, while the p-CMS starts at 2 points. C-CMS scoring was therefore modified by recoding 0 points to 2 points.

The item 'pain' in the c-CMS is scored on a scale of 0 (no), 5 (mild), 10 (moderate) or 15 points (severe). For the p-CMS, 'pain' is the average of question a1 (similarly scored to the c-CMS) and question a2 (a Numeric Rating Scale from 0 (no pain) to 15 (maximum pain), of which the inverse is taken). We did not make any modifications to this item, but for the percentage agreement analysis and weighted kappa analysis (described below) it was necessary to use only item a1 of the p-CMS instead of the total item 'pain'.

We calculated the total c-CMS with the modifications as well, and performed our main analyses on the modified c-CMS scoring. The original scoring of the c-CMS was used in additional analyses.

Agreement between c-CMS and p-CMS

To assess the interchangeability of the two methods, the most appropriate measure is agreement[31, 32]. After all, for interchangeability, the most pressing question for the clinician is, 'If the patient rates themselves, would they fill in the same answer as I would?'. If the answer is yes, the score is interchangeable. However, if the answer is no, the clinician would need to repeat the measurement themselves.

We studied agreement for each separate item as well as for the total score of both versions. For the ordinal items 'pain', 'work', 'recreational activities', 'undisturbed sleep', 'ability to position hand in space' and the mobility items, we calculated the percentage exact agreement: the percentage of cases in which the patient gave the exact same score as the clinician. For 'pain', only item a1 from the p-CMS was used.

For ordinal items with five or six answer categories (i.e., 'ability to position hand in space' and the mobility items), we also calculated the percentage adjacent agreement: the percentage of cases in which the scores were the same or differed by only one level (either up or down).

For the items 'strength' and for the total score, we calculated the LoAs using the Bland-Altman method[18]. The LoAs give a range within which 95% of the differences between the measurements are expected to lie, and it is clinically intuitive to interpret.

Reliability of c-CMS versus p-CMS

To directly compare the performance of the Dutch p-CMS to the original English p-CMS[16] and to p-CMS versions in other languages[19], we emulated their reliability analyses by calculating weighted kappa for the ordinal items. We only used the modified c-CMS scoring for the items 'work', 'recreational activities' and 'ability to position hand in space'. For the item 'strength' and the total score, we also calculated the intraclass correlation coefficient (ICC) using a two-may random-effects model, absolute agreement (ICC(2,1))[14].

Agreement for test-retest reproducibility of c-CMS and of p-CMS

To assess the test-retest reproducibility of both the c-CMS and the p-CMS, we repeated the agreement analyses described above on the

repeated measurements (i.e. T0 and T1) of each instrument. In addition, we calculated the standard error of measurement ($SEM_{consistency} = \sqrt{\sigma^2_{error}}$) for the test-retest reproducibility of c-CMS and p-CMS, as an indication of the measurement error of a score with respect to the 'true' score[32, 33]. The value of σ^2_{error} was derived from a two-way random effects model, consistency (ICC(3,1))[14]. Finally, we used the SEM to determine the smallest detectable change (SDC) for the c-CMS and p-CMS by calculating SDC = 1.96 x √2 x SEM[32].

The planned measurement interval between T0 and T1 was 14 days to reduce possible recall bias at T1. However, shoulder function could change within those 14 days. For the test-retest analysis, we therefore added two sensitivity analyses: only on patients who answered 'a little deterioration', 'no change' or 'a little improvement' on the T1 anchor question (the 'less strict' sensitivity analysis), and only on patients who answered 'no change' on the T1 anchor question (the 'more strict' sensitivity analysis).

Sample size

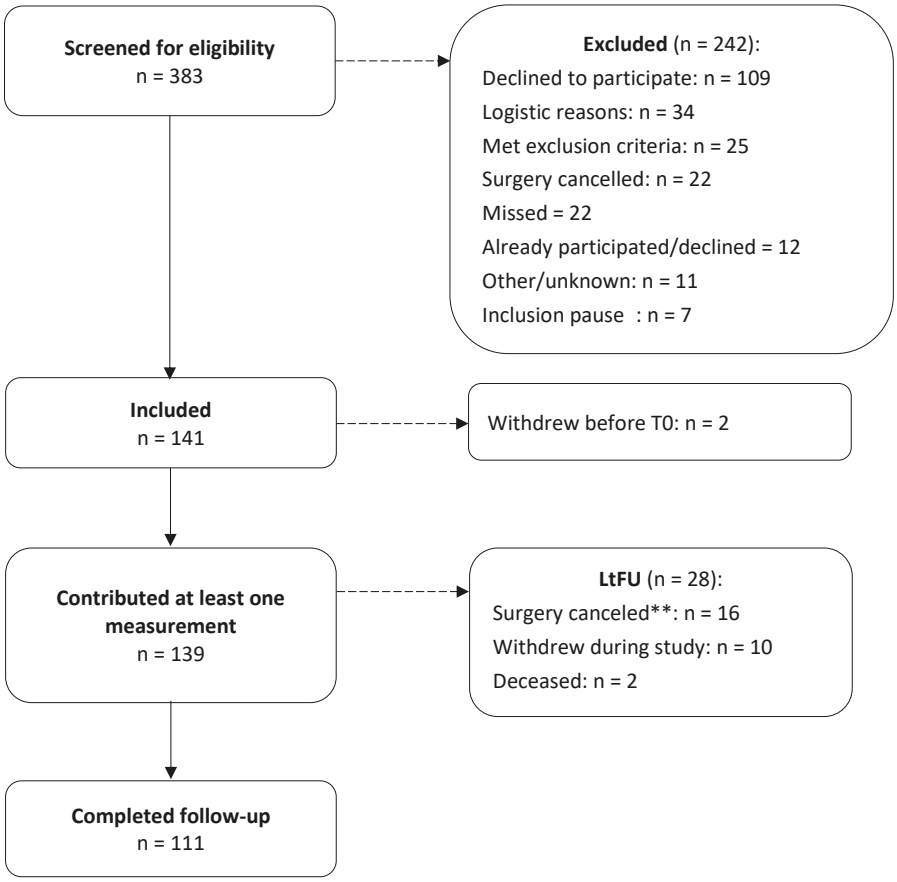
The guidelines published by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative advocate a sample size of 100 subjects for studies that evaluate measurement properties[21, 28]. To account for expected loss to follow-up, we therefore aimed to include at least 125 subjects.

Results

Study sample

Of the 383 screened patients, 141 provided informed consent and were included in the study. Two subjects withdrew before any measurements were taken, leaving 139 subjects for analysis. See Figure 1 for complete study flow details.

The majority of subjects were female (75.5%), the mean age was 69.7 (SD 9.6) years and anatomic Total Shoulder Arthroplasty (aTSA) was performed equally often as reverse Total Shoulder Arthroplasty (rTSA). For more details, see Table 1. Details on the distribution of CMS scores and measurement intervals are in Table 2.



**: The transition from Reinier de Graaf Hospital and Haga Hospital into Reinier Haga Orthopedic Center (RHOC) resulted in an inclusion pause of approximately two months for these centers*

***: When surgery was cancelled measurements at T2 were always missing, but patients could still contribute to T0 and/or T1*

Fig. 1 Study flowchart

Table 1. Patient characteristics (n = 139)

Gender*	
Male	34 (24.5%)
Female	105 (75.5%)
Age[T]	**69.7 ± 9.6**
Preoperative diagnosis*	
Primary osteoarthritis	86 (61.9%)
Cuff tear arthropathy	32 (23.0%)
Avascular necrosis	9 (6.5%)
Secondary osteoarthritis (fracture sequalae)	6 (4.3%)
Massive irreparable cuff rupture	4 (2.9%)
Other	2 (1.4%)
Primary or revision*	
Primary arthroplasty	135 (97.1%)
Revision arthroplasty	4 (2.9%)
Type of prosthesis*	
aTSA	56 (40.3%)
rTSA	56 (40.3%)
HA	7 (5.0%)
No surgery	20 (14.4%)

aTSA = anatomic Total Shoulder Arthroplasty, rTSA = reverse Total Shoulder Arthroplasty, HA = hemiarthroplasty

* Values are given as *n (%)*

T Values are given as *mean ± standard deviation*

Table 2 CMS scores and measurement intervals

Clinician CMS[T]	
T0	31.3 ± 12.6
T1	31.1 ± 12.5
T2	59.3 ± 16.0
Patient CMS[T]	
T0	27.1 ± 13.1
T1	26.2 ± 13.2
T2	59.1 ± 17.3
Intervals[¥] *(in days)*	
Clinician T0 – T1	14 [11 – 23]
Patient T0 – T1	14 [13 – 18]
Clinician vs patient T0	2 [1 – 6]
Clinician vs patient T2	2 [1 – 4]

CMS = Constant Murley score

T Values are given as *mean ± standard deviation*

¥ Values are given as *median [interquartile range]*

Agreement between c-CMS and p-CMS

The percentage exact agreement between the c-CMS and the p-CMS ranged from 32% to 67%, with the objective assessment (i.e. mobility items) having lower exact agreement than the subjective assessment. The percentage adjacent agreement ranged from 67% to 84%. Table 3 shows the percentage agreement for each ordinal item.

In the Bland-Altman analysis, the strength assessment and the total score had a mean bias of 2.6 lbs and 1.1 points, respectively. Figures 2a through 2d show the Bland-Altman plots with corresponding LoAs and 95% CIs. All exact values can be found in Table A-I in the Appendix.

Reliability of c-CMS versus p-CMS

Our weighted kappa values were lower than those found by Levy et al.[16], showing moderate agreement at most. Only the item 'Pain' showed substantial agreement. The ICCs for strength and total score showed good reliability for these scores. The exact results of our reliability analysis can be found in Table 3.

Test-retest reproducibility of c-CMS and of p-CMS

The percentage agreement, SEM and SDC of both the c-CMS and p-CMS, calculated from the repeated measurements at T0 and T1, are presented in Table 3. Figures 3 and 4 show the Bland-Altman plots of both versions. There were no large differences in test-retest reproducibility between the c-CMS and p-CMS. Again, the subjective assessment showed higher agreement than the objective assessment in both CMS versions. For the total score, the mean bias and LoAs were slightly smaller for the c-CMS (Figures 3b) than for the p-CMS (Figures 4b). Exact values of the mean biases, LoAs and corresponding CIs are presented in Table A-I in the Appendix.

Table 4 presents the descriptive statistics for the anchor questions. The planned sensitivity analyses did not substantively change the test-retest reproducibility of the c-CMS or p-CMS, although the LoAs of the total p-CMS narrowed slightly. Sensitivity analysis results are presented in Table A-II and Figures A-1 through A-4 in the Appendix.

Table 3. Agreement within clinician CMS, within patient CMS, and between clinician and patient CMS

	Clinician CMS vs. Patient CMS n = 224		Test – retest Clinician CMS n = 112	Test – retest Patient CMS n = 132
Separate CMS items	Agreement	K_w	Agreement	
Subjective assessment				
Pain¥		0.83 (0.79 – 0.87)		
Exact agreement	65%		67%	63%
Work		0.57 (0.49 – 0.64)		
Exact agreement – modified*	59%			
Exact agreement – original**	35%		71%	77%
Recreational activities		0.41 (0.31 – 0.52)		
Exact agreement – modified*	51%			
Exact agreement – original**	24%		64%	75%
Undisturbed sleep		0.62 (0.52 – 0.72)		
Exact agreement	67%		72%	77%
Ability to position hand in space		0.65 (0.56 – 0.73)		
Exact agreement – modified*	43%			
Exact agreement - original**	42%		46%	53%
Adjacent agreement – modified*	78%			
Adjacent agreement – original**	78%		81%	87%
Objective assessment				
Forward elevation		0.67 (0.59 – 0.74)		
Exact agreement	35%		49%	52%
Adjacent agreement	84%		89%	88%
Lateral elevation		0.63 (0.55 – 0.71)		
Exact agreement	37%		65%	56%
Adjacent agreement	83%		93%	89%
Functional external rotation		0.57 (0.47 – 0.66)		
Exact agreement	32%		54%	53%
Adjacent agreement	67%		85%	80%

	Clinician CMS vs. Patient CMS n = 224		Test – retest Clinician CMS n = 112	Test – retest Patient CMS n = 132
Functional internal rotation		0.44 (0.33 – 0.55)		
Exact agreement	36%		57%	52%
Adjacent agreement	79%		94%	95%
Strength				
ICC		0.970	-	-
$SEM_{consistency}$	-		2.0 lb	1.9 lb
CMS total score				
ICC		0.875* 0.874**		
$SEM_{consistency}$			6 points*,**	7 points
SDC			16 points*,**	18 points

CMS = Constant Murley score, ICC = intraclass correlation coefficient, K_w = weighted kappa, SDC = Smallest Detectable Change SEM = Standard Error of Measurement
Agreement is presented as *percentage exact agreement* or *percentage adjacent agreement*
Reliability is presented as K_w *(95% CI) or ICC*
¥ For the p-CMS, only question a1 is used for the percentage agreement and K_w
* = modified CMS scoring method
** = original CMS scoring method

5

Table 4. Descriptive statistics for anchor questions at T1 and T2

	T1	T2
Very much deteriorated	5 (3.5%)	-
Much deteriorated	11 (7.8%)	5 (3.5%)
A little deteriorated	51 (36.2%)	4 (2.8%)
Unchanged	63 (44.7%)	4 (2.8%)
A little improved	2 (1.4%)	13 (9.2%)
Much improved	-	46 (32.6%)
Very much improved	-	36 (25.5%)
NA	9 (6.4%)	33 (23.4%)

NA = not available

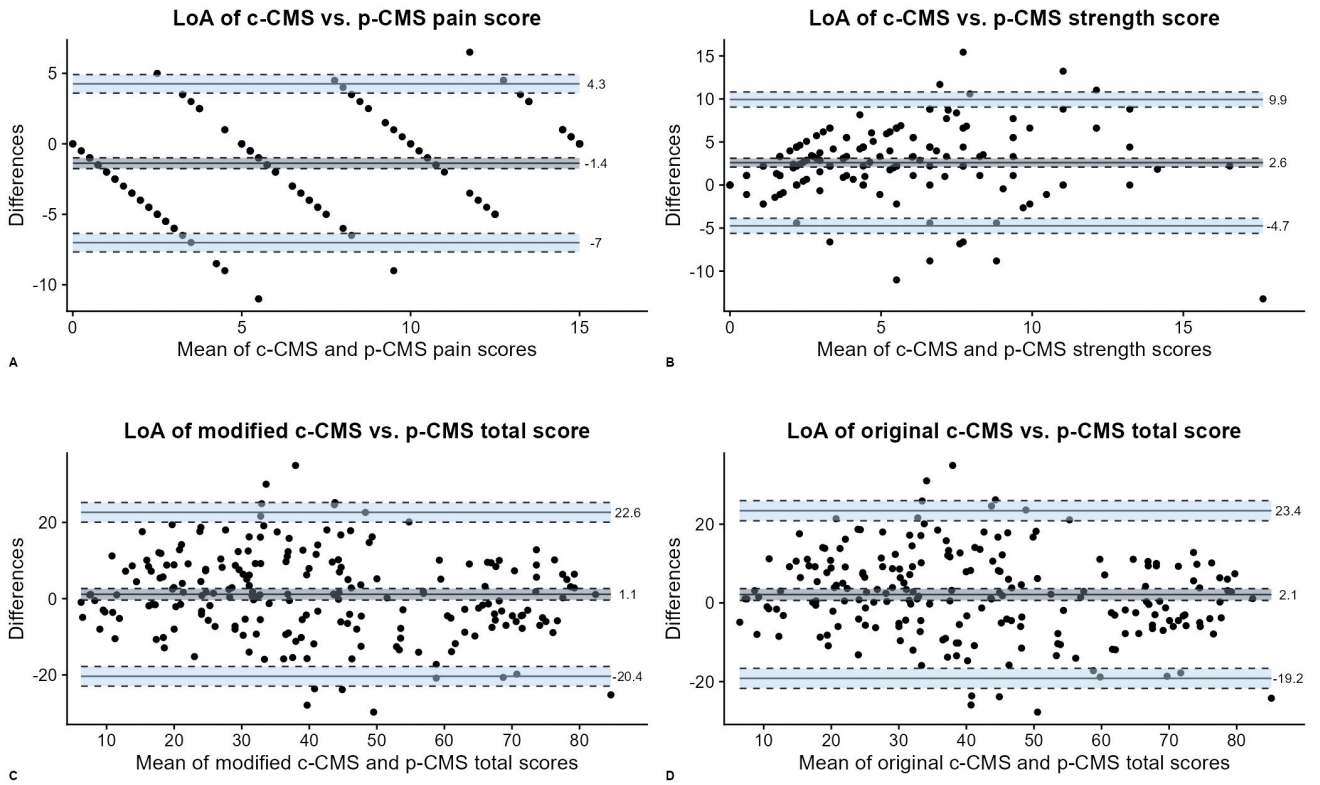


Fig. 2 Limits of Agreement for the comparison of c-CMS with p-CMS for a) pain score, b) strength score, c) modified c-CMS total score and d) original c-CMS total score

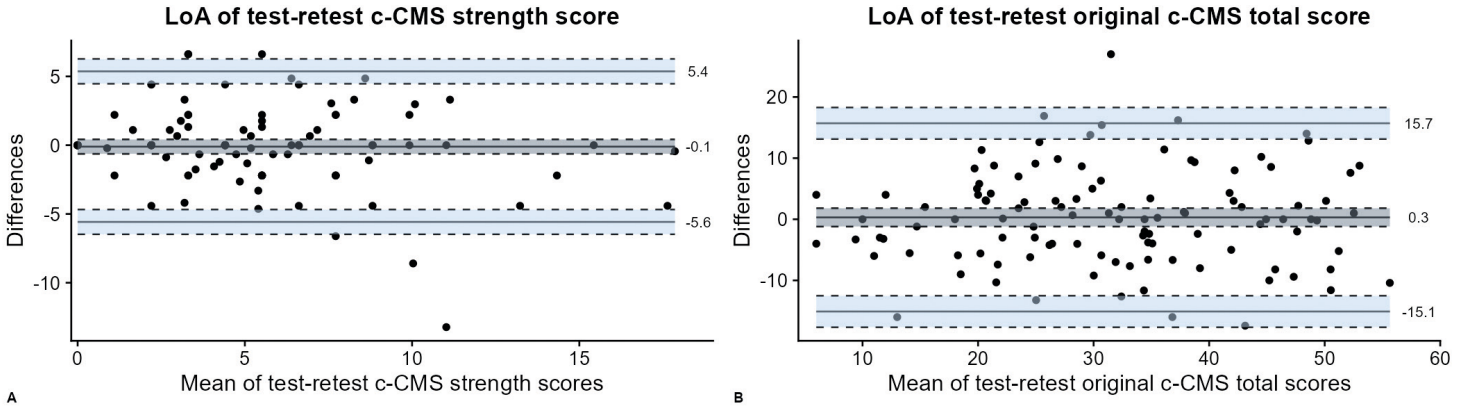


Fig. 3 Limits of Agreement for test-retest reproducibility of the c-CMS for a) strength score and b) original c-CMS total score

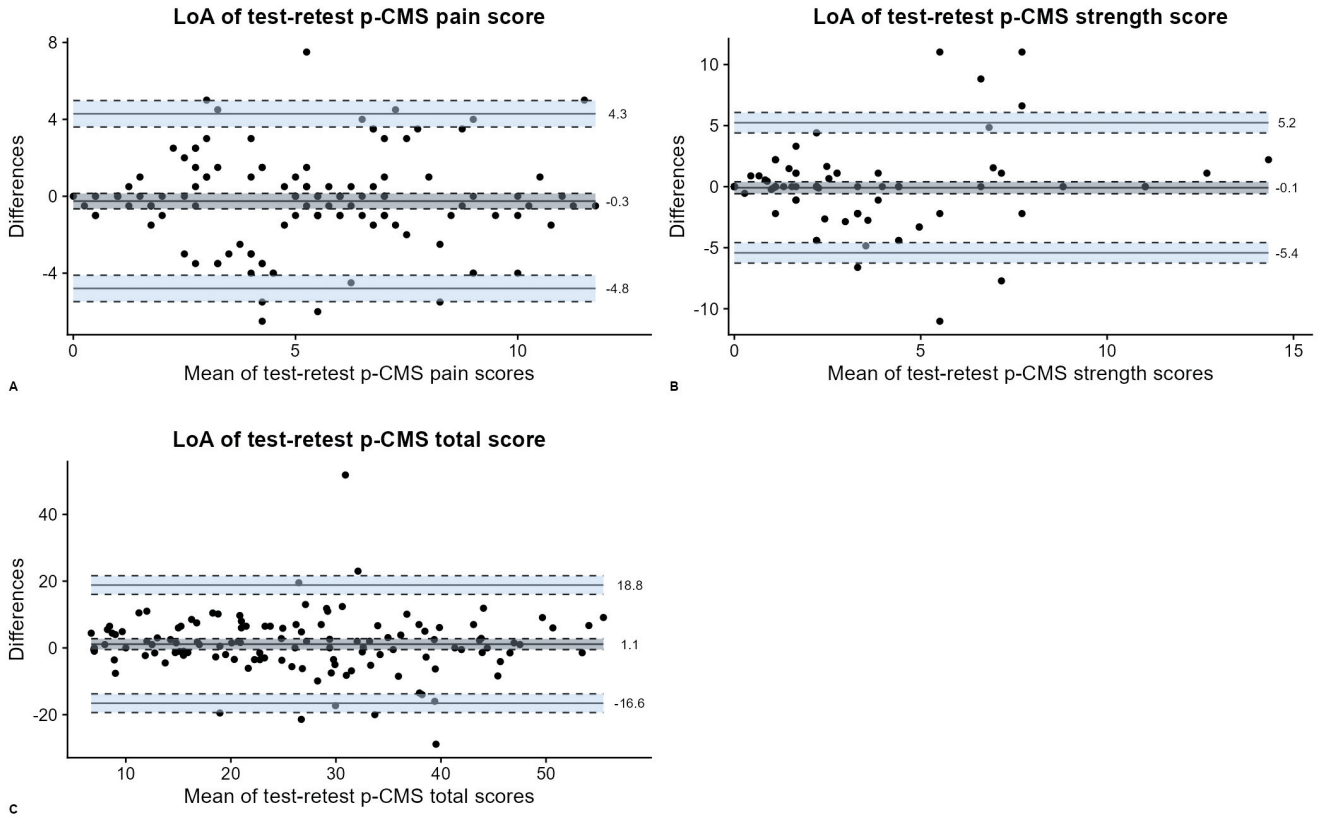


Fig. 4 Limits of Agreement for test-retest reproducibility of the p-CMS a) pain score, b) strength score and c) total score

Discussion

Our study shows that agreement between the c-CMS and p-CMS in patients undergoing shoulder arthroplasty is insufficient to warrant direct interchangeability. In addition, the test-retest reproducibility of both the c-CMS and p-CMS is poor, with SDCs exceeding 10% of the 0-100 measurement scale.

First, we address whether the two methods are interchangeable by examining the wide LoAs for the total score of the c-CMS compared to the p-CMS. One method can only reliably replace the other if the measurement error (LoA or SDC) is smaller than the minimal clinically important difference (MCID)[1, 11, 27]. After all, if the measurement error is larger than the MCID, differences in outcomes could be attributed solely to differences in methods, not to true clinically relevant change. In our study, the LoAs for c-CMS versus p-CMS total scores (upper LoA 22.6, lower LoA -20.4) are much wider than the MCIDs of the CMS total score, even when accounting for CIs around the estimated LoAs (see Table A-I in the Appendix); reported MCIDs in SA literature range from -0.3 to 12.8[13, 25, 34].

This is especially problematic when comparing a new p-CMS measurement from one individual to earlier-obtained c-CMS scores from the same individual. A difference that reaches the MCID may imply a relevant change, whereas in reality it could simply be due to measurement error. The same holds true when comparing p-CMS measurements with c-CMS values from other scientific studies.

We therefore cannot encourage the interchangeable use of the Dutch p-CMS with the c-CMS, even though previous studies concluded that other p-CMS versions performed well enough to warrant interchangeability.

Compared to Levy et al.[16], our LoAs are much wider. This may be due to differences in populations: Levy et al.[16] included patients with a wide variety of shoulder conditions and surgical procedures, whereas we included only SA patients. More importantly, subjects in Levy et al.[16] had a mean age of 54 years. In our study, the mean age was almost 70 years. Self-assessment may be easier for younger patients or those with other shoulder conditions or surgeries than SA.

In contrast, our results do support those of Mısırlıoğlu et al.[19]: we found similar low weighted kappa values.

For the test-retest reproducibility of the p-CMS, we again found wide LoAs for the p-CMS total score. Just as with the comparison between p-CMS and c-CMS, the LoAs were greater than the aforementioned MCIDs. When performing the sensitivity analyses, the LoAs became narrower and approximated the largest MCID of 12.8 points[13]. However, in the sensitivity analyses the SDC remained high at 13 points or more, still above the MCID, indicating poor test-retest reproducibility.

For the test-retest reproducibility of the c-CMS, we found similar poor results: an SDC of at least 16 points and LoAs exceeding the MCIDs. Our LoAs align with the LoAs for the c-CMS reported by other authors[2, 20], although our SDC was slightly larger than that reported by Moeller et al.[20] (SDC = 13.1). Given the wide LoAs from our study and previous studies, we argue that the test-retest reproducibility of the c-CMS is suboptimal.

This suboptimal reproducibility of the c-CMS raises concerns about its widespread use in both clinical practice and scientific research and, at the same time, may create an opportunity to use the p-CMS despite its lack of interchangeability with the c-CMS.

On the one hand, one could argue that the c-CMS is of limited value in evaluating change in a patient's shoulder function accurately. Since the LoAs and SDCs encompass SA-specific MCIDs, the question is whether we can actually measure relevant change within patients. Furthermore, the greatest perceived benefit of the c-CMS is the objective evaluation of strength and ROM. However, the benefit is limited if we are not able to reproduce these domains.

On the other hand, despite suboptimal test-retest reproducibility, in practice, the c-CMS remains widely accepted for evaluating shoulder function. Given the comparable reproducibility of the p-CMS to the c-CMS, although poor, one could argue that the p-CMS is similarly acceptable. In that view, the p-CMS could be implemented in standard care to alleviate time constraints during consultations by having patients complete the p-CMS beforehand. In the context of scientific studies, the p-CMS could

eliminate logistical barriers to collect mobility and strength data by allowing patients to complete it remotely. However, it is essential to acknowledge that using the p-CMS would have severe limitations, as described in the section above on interchangeability. The potential use of the p-CMS comes with a serious trade-off.

Recommendations on the use of the c-CMS and p-CMS

As stated above, we cannot encourage the interchangeable use of the c-CMS and the Dutch p-CMS. In addition, we strongly urge the orthopedic community to critically consider whether the benefit of the objective evaluation in the CMS actually outweighs its drawbacks, given its poor reproducibility. More specifically, we argue that including range of motion and strength is no longer of major benefit.

Range of motion and strength are primarily proxies for assessing whether patients can use their shoulders for daily tasks. This approach was especially valuable when the CMS was introduced in 1987[7], at a time when outcome evaluations focused chiefly on clinical measures such as prosthesis survival and range of motion[30]. However, over the last decades, the use of PROMs has increased markedly[12, 15], because they focus on outcomes that matter to patients (ranging from joint-specific symptoms to functional limitations and health-related quality of life)[6].

Many shoulder-specific PROMs currently in use, such as the Oxford Shoulder Score (OSS[9]) and the Simple Shoulder Test (STT[17]), include questions about functional activities that require a certain range of motion and strength. Although studies assessing agreement parameters for these PROMs are scarce (most studies use relative reliability measures), reproducibility appears better for these PROMs than for the c-CMS[10, 26, 29]. Given the poor reproducibility of the c-CMS, we argue that it has become obsolete. Consequently, we likewise do not advocate the use of the Dutch p-CMS over existing shoulder-specific PROMs, either.

Strengths and limitations

Just as every other study, ours has both strengths and limitations. The c-CMS protocol was not standardized for the assessors in this study. We

purposefully chose not to standardize to match clinical practice; Rocourt et al.[24] noted large discrepancies in how test protocols were defined and performed in different European hospitals. On the other hand, Blonna et al.[2] found higher reliability and reproducibility with a standardized version of the CMS than with a non-standardized version. Although the lack of standardization could thus have reduced our agreement results for both interchangeability and test-retest outcomes, our results are representative of use in clinical practice and in studies that use c-CMS scores collected in standard care.

A major strength of our study is the use of absolute agreement parameters instead of relative reliability parameters. Agreement and reliability parameters have different uses and interpretations[31,32]. Agreement assesses how close scores for repeated measurements are (both for intra- and interrater agreement) and thus concerns measurement error in isolation.

Conversely, reliability assesses how well patients can be distinguished from one another despite measurement error; this error is thereby related to the variability within a sample. The conceptual differences between agreement and reliability explain why the ICC and weighted kappa (relative reliability parameters) can be high while absolute agreement is poor: in such cases, the variability between persons is large relative to the measurement error. The conceptual differences also mean that reliability parameters are suited for instruments with discriminate purposes, whereas agreement parameters are suited for instruments with evaluative purposes. The CMS is mainly used to evaluate changes in shoulder function; it is not used discriminatively to, for example, decide which patients need surgery. Subsequently, absolute agreement parameters are the most appropriate measures for assessing interchangeability and reproducibility of the c-CMS and p-CMS.

Conclusion

The Dutch p-CMS cannot be used interchangeably with the c-CMS. In addition, we argue that the CMS has become obsolete for quantifying shoulder functionality, due to the availability of many shoulder-specific PROMs.

Supplementary materials

The supplementary materials can be viewed online at
https://ars.els-cdn.com/content/image/1-s2.0-S2471549226000341-mmc1.pdf

Appendix 1

References

1. Barten JA, Pisters MF, Huisman PA, Takken T, Veenhof C. Measurement properties of patient-specific instruments measuring physical function. J Clin Epidemiol 2012;**65**(6):590-601. doi:10.1016/j.jclinepi.2011.12.005
2. Blonna D, Scelsi M, Marini E, Bellato E, Tellini A, Rossi R, et al. Can we improve the reliability of the Constant-Murley score? J Shoulder Elbow Surg 2012;**21**(1):4-12. doi:10.1016/j.jse.2011.07.014
3. Boehm D, Wollmerstedt N, Doesch M, Handwerker M, Mehling E, Gohlke F. [Development of a questionnaire based on the Constant-Murley-Score for self-evaluation of shoulder function by patients]. Unfallchirurg 2004;**107**(5):397-402. doi:10.1007/s00113-004-0757-3
4. Castor EDC. Castor Electronic Data Capture; 2019. Available from: https://castoredc.com. [Accessed August 28, 2019.
5. Chelli M, Levy Y, Lavoué V, Clowez G, Gonzalez JF, Boileau P. The "Auto-Constant": Can we estimate the Constant-Murley score with a self-administered questionnaire? A pilot study. Orthop Traumatol Surg Res 2019;**105**(2):251-6. doi:10.1016/j.otsr.2018.11.023
6. Ciani O, Federici CB. Value Lies in the Eye of the Patients: The Why, What, and How of Patient-reported Outcomes Measures. Clin Ther 2020;**42**(1):25-33. doi:10.1016/j.clinthera.2019.11.016
7. Constant C, Murley A. A clinical method of functional assessment of the shoulder. Clinical orthopaedics and related research 1987(214):160-4.
8. Constant CR, Gerber C, Emery RJ, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. Journal of shoulder and elbow surgery 2008;**17**(2):355-61.
9. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br 1996;**78**(4):593-600.
10. Frich LH, Noergaard PM, Brorson S. Validation of the Danish version of Oxford Shoulder Score. Dan Med Bull 2011;**58**(11):A4335.
11. Grunow JJ, Hartmann L, Ulm B, Maechler M, Blobner M, Seidenspinner K, et al. Reliability of pre-admission patient-reported outcome measures postoperatively assessed via proxies: a prospective, multicenter observational study. Crit Care 2025;**29**(1):200. doi:10.1186/s13054-025-05431-6
12. Jin H, He M, Xie W, Xiong Z, Deng Z, Li Y. Research Trends of Patient-Reported Outcome Measures in Orthopedic Medical Practices: A Bibliometric and Visualized Study. Medicina (Kaunas) 2023;**59**(9). doi:10.3390/medicina59091664
13. 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**(7):1371-83. doi:10.1097/corr.0000000000002164
14. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coeffici ents for reliability research. J Chiropr Med 2016;**15**(2):155-63.
15. Lakey E, Hunt KJ. Patient-Reported Outcomes in Foot and Ankle Orthopedics. Foot Ankle Orthop 2019;**4**(3):2473011419852930. doi:10.1177/2473011419852930
16. Levy O, Haddo O, Massoud S, Mullett H, Atoun E. A patient-derived Constant-Murley score is comparable to a clinician-derived score. Clin Orthop Relat Res 2014;**472**(1):294-303. doi:10.1007/s11999-013-3249-3
17. Lippitt S, Matsen F, Fu F, Hawkins R. A practical tool for evaluating function: the Simple Shoulder Test. The shoulder: a balance of mobility and stability Rosemont (IL): American Academy of Orthopaedic Surgeons 1993.

18. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of cli nical measurement. Lancet 1986;**327**(8476):307-10.
19. Mısırlıoğlu T, Eren İ, Taşkıran Ö, Günerbüyük C, Birsel O, Canbulat N, Demirhan M. Turkish version of the patient-based Constant-Murley Score: Its cross- cultural adaptation, validity, reliability and comparison with the cli nician-based version. Turk J Phys Med Rehabil 2022;**68**(3):336-47.
20. Moeller AD, Thorsen RR, Torabi TP, Bjoerkman A-SD, Christensen EH, Maribo T, Christiansen DH. The Danish version of the modified Constant-Murley shoulder score: rel iability, agreement, and construct validity. J Orthop Sports Phys Ther 2014;**44**(5):336-40.
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of life research 2010;**19**(4):539-49.
22. Mosher ZA, Ewing MA, Collins CS, Young PG, Brabston EW, Momaya AM, et al. Usage Trends of Patient-reported Outcome Measures in Shoulder Literature. J Am Acad Orthop Surg 2020;**28**(17):e774-e81. doi:10.5435/jaaos-d-19-00455
23. R. Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2025.
24. Rocourt MHH, Radlinger L, Kalberer F, Sanavi S, Schmid NS, Leunig M, Hertel R. Evaluation of intratester and intertester reliability of the Constant- Murley shoulder assessment. J Shoulder Elbow Surg 2008;**17**(2):364-9.
25. Simovitch R, Flurin P-H, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the minimal clinically important difference. Journal of shoulder and elbow surgery 2018;**27**(2):298-305.
26. St-Pierre C, Desmeules F, Dionne CE, Frémont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. Disabil Rehabil 2016;**38**(2):103-22. doi:10.3109/09638288.2015.1027004
27. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;**60**(1):34-42. doi:10.1016/j.jclinepi.2006.03.012
28. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of life research 2012;**21**(4):651-7.
29. van Kampen DA, Willems WJ, van Beers LW, Castelein RM, Scholtes VA, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res 2013;8:40. doi:10.1186/1749-799x-8-40
30. Van Lieshout EMM, Wijffels MME. Patient-reported outcomes: Which ones are most relevant? Injury 2020;51 Suppl 2:S37-s42. doi:10.1016/j.injury.2019.10.073
31. Vet HCW, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's κ. BMJ 2013;**346**(apr12 1):f2125.
32. Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;**59**(10):1033-9.
33. Weir JP. Quantifying test-retest reliability using the intraclass correlation c oefficient and the SEM. J Strength Cond Res 2005;**19**(1):231-40.

34. Zhou A, Xu S, Yew KSA, Lie DTT. Minimal Clinically Important Differences for Oxford, Constant, and University of California Los Angeles Shoulder Scores After Reverse Shoulder Arthroplasty to Allow Interpretation of Patient-Reported Outcome Measures and Future Statistical Power Analyses. Arthroscopy 2023;**39**(6):1405-14. doi:10.1016/j.arthro.2022.12.027

Declarations

Acknowledgements

We would like to thank R. Bazuin (RHOC) for her efforts in including patients and helping with data-collection. Furthermore, we would like to thank R. Janssen, PA (OLVG), Ilse Bouman, PA (Haga hospital), J.A. van der Linde, MD PhD (RHOC), P.F. Flikweert, MD PhD (RHOC/Haga hospital), F. van Eijk, MD PhD (Haga hospital) and M. van den Bekerom, MD Prof (OLVG) for their efforts in including patients and performing c-CMS measurements.

Funding

This work was supported by the Scientific Research Council / the Board of Director of the Reinier de Graaf Hospital, Delft, The Netherlands (grant no. 621608). The funding source was not involved in the study design, collection analysis and interpretation of the data, writing of the report and decision to submit the article for publication.

6

General discussion

General discussion

Shoulder arthroplasty (SA) is a successful intervention to relieve pain and improve function for patients who suffer from degenerative shoulder disorders. It can be seen as a life-altering event[1] but its prospect can also cause uncertainty and preoperative anxiety[2,3], which in turn may lead to poorer outcomes[4].

The origin of this PhD thesis lies in the desire to improve the patient's experience of the shoulder arthroplasty journey. Part of this improvement is to enrich the orthopedic surgeon's understanding of patient values and preoperative expectancies of the surgery, in order to enhance postoperative outcomes. These efforts specifically aim to facilitate better-aligned communication between the surgeon and the patient about the surgery and, ultimately, to enable the patient to feel confident, supported and assured throughout the pre- and postoperative period.

Despite the relevance of all these topics, it was necessary to focus this thesis on specific aspects of patient counseling in order to conduct in-depth research. Therefore, the studies in this thesis provide answers on several of these aspects.

This general discussion begins by addressing the limitations and considerations across three main themes: 1) what we can tell patients about their outcomes, 2) patients' expectations of SA and 3) measuring change in shoulder function. This is followed by practical suggestions for clinicians to adapt their daily practice based on our study findings. The general discussion concludes with reflections on additional steps, such as validation and implementation studies, that are needed to reach our goal of optimally guiding patients through their shoulder arthroplasty journey.

Part I – What can we tell patients about their outcomes after shoulder arthroplasty?

Patient education – prognostic factors

We know from previous research[5] that patients desire informational support from their surgeon.

For orthopedic surgeons to tailor their advice and expectation management to individual SA patients, they need information on patient-related prognostic factors for functional and pain-related outcomes after SA. From our systematic review, a notable finding was that the strongest predictors, those clearly associated with worse or better outcomes, were each examined in only a single study. In contrast, other prognostic factors analyzed across multiple studies showed no consistent direction of effect and often allowed for neutral outcomes. Unfortunately, our review was not able to detect strong prognostic factors, i.e. factors confirmed in multiple studies.

Many studies in our systematic review exhibited severe limitations in their reporting, which prevented us from conducting additional meta-analyses: although several factors were included in multivariable analyses across multiple studies, often only results for the predictor of interest were reported. Complete reporting is essential to uphold the high standards of scientific integrity. Moreover, researchers have an ethical responsibility to maximize the value derived from the time and effort sacrificed by study participants. Having the level of evidence downgraded or being unable to contribute to the aggregation of evidence due to reporting bias fails to fulfill this responsibility.

However, incomplete or selective reporting was not the only issue; many included studies also used inconsistent language. For instance, they would mention correcting for confounders yet report the independent predictive effect of a prognostic factor[6]. This highlights a deeper problem in biomedical research: the distinction between causal and predictive modeling is not always clear.

Inconsistent language and the underlying lack of knowledge of conceptual differences between causal and predictive research hampers the design

of high-quality studies from which valid inferences can be drawn. It also impedes readers' accurate interpretation of results. For example, mixing causal and predictive terminology may lead readers to wrongfully conclude that altering certain practices will affect outcomes.

While researching separate prognostic factors is one approach to tailor preoperative education to individual patients, another approach is developing and validating prediction models. After all, prediction models use multiple prognostic factors in combination to predict individual risk. Karimi et al.[7] found in their systematic review that machine learning models could predict clinical improvement in PROMs after TSA with 93% to 99% accuracy. However, none of the models included in this systematic review appear to have been externally validated. We are also not aware of any impact studies to investigate the influence of the models on decision making and patient outcomes. These are essential steps in prognostic model research, but unfortunately often skipped in favor of developing new models[8-10].

Patient education – pain trajectories

We also utilized yet another approach to contribute to better patient education, by examining patients' pain recovery trajectories during the first eight postoperative weeks using Latent Growth Curve Modeling (LGCM). We found six distinct trajectories that, from a clinical standpoint, could be further categorized into two subgroups: patients with a faster recovery (low pain scores after two weeks) and patients with a slower recovery (with pain scores needing up to eight weeks to reach similar low levels).

The relevance of this study is supported by several authors. Furtado et al.[11] found that 83% of SA patients wanted preoperative education programs to include experiences from previous patients. Similarly, Goldsmith et al.[5] reported that TKA patients desired insights into others' experiences along with realistic forewarning about the severity and longevity of postoperative pain. Kennedy et al.[12] found that TKA and THA patients wanted information on the expected levels of postoperative pain. Furthermore, this informational need is not felt by only orthopedic patients; in a study by Mavridou et al.[13], 76% of patients undergoing surgery (of whom only a small number underwent orthopedic surgery) wanted information on how long postoperative pain would last.

When patients do not receive sufficient preoperative information about what to expect, uncertainty can arise. Kagan et al.[4] define uncertainty as "a dynamic state in which there is a perception of being unable to assign probabilities for outcomes that, in turn, prompts a discomforting, uneasy sensation." They found that preoperative anxiety and uncertainty were related to worse mental and physical recovery after elective arthroplasty. In contrast, adequate preoperative education can reduce anxiety levels[14-16]. This is important, as several studies have shown that preoperative anxiety predicts postoperative pain intensity after joint arthroplasty[17-21].

Yet, prediction is not the same as causation: lowering preoperative anxiety may not necessarily lead to better outcomes. Even so, some authors have suggested potential causal or bidirectional relationships[22,23]. Moreover, Ho et al.[24] developed a patient-specific integrated education program (IEP) for TKA patients and compared it to usual care. Patients in the IEP group reported lower state-anxiety scores on the day of admission and at several postoperative time points, along with better pain and patient-reported outcome results after surgery, compared to the usual care group.

The trajectories we discerned are thus valuable for adding more detail to the preoperative education of SA patients, although an IEP such as Ho et al.'s[24] entails much more than just information on pain recovery. Furthermore, uncertainty about early postoperative pain recovery is only one domain that can lead to preoperative anxiety. Worries might also arise on surgical complications, other aspects of the recovery period, other outcomes, and the hospital period itself[25,26]. While a few of these aspects are required to discuss according to the Dutch Medical Treatment Contracts Act (WGBO) (e.g. complications), all issues should be addressed in conjunction with each other for preoperative education programs to have optimal effect.

Part II – Patients' expectations of shoulder arthroplasty

As discussed previously, an important goal of preoperative education programs is to reduce anxiety and uncertainty. Kagan et al.'s[4] definition of uncertainty, more specifically the part "being unable to assign probabilities for outcomes", directly links to the field of patient expectations; defining expectations as probabilistic (what a patient realistically expects) is one

of the two main paradigms to measure expectations, the other paradigm being value-based (what is important to a patient)[27,28]. In our research, we studied both because the distinction between probabilistic and value-based expectations is crucial, especially since some expectation questionnaires (such as those from the Hospital for Special Surgery (HSS)) exist in both versions. One asks, "How important are these expectations in the treatment of your [condition]?" (value-based), while the other asks, "How much improvement do you expect from the treatment of your [condition]?" (probabilistic). Unfortunately, primary studies often fail to clarify which version was used, sometimes omitting the question type altogether or describing one type while citing the other.

Since preoperative counseling and educational programs often aim to set realistic expectations[29-31] and thus target probabilistic expectations, their effectiveness cannot be defined using value-based instruments. Counseling the patient on how likely outcomes are, will not necessarily change *how important* these expectations are to the patient. Conversely, how important certain expectations are to the patient does not capture how the patient perceives the probability of these outcomes to occur. For example, Lawrence et al.[32] concluded that expectations of shoulder surgery were not altered after preoperative counseling by the surgeon. However, they measured value-based expectations and not probabilistic ones, capturing only the rated importance of the expectations.

Nonetheless, value-based expectations are still equally important: they guide research and clinical conversations towards what matters most to patients.

The distinction between probabilistic and value-based paradigms is not the only source of variation in expectation research. The definition of and methodology to measure expectation fulfilment and satisfaction can also differ between studies. Orthopedic literature reveals two main approaches to measure fulfilment. Some studies query patients' current status or degree of progress without first recalling their preoperative expectations[33-35], which is what we did in our study in Chapter 4. This approach minimizes recall bias and cognitive burden, but will not capture what patients themselves *think* of fulfilment. In contrast, other studies explicitly ask patients to what extent their preoperative expectations were met, sometimes by repeating the original expectations verbatim[36-39]. While

this method directly assesses patients' subjective perception of fulfillment, it risks introducing substantial recall bias.

With respect to satisfaction, we focused on how satisfied patients were with the functional outcome after surgery. Although this is the most logical type of satisfaction given that the questionnaires are outcome-oriented, it is a different type than satisfaction with process of or experience with received care; our conclusion that expectation fulfilment or improvement is related to satisfaction therefore cannot be generalized to satisfaction with process or experience measures.

As a final remark it is worth repeating that our study is the first to actually quantify the relationship between expectation fulfillment and satisfaction in SA patients, moving beyond mere correlations as seen in Green et al.[40]. Unfortunately, correlation analyses are still too prevalent in orthopedic literature. Correlation analyses are rarely informative in this field and should be reserved for hypothesis-generating studies; orthopedic research typically starts with clear hypotheses, for which effect estimates with uncertainty measures offer far greater clinical value.

Part III – Measuring change in shoulder function of shoulder arthroplasty patients

Besides asking whether expectations have been fulfilled, another way to measure change in the patient's status is to use validated outcome measures that directly capture relevant outcomes. One of the most often used measurement instruments in shoulder arthroplasty research to evaluate change over time is the Constant Murley score (CMS)[41-45]. However, the CMS is time-consuming to perform during consultations[46]. Although not explicitly documented in literature, the amount of time involved may introduce substantial bias in retrospective studies that use prospectively gathered data from standard care, as missing CMS scores could be associated with specific patient characteristics. For example, patients exhibiting higher anxiety or catastrophizing may require more time to have questions answered or receive reassurance during their appointment[47], leaving insufficient time for the orthopedic surgeon to complete the CMS. Having a patient-based version of the CMS might reduce the number of missing scores, which could benefit future retrospective studies. We have

therefore sought to translate an English-language patient-based version of the score (p-CMS) to Dutch and to investigate whether this Dutch version is interchangeable with the original, clinician-based score (c-CMS).

Unfortunately, we had to conclude that the Dutch p-CMS cannot be used interchangeably with the c-CMS. In fact, the test-retest reliability of not only the p-CMS but also the c-CMS was poor enough for us to advocate for discontinuing the CMS to quantify shoulder function in favor of better reproducible PROMs.

This may understandably evoke considerable resistance among orthopedic surgeons or other healthcare providers who use the CMS. The CMS does indeed have high face validity; someone needs sufficient range of motion and strength to perform daily activities and it is therefore valuable to consider these outcomes when evaluating a patient. It is important to note, however, that advocating for discontinuing the CMS to quantify shoulder function does not mean abandoning shoulder function assessments altogether during consultations and follow-up. Still, besides recording range of motion or strength, it is essential to observe and sense how patients function with their shoulder, including noticing bodily tension, compensatory movements, and associated thoughts or fears. This should remain an essential part of preoperative and postoperative face-to-face appointments by all means.

Changes in daily practice

Based on the research outcomes in this thesis, orthopedic surgeons can already implement several changes in their daily practice. First and foremost, they can use figures of the trajectories to show in an intuitive manner how pain recovery progresses for many patients during the first eight weeks. These figures can be used preoperatively to mentally prepare patients and adjust expectations, and postoperatively to reassure them that, even if they experience higher initial pain levels, all trajectories converge to low pain scores by eight weeks.

They can also recognize the difference between probabilistic and value-based expectations; letting value-based expectations direct the conversation straight to what is relevant for the patient and using

probabilistic expectations to gauge whether additional patient education is needed to correct unrealistic ones. The bar charts in Chapter 4 may serve as visual aids. To free up time to discuss these matters, surgeons can decide to discontinue the use of the Constant-Murley score in clinical practice.

Finally, the surgeon can discuss factors from the systematic review during preoperative education, albeit cautiously as many have only been investigated in few studies and thus remain uncertain.

Future perspectives

This thesis represents a step in the right direction towards optimally guiding patients through their shoulder arthroplasty journey. However, to fully realize this ambitious goal, more steps are clearly needed. Research on two domains can contribute towards further personalization of preoperative education targeted towards SA patients: better individualized prediction of outcomes and patient segmentation followed by customization.

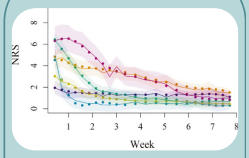
To achieve better individualized prediction of outcomes after surgery, future studies should externally validate existing prediction models as well as the early pain recovery trajectories identified in this thesis. If external validation reveals shortcomings in these models, attempts to improve them are warranted. The rapidly growing possibilities within statistical methods and large language models may play an important role here. Should external validation confirm their promise, integrating these models into a comprehensive preoperative education program is valuable. Tools like decision aids, central to value-based and patient-centered care, stand to benefit greatly from such predictions.

Nevertheless, the right content alone is insufficient; equally important are design elements like timing, delivery mode and tone of voice, which may vary across patient types. Patient involvement in program development is thus essential. Future studies could segment the SA population into subgroups sharing similar clinical, psychological, and communication characteristics, tailoring education programs accordingly[48-50]. Digital solutions, such as patient journey apps, are particularly well-suited to additional customization.

Finally, once promising personalized education programs have been designed, impact studies are needed to assess their effects on key outcomes such as pain, function, satisfaction, and patient reported experience measures. For such programs to stand the test of time, their benefits must outweigh the investments in technology, personnel, and training.

IMPROVING THE PATIENT'S EXPERIENCE IN SHOULDER ARTHROPLASTY – TOWARDS MORE PATIENT-SPECIFIC COUNSELING AND EXPECTATION MANAGEMENT

Changes in daily practice to implement today



Visual trajectories improve understanding and reassurance

Use before surgery to set expectations

Use after surgery to reassure patients

Most patients reach low pain scores after 2 weeks

Create time for meaningful discussion

Discontinue routine use of Constant-Murley score

Free up time for patient-centered communication

Different expectations require different conversations

Use value-based expectations to guide the conversation

Use probabilistic expectations to correct unrealistic beliefs

Discuss risk factors carefully

Many risk factors have only been investigated in single studies

Previous shoulder surgery may predict worse medium-long term functional outcome

Future research should focus on...

...more accurate predictions to enable personalized education

Improve and validate prediction models

Integrate into decision aids

...different educational needs for different types of patients

Segment shoulder arthroplasty patients into subgroups with similar clinical, psychological and communication characteristics

Tailor patient education to each subgroup and utilize digital patient journey tools

... proving real benefit of education programs

Study impact on functional measures as well as satisfaction and experience measures

Benefits must outweigh investments in technology, personnel and training

References

1. Gustafsson BÅ, Ekman SL, Ponzer S, Heikkilä K. The hip and knee replacement operation: An extensive life event. *Scandinavian Journal of Caring Sciences* 2010; **24**(4): 663-70.
2. Buus AAØ, Hejlsen OK, Dorisdatter Bjørnes C, Laugesen B. Experiences of pre- and postoperative information among patients undergoing knee arthroplasty: a systematic review and narrative synthesis. *Disability and Rehabilitation* 2021; **43**(2): 150-62.
3. Cronström A, Jönsson T, Limbäck G, Ljung M, Ståhl C, Östlind E. Stairway to heaven via the highway to hell: a qualitative study on patients' experience of knee joint replacement surgery. *J Orthop Surg Res* 2025; **20**(1): 570.
4. Kagan I, Bar-Tal Y. The effect of preoperative uncertainty and anxiety on short-term recovery after elective arthroplasty. *J Clin Nurs* 2008; **17**(5): 576-83.
5. Goldsmith LJ, Suryaprakash N, Randall E, et al. The importance of informational, clinical and personal support in patient experience with total knee replacement: a qualitative investigation. *BMC Musculoskelet Disord* 2017; **18**(1): 127.
6. Moverman MA, Puzzitiello RN, Pagani NR, et al. Functional somatic syndromes are associated with suboptimal outcomes and high cost after shoulder arthroplasty. *Journal of Shoulder and Elbow Surgery* 2022; **31**(1): 48-55.
7. Karimi AH, Langberg J, Malige A, Rahman O, Abboud JA, Stone MA. Accuracy of machine learning to predict the outcomes of shoulder arthroplasty: a systematic review. *Arthroplasty* 2024; **6**(1): 26-.
8. Collins GS, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ* 2024; **384**: e074819-e.
9. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; **98**(9): 691-8.
10. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Medicine* 2013; **10**(2): e1001381-e.
11. Furtado R, MacDermid JC, Bryant D, Faber KJ, Drosdowech DS, Athwal GS. Balancing clinician and patient priorities for total shoulder replacement preoperative education programs. *Patient Education and Counseling* 2023; **112**.
12. Kennedy D, Wainwright A, Pereira L, Robarts S, Dickson P, Christian J, Webster F. A qualitative study of patient education needs for hip and knee replacement. *BMC Musculoskeletal Disorders* 2017; **18**(1): 413-.
13. Mavridou P, Manataki A, Arnaoutoglou E, Damigos D. A Survey of Patients' Preoperative Need for Information About Postoperative Pain—Effect of Previous Surgery Experience. *Journal of PeriAnesthesia Nursing* 2017; **32**(5): 438-44.
14. Althobiti E, Almashi A, Albawinei A, Alnashri F, Alsubaiea F, Al Nagshabandi E, Banakhar M. Effect of Preoperative Education on Patient Anxiety Level: A Scoping Review. *Evidence-Based Nursing Research* 2020; **2**(1): 10-.
15. Aydin D, Klit J, Jacobsen S, Troelsen A, Husted H. No major effects of preoperative education in patients undergoing hip or knee replacement--a systematic review. *Danish medical journal* 2015; **62**(7).
16. Longo UG, De Salvatore S, Rosati C, et al. The Impact of Preoperative Education on Knee and Hip Replacement: A Systematic Review. *Osteology* 2023; **3**(3): 94-112.
17. Çalışkan E, Aksoy N. The Relationship Between Preoperative Anxiety Level and Postoperative Pain Outcomes in Total Hip and Knee Replacement Surgery: A Cross-sectional Study. *Journal of PeriAnesthesia Nursing* 2025; **40**(1): 76-82.

18. Fillingham YA, Hanson TM, Leinweber KA, Lucas AP, Jevsevar DS. Generalized Anxiety Disorder: A Modifiable Risk Factor for Pain Catastrophizing After Total Joint Arthroplasty. *The Journal of Arthroplasty* 2021; **36**(7): S179-S83.
19. Hernández C, Díaz-Heredia J, Berraquero ML, Crespo P, Loza E, Ruiz Ibán MÁ. Pre-operative Predictive Factors of Post-operative Pain in Patients With Hip or Knee Arthroplasty: A Systematic Review. *Reumatología Clínica (English Edition)* 2015; **11**(6): 361-80.
20. Rolfson O, Dahlberg LE, Nilsson Jå, Malchau H, Garellick G. Variables determining outcome in total hip replacement surgery. *The Journal of Bone and Joint Surgery British volume* 2009; **91-B**(2): 157-61.
21. Wood TJ, Gazendam AM, Kabali CB, Hamilton Arthroplasty G. Postoperative Outcomes Following Total Hip and Knee Arthroplasty in Patients with Pain Catastrophizing, Anxiety, or Depression. *The Journal of arthroplasty* 2021; **36**(6): 1908-14.
22. Granot M, Ferber SG. The Roles of Pain Catastrophizing and Anxiety in the Prediction of Postoperative Pain Intensity. *The Clinical Journal of Pain* 2005; **21**(5): 439-45.
23. Hassett AL, Marshall E, Bailey AM, et al. Changes in Anxiety and Depression Are Mediated by Changes in Pain Severity in Patients Undergoing Lower-Extremity Total Joint Arthroplasty. *Regional Anesthesia and Pain Medicine* 2018; **43**(1): 14-8.
24. Ho C-J, Chen Y-T, Wu H-L, Huang H-T, Lin S-Y. The Effects of a Patient-Specific Integrated Education Program on Pain, Perioperative Anxiety, and Functional Recovery following Total Knee Replacement. *Journal of personalized medicine* 2022; **12**(5).
25. Bryan JP. Concerns Related to Preoperative Anxiety in the Total Joint Arthroplasty Patient. *JBJS Journal of Orthopaedics for Physician Assistants* 2021; **9**(4).
26. King A, Bartley J, Johanson DL, Broadbent E. Components of preoperative anxiety: A qualitative study. *Journal of Health Psychology* 2019; **24**(13): 1897-908.
27. Kravitz RL. Patients' expectations for medical care: An expanded formulation based on review of the literature. *Medical Care Research and Review* 1996; **53**(1): 3-27.
28. Laferton JAC, Kube T, Salzmann S, Auer CJ, Shedden-Mora MC. Patients' expectations regarding medical treatment: a critical review of concepts and their assessment. *frontiersinorgJAC Laferton, T Kube, S Salzmann, CJ Auer, MC Shedden-MoraFrontiers in psychology, 2017·frontiersinorg* 2017; **8**(FEB).
29. Lee SH, Yoo HJ, Nam HS, Ho JP, Lee YS. The preoperative education on realistic expectations does not continually improve patients' satisfaction after total knee arthroplasty? A randomized controlled trial with serial assessment. *Knee Surgery, Sports Traumatology, Arthroscopy* 2025; **33**(11): 3853-65.
30. Nam HS, Yoo HJ, Ho JPY, Kim YB, Lee YS. Preoperative education on realistic expectations improves the satisfaction of patients with central sensitization after total knee arthroplasty: a randomized-controlled trial. *Knee Surg Sports Traumatol Arthrosc* 2023; **31**(11): 4705-15.
31. Tolk JJ, Janssen RPA, Haanstra TM, Van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: A randomized controlled trial: The EKSPECT study. *Bone and Joint Journal* 2021; **103-B**(4): 619-26.
32. Lawrence C, Zmistowski BM, Lazarus M, Abboud J, Williams G, Namdari S. Expectations of Shoulder Surgery Are Not Altered by Surgeon Counseling of the Patient. *Joints* 2017; **5**(3): 133-7.
33. Kästner A, Ng Kuet Leong VSC, Petzke F, Budde S, Przemeck M, Müller M, Erlenwein J. The virtue of optimistic realism - expectation fulfillment predicts patient-rated global effectiveness of total hip arthroplasty. *BMC Musculoskeletal Disorders* 2021; **22**(1).

34. Mancuso CA. Editorial Commentary: Assessing Outcomes in Terms of Fulfillment of Patient Expectations Is Complementary to Traditional Measures Including Satisfaction. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2022; **38**(6): 1876-8.
35. Mancuso CA, Duculan R, Cammisa FP, Sama AA, Hughes AP, Lebl DR, Girardi FP. Proportion of Expectations Fulfilled. *Spine* 2016; **41**(11): 963-70.
36. Deakin AH, Smith MA, Wallace DT, Smith EJ, Sarungi M. Fulfilment of preoperative expectations and postoperative patient satisfaction after total knee replacement. A prospective analysis of 200 patients. *Knee* 2019; **26**(6): 1403-12.
37. Henry LE, Aneizi A, Nadarajah V, et al. Preoperative expectations and early postoperative met expectations of extremity orthopaedic surgery. *Journal of Clinical Orthopaedics and Trauma* 2020; **11**: S829-S36.
38. Lim WZ, Noviani M, Yeo S-J, et al. Expectation fulfillment is associated with good outcomes and patient satisfaction after knee arthroplasty: a prospective study in a multi-ethnic Asian population. *Scientific Reports* 2025; **15**(1): 15133-.
39. Lützner C, Beyer F, David L, Lützner J. Fulfilment of patients' mandatory expectations are crucial for satisfaction: a study amongst 352 patients after total knee arthroplasty (TKA). *Knee Surgery, Sports Traumatology, Arthroscopy* 2023; **31**(9): 3755-64.
40. Green A, Neradilek MB, Thompson KM, Mayer A. Expectations affect outcome and satisfaction after anatomic total shoulder arthroplasty. Seminars in Arthroplasty: JSES; 2020: Elsevier; 2020. p. 297-307.
41. Constant C, Murley A. A clinical method of functional assessment of the shoulder. *Clinical orthopaedics and related research* 1987; (214): 160-4.
42. Constant CR, Gerber C, Emery RJ, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. *Journal of shoulder and elbow surgery* 2008; **17**(2): 355-61.
43. Cronin KJ, Magnuson JA, Murphy ML, Unger RZ, Jacobs CA, Blake MH. Responsiveness of patient-reported outcomes in shoulder arthroplasty: what are we actually measuring? *Journal of Shoulder and Elbow Surgery* 2021; **30**(5): 1174-80.
44. Hijji FY, Cheslik TG, Schneider AD, Schach BM, Venkatarayappa I. The popularity of outcome measures used in shoulder arthroplasty literature. *Shoulder & Elbow* 2021; **13**(3): 237-47.
45. Sims MT, Detweiler BN, Scott JT, Howard BM, Detten GR, Vassar M. Inconsistent selection of outcomes and measurement devices found in shoulder arthroplasty research: An analysis of studies on ClinicalTrials.gov. *PLOS ONE* 2017; **12**(11): e0187865-e.
46. Kohn D, Geyer M. The subjective shoulder rating system. *Arch Orthop Trauma Surg* 1997; **116**(6-7): 324-8.
47. van Maren K, Brown LE, Cremers T, Khatiri MZ, Ring D, Fatehi A. In orthopaedic speciality care, longer explanations are not more caring or more satisfying. *Clinical Orthopaedics and Related Research®* 2021; **479**(12): 2601-7.
48. Dekkers T, Melles M, Vehmeijer SBW, de Ridder H. Effects of Information Architecture on the Effectiveness and User Experience of Web-Based Patient Education in Middle-Aged and Older Adults: Online Randomized Experiment. *J Med Internet Res* 2021; **23**(3): e15846.
49. Groeneveld B, Melles M, Vehmeijer S, Mathijssen N, Dekkers T, Goossens R. Developing digital applications for tailored communication in orthopaedics using a Research through Design approach. *Digit Health* 2019; **5**: 2055207618824919.
50. Groeneveld BS, Dekkers T, Mathijssen NMC, Vehmeijer SBW, Melles M, Goossens RHM. Communication Preferences in Total Joint Arthroplasty: Exploring the Patient Experience Through Generative Research. *Orthop Nurs* 2020; **39**(5): 292-302.

A

Appendices

ENGLISH SUMMARY

A well-functioning shoulder joint is essential for daily activities such as self-care, household or work-related tasks, and recreational activities. When disorders cause pain and loss of function, these activities become more difficult or sometimes even impossible to perform. Common shoulder conditions include osteoarthrosis (OA) and cuff tear arthropathy (CTA), more often found in older adults. When non-surgical treatment of OA or CTA provides insufficient improvement, shoulder arthroplasty (SA) may provide a solution. Depending on the indication, a hemiarthroplasty, an anatomic total shoulder arthroplasty, or a reverse total shoulder arthroplasty is performed.

Although SA is generally successful in relieving pain and improving function, it is still a major surgical procedure. Patients experience uncertainty about the operation itself, the postoperative course, and the outcomes. The aim of this thesis was therefore to improve the patient experience during the preoperative and early postoperative phases of the SA journey.

Part I of the thesis focused on the content of preoperative education: what can we tell patients about their outcomes? **Part II** examined preoperative expectations of SA patients, with attention to two paradigms for measuring expectations. **Part III** evaluated whether a widely used instrument for assessing shoulder function could be replaced by a patient-reported version.

Part I – What can we tell patients about their outcomes after shoulder arthroplasty?

Chapter 2 presents a systematic review of patient-related prognostic factors for functional recovery and pain after SA. Across 33 longitudinal cohort studies (over 6,900 patients), 52 potential prognostic factors were identified. Meta-analyses demonstrated a clear predictive effect only for a history of previous shoulder surgery, which was associated with worse functional outcomes in the mid to long term. Other factors were reported in only a single study, or meta-analyses yielded results consistent with the possibility of no predictive effect. These findings support surgeons in tailoring preoperative counseling and expectation management to the individual patient.

Chapter 3 analyzed day-to-day pain trajectories during the first eight weeks after SA, a period often overlooked in previous research. Latent Growth Curve Modeling in 230 patients identified six distinct recovery trajectories, ranging from slower to faster recovery. The majority belonged to a "Faster" group, achieving low pain scores within two weeks after surgery. A smaller "Slower" group showed a more gradual decline in pain and more often had higher ASA scores, poorer preoperative function, and less participation in paid or unpaid work. Nevertheless, they reached similarly low pain levels by the eighth postoperative week. These findings enable evidence-based reassurance in the preoperative or early postoperative phase: approximately five out of six patients can expect minimal pain within two weeks after surgery.

Part II – Patients' expectations regarding shoulder arthroplasty

Using the same cohort as in **Chapter 3**, **Chapter 4** examined the relationship between preoperative expectations and postoperative satisfaction, distinguishing between probabilistic and value-based expectations. Approximately 80% of probabilistic expectations were completely fulfilled. For "very important" value-based domains, complete improvement was less frequent, but almost no patients reported no improvement at all. A higher degree of expectation fulfilment or improvement was associated with greater odds of being "very satisfied," even after adjusting for optimism. These findings align with research in knee arthroplasty, demonstrating that fulfillment of expectations is a key determinant of satisfaction. Value-based expectations provide orthopedic surgeons with starting points to discuss what is important to the patient, while probabilistic expectations help to identify and adjust unrealistic expectations.

Part III – Measuring change in shoulder function of shoulder arthroplasty patients

One of the most widely used instruments for assessing shoulder function is the Constant-Murley Score (CMS), although it is time-consuming. In **Chapter 5**, the clinician-based CMS and the Dutch translation of a patient-based version were compared in 139 SA patients. The limits of agreement between the two versions exceeded the minimal clinically important differences for SA patients, indicating that the patient-based CMS cannot

replace the clinician-based CMS. Moreover, the reproducibility of both versions was poor. Given the availability of more reproducible patient-reported outcome measures (PROMs), such as the Oxford Shoulder Score, we therefore discourage continued use of the CMS.

The findings of this thesis provide practical guidance for improving preoperative and early postoperative counseling. Future research may further individualize care, for example by validating prediction models and identifying subgroups with comparable clinical, psychological, and communication profiles. Impact studies are needed to evaluate the effects of such tailored educational programs on key outcomes.

Nederlandse samenvatting

Een goed functionerend schoudergewricht is essentieel voor dagelijkse activiteiten zoals zelfzorg, huishoudelijke of werkgerelateerde taken en recreatieve activiteiten. Bij aandoeningen die pijn en functieverlies veroorzaken, worden deze activiteiten moeilijker of soms zelfs onmogelijk om uit te voeren. Artrose (OA) en cuff tear arthropathie (CTA) zijn twee van zulke schouderaandoeningen die vaak voorkomen, voornamelijk bij oudere patiënten. Wanneer conservatieve behandeling van OA en CTA onvoldoende effect heeft, kan schouderarthroplastiek (SA) uitkomst bieden. Afhankelijk van de indicatie wordt gekozen voor een hemiprothese, een anatomische totale schouderprothese of een reverse totale schouderprothese.

Over het algemeen zorgt SA vaak voor minder pijn en een verbetering van de schouderfunctie. Het is echter een ingrijpende operatie. Patiënten ervaren onzekerheid over de operatie zelf, het postoperatieve beloop en de uitkomsten. Het doel van dit proefschrift was daarom om de patiëntervaring in de preoperatieve en vroege postoperatieve fase van het SA-traject te verbeteren.

Deel I van het proefschrift richtte zich op de inhoud van de preoperatieve voorlichting: wat kunnen we patiënten vertellen over hun uitkomsten? **Deel II** onderzocht preoperatieve verwachtingen van SA-patiënten, met aandacht voor twee paradigma's om verwachtingen te meten. **Deel III** evalueerde of een veelgebruikt meetinstrument voor schouderfunctie kon worden vervangen door een patiëntgerapporteerde variant.

Deel I – Wat kunnen we patiënten vertellen over hun uitkomsten na schouderarthroplastiek?

Hoofdstuk 2 beschrijft een systematische review naar patiëntgerelateerde prognostische factoren voor functioneel herstel en pijn na SA. In 33 longitudinale cohortstudies (ruim 6.900 patiënten) werden 52 potentiële prognostische factoren geïdentificeerd. Meta-analyses toonden alleen een duidelijk voorspellend effect van een eerdere schouderoperatie op slechtere functionele uitkomsten op middellange tot lange termijn. Andere

factoren werden elk slechts in één studie gerapporteerd, of meta-analyses toonden resultaten die ook de mogelijkheid van geen voorspellend effect op uitkomsten lieten zien. Deze bevindingen ondersteunen chirurgen bij het afstemmen van de voorlichting en het verwachtingsmanagement op hun individuele patiënt.

Hoofdstuk 3 analyseerde het dag-tot-dag pijnherstel in de eerste acht weken na SA, een periode die vaak over het hoofd is gezien in eerdere studies. Met Latent Growth Curve Modeling bij 230 patiënten werden zes hersteltrajecten onderscheiden, variërend van een langzamer tot een sneller herstel. De meerderheid behoorde tot een 'Snellere' groep die binnen twee weken lage pijnscores bereikte. Een kleinere 'Langzamere' groep liet een geleidelijker herstel zien en kenmerkte zich vaker door hogere ASA-scores, slechtere preoperatieve functie en minder vaak (on) betaald werk. Ook zij bereikten echter rond acht weken na de operatie vergelijkbaar lage pijnniveaus. Deze resultaten maken evidence-based geruststelling in de preoperatieve of vroege postoperatieve fase mogelijk: circa vijf op de zes patiënten ervaren binnen twee weken na de operatie minimale pijn.

Deel II – Verwachtingen van patiënten ten aanzien van schouderarthroplastiek

Met hetzelfde cohort als in **Hoofdstuk 3**, onderzochten we in **Hoofdstuk 4** de relatie tussen preoperatieve verwachtingen en postoperatieve tevredenheid, waarbij we onderscheid hebben gemaakt tussen probabilistische en waardegedreven verwachtingen. Ongeveer 80% van de probabilistische verwachtingen werd volledig vervuld. Voor 'zeer belangrijke' waardegedreven domeinen was volledige verbetering minder frequent, maar vrijwel niemand rapporteerde geen enkele verbetering. Een hogere mate van vervulling of verbetering van verwachtingen ging gepaard met een grotere kans op 'zeer tevreden' zijn, ook na correctie voor optimisme. Dit bevestigt bevindingen uit onderzoek naar knieprotheses: het vervullen van verwachtingen is een belangrijke determinant van tevredenheid. Waardegedreven verwachtingen bieden de orthopedisch chirurg aanknopingspunten om het gesprek aan te gaan over wat de patiënt belangrijk vindt, terwijl probabilistische verwachtingen helpen bij het bijstellen van onrealistische verwachtingen.

Deel III – Meten van verandering in schouderfunctie bij patiënten met een schouderprothese

Een van de meest gebruikte instrumenten om schouderfunctie te evalueren is de Constant-Murley Score (CMS), maar dit is een tijdrovend instrument. In Hoofdstuk 5 werden de artsversie van de CMS en de Nederlandse vertaling van een patiëntversie vergeleken bij 139 SA-patiënten. De grenzen van overeenstemming overschreden de minimaal klinisch relevante verschillen voor SA-patiënten, wat betekent dat de patiënt versie van de CMS de artsversie niet kan vervangen. Bovendien bleek de reproduceerbaarheid van beide versies beperkt. Gezien de beschikbaarheid van beter reproduceerbare PROMs, zoals de Oxford Shoulder Score, raden we daarom het gebruik van de CMS als meetinstrument af.

De bevindingen uit dit proefschrift bieden handvatten voor betere preoperatieve en vroege postoperatieve voorlichting. Toekomstig onderzoek kan bijdragen aan verdere individualisering, bijvoorbeeld door predictiemodellen te valideren en subgroepen met vergelijkbare klinische, psychologische en communicatieve profielen te identificeren. Impactstudies zijn nodig om het effect van dergelijke op maat gemaakte educatieprogramma's op relevante uitkomsten vast te stellen.

List of publications

In this thesis

Patient-related prognostic factors for function and pain after shoulder arthroplasty: a systematic review.
Hesseling B, Prinsze N, Jamaludin F, Perry SI, Eygendaal D, Mathijssen NM, Snoeker BA.
Systematic reviews. 2024 Nov 22;13(1):286.

Trajectories of pain recovery during the first 8 weeks after shoulder arthroplasty: results from the shoulder diary study using latent growth curve modeling.
Hesseling B, Snoeker BA, Ten Brinke B, Slot K, Zwitser EW, van Rosmalen J, Eygendaal D, Mathijssen NM.
Journal of Shoulder and Elbow Surgery. 2026 Feb 1;35(2):438-53.

Complete fulfillment of expectations is associated with greater satisfaction after shoulder arthroplasty: results from a prospective multicenter cohort study.
Hesseling B, Berends Y, van der Linden S, Tolk JJ, van der Linde JA, Visser CP, Slot K, Eygendaal D, Snoeker BA, Mathijssen NM.
JSES International. 2026 May 1;10(3).

The Dutch patient-based Constant-Murley score cannot be used interchangeably with the clinician-based Constant-Murley score in shoulder arthroplasty patients.
Hesseling B, Perry SI, Hoelen MA, van Deurzen DF, Eygendaal D, Mathijssen NM, Snoeker BA.
Journal of Shoulder and Elbow Arthroplasty. 2026 Sep 1;10(3).

A

Not in this thesis

Optimizing the dose of local infiltration analgesia and gabapentin for total knee arthroplasty, a randomized single blind trial in 128 patients.
van Haagen MH, Verburg H, **Hesseling B**, Coors L, van Dasselaar NT, Langendijk PN, Mathijssen NM.
The Knee. 2018 Jan 1;25(1):153.

Reply to Kendall et al.
van Haagen MH, Verburg H, **Hesseling B**, Coors L, van Dasselaar NT, Langendijk PN, Mathijssen NM.
The Knee. 2018 Mar 1;25(2):348-9.

Fast starters, slow starters, and late dippers: trajectories of patient-reported outcomes after total hip arthroplasty: results from a Dutch nationwide database.
Hesseling B, Mathijssen NM, Van Steenbergen LN, Melles M, Vehmeijer SB, Porsius JT.
JBJS. 2019 Dec 18;101(24):2175-86.

The correlation of shoe size and component size of primary total knee arthroplasty.
van Egmond JC, Verburg H, **Hesseling B**, Mathijssen NM.
The Journal of Knee Surgery. 2020 Mar;33(03):260-4.

Three distinct recovery patterns following primary total knee arthroplasty: Dutch arthroplasty register study of 809 patients.
Van Egmond JC, **Hesseling B**, Melles M, Vehmeijer SB, van Steenbergen LN, Mathijssen NM, Porsius JT.
Knee Surgery, Sports Traumatology, Arthroscopy. 2021 Feb;29(2):529-39.

Safe upper limits of serum cobalt and chromium levels for a metal-on-metal total hip bearing: a 10-year follow-up study.
Koper MC, **Hesseling B**, Tuinebreijer WE, van der Linden H, Mathijssen NM.
The Journal of Arthroplasty. 2021 Jun 1;36(6):2080-6.

Short-term functional outcome after fast-track primary total knee arthroplasty: analysis of 623 patients.
Van Egmond JC, **Hesseling B**, Verburg H, Mathijssen NM.
Acta Orthopaedica. 2021 Oct 8;92(5):602-7.

Hip arthroscopy for femoroacetabular impingement syndrome results in 2 recovery patterns based on preoperative pain and on arthritis: improvers and non-improvers.
Röling MA, **Hesseling B**, Mathijssen NM, Bloem RM.
Arthroscopy, Sports Medicine, and Rehabilitation. 2021 Oct;3(5):e1481-90.

Validation of the Dutch version of the Hip Outcome Score; validity, reliability, and responsiveness in patients with femoroacetabular impingement syndrome.
Röling MA, **Hesseling B**, Jansen SP, Bloem RM, Mathijssen NM.
Journal of Hip Preservation Surgery. 2021 Aug 1;8(3):298-304.

Early fixation of the humeral component in stemless total shoulder arthroplasty: a radiostereometric and clinical study with 24-month follow-up.
Ten Brinke B, **Hesseling B**, Eygendaal D, Hoelen MA, Mathijssen NM.
The bone & joint journal. 2022 Jan 1;104(1):76-82.

Early migration in unicompartmental knee arthroplasty: a radiostereometric study of 26 patients with 24 months of follow-up.
Pasma JH, **Hesseling B**, de Esch N, Verburg H, D NIESTEN D, Mathijssen NM.
Acta Orthopaedica. 2022 Dec 19;93:914-21.

Revision rates and patient-reported outcomes after shoulder hemiarthroplasty vs. anatomic total shoulder arthroplasty in patients younger than 60 years: a cohort study with data from the Dutch Arthroplasty Register (LROI)
Steenbeek ED, **Hesseling B**, Spekenbrink-Spooren A, Van der Linde JA. 23.
Journal of Shoulder and Elbow Surgery. 2025 Feb 10;34(9):2114-21.

Migration in unicompartmental knee arthroplasty with the Persona Partial Knee: a cohort study of 26 patients using radiostereometry with 60 months of follow-up.
Pasma JH, **Hesseling B**, de Esch N, Verburg H, Niesten DD, Mathijssen NM.
Acta Orthopaedica. 2026 Jan 3;97:1-8.

The effect of suturing techniques on aberrant wound recovery after total hip arthroplasty: study protocol for a prospective, randomized trial (HIP-STITCH).
Miedema BW, Vehmeijer SB, van der Linden H, **Hesseling B**.
Bone & Joint Open. 2026 May 11;7(5):619-26.

Migration in large femoral head ceramic-on-ceramic monoblock acetabular component compared with modular ceramic-on-polyethylene acetabular component in total hip arthroplasty using radiostereometric analysis with 5 years' follow-up: a randomized controlled trial.
Caner IÇ, **Hesseling B**, Pasma JH, Ian BL, Bloem RM, Mathijssen NM.
Acta Orthopaedica. 2026 Jul 1;97:448-55.

Cement compression device for cemented total knee arthroplasty: a randomized controlled trial.
Brouwer AJ, **Hesseling B**, Verburg H, Goossens RH, Mathijssen NM.
Bone & Joint Open. 2026 Jul 16;7(7):944-51.

Submitted/under review

Expectation fulfilment in rotator cuff repair.
van Moorsel L, **Hesseling B**, van der Linde JA.
Under review at Shoulder and Elbow.

Migration patterns of the humeral component of the stemless Simpliciti Shoulder System in stemless anatomic total shoulder arthroplasty A radiostereometric and clinical study with 60-month follow-up.
De Bruin LJ, ten Brinke B, Oomen R, Hoelen M, **Hesseling B**, Mathijssen NM.
Accepted at Journal of Shoulder and Elbow Surgery.

Portfolio

Personal details	
Name PhD student	Drs. Brechtje Hesseling
PhD period	2015 - 2026
Department	Orthopedics and Sports Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
Promotor	Prof. dr. D. Eygendaal
Co-promotores	Dr. N.M.C. Mathijssen, Dr. B.A.M. Snoeker

General courses	Year	ECTS
GCP-WMO registration	2016	0.1
Basic concepts of qualitative research in health care	2017	5
Masterclass 'Proefpersoneninformatie: tips & tricks rondom informeren'	2017	0.25
RSA course	2018	0.25
Expertdag 'Ethische dilemma's en privacy in wetenschappelijk onderzoek'	2018	0.2
Data Science: Statistical Programming with R	2019	1.5
GCP-WMO re-registration	2019	0.1
Data Science: Multiple Imputation in Practice	2021	1.5
Systematic Review and Meta-Analysis of Prognosis Studies	2022	0.8
GCP-WMO re-registration	2022	0.1
PhD Introduction session	2023	0.2
Scientific Integrity	2023	0.3
Structural Equation Modeling in R using lavaan	2024	1.5
Causal Mediation Analysis	2024	1.4
Survival Analysis	2024	1.5
GCP-WMO re-registration	2025	0.1
Repeated Measurements	2025	1.5

Personal development	Year	ECTS
De Baak – Ik en de Anderen	2022	2.6
Personal Leadership & Communication	2023	1
De Baak – De Informele Leider	2025	1.9
Podium presentations	**Year**	**ECTS**
Pain Science in Motion IV congress, Las Vegas, United States *"Detailed pain recovery trajectories during the first eight weeks after shoulder arthroplasty"*	2024	0.8
Combined NOV/NOF congress, Rotterdam *"Detailed pain recovery trajectories during the first eight weeks after shoulder arthroplasty"*	2024	1.1
Finale wetenschapsprijs Werkgroep Schouder & Elleboog *"Vroege pijntrajecten na een schouderprothese"*	2025	0.5
Wetenschapsavond Reinier de Graaf ziekenhuis, Delft *"Volledig vervulde verwachtingen zijn gerelateerd aan hogere tevredenheid na een schouderprothese"*	2025	0.5
Wetenschapsmiddag HagaZiekenhuis, Den Haag *"Volledig vervulde verwachtingen zijn gerelateerd aan hogere tevredenheid na een schouderprothese"*	2025	0.5
Finale wetenschapsprijs Werkgroep Schouder & Elleboog *"Is een patiëntversie inwisselbaar met de artsversie van de Constant Murley Score bij schouderprothesiologie?"*	2026	0.5
Wetenschapsavond Reinier de Graaf ziekenhuis, Delft *"Is een patiëntversie inwisselbaar met de artsversie van de Constant Murley Score bij schouderprothesiologie?"*	2026	0.5

Poster presentations	Year	ECTS
Pain Science in Motion III congress, Savona, Italy *"The Shoulder Diary - The role of expectations in patient reported outcomes following shoulder arthroplasty"*	2019	0.8
Wetenschapsavond Reinier de Graaf ziekenhuis, Delft *"Detailed pain recovery trajectories during the first eight weeks after shoulder arthroplasty"*	2024	0.3
SECEC congress 2025, Rotterdam *"Complete fulfillment of expectations is associated with greater satisfaction after shoulder arthroplasty"*	2025	1.1

Awards	Year
2nd Runner up 'Audience award for best oral presentation', Pain Science in Motion IV congress	2024
Nominee 2025 Science award Werkgroep Schouder & Elleboog	2025
Nominee 2026 Science award Werkgroep Schouder & Elleboog	2026
Winner 'Jury award' Wetenschapsavond Reinier de Graaf ziekenhuis	2026
Winner 'Audience award' Wetenschapsavond Reinier de Graaf ziekenhuis	2026

Teaching activities	Year	ECTS
Wake Up Science, Reinier Haga Orthopedic Center	2020 - now	2
Supervision Masterstudent Evidence Based Practice in Health Care – Nisa Prinsze	2021 – 2023	2
Supervising students during internships (bachelor and master, university and university of applied sciences)	2022 – now	1
Lessons 'Clinical Epidemiology' and 'Statistics' at Breederode Hogeschool, Opleiding tot research verpleegkundige / clinical research coordinator	2022-2023	1.5

Curriculum Vitae



Brechtje Hesseling werd op 9 maart 1984 geboren in Groningen en op haar vierde verhuisde het gezin naar Delft. Na het behalen van haar VWO-diploma ging ze Fysiotherapie studeren aan de Hogeschool Leiden. Tijdens deze studie ontdekte ze al de aantrekkingskracht van wetenschappelijk onderzoek, maar na haar afstuderen in 2006 wilde ze eerst werkervaring opdoen in de praktijk. Als fysiotherapeut in de eerste lijn ontwikkelde ze steeds meer interesse in het schoudergewricht en begeleidde ze regelmatig patiënten na het krijgen van een schouderprothese. In 2010 trouwde ze met haar man Friso.

De affiniteit met wetenschappelijk onderzoek bleef al die tijd aanwezig en in 2011 begon Brechtje aan haar opleiding tot klinisch epidemioloog bij de master 'Evidence Based Practice in Health Care' (EBPiHC) in het AMC te Amsterdam. Voor haar afstuderen schreef ze een systematic review die onderdeel was van het promotietraject van haar begeleider, Barbara Snoeker. Hiermee studeerde ze in 2013 cum laude af.

Via een studiegenoot kwam ze in 2014 in contact met Nina Mathijssen, onderzoeker bij de afdeling Orthopedie van het Reinier de Graaf ziekenhuis in Delft, om de mogelijkheden voor een eigen promotietraject te verkennen. Het werd een buitenpromotie, met veel mogelijkheden om zelf de insteek te bepalen.

In 2015, na de geboorte van haar dochter Janthe, startte Brechtje definitief met haar promotietraject met begeleiding vanuit de opleiding EBPiHC door promotoren prof.dr. Cees Lucas en prof.dr. Mirjam Sprangers. Twee jaar later stopte ze met haar werk als fysiotherapeut en stapte ze volledig over naar haar rol als onderzoeker bij de afdeling Orthopedie in het Reinier de Graaf Ziekenhuis en later in het Reinier Haga Orthopedisch Centrum (RHOC). In 2023 werd de rol van promotor overgenomen door prof.dr. Denise Eygendaal en zette Brechtje haar promotie voort bij de afdeling Orthopedie en Sportgeneeskunde in het Erasmus MC, Rotterdam.

Onder begeleiding van haar promotor en co-promotoren dr. Barbara Snoeker en dr. Nina Mathijssen deed Brechtje onderzoek naar welke factoren het herstel en de tevredenheid na schouderarthroplastiek beïnvloeden en naar hoe patiënten het vroege herstel na de operatie ervaren. Ze is en blijft in het RHOC betrokken bij veel verschillende wetenschappelijke onderzoeken en begeleidt daar ook studenten en externe auteurs bij het opzetten van onderzoek, doen van data-analyses en het schrijven van wetenschappelijke artikelen.

Dankwoord

Het schrijven van dit proefschrift is een hele reis geweest, met ups en downs. Die reis zou niet mogelijk zijn geweest (of op zijn minst een heel stuk moeilijker!) zonder de bijdragen, hulp en steun van velen. In dit dankwoord doe ik mijn best om recht te doen aan ieders bijdrage. En omdat ik nu eenmaal niet kort en bondig ben, neem ik er de ruimte voor.

Mijn co-promotoren: dr. Nina Mathijssen en dr. Barbara Snoeker

Beste **Nina**, via een studiegenoot kwam ik in 2014 bij jou terecht met de vraag of jullie misschien nog een onderzoek hadden zonder onderzoeker, omdat ik een onderzoeker zonder onderzoek was. Je gaf me veel vrijheid om een eigen richting te zoeken, waaruit uiteindelijk dit mooie proefschrift is ontstaan. Je toonde veel vertrouwen in me, mede door me vanaf 2016 een plek als onderzoeker in het Reinier de Graaf ziekenhuis te bieden die is uitgegroeid tot mijn huidige rol in het RHOC. Dankjewel!

Beste **Barbara**, vanaf het begin van mijn master Evidence Based Practice in Health Care in 2011 was je mijn werkgroepdocent en heb ik heel veel van je geleerd. Nadat ik met mijn afstudeerproject bij mocht dragen aan jouw eigen promotieonderzoek, werd je samen met Nina mijn co-promotor. Je bleef me uitdagen om kritisch na te denken en focus te houden in mijn onderzoek. Ook aan jou heel veel dank!

Zonder jullie aanhoudende steun, aanmoediging en begeleiding was dit proefschrift er niet gekomen. Jullie hebben me geholpen om enorm te groeien. Niet alleen als onderzoeker, maar ook als persoon.

Mijn promotor: professor dr. Eygendaal

Beste **Denise**, halverwege 2023 werd je mijn promotor en wat was het een zachte landing. Je positieve insteek en enthousiasme hebben me goed gedaan. Met hernieuwde energie ben ik daarna de (lange) eindsprint in gegaan. Bedankt voor je vertrouwen in mij; het heeft tot een mooi resultaat geleid.

Geachte leden van de beoordelingscommissie, **Prof.dr. Bierma – Zeinstra**, **Prof.dr. Vliet Vlieland** en **dr. A. van Noort**. Hartelijk dank voor het beoordelen van dit proefschrift en uw aanwezigheid tijdens de verdediging. Ook dank aan **de opponenten** tijdens de verdediging van dit proefschrift voor hun deelname en tijd.

Nisa Prinsze, de systematic review in dit boekje was jouw masterthesis. Bedankt voor je harde werk; het was een flinke puzzel om alle prognostische factoren logisch te ordenen! **Sander Perry**, jouw begeleiding bij de masterthesis en je klinimetrische kennis bij de Constant studie waren onmisbaar.

Ook veel dank aan mijn andere co-auteurs: **Youri Berends**, **Bart ten Brinke**, **Derek van Deurzen, Max Hoelen, Faridi Jamaludin**, **Just v.d. Linde**, **Sophie v.d. Linden**, **Karin Slot**, **Joost van Rosmalen**, **Jaap Tolk**, **Cornelis Visser** en **Eline Zwitser**. Jullie inhoudelijke kennis en feedback hebben de manuscripten naar een hoog niveau getild.

Geen co-auteurs maar wel onmisbaar, want zonder jullie hulp zouden er geen patiënten zijn geïncludeerd: **Roos Bazuin**, **Michel v.d. Bekerom**, **Joyce Benner**, **Ilse Bouman**, **Roel Janssens, Paulien van Kampen**, **Elga Meurs** en **Ysbrand Ram**.

Alle **patiënten** die deel hebben genomen aan mijn onderzoeken: ontzettend bedankt voor jullie tijd en geduld om al mijn vragenlijsten in te vullen, en om soms extra terug te komen naar het ziekenhuis voor metingen.

Lieve collega's van het KWIK: **Arlisia**, **Beau, Carin**, **Ian**, **Ilse**, **Jantsje**, **Marjolein**, **Nicole**, **Petra**, **Rianne** en **Roos**, en oud-collega's **Maryse** en **Petra**. Bedankt voor de gezellige momenten, goede gesprekken, teamuitjes, aanmoediging en natuurlijk praktische hulp bij de uitvoering van mijn onderzoeken. Ik ben blij dat ik het afronden van mijn promotietraject met jullie kan vieren.

Gerald, ook jij bedankt voor de kans die ik heb gekregen bij dat eerste gesprek in 2014!

Ook dank aan **alle RHOC (oud)collega's**, zonder jullie zou het RHOC niet zo'n fijne plek zijn om te werken en zou er ook geen onderzoek mogelijk zijn. Speciale shout-out naar (**oud)collega's** van het RdGG waar het ooit begon: **Janet**, **Petra, Suzanne** en **Wendy**!

Aan alle mede-onderzoekers in kamer NC-424 in het Erasmus MC. Ik heb het al heel vaak gezegd, maar ik zeg het gewoon nog een keer: heel erg bedankt voor de warme ontvangst in jullie groep, ik heb me vanaf het begin af aan welkom gevoeld. Af en toe sparren over methodologie en statistiek, praktische problemen oplossen en vooral ook veel gezelligheid. Ook jullie feedback op presentaties heb ik zeer op prijs gesteld. Bedankt **Britt**, **Camille, Delong**, **Eline**, **Floor**, **Floris**, **Irene**, **Jinchi**, **Joshua**, **Kristel**, **Laura**, **Lea**, **Leontien**, **Niels**, **Merel**, **Mark**, **Maud**, **Michiel**, **Noortje**, **Rebecca, Roos**, **Sakis**, **Shreemandjari**, **Veerle** en **Zhongcan**. Een speciaal bedankje aan **Fleur** en **Myrthe**, onder jullie vleugels was ik razendsnel wegwijs in alle afspraken, systemen en processen. **Derek**, bedankt voor alle 3D printjes en ik hoop een keer bij je in de leer te mogen komen.

Max en **Erin**, bedankt voor jullie begeleiding en feedback tijdens de research department meetings.

En natuurlijk veel dank aan **Simone**, je was mijn houvast in alles omtrent de procedures voor het afronden van mijn promotie!

Deze reis was niet begonnen zonder **prof.dr. Cees Lucas** en **prof.dr. Mirjam Sprangers**. Bedankt voor de begeleiding in de eerste jaren van mijn promotietraject.

Niet alleen collega's, maar ook **vrienden** en **familie** hebben eraan bijgedragen dat dit proefschrift er is gekomen.

Alle **buren** van de GKS en VK, en in het bijzonder **Marloes**, **Dominik**, **Carolien**, **Hester** en **Femke**, bedankt voor jullie interesse in mijn promotietraject, jullie luisterend oor en de gezelligheid op de straatfeesten.

Lieve **Pim** en **Camilla**, jullie vriendschap is me heel veel waard. Dankjewel dat jullie er altijd voor ons zijn, dat jullie me opbeuren wanneer ik dat

goed kan gebruiken maar me ook de ruimte geven om me terug te trekken wanneer ik dát juist even nodig heb. En wat fijn dat Thomas, Anouk en Janthe zo goed met elkaar kunnen opschieten. Waar een goed georganiseerde blind date al niet toe kan leiden!

Ook mijn schoonfamilie bedank ik graag. **Walter** en **Joke**, **Olaf**, **Herma**, **Sophie**, **Iris** en **Niels**, en **Marc** en **Susanne**: jullie bleven altijd geïnteresseerd in hoe het ging en hoe ver ik was met alle onderzoeken. En in de tussentijd is Susanne ook nog gepromoveerd, zo telt de familie steeds meer doctoren.

Maartje, lieve grote zus. Wat ben je in de afgelopen jaren toch een hoop avonturen aan gegaan. Mooie reizen gemaakt, ashrams bezocht, nieuwe opleidingen en trainingen gedaan, naar Zwitserland verhuisd en een kampeerbusje gekocht. Het is altijd gezellig als jij en **Jan** op de camping in Duitsland of Oostenrijk aansluiten met de bus. Ik geniet ervan om mooie vondsten met je te delen, of het nou mooie pennen, lekkere essentiële oliën of leuke quotes zijn. Bedankt dat je altijd tijd voor me maakt, hoe druk je het ook hebt. Ik ben blij dat je mijn paranimf wil zijn.

Pap en **mam**, waar zou ik zijn zonder jullie. Jullie hebben me geleerd om altijd mijn best te doen, maar als het plan verandert is het ook oké. Toen ik na mijn burn-out bedacht dat ik misschien wel onderzoeker in plaats van fysiotherapeut wilde worden, maakten jullie het mogelijk om de master te volgen. Ook daarna zijn jullie me blijven steunen. Met aanmoediging, soms met enige zorgen of ik niet teveel hooi op mijn vork nam, met extra oppassen op Janthe en het uitlenen van jullie huis voor schrijfweekenden. Ontzettend bedankt voor alles wat jullie hebben gedaan en nog steeds doen.

Mijn allerliefste **Janthe**, wat ben ik blij dat jij mijn dochter bent! Zo heerlijk creatief, nieuwsgierig en oprecht geïnteresseerd in mensen *('Mama, als je onderzoeker bent, betekent dat dan dat je ook goed onder tafels en stoelen zoekt enzo?'*, toen je nog veel kleiner was). Dankzij jou ben ik gaan dansen, ik vind het heerlijk om dat met jou te delen. Bedankt voor al je geduld wanneer ik weer eens een weekend of week wegging om aan mijn 'extra diploma' te werken. Ik hou ontzettend veel van jou en kijk ernaar uit om nog heel veel samen te beleven!

Tot slot, lieve **Friso**. Toen we elkaar in 2005 ontmoetten was jij zelf aan het promoveren, al zag ik toen nog niet goed wat dat precies inhield. Wat heb je mij in alle jaren ontzettend gesteund. Niet alleen in mijn promotietraject, maar ook in mijn persoonlijke groei; die twee gingen hand in hand. Als ik door de bomen het bos niet meer zag, hielp jij me uitzoomen. Hielp je feedback te vertalen zodat ik de kern daarvan zag in plaats van in paniek te raken. En hielp je me om argumenten steviger te maken door ze juist op structuur te ontleden, en niet op de inhoud. Je gaf me de ruimte om (aan het eind regelmatig) een weekend of week weg te gaan, waarin ik kon knallen met analyseren of schrijven. Het is niet meer dan logisch dat je tijdens mijn verdediging als paranimf achter me staat en wat vind ik het fijn om met jou op die manier deze reis af te sluiten. Ontzettend bedankt voor al het geduld dat je hebt gehad, alle fijne gesprekken die we kunnen hebben en dat ik er bij jou gewoon mag zijn. Ik hou van je.

A