Combining clinical information and Patient Reported Outcome Measures

in Orthopaedic Surgery and Sports Medicine

Derk van Kampen

Combining clinical information and Patient Reported Outcome Measures in Orthopaedic Surgery and Sports Medicine

Derk Anton van Kampen

PhD thesis, Utrecht University, the Netherlands

The studies described in this thesis were carried out at the department of Orthopaedic Surgery and Traumatology of the Onze Lieve Vrouwe Gasthuis (OLVG), Amsterdam, the Netherlands and the department of Orthopaedic Surgery and Sports Medicine, University of Pittsburgh Medical Center (UPMC), Pittsburgh, USA.

The research carried out in this thesis was supported by European Society for Surgery of the Shoulder and the Elbow (2009 SECEC/ ESSSE Research Grant), Anna Foundation (2009 Research Grant), and the Marti-Keuning Eckhardt Foundation (2009 Research Grant).

The publication of this thesis was supported by the Anna Foundation, Bauerfeind Benelux, BISLIFE, Boehringer Ingelheim, Link-Lima, Marti-Keuning Eckhardt Foundation, Nederlandse Orthopaedische Vereniging, Pro-Motion Medical, Reumafonds, Teaching Hospital OLVG, Tornier and Wellspect Healthcare.

Cover design: Elske Sieswerda & Optima

ISBN: 978-94-6169-391-4

Combining clinical information and Patient Reported Outcome Measures in Orthopaedic Surgery and Sports Medicine

Het combineren van patiëntinformatie en patiënt-gerapporteerde uitkomstmaten in de orthopedische chirurgie en sportgeneeskunde

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht,
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
donderdag 6 juni 2013 des middags te 2.30 uur door

Derk Anton van Kampen

geboren op 15 maart 1981
te Enschede

Promotor:	Prof. dr. R.M. Castelein
Co-promotoren:	Dr. W.J. Willems
	Dr. C.B. Terwee

TABLE OF CONTENTS

Chapter 1

Introduction and aims

September 23, 1999

"Contact with the Mars Climate Orbiter, a NASA robotic space probe intended to study Martian climate and atmosphere, was lost as it entered its orbit around the red planet. The mission had turned into an unprecedented fiasco as the module disintegrated and crashed on the surface of the planet. The cause: miscommunication about the measurement standard used: Imperial units (pounds-second, lbf x s) instead of Metric units (newton-seconds, N x s)."

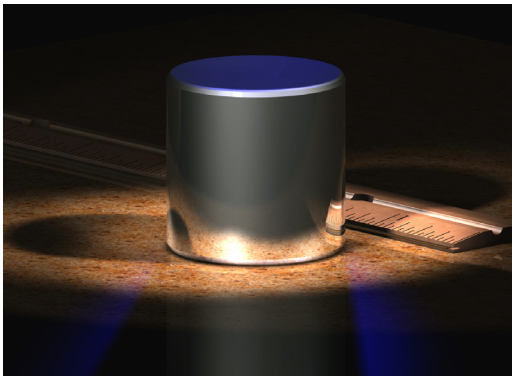
Measurements are the foundation of science. The earliest measurement instruments were practical, namely one's own foot and thumb. Both were used to measure distances, but variations in size led to differences in measurements, making the measurements unreliable.

In 1889, during the international Conférence Générale des Poids et Mesures at the Bureau International des Poids et Mesures, the length of the meter was officially and internationally established, and an international prototype of the meter was created for use as a gold standard (Figure I).[21,44] In the same year, the Bureau International des Poids et Mesures also declared that the prototype kilo, developed in 1799 of 90% platinum and 10% iridium, was to be the international standard for the kilogram (Figure II).[26,27,41] The international consensus to utilize established standards has been very important for the implementation of measurement instruments, as their limited variations makes measurements throughout the world more uniform and reliable.

Figure I.

Proper measurement instruments continue to be developed for use in medical science and clinical practice. As in other areas of science, reliable and valid measurements are important for reporting and comparing outcomes in the field of medicine. Reliable and valid measurements are the basic tools needed for establishing a diagnosis, for estimating a prognosis and for evaluating treatment efficacy. In this context, a measurement instrument can be a single

Figure II.



question, a questionnaire, a score obtained via physical examination, a laboratory measurement, a score obtained through observation of an image etcetera.

Nowadays, there is a strong increase in the use of highly sophisticated radiological measurement instruments like: MRI-, PET- and CT-scanners to help determine diagnoses. However, even in this era of increasing healthcare technology, more traditional measurement tools such as patient characteristics, history/anamnesis, physical examination and, if necessary, additional diagnostic tests remain important for making the right diagnosis.

Measurement instruments are not just useful for establishing a diagnosis; they can also be used to determine outcome measures for evaluating treatment effects, such as the infection rate or the revision rate after joint arthroplasty. Historically, emphasis has been placed on the judgement of the clinician in determining the results of treatment; fortunately, in more recent years, the patient's perspective of the outcome has received increased attention. Patient Reported Outcomes Measures (termed 'PROMs' and also known as 'PROs') specifically reflect the patient's perspective. As such, PROMs have gained widespread recognition for evaluating treatment in both clinical practice and research.[19,30,37] PROMs are measurement instruments, often questionnaires, that evaluate some aspect of a patient's health status exclusively from the patient's perspective without any interpretation of the response by anyone else.[28] Not only do PROMs provide unfiltered data that represent the patient's point of view, they also facilitate shared decision-making and, when the PROMs are standardized, they can potentially be used as benchmarks for comparing the results of treatment between centres.[33] Thus, PROMs are increasingly being used as primary outcomes in clinical studies.

Because of a worldwide increase in cost of healthcare in an arena with limited resources, there is an increasing need to demonstrate that the highest possible quality of care is delivered at the lowest cost. PROMs are marked by the US Food and Drug Administration a very useful measurement tools for this purpose.[28]

One of the pioneers of systematic treatment evaluation was the famous surgeon Dr. E. A. Codman. At the beginning of the 20th century, Codman proposed the 'end result idea' which involved a systematic analysis of outcomes of treatment by long-term follow-up with the aim of improving clinical practice.[7,8,13] In Codman's own words, the 'end result idea' is "merely the common-sense notion that every hospital should follow every patient it treats, long enough to determine whether or not the treatment has been successful, and then to inquire 'if not, why not?' with a view to preventing a similar failure in the future."[13]

Many of the PROMs that are currently used for evaluating the shoulder were developed in the 1990s. For example, the Oxford Shoulder Score and Oxford Shoulder Instability Score was developed in Oxford, England under the guidance of A. Carr, and the Simple Shoulder Test was developed in Seattle, Washington (USA) by Matsen et al.[10,11,23] The number of PROMs used for shoulder complaints has increased substantially since then, with *The AO Handbook of Musculoskeletal Outcome Measures and Instruments* (2009) reporting 28 PROMs for the shoulder.[36] Rigorous evaluation revealed large variations in the quality of these instruments.[36]

At the international level, the Constant-Murley score (CM) for the shoulder is the most frequently used clinician-based outcome score and is recommended by the European Society of Shoulder & Elbow Surgery (SECEC/ESSE) for treatment evaluation.[31] The CM includes a pain score, a functional assessment, a range-of-motion measure and a strength measure.[9,39,43] Notably, clinician-based outcomes such as the CM have some important limitations. First, CM carries the risk of low reproducibility, because of a lack of standardisation for performance tests like the muscle strength test.[1,6,31,38] Second, and even more important, clinician-based outcome scores do not always reflect the patient's perspective regarding important complaints. Third, clinician-based outcomes such as the CM can only be used in a clinical setting. Clinical follow-up in longitudinal studies is expensive and represents a greater burden for the patients.

An advantage to using PROMs instead of clinician-based outcomes is that PROMs can be completed at the patient's home or, for example, in the waiting room of an outpatient clinic. This facilitates the use of PROMs and is less time consuming for clinicians. According to recent reviews,[2,12] the most commonly used PROMs for general shoulder complaints include the Disability of the Arm, Shoulder and Hand (DASH),[18] the Simple Shoulder Test (SST),[23,40] the Oxford Shoulder Score (OSS)[3,10] and the Shoulder Pain and Disability Index (SPADI).[32] For traumatic anterior shoulder instability, the Western Ontario Shoulder Instability index (WOSI) is the most studied and hence the most widely accepted PROM.[2,5,22,34] However, before these PROMs can be used for research and patient evaluation in the Netherlands, the Dutch language versions must first be validated.

OBJECTIVE

The main objective of this thesis was to study the added value of patient information and PROMs for patient evaluation in orthopaedic surgery and sports medicine. Towards this end, the thesis is divided into three parts.

Part I: Patient Reported Outcome Measures (PROMs). In the first part of the thesis, the measurement properties of a commonly-used shoulder-specific PROM (the Simple Shoulder Test) were evaluated in a Dutch population of patients with shoulder problems. In addition, we determined the score improvements, in points, that should be considered Minimal Clinically Important Change (MIC or MCIC) for four commonly used shoulder-specific PROMs.

Part II: The diagnostic value of combining patient information and clinical tests. The second part of the thesis evaluated the diagnostic value of many individual clinical tests for shoulder complaints. By combining information obtained from patients with information from multiple clinical tests in a prediction model, we aimed to increase the tests' diagnostic value for diagnosing rotator cuff tears and traumatic anterior shoulder instability.

Part III: Application of measurement instruments in clinical practice. The third part of the thesis first reports long-term outcome after arthroscopic shoulder stabilization based on failure rate (re-dislocation) and PROMs. Secondly, this part of the thesis evaluated the diagnostic value of patient self-reported symptoms used in conjunction with neurocognitive test results to identify and evaluate brain concussion.

THESIS OVERVIEW AND RESEARCH QUESTIONS TO BE ADDRESSED

Chapter 1: General Introduction

PART I: PATIENT REPORTED OUTCOME MEASURES (PROMS)

Chapter 2: Validation of the Dutch Version of the Simple Shoulder Test

The Simple Shoulder Test (SST) is a simple and short PROM that evaluates patients with shoulder complaints.[23] It is commonly used in clinical practice internationally as well as in clinical research. Until now, there was no officially validated Dutch version of the SST. Here we provided a Dutch-language translation and validation of the SST. This version of the SST is potentially a culturally equivalent instrument that will allow direct comparison of national and international study results.[16]

The aims of the study presented in Chapter 2 are translation of the SST into Dutch and evaluation of its measurement properties based on recently published guidelines from the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) group.[25]

Question 1: Is the Dutch version of the Simple Shoulder Test a valid and reliable tool that can be used in Dutch shoulder patients?

Chapter 3: Interpretation of changes in the scores of four commonly used shoulder questionnaires: Simple Shoulder Test (SST), DASH, *Quick*DASH and the Oxford Shoulder Score (OSS)

The SST, DASH, *Quick*DASH and the OSS are among the most commonly used PROMs for the shoulder. To evaluate treatment effects, it is important to know how many points a patient must improve on a PROM to demonstrate a clinically relevant change. This information can help clinicians interpret treatment results from the patient's perspective, both in clinical practice and for research purposes.

The aim of the study presented in Chapter 3 was to determine the measurement error as expressed as the Smallest Detectable Change (SDC) and the MIC of four commonly used shoulder PROMs (SST, DASH, *Quick*DASH and OSS).

Question 2: How many points should an individual patient improve on a PROM to experience a clinically relevant change?

PART II: THE DIAGNOSTIC VALUE OF COMBINING PATIENT INFORMATION AND CLINICAL TESTS

Chapter 4: The diagnostic value of the combination of patient characteristics, history and clinical shoulder tests for the diagnosis of rotator cuff tear

A rotator cuff tear can be difficult to diagnose. There are many different clinical shoulder tests for rotator cuff tears, but none have high diagnostic value.[17] Combining patient characteristics, history and multiple clinical tests has the potential to increase the diagnostic value of detecting a rotator cuff tear.

The aims of the study presented in Chapter 4 were to determine the diagnostic values of individual clinical tests and to develop a prediction model that combined patient characteristics, history and clinical test results to improve the diagnosis of rotator cuff tears using magnetic resonance arthrography (MRA) as the reference standard.

Question 3: Can the diagnostic value for detecting a rotator cuff tear be improved by combining patient characteristics, history and multiple clinical tests?

Chapter 5: The diagnostic value of the combination of patient characteristics, history and clinical tests for traumatic anterior shoulder instability

Careful patient history and physical examination are the cornerstones of diagnosing anterior shoulder instability. Sometimes, however, the diagnosis is not very clear because of a non-specific medical history and physical examination results that are inconclusive. Previous studies have focused on the diagnostic value of individual clinical tests,[17] but in clinical practice a single clinical test is often not sufficient for diagnosing a particular problem. No previous studies have looked at the value of combining patient characteristics and history with multiple clinical tests for diagnosing anterior shoulder instability.

The aims of the study presented in Chapter 5 were to determine the diagnostic values of 6 clinical tests for anterior shoulder instability and to develop a prediction model that combines patient characteristics, history and clinical tests to improve the diagnostic value.

Question 4: Can the diagnostic value for detecting traumatic anterior shoulder instability be improved by combining patient characteristics, history and multiple clinical tests?

PART III: APPLICATION OF MEASUREMENT INSTRUMENTS IN CLINICAL PRACTICE

Chapter 6: Long-term results after arthroscopic shoulder stabilization using suture anchors: an 8- to 10-year follow-up

Anterior shoulder instability is a common problem. The incidence is reported to be around 38 per 100 000 persons per year.[35,42] For the Netherlands, which has 16 million inhabitants, this is around 5120 shoulder dislocations a year. The dislocation is directed anteriorly in more than 90% of the cases.[15] In many patients, surgical stabilisation is required for recurrent instability despite primary conservative treatment. Traditionally, this has been accomplished by open surgery, in many cases with favourable results.[4,29] In more recent years, because of concerns over soft tissue damage and in the wake of an increasing general emphasis on less invasive procedures, many patients are treated arthroscopically. Arthroscopic treatment is intended to result in less soft issue trauma, smaller surgical scars, less postoperative pain and increased functionality. Unfortunately, there are not many long-term results that evaluate this arthroscopic technique in terms of both recurrence rate and subjective shoulder function.[14]

The study presented in Chapter 6 aimed to prospectively evaluate long-term surgical outcomes after arthroscopic shoulder stabilization in patients with traumatic anterior shoulder instability using absorbable suture anchors with re-dislocation as the primary outcome and subjective shoulder function as the secondary outcome.

Question 5: What are the long-term results of arthroscopic shoulder stabilization?

Chapter 7: The added value of neurocognitive testing following sports-related concussion

Another important area that would benefit from more standardised testing of the patient's condition is the field of sports-related brain concussion. The diagnosis and subsequent management of this condition has traditionally relied heavily on the athlete's self-reported symptoms, but there is accumulating evidence that clinician-based outcomes have added value.[24] There is a clear need for better criteria for diagnosing concussion and for developing appropriate return-to-play recommendations. Both patient- and clinician-based information should be considered when making the diagnosis and recommendations. Reliance on the athlete's self-reported symptoms may result in potential exposure to additional injury if the athlete returns to play too early; it may also result in the athlete postponing the return to sports to a later time than is medically required.[20] Neurocognitive testing is already widely used in the USA at the high school, collegiate and professional levels of sport participation by having the athletes participate in baseline- and post-concussion evaluations. These measures may be helpful for evaluating the symptoms and cognitive functioning of the affected athlete and may allow the use of more objective criteria for determining subsequent management of the injury.

The aim of the study presented in Chapter 7 was to evaluate the diagnostic value of players' self-reported symptoms to detect concussion when used in combination with neurocognitive scores (comparing: pre-concussion and post-concussion scores) in a group of high school and college athletes.

Question 6: Are neurocognitive testing results a valuable adjunct to a patient's self-reported symptoms for detecting post-concussive abnormalities after sports-related brain concussions?

REFERENCES

1. Allom R, Colegate-Stone T, Gee M, Ismail M, Sinha J. Outcome analysis of surgery for disorders of the rotator cuff: a comparison of subjective and objective scoring tools. J Bone Joint Surg Br. 2009 Mar; 91(3):367–373. doi:10.1302/0301-620X.91B3.20714
2. Angst F, Schwyzer H-K, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and Its Short Version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society Standardized Shoulder. Arthritis Care Res. 2011 Nov 7;63(S11):S174–S188. doi:10.1002/acr.20630
3. Berendes T, Pilot P, Willems WJ, Verburg H, Slaa te R. Validation of the Dutch version of the Oxford Shoulder Score. J Shoulder Elbow Surg. 2010 Sep;19(6):829–836. doi:10.1016/j.jse.2010.01.017
4. Berendes TD, Wolterbeek R, Pilot P, Verburg H, Slaa te RL. The open modified Bankart procedure: outcome at follow-up of 10 to 15 years. J Bone Joint Surg Br. 2007 Aug;89(8):1064–1068. doi: 10.1302/0301-620X.89B8.19280
5. Bot SDM. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis. 2004 Apr 1;63(4):335–341. doi:10.1136/ard.2003.007724
6. Christie A, Hagen KB, Mowinckel P, Dagfinrud H. Methodological properties of six shoulder disability measures in patients with rheumatic diseases referred for shoulder surgery. J Shoulder Elbow Surg. 2009 Jan;18(1):89–95. doi:10.1016/j.jse.2008.07.008
7. Codman EA, Chipman WW, Clark JG, Kanavel AB. Codman: Standardisation of hospitals: report of the committee appointed by the Clinical Congress of Surgeons of North America. Trans Clin Cong Surg North Am. 1913;
8. Codman EA. The Shoulder. Krieger Publishing Company; 1934.
9. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin. Orthop. Relat. Res. 1987 Jan;(214):160–164.
10. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br. 1996 Jul;78(4):593–600.
11. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. J Bone Joint Surg Br. 1999 May;81(3):420–426.
12. Desai AS, Dramis A, Hearnden AJ. Critical appraisal of subjective outcome measures used in the assessment of shoulder disability. Ann R Coll Surg Engl. 2010 Jan;92(1):9–13. doi:10.1308/003588410X12518836440522
13. Donabedian A. The end results of health care: Ernest Codman's contribution to quality assessment and beyond. 1989.
14. Freedman KB. Open Bankart Repair Versus Arthroscopic Repair With Transglenoid Sutures or Bioabsorbable Tacks for Recurrent Anterior Instability of the Shoulder: A Meta-analysis. Am J Sports Med. 2004 Sep 1;32(6):1520–1527. doi:10.1177/0363546504265188
15. Goss TP. Anterior glenohumeral instability. Orthopedics. 1988 Jan;11(1):87–95.
16. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993 Dec;46(12):1417–1432.
17. Hegedus EJ. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med. 2012 Jul 7;doi:10.1136/bjsports-2012-091066
18. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). The Upper Extremity Collaborative Group (UECG). Am. J. Ind. Med. 1996 Jun;29(6):602–608.

19. Jackowski D, Guyatt G. A guide to health measurement. Clin. Orthop. Relat. Res. 2003 Aug;(413):80–89. doi:10.1097/01.blo.0000079771.06654.13
20. Kelly JP, Rosenberg JH. Diagnosis and management of concussion in sports. Neurology. 1997 Mar;48(3): 575–580.
21. Khanna D, Krishnan E, Dewitt EM, Khanna PP, Spiegel B, Hays RD. Patient-Reported Outcomes Measurement Information System (PROMIS®) — The future of measuring patient reported outcomes in rheumatology. Arthritis Care Res. 2011 Nov 7;63 Suppl 11(S11):S486–S490. doi:10.1002/acr.20581
22. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med. 1998 Nov;26(6):764–772.
23. Lippitt SB, Harryman DT, matsen FA. The Shoulder: a balance of mobility and stability. American Academy of Orthopedic Surgeons Rosemont IL; 1993.
24. Lovell MR, Collins MW, Maroon J. Inaccuracy of symptom reporting following concussion in athletes. Medicine and Science in Sports Exercise. 2002 Nov 19;(34):S298.
25. 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. Qual Life Res. 2010 May;19(4):539–549. doi:10.1007/s11136-010-9606-8
26. Murrell GA, Walton JR. Diagnosis of rotator cuff tears. Lancet. 2001 Mar 10;357(9258):769–770. doi: 10.1016/S0140-6736(00)04161-1
27. Park HB, Yokota A, Gill HS, Rassi El G, McFarland EG. Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. J Bone Joint Surg Am. 2005 Jul;87(7):1446–1455. doi:10.2106/JBJS.D.02335
28. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health. 2007 Nov;10 Suppl 2:S125–37. doi:10.1111/j.1524-4733.2007.00275.x
29. Pelet S, Jolles BM, Farron A. Bankart repair for recurrent anterior glenohumeral instability: results at twenty-nine years' follow-up. J Shoulder Elbow Surg. 2006 Mar;15(2):203–207. doi:10.1016/j.jse.2005.06.011
30. Poolman RW, Swiontkowski MF, Fairbank JCT, Schemitsch EH, Sprague S, de Vet HCW. Outcome instruments: rationale for their use. The Journal of Bone and Joint Surgery. 2009 May;91 Suppl 3:41–49. doi:10.2106/JBJS.H.01551
31. Pynsent PB, Fairbank JCT, Carr A. Outcome measures in orthopaedics. Butterworth-Heinemann; 1993.
32. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res. 1991 Dec;4(4):143–149.
33. Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011 Jul;93(7):867–875. doi:10.1302/0301-620X.93B7.25737
34. Rouleau DM, Farber K, MacDermid JC. Systematic review of patient-administered shoulder functional scores on instability. J Shoulder Elbow Surg. 2010 Dec 1;19(8):1121–1128. doi:10.1016/j.jse.2010.07.003
35. Slaa te RL, Wijffels MPJM, Brand R, Marti RK. The prognosis following acute primary glenohumeral dislocation. J Bone Joint Surg Br. 2004 Jan;86(1):58–64. doi:10.1302/0301-620X.86B1.13695
36. Suk M, Hanson B, Norvell DC, Helfet DL. Musculoskeletal Outcomes Measures and Instruments. Thieme Medical Pub; 2009.
37. Swiontkowski MF, Buckwalter JA, Keller RB, Haralson R. The outcomes movement in orthopaedic surgery: where we are and where we should go. J Bone Joint Surg Am. 1999 May;81(5):732–740.

38. Terwee CB, de Winter AF, Scholten RJ, Jans MP, Devillé W, van Schaardenburg D, et al. Interobserver reproducibility of the visual estimation of range of motion of the shoulder. YAPMR. 2005 Jul;86(7): 1356–1361.

39. van de Water ATM, Shields N, Taylor NF. Outcome measures in the management of proximal humeral fractures: a systematic review of their use and psychometric properties. J Shoulder Elbow Surg. 2011 Mar;20(2):333–343. doi:10.1016/j.jse.2010.10.028

40. van Kampen DA, van Beers LWAH, Scholtes VAB, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg. 2012 Jun;21(6):808–814. doi:10.1016/j.jse.2011.09.026

41. Comptes Rendes de la 1e (CGPM) Conférence Générale des Poids et Mesures. 1889 Jan 22;

42. Richtlijn: Schouderluxatie. 2005 Nov 12;

43. Richtlijn diagnostiek en behandeling van het subacromiaal pijnsyndroom. 2012 Nov 30;:1–172.

44. Bureau International des Poids et Mesures. http://www.bipm.org.

Chapter 2

Validation of the Dutch version of the Simple Shoulder Test

DA Van Kampen
LWAH van Beers
VAB Scholtes
CB Terwee
WJ Willems

Journal of Shoulder and Elbow Surgery (JSES) dec 2011.
Vol 21, Issue 6, June 2012, 808–814

ABSTRACT

Background: The Simple Shoulder Test (SST) is an internationally-used patient reported outcome (PRO) for clinical practice and research purposes. It is developed for measuring functional limitations of the affected shoulder in patients with shoulder dysfunction and contains 12 questions (YES/NO). The purpose of this study was to create a Dutch translation of the SST and to assess the reliability and validity.

Methods: The SST was translated into Dutch using forward and backward translations. A consecutive cohort of patients with shoulder problems visiting an orthopedic clinic completed the Dutch version of the SST twice within 28 days. Additionally the Dutch validated versions of the Disabilities of the Arm, Shoulder and Hand (DASH), Oxford Shoulder score (OSS) and the Constant-Murley shoulder assessment (CM) were completed for assessing construct validity.

Results: One hundred and ten patients with a mean age of 39y (SD14), 72% male, completed the questionnaires. The internal consistency was high (Cronbach's alpha 0.78) . The test- retest reliability was very good (ICC 0.92, n=51). The measurement error expressed in the Standard Error of Measurement was 1.18 and the Smallest Detectable Change was 3.3 on a scale from 0 to 12. The construct validity was supported by expected high correlations between the Dutch version of the SST and the DASH (r = -0.74) and between the SST and the OSS (r=-0.74) and an expected moderate correlation between the SST and the CM (r=0.59).

Conclusion: The Dutch version of the SST seems to be a reliable and valid instrument for evaluating functional limitations in patients with shoulder complaints.

INTRODUCTION

There are several measurement tools for the evaluation of shoulder function. Distinctions can be made between clinician reported and patient reported outcomes (PRO). Although clinician reported outcomes have always been important, PRO's specifically capture the patient's perspective.[15,24,33] As such, they have gained a widespread recognition in the evaluation of treatment in clinical practice and research.[24,33] Currently, the use of PRO's in clinical practice and research is increasingly dictated by national governments, insurance companies, and the US Food and Drug Administration (FDA).[23] There are many different PRO's available. They vary in content, purpose (discrimination or evaluation), and quality (i.e. measurement properties). There are PROs measuring general health, like the Short Form 36 health survey (SF-36) and disease-specific PROs, like the Western Ontario Shoulder Instability Index (WOSI) for patient with shoulder instability, or body part specific PROs, like the Simple Shoulder test (SST).[16,17,39] Which one to choose not only depends on their content, but also on their measurement properties: a PRO should be valid and reliable. However, a systematic review of shoulder disability questionnaires showed that only few measures have adequate quality.[5] One of the most often used shoulder instruments is the Simple Shoulder Test (SST). The SST is a simple and short PROs measuring functional limitations of the affected shoulder in people with shoulder dysfunction. It can be completed within 3 minutes and patients can do this at home, making it very practical for the busy clinician.[17] The SST has shown to have good test-retest reliability (ICC 0.99).[3] Hence, for discriminative purposes the SST is suggested for use in patients with shoulder complaints in general. Until now the SST has only been validated in English.[29] The translation and validation of internationally used PROs will lead to culturally equivalent instruments which will allow direct comparison of national and international study results.[14] The purpose of this study was to translate the SST into Dutch and to evaluate its measurement properties based on recently published guidelines from the consensus-based standards for the selection of health measurement instruments (COSMIN) group.[20] In this study we determined the internal consistency, reliability, construct validity and floor and ceiling effect of the Dutch version of the SST in patients with shoulder complaints in secondary care.

MATERIAL AND METHODS

Translation procedure

First, the original English version of the SST was translated into Dutch by three native Dutch speaking, medically educated, translators independently of each other. Subsequently, they created a consensus version (the Dutch SST) which was also checked for possible cross-culture differences. Subsequently, forty patients with shoulder complaints were asked to assess the comprehensibility of the Dutch SST, by looking at the understandability and logic of the

independent questions.[37] These forty patients were not included in the current study. Finally the Dutch SST was translated back to English by a professional translator. The back-translated version was compared to the original English version.[17]

Validation process

Second, we assessed the measurement properties (reliability, validity, responsiveness) in the Dutch population. Institutional approval was wavered by our local ethical committee and written signed informed consent was obtained from all participants. A prospective cohort of patients was recruited consecutively between February 2009 and October 2010 at the Orthopedic outpatient clinic of the Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands, by one orthopedic surgeon (WJW). Inclusion criteria were: 16 years or older and having shoulder problems as diagnosed by the orthopedic surgeon (WJW). Exclusion criteria were: fractures, frozen shoulder and problems with reading and understanding the Dutch language. We aimed to include at least 100 patients, which is considered adequate for assessing measurement properties.[35] All patients received an email with a link to the web based PRO or could alternatively receive the paper version. The web-based and paper version of the PRO were identical. Every question was obligated to answer, so there could not be any missing values.

The Simple Shoulder test (SST)

The SST is a body part specific PRO, measuring functional limitations of the affected shoulder in patients with shoulder dysfunction. It was originally developed in the US for evaluating patients with common shoulder problems like rotator cuff tear, impingement, instability and degenerative arthritis.[17] The SST consists of 12 questions with dichotomous response options. For each question, the patient indicates if he/she is able or not able to do the activity. The scores are summarized into a total score, which ranges from 0 (worst) to 12 (best) shoulder functioning. Missing data were treated as follows; one or two missing values were substituted with the average value for the other items. If more than two items were missing, the response to this questionnaire was considered invalid and no total score was calculated.

Oxford Shoulder Score (OSS)

The OSS is a body part specific PRO. It contains 12 items regarding pain and shoulder function. It was developed in the UK.[7] There are 5 response options for every question, corresponding to a score ranging from 1 (least difficulties) to 5 (most difficulties). Scores are summarized into a total score, with a range from 12 (best) to 60 (worst) shoulder functioning. This questionnaire has been validated in shoulder patients with impingement, arthritis and frozen shoulder. The OSS has been validated in Dutch shoulder patients.[4]

Disabilities of the Arm, Shoulder and Hand (DASH)

The DASH is a body part specific PRO. It contains 30 questions and was designed to measure physical functioning and symptoms in patients with musculoskeletal disorders of the upper

limb. It was developed in the US. Items are summarized into a total score ranging from 0 (excellent) to 100 (worst). The measurement properties have been assessed in patients with disorders of the upper limp: shoulder, elbow, wrist and hand.[2] The DASH has been shown to be reliable, valid and responsive in patients with shoulder disability.[8] This instrument has been validated in English and Dutch for patients with a disorder of the upper limb.[38]

Short Form Health Survey (SF-36)

The SF-36 is a general PRO. It is a short-form health survey with 36 questions. It is widely used to assess the general health of patients with all kind of disorders. It provides scores on 8 dimensions (subscales): physical functioning, social functioning, role limitations caused by physical problems, role limitations caused by emotional problems, general mental health, vitality, social functioning, bodily pain, and perception of general health. Each subscale has a minimum score of 0 points (worst) and a maximum score of 100 points (best).[39] The SF-36 has been translated and validated in the Dutch language.[1,30] The SF 36 is the most widely evaluated PRO for assessing general health.[11] Previous studies have also validated the SF 36 specifically for shoulder complaints.[12,22]

Constant-Murley shoulder assessment (CM)

The CM functional assessment of the shoulder is a combined clinician- and patient reported outcome score.[6] It includes a pain score, functional assessment, range of motion and strength measures. This generates a score between 0 (poor) and 100 (excellent). The CM can only be used in a clinical setting. The CM is the most often used assessment score in shoulder research.[36] The CM has been validated for patient with all common shoulder complaints like: rotator cuff tear, impingement, instability, degenerative arthritis, and has shown good measurement properties.[28] There is no Dutch validated version of the CM. We used the original English version of the CM. The CM was assessed by the treating orthopedic surgeon (WJW).

Assessment of measurement properties

Internal consistency

Internal consistency is the interrelatedness among the items in a scale[21]. Different items in an instrument may ask the same questions in a slightly different manner to reliably capture the respondent's opinion or level of function. The Cronbach´s alpha is considered an adequate measure of internal consistency, when it has been shown that the (sub)scale is unidimensional (e.g. by factor analysis). A low Cronbach's alpha indicates a lack of correlation between the items in a scale, which makes summarizing the items unjustified. A very high Cronbach alpha (>0.95) reflects high correlations among the items in the scale, which indicate redundancy of one or more items.[25] We used the widely accepted cut off for the Cronbach´s alpha of 0.7 or higher.[18]

Reliability

Reliability is the proportion of the total variance in the measurements which is due to true differences between patients. This refers to the degree to which the measurement instrument is free from measurement error, and estimates the extent to which scores for patients who have not changed are the same for repeated measurements at different time point.[21,31] Test-retest reliability is assessed by completing a PRO on two occasions. The time interval between the test and retest must be long enough to prevent recollection of the previous answers and short enough to prevent that a real change in the construct to be measured has occurred. In this study we assumed that there would be no real change in patient's functioning within an interval of 1-4 weeks.

Measurement error

Measurement error is the systematic and random error of a patient´s score that is not attributed to true changes in the construct to be measured.[21] Measurement error can be expressed as the Standard Error of Measurement (SEM) or the Smallest Detectable Change (SDC). These calculations are expressed in the unit of measurement of the scale of the PRO. SEM represents the standard deviation of repeated measures in one patient. The SDC represents the minimal change that a patient has to show on the scale to ensure that the observed change is real and not just measurement error.

Construct validity

Validity is the degree to which a PRO instrument measures the construct it is supposed to measure. In the absence of a gold standard, construct validity refers to the extent to which a particular measure relates to other measures, based on theoretically derived hypotheses for the constructs that are being measured. We used the OSS, DASH, SF-36 and the CM for assessing the construct validity of the SST. Before starting the study we formulated hypotheses for the minimal level of validity, presented in Table I. These hypotheses were based on clinical experience, knowledge about the various PROs and consensus among the study investigators. We expected the highest correlation of the SST with the OSS, because these questionnaires measure the same construct, addressing specifically the shoulder. The DASH looks at the whole upper extremity, so we expected a high correlation with the SST, but less than the correlation of the SST with a shoulder specific PRO. The SF-36 is a general PRO so we expected a correlation with the SST between 0.50 and 0.70. The CM is a combined clinician- and patient based questionnaire and focusing stronger on range of motion and muscle strength, and therefore we expected a lower correlation with the SST.

Table I. Predetermined hypotheses for testing the validity of the Dutch version of the Simple shoulder test

1.	A correlation of at least -0.7 was expected between the SST and the Oxford shoulder score.
2.	A correlation of at least -0.6 was expected between the SST and the DASH.
3.	A correlation between 0.5-0.7 was expected between the SST and the SF-36 subscale physical function
4.	The correlation between the SST and the SF-36 physical function should at least be 0.1 higher than the correlation between the SST and the other subscales of the SF-36.
5.	A correlation of at least 0.5 was expected between the SST and the Constant score.

Floor and ceiling effects

The presence of floor or ceiling effects may have a negative effect on the quality of the instrument. If a group of patients scores primarily in the extremes the responsiveness may be limited. Floor or ceiling effects were considered to be present if >15 % of the respondents achieved the minimum or maximum possible score.[19]

Statistical analyses

We assessed the internal consistency by performing factor analysis and by calculating the Cronbach's alpha. Confirmatory factor analysis for categorical items was performed in Mplus using the method of weighted least squares with mean and variance adjustment (WLSMV). We examined factor loadings and model fit. Factor loadings represent the correlation between the items in the questionnaire and the factors (the underlying dimensions). Analogous to Pearson's r, the squared factor loading is the percent of variance in that indicator variable explained by the factor. Factor loadings are generally considered to be meaningful when they exceed 0.30 or 0.40.[10] We considered factor loadings of at least 0.50 appropriate. The Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA) were used as measures for model fit. A CFI and TLI of > 0.95 and a RMSEA of < 0.05 were considered as adequate fit. For a moderate fit, values > 0.90 and < 0.08 were used. Reliability was assessed by calculating the Intraclass Correlation Coefficient (ICC) with a 95% confidence interval (CI). A two-way mixed effects model for absolute agreement was used. An ICC> 0.8 was considered good. The SEM was calculated from the square root of the variance between the measurements and the error variance of the ICC. The SDC was calculated as $1.96*\sqrt{2}*SEM$. Validity was tested by calculating Pearson's correlation coefficients. These analyses were carried out using the Statistical Package for the Social Sciences (SPSS; Gorinchem, the Netherlands), version 18.0.

RESULTS

Forward and backward translation of the SST revealed no problems or language difficulties. Because of cross-cultural differences between the US and the Netherlands, questions 9 and 10, asking about the ability to throw a softball a certain distance, were adapted to the Dutch situation. In the Netherlands not many people play softball. Tennis is a much more common sport, so we changed the softball into a tennis ball and we reduced the distance, based on the different weight of the soft- and tennis bal. The pilot study with 40 subjects, reported easy administration and no understandability problems. Full text version of the Dutch version of the SST is available in the Appendix.

One-hundred and ten consecutive patients with shoulder complaints were included in the study and they all completed the questionnaires, we had no patients who refused to participate. There were no missing values. Out of the 110 patients only 1 patient used a paper version. Ta-

ble II shows the demographic data of all patients included in this study. The last fifty-five patients of this consecutive cohort were asked to complete the questionnaires a second time, to assess test-retest reliability, of which three failed to do so and had to be excluded. One patient had too much time between test and re-test (57 days), and was therefore also excluded from the test re-test reliability analysis. There were no substantial differences between both groups.

Table II. Demographic data of the study groups.

	Initial assessment	Second assessment (test-retest)
N	110	51
Mean age [range]	39 years [16-71]	39 years [16-69]
Gender	72% male	73% male
Left shoulder - right shoulder – both (n)	38 – 69 – 3	16 – 35 - 0
Diagnose (n)		
Impingement	10	6
Rotator cuff tear	32	12
SLAP lesion	17	10
Instability	48	22
Tendinitis of the biceps	3	1

Internal consistency

The results of the confirmatory factor analysis are shown in Table III. A 1-factor model fitted the data moderately. CFI was 0.943, TLI was 0.931 and RMSEA was 0.068. Items 1, 2 and 12 had relatively low factor loadings (<0.50). The Cronbach's alpha was 0.78, which indicates a good internal consistency. Removing items from the questionnaire did not result in a higher Cronbach's alpha.

Table III. Factor loadings of the SST

Question	Estimate	S.E.
1	0.50	0.11
2	0.39	0.12
3	0.87	0.07
4	0.89	0.06
5	0.70	0.11
6	0.82	0.08
7	0.76	0.07
8	0.54	0.10
9	0.80	0.08
10	0.65	0.10
11	0.84	0.06
12	0.43	0.12

The mean time between completing the first and second questionnaire was 13 days. Table IV shows the distributions of the data at test and retest and the reliability of the SST. The scores on the SST at retest were a little higher (mean 0.1) than at the first test. The ICC was 0.92, which indicates excellent test-retest reliability. The measurement error expressed in SEM was 1.18. The measurement error expressed in SDC was 3.3 on a scale of 0 to 12.

Table IV. Test – retest reliability and descriptive statistics of the Simple Shoulder Test for 51 patients who completed two PROs

mean (SD)			**SEM**	**SDC**	**ICC (95% CI)**
Baseline	Retest	Change			
8.4 (2.8)	8.6 (2.8)	0.1 (1.3)	1.18	3.27	0.92 (0.86-0.95)

SD, standard deviation; SEM, standard error of measurement; SDC, smallest detectable change; ICC, intra-class correlation coefficient; CI, confidence interval

Construct validity

Correlations between the SST and the OSS, the DASH, the SF-36 subscales and the CM, respectively are presented in Table V. The SST correlation with the subscale physical function was at least a 0.1 higher than the correlation of the SST with the other subscales of the SF-36, except for the subscale bodily pain, which had a comparable correlation. All other hypotheses were confirmed.

Table V. Validity of the Simple Shoulder Test expressed by Pearson correlation.

	Pearson correlation with the SST
Oxford Shoulder score	-0.74
DASH	-0.74
SF-36 sub scales	
BP	0.51
PF	0.56
SF	0.01
RF	0.32
RE	0.18
MH	0.18
VT	0.22
GH	0.16
Constant-Murley	0.59

Abbreviations: BP, bodily pain; PF, physical function; SF, social function; RF, role limitations because of physical problems; RE, role limitations because of emotional problems; MH, mental health; VT, vitality; GH, general health perception.

Floor and ceiling effects

There were no floor or ceiling effects: 1,8 percent of all patients reported the worst possible score (0) and 13.6 percent reported the best possible score (12) on the SST.

DISCUSSION

This study shows that the Dutch version of the SST is a valid, reliable and internally consistent instrument for the assessment of patients with shoulders complaints, and has no floor or ceiling effects. In determining the construct validity all except one of our predetermined hypotheses were confirmed. The SST correlated highest with the OSS, DASH and the SF36 subscale physical functioning. The SST correlated lower with the CM (0.59). These results are comparable with a similar validation study of the OSS by Berendes et al.[4] Lower correlations between these two PRO´s (SST/OSS) and the CM can be explained by the fact that the CM is mostly clinician based and focuses more on physical symptoms, like shoulder movements and pain, and less on physical functioning. Furthermore, we found only a minimal difference between the correlation of the SST with the SF-36 subscale physical functioning and the correlation with the SF-36 subscale bodily pain. These results are comparable with the results of Beaton et al., showing correlations of 0.58 and 0.62, respectively.[3] We hypothesized a difference of at least 0.1, based on the fact the SST focuses more on function than on pain. The fact that we did not find this difference is possibly due to the fact that the questions being asked in the SST do not differ, strong enough, between loss of function and pain related loss of functioning. This a common problem in disease-specific PROs, as was also shown in the Western Ontario and McMaster University Osteoarthritis Index (WOMAC); the most commonly used PRO for Knee osteoarthritis.[32] The questions used in these PROs do not specify enough if the limitation in functioning is due to less function or if it is due to pain.

In determining reliability, we found a high ICC (0.92), which is comparable to other studies.[3,13] In addition, we determined the measurement error in terms of SEM and SDC. The found a SEM of 1.18 points, and a SDC of 3.3 points. This SDC means that if you want to determine a treatment effect, you need to find a difference of at least 3.3 points in an individual patient to make sure that the difference is not due to random error. So an observed change on a scale must be larger than the SDC to ensure a real change.[9] An instrument is useful if the SDC is smaller than the Minimal Important Change (MIC). Only one study looked at MIC for the SST. Tashjian et al, showed a MIC of 2.05 for the SST.[34] So, if you find a change of 3 points on the SST in a individual patient, this is larger than the MIC but it could still be due to measurement error. Determining the MIC for the Dutch version of the SST is important and part of our future research.

Finally, factor analysis showed unidimensionality of the SST. We did find a lower factor loading for questions 1, 2 and 12 than for the other questions. Analyzing questions 1 and 2 showed that these questions focus more on pain instead of function. Question 12 is more

general, asking if the patient is able to perform his/her work, which may be effected by other factors than shoulder functioning. This probably explains the lower factor loading of these three questions. The Cronbach's alpha of 0.78 is comparable with the results of Roddey et al., who found a Cronbach´s alpha of 0.85.[27]

A strong point of our study is the fact that we had no missing data. This was a clear advantage of web-based administration of the questionnaires. On the website patients could not continue when items were not completed. Furthermore, all patients completed the SST the first time, and for the retest, we only had a 5% lost to follow up. Finally, we feel that our validation methods are very robust and could be used as an example for future PRO validation studies. A limitation of the study was that we did not correlate the SST with more shoulder specific PROs like the Shoulder Pain and Disability index (SPADI).[26] Although it would be interesting to correlate the SST to more questionnaires, it would increase the burden on the patients and we think that this study included the most important PROs for this specific validation. Based on a comparison with previously published results, with comparable internal consistency, reliability and construct validity, we conclude that the Dutch version of the SST seems to be a culturally equivalent instrument which can be used for direct comparison of national and international study results.

Conclusion

The Dutch version of the SST is user-friendly and can easily be administered web-based but also on paper. Eighty-six percent of our predefined hypotheses about the construct validity could be confirmed. We found high reliability (ICC 0.92) and a high internal consistency (Cronbach's alpha 0.78). Therefore we consider the SST a valid and reliable instrument suitable for monitoring groups of patients. The SST can be used for clinical trials and cross country comparisons. We recommend the use of the Dutch translation of the SST for evaluating groups of patients with shoulder complaints.

REFERENCES

1. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 1998;51:1055-68. S0895-4356(98)00097-3
2. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. J Hand Ther 2001;14:128-46.
3. Beaton DE, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg 1998;7:565-72.
4. 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;19:829-36. doi: 10.1016/j.jse.2010.01.017
5. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis 2004;63:335-41. doi: 10.1136/ard.2003.007724
6. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res 1987:160-4.
7. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br 1996;78:593-600.
8. Desai AS, Dramis A, Hearnden AJ. Critical appraisal of subjective outcome measures used in the assessment of shoulder disability. Ann R Coll Surg Engl 2010;92:9-13. doi: 10.1308/003588410X12518836440522
9. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Phys Ther 1994;74:777-88.
10. Floyd FJ WK. Factor analysis in the development and refinement of clinical assessment instruments. Psychol Assess 1995;7:286-99.
11. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 2002;324:1417. doi: 10.1136/bmj.324.7351.1417
12. Gartsman GM, Brinker MR, Khan M, Karahan M. Self-assessment of general health status in patients with five common shoulder conditions. J Shoulder Elbow Surg 1998;7:228-37.
13. Godfrey J, Hamman R, Lowenstein S, Briggs K, Kocher M. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. J Shoulder Elbow Surg 2007; 16:260-7. doi: 10.1016/j.jse.2006.07.003
14. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417-32.
15. Jackowski D, Guyatt G. A guide to health measurement. Clin Orthop Relat Res 2003:80-9. doi: 10.1097/01.blo.0000079771.06654.13
16. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med 1998;26:764-72.
17. Lippitt SB HD, Matsen FA. A practical tool for evaluating function: The simple shoulder test. In The Shoulder: A Balance of Mobility and Stability.: American Academy of Orthopaedic Surgeons; 1993.
18. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clin Ther 1996;18:979-92.
19. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293-307.

20. 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. Qual Life Res 2010;19:539-49. doi: 10.1007/s11136-010-9606-8

21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737-45. doi: 10.1016/j.jclinepi.2010.02.006

22. Ostor AJ, Richards CA, Prevost AT, Speed CA, Hazleman BL. Diagnosis and relation to general health of shoulder disorders presenting to primary care. Rheumatology (Oxford) 2005;44:800-5. doi: 10.1093/rheumatology/keh598

23. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10 Suppl 2:S125-37. doi: 10.1111/j.1524-4733.2007.00275.x

24. Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch EH, Sprague S, de Vet HC. Outcome instruments: rationale for their use. J Bone Joint Surg Am 2009;91 Suppl 3:41-9. doi: 10.2106/JBJS.H.01551

25. Pynsent PB. Choosing an outcome measure. J Bone Joint Surg Br 2001;83:792-4.

26. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res 1991;4:143-9.

27. Roddey TS, Olson SL, Cook KF, Gartsman GM, Hanten W. Comparison of the University of California-Los Angeles Shoulder Scale and the Simple Shoulder Test with the shoulder pain and disability index: single-administration reliability and validity. Phys Ther 2000;80:759-68.

28. Roy JS, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. J Shoulder Elbow Surg 2010;19:157-64. doi: 10.1016/j.jse.2009.04.008

29. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. Arthritis Rheum 2009;61:623-32. doi: 10.1002/art.24396

30. Salaffi F, De Angelis R, Stancati A, Grassi W. Health-related quality of life in multiple musculoskeletal conditions: a cross-sectional population based epidemiological study. II. The MAPPING study. Clin Exp Rheumatol 2005;23:829-39.

31. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury 2011;42:236-40. doi: 10.1016/j.injury.2010.11.042

32. Stratford PW, Kennedy DM. Does parallel item content on WOMAC's pain and function subscales limit its ability to detect change in functional status? BMC Musculoskelet Disord 2004;5:17. doi: 10.1186/1471-2474-5-17

33. Swiontkowski MF, Buckwalter JA, Keller RB, Haralson R. The outcomes movement in orthopaedic surgery: where we are and where we should go. J Bone Joint Surg Am 1999;81:732-40.

34. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am 2010;92:296-303. doi: 10.2106/JBJS.H.01296

35. 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. Qual Life Res 2011. doi: 10.1007/s11136-011-9960-1

36. van de Water AT, Shields N, Taylor NF. Outcome measures in the management of proximal humeral fractures: a systematic review of their use and psychometric properties. J Shoulder Elbow Surg 2011;20:333-43. doi: 10.1016/j.jse.2010.10.028

37. van Laarhoven HA TSR, De Boer LM, Willems WJ. . De Nederlandse Simpele Schouder Test, een efficiënte wijze van schouderfunctie bepaling. NTvO 2001;8:5-8.

38. 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;15:347-54. doi: 10.1016/S0894-1130(02)80006-0

39. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.

Chapter 3

Interpretation of changes in the scores of four commonly used shoulder questionnaires: Simple Shoulder Test, DASH, *Quick*DASH, and the Oxford Shoulder Score

DA van Kampen
WJ Willems
LWAH van Beers
RM Castelein
VAB Scholtes
CB Terwee

Submitted for journal publication

ABSTRACT

Background: Improved interpretation of orthopedic treatment effects is needed. Pre- and post-treatment evaluation with patient-reported outcome measures (PROMs) can be used to determine treatment effects by calculating change scores. The smallest detectable change (SDC) and minimal important change (MIC) are important benchmarks for interpreting change scores. Here, we determined the SDC and MIC for four shoulder-related PROMs: the Simple Shoulder Test (SST); Disabilities of the Arm, Shoulder, and Hand (DASH and *Quick*DASH); and Oxford Shoulder Score (OSS).

Methods: We recruited 164 consecutive patients with shoulder problems who visited an orthopedic clinic. Patients completed the SST, DASH, and OSS at their first visit, 2 weeks later, and at 6 months post-treatment; *Quick*DASH scores were calculated from DASH scores. The SDC was calculated with a test re-test protocol (baseline to 2 weeks). For the MIC, change scores (baseline to 6-months post-treatment) were calculated in seven patient subgroups, according to an additional self-administered ranking of change over time (anchor-based mean change technique). The MIC was defined as the average of the "slightly improved" anchor score group.

Results: The SDC was 2.6 for SST, 16.0 for DASH, 16.4 for *Quick*DASH, and 6.0 for OSS. The MIC change score was 2.2 for SST, 12.4 for DASH, 13.4 for *Quick*DASH, and 6.0 for OSS.

Conclusions: Based on SDC and MIC, to indicate clinically relevant change, change scores should exceed 2.6 for SST, 16 for DASH and *Quick*DASH, and 6 for OSS.

INTRODUCTION

Shoulder pain is the third most common musculoskeletal complaint, after back and knee pain.[37] It is associated with considerable disability for the patient and costs to society. Depending on the diagnosis, many different surgical and non-surgical treatment modalities have been described. In research and clinical practice, determining whether a treatment results in meaningful improvement of symptoms requires the use of high-quality measurement tools.

Over the past decade, there has been a shift in interest from pathophysiological measurements to measuring patient-perceived health. This has resulted in increased use of patient-reported outcome measures (PROMs, also known as PROs). PROMs are self-evaluated measurements of any aspect of a patient's health status, without interpretation of the patient's response by a clinician or anyone else.[25] PROMs are often questionnaires specifically evaluating pain and function from the patient's perspective. The quality of a PROM can be determined by assessing the measurement properties of the instrument. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative provides a checklist of standards for assessing the measurement properties of validity, reliability, and responsiveness.[21,22] This list does not include interpretability, which is a very important attribute of a questionnaire used in daily clinical practice. Interpretability refers to what a PROM score means; for example, a given score can be interpreted by providing reference data from the general population.

Interpretability is also important in regard to change scores; it is important to know when it can be said that a patient has improved. With many PROMs, change scores are often difficult or impossible to interpret, simply because we do not know exactly what a given difference in score means. Interpreting change in PROM scores requires two benchmarks: the measurement error, expressed as the Smallest Detectable Change (SDC), and the Minimal Important Change (MIC). The SDC is a measure of the variation in a scale due to measurement error. Thus, a change score can only be considered to represent a real change if it is larger than the SDC. The MIC is defined as the smallest measured change score that patients perceive to be important.[12] If the SDC is smaller than the MIC, it is possible to distinguish a clinically important change from measurement error with a large amount of certainty. However, this is much more difficult if the SDC is larger than the MIC, since there is a considerable chance that the observed change is caused by measurement error.[33]

Both the SDC and MIC are expressed using the same units as the original measure, and thus these numbers have considerable value for clinical use. Using these two benchmarks to interpret change scores is particularly beneficial when PROMs are applied in individual patients, such as in clinical practice. On a group level, knowledge of the MIC will also provide clinicians with better options for interpreting study results. The MIC can be used to calculate the percentage of patients who report a change greater than the MIC (responders) in each arm of a trial, and these percentages of responders can be compared.[29] Researchers can also use the

SDC and the MIC on a group level to calculate an adequate sample size or to perform power analyses, as described by Terwee et al.[33]

Few studies have assessed measurement error (SDC) and interpretability (MIC) of body part-specific PROMs for patients with shoulder problems.[4,14,20,26,28,30] Therefore, the present study aimed to determine the SDC and MIC of four commonly used shoulder PROMs: the Disabilities of the Arm, Shoulder, and Hand (DASH); the Shortened Disabilities of the Arm, Shoulder, and Hand Questionnaire (*Quick*DASH); the Simple Shoulder Test (SST); and the Oxford Shoulder Score (OSS).

MATERIALS AND METHODS

A prospective cohort of patients with shoulder complaints was consecutively recruited between February 2009 and December 2011 by one orthopedic surgeon (W.J.W.) at the orthopedic outpatient clinic of the Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands. Inclusion criteria were age of 16 years or older and the presence of shoulder problems as diagnosed by the orthopedic surgeon (W.J.W.). Both surgical and non-surgical patients were included. Exclusion criteria were fractures, frozen shoulder, and problems with reading and understanding the Dutch language. Institutional approval was obtained by our local ethical committee, and written informed consent was obtained from all participants.

Measurements

Using a web-based system at home, the patients completed an online questionnaire containing the four different body part-specific PROMs at three different time-points: T1 (baseline), T2 (2 weeks after baseline), and T3 (6-month follow-up). The given questionnaires were identical at all three time-points, except for two anchor questions added at T3 (see outcome measures for details). The whole cohort was invited to complete the questionnaire at time-points T1 and T3, whereas only a subset of the cohort was also asked to complete the questionnaire at time-point T2; this was done to limit the response burden. According to international guidelines, a minimum of 50 patients is considered adequate for assessing measurement properties.[31] Since the risk of participant loss to follow-up increased after several months, we included at least 150 patients at baseline. The subset for T2, used to determine the measurement error, was predetermined at 100 patients. The online questionnaire required an answer for each question, such that there could not be any missing values.

Outcome Measures

Simple Shoulder Test (SST)

The SST measures functional limitations of the affected shoulder in patients with shoulder dysfunction. It was originally developed in the United States by Matsen et al,[19] for evaluating

patients with common shoulder problems. The SST consists of 12 questions with dichotomous response options; for each question, patients indicate if they are able or unable to perform an activity. The scores of the questions are summarized, with the total score ranging from 0 (worst) to 12 (excellent). The SST has been validated in patients with shoulder complaints,[2,19] including Dutch shoulder patients.[34]

Disabilities of the Arm, Shoulder, and Hand (DASH)

The DASH was developed in the United States by Hudak et al.[17] It is a 30-item, patient-reported questionnaire designed to measure physical functioning and symptoms in people with musculoskeletal disorders of the upper limbs.[17] Items are summarized into a total score, ranging from 0 (excellent) to 100 (worst). The measurement properties have been assessed in patients with disorders of the shoulder, elbow, wrist, and hand.[3] The recent review by Desai et al,[13] showed that the DASH is reliable, valid, and responsive in patients with shoulder disability, and this instrument has been validated in Dutch patients with an upper limb disorder.[35]

Shortened Disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH)

The *Quick*DASH is the short version of the original DASH. It was developed by Beaton et al,[5] it contains 11 of the original 30 items, and the score range is from 0 (excellent) to 100 (worst). The measurement properties are comparable with the DASH and have been evaluated in patients with upper extremity disorders.[15] Here, we computed the *Quick*DASH score from the responses to the full DASH questionnaire.

Oxford Shoulder Score (OSS)

The OSS was developed in Oxford (UK) by Dawson et al,[9] for patients with shoulder problems. It contains 12 items related to pain and shoulder function. There are five response options for each question, corresponding to a score ranging from 1 (least difficult) to 5 (most difficult). Scores of the 12 questions are summarized into a total score that ranges from 12 (excellent) to 60 (worst). The OSS has been validated in patients with shoulder complaints,[7,9,13] including in Dutch shoulder patients.[6]

Anchors

An anchor is a global rating scale in which patients are asked, in a single question at follow-up, to indicate how much their function (functional anchor) or pain (pain anchor) has changed since baseline.[10,12] The response options are as follows: completely recovered, much improved, slightly improved, unchanged, slightly worse, much worse, and worse than ever.

Statistical analysis

Smallest detectable change (measurement error)

Measurement error is the systematic and random error of a patient's score that is not attributed to true changes in the measured construct.[11,12,23] Data from T1 and T2 were used to determine the measurement error. We assumed that there would be no real change in a patient's functioning within a 2-week interval (range, 1 to 4 weeks). Measurement error can be expressed as the standard error of measurement (SEM) or the smallest detectable change (SDC). The SEM represents the standard deviation of repeated measures in one patient, and was calculated from the square root of the variance between the measurements and the error variance of the ICC (√(VarError. + VarOccasion)). The SDC represents the minimal change that a patient must show on the scale to ensure that the observed change is real and not just measurement error. The SDC was calculated as $1.96 \times \sqrt{2} \times$ SEM. These values were expressed in the unit of measurement of the PROM scale.

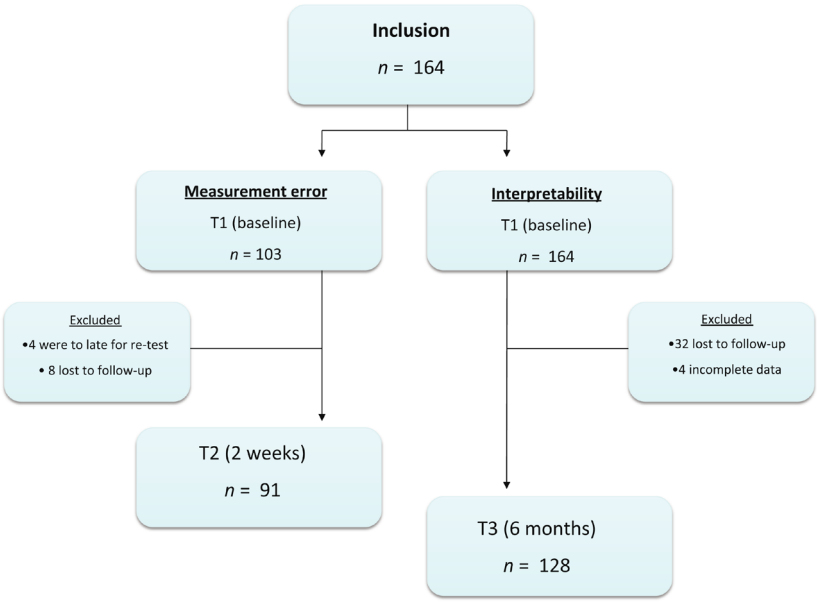
Minimal important change

The change scores on the questionnaires were calculated by subtracting each patient's T3 (6 month) score from the T1 (baseline) score, and were then used to determine the minimal important change (MIC) using an anchor-based mean change score technique.[8] The anchor scores were used to categorize patients into seven subgroups, varying from completely recovered to worse than ever. Change scores were calculated in each of the seven subgroups. The MIC was defined as the mean change score in the subcategory of patients who were "slightly improved" according to the anchor scores.[12] The SST, DASH, and *Quick*DASH primarily assess shoulder function; therefore, we compared these change scores only to the functional anchor. The OSS includes questions on both pain and function; therefore, we compared the OSS change score with both the pain and functional anchors.

RESULTS

Figure I illustrates the flow of the patients through the study. We asked 164 consecutive patients with shoulder complaints to participate in this study. None refused to participate; thus, the initial response rate at T1 was 100%. Of these, 103 patients were sent the questionnaire at T2. A total of 95 completed the questionnaire at T2; however, only 91 of these could be analyzed since four patients submitted this questionnaire after the maximum period of 4 weeks (response rate for measurement error: 89%). Of all 164 patients, 132 patients completed the questionnaire at T3 (6-month follow-up). Of these, 128 could be analyzed since four patients did not answer the anchor questions on function and pain (response rate for interpretability: 78%). The demographic data are presented in Table I. At the 6-month evaluation, 53% of the patients were treated surgically.

Figure I. Flowchart



n, the number of patients evaluated

Table I. Demographic data

		Baseline	SDC analysis	MIC analysis
Number		164	91	128
Mean Age, y	[range]	41 [16-76]	39 [16-76]	39 [16-76]
Gender	M	115 (70%)	62 (68%)	59 (69%)
	F	49 (30%)	29 (32%)	39 (31%)
Side	L	59 (36%)	29 (32%)	45 (35%)
	R	101 (62%)	62 (68%)	81 (63%)
	B	4 (2%)	-	2 (2%)
Diagnosis	Impingement syndrome	18 (11%)	11 (12%)	14 (11%)
	Rotator cuff tear	39 (24%)	21 (23%)	29 (23%)
	SLAP lesion	25 (15%)	15 (17%)	21 (16%)
	Anterior instability	75 (46%)	41 (45%)	58 (45%)
	Tendinitis biceps	7 (4%)	3 (3%)	6 (5%)

Data given as numbers (percentages), unless otherwise stated. M, male; F, female; L, left; R, right; B, both; SLAP, superior labral tear from anterior to posterior.

Smallest detectable change (measurement error)

The 91 patients who completed questionnaires at both T1 and T2, did so within a mean time period of 12.8 days (SD, 5.5). Table II shows the raw scores from T1 and T2 and the change scores. The SDC was 2.6 for the SST, 16.0 for the DASH, 16.4 for the *Quick*DASH, and 6.0 for the OSS (Table II).

Table II. PROMs characteristics, and scores at baseline and follow-up

		SST	DASH	*Quick*DASH	OSS
Minimum score		0 (worst)	0 (excellent)	0 (excellent)	12 (excellent)
Maximum score		12 (excellent)	100 (worst)	100 (worst)	60 (worst)
T1: mean (SD)		8.5 (2.8)	24.4 (16.0)	25.5 (17.4)	24.4 (7.5)
T2: mean (SD)		8.7 (2.7)	22.5 (14.9)	23.9 (16.1)	23.4 (7.2)
T3: mean (SD)		9.8 (2.5)	16.9 (13.9)	17.1 (14.7)	20.8 (6.5)
Change score T1-T2, mean (SD)		0.01 (1.4)	−0.7 (8.3)	−0.24 (8.7)	−0.6 (3.1)
Change score T1-T3, mean (SD)		1.3 (2.7)	−6.9 (13.8)	−7.9 (15.3)	−3.5 (6.6)
SDC		2.6	16.0	16.4	6.0
MIC	Functional anchor	2.2	12.4	13.4	6.0
	Pain anchor				4.7

T1, baseline; T2, 2 weeks; T3, 6 months; SDC, smallest detectable change; MIC, minimal important change.

Minimal important change

The mean change scores per subgroup based on the functional and pain anchors are presented in Tables III and IV, respectively. From these data, we used the mean change score of the "slightly improved" group to determine the MIC. The MIC for function was 2.2 for the SST, 12.4 for the DASH, 13.4 for the *Quick*DASH, and 6.0 for the OSS. The MIC for pain was only calculated for the OSS, and was 4.7. The MIC data are presented in Table II.

Table III. Mean change score of the four PROMs according to the functional anchor

Functional anchor		**Mean change score (SD)**			
	n	SST	DASH	*Quick*DASH	OSS
Completely recovered	14	2.9 (1.8)	−13.2 (9.6)	−13.5 (11.6)	−6.0 (4.6)
Much improved	37	2.9 (2.2)	−15.6 (13.3)	−17.9 (15.0)	−7.2 (6.8)
Slightly improved	**23**	**2.2 (2.7)**	**−12.4 (11.7)**	**−13.4 (12.7)**	**−6.0 (5.3)**
Unchanged	43	−.1 (1.5)	0.3 (9.6)	−0.1 (10.8)	−1.0 (4.2)
Slightly worse	5	−0.6 (0.9)	9.8 (6.2)	6.8 (3.6)	4.2 (4.6)
Much worse	6	−4.0 (2.8)	13.1 (9.6)	14.8 (10.3)	8.3 (5.4)
Worse than ever	0				

The "slightly improved" group (in bold) was used for the MIC calculation in the present study; MIC, minimal important change.

Table IV. Mean change score of the Oxford Shoulder Score according to the pain anchor

Pain anchor	Mean change score (SD)	
	n	OSS
Completely recovered	15	−4.7 (4.2)
Much improved	40	−7.4 (6.8)
Slightly improved	**22**	**−4.7 (6.1)**
Unchanged	39	−1.6 (3.9)
Slightly worse	7	4.6 (4.5)
Much worse	5	8.2 (6.1)
Worse than ever	0	

The "slightly improved" group (in bold) was used for the MIC calculation in the present study; MIC, minimal important change.

Table V. Overview of previous published SDC and MIC values for the SST, DASH, *Quick*DASH, and OSS

PROMs	Study	*n*	SDC	MIC
SST	Tashijan[29]	81	n.m.	2.8
DASH	Schmitt[a,27]	53	14.6[a]	10.2
	Beaton[4]	361	10.7	11.5
	Gummeson[14]	109	n.m	10
***Quick*DASH**	Mintken[ab,19]	101	13.3[ab]	8.2
	Polson[c,25]	35	n.m.	13.1[c]
OSS	-	-	-	-

SDC, smallest detectable change; MIC, minimal important change.
[a]Data recalculated to a 95% interval.
[b]Mintken used the unchanged group at follow-up, test re-test.
[c]Polson used the much-improved group for the MIC calculation, we used the minimally improved group in this table n.m., not mentioned.

DISCUSSION

Monitoring the effects of treatment is of well-recognized importance and is the foundation of modern evidence-based health care. Smallest detectable change (SDC) and minimal import change (MIC) can be used as benchmarks for the interpretability of a PROM to determine whether the observed change is beneficial to the patients. Here, we determined the SDC and MIC of four commonly used shoulder PROMs in a heterogeneous group of shoulder patients. We found an SDC of 2.6 and a MIC of 2.2 for the SST, an SDC of 16.0 and a MIC of 12.4 for the DASH, and an SDC 16.4 and a MIC of 13.4 for the *Quick*DASH. For the OSS, we found an SDC of 6.0 and MIC values of 6.0 and 4.7 for function and pain, respectively. Overall, the SDC was slightly larger than the MIC for all four PROMs.

To determine whether a change score on an individual patient level is clinically important and not just measurement error, the SDC score must not exceed the MIC change score.[33] In our study, all PROMs had an SDC that was slightly larger than the MIC. This means that if an individual patient has a change score as large as the MIC, we cannot be 95% sure that this change is not due to measurement error. Or in other words, the risk of measurement error is larger than 5%. However, as the differences between the SDC and the MIC were rather small, we think that these four PROMs are suitable for use in clinical practice and research.

Although the observed differences between SDC and MIC were very small, it is desirable to find ways to minimize the SDC. One way of decreasing the SDC in a clinical setting is by averaging multiple measurements (i.e., repeated measurements at one point in time) in order to decrease the measurement error. However, this is difficult using questionnaires because it is a burden for patients and there is a high risk of recall bias. It might also be possible to improve the quality of the questionnaires by adding extra questions or improving the wording of questions.

The observed difference between SDC and MIC is less problematic in research because mean scores of groups of patients are used instead of individual patient scores; therefore the measurement error should be calculated for a mean score instead of for a single score. The SDC of a mean score is much smaller (by a factor of the square root of the sample size) than the SDC of a single score.[11]

Table V presents an overview of the previously reported measurement error (SDC) and MIC of the PROMs evaluated in this paper.[4,14,20,26,28,30] Our results for the SST are comparable with the results published by Tashijan et al,[30] who determined the MIC in 81 patients with rotator cuff tears. Although they used a comparable anchor-based mean change score method, they determined the MIC by subtracting the change score of the "unchanged group" from that of the "slightly improved" group. While there is no consensus on whether this subtraction should be performed, Hays et al[16] have argued that if the mean change in the "unchanged" group is 2 points and the mean change in the "slightly improved" group is 4 points, this means that a 2-point change is insufficient and that it takes a greater change of 4 points to constitute a MIC.[16] We agree with Hays et al[16] that the "unchanged" change score should not be subtracted from the "slightly improved" change score.

Our results for the DASH were comparable with results found in the literature. Schmitt et al[28] used the anchor-based mean change method to analyze a heterogeneous group of 53 shoulder patients, and found an SEM of 5.22 and a MIC of 10.2. They used a 90% interval for the SDC calculation. To improve comparability, we recalculated their data to a 95% interval, resulting in an SDC of 14.6. Beaton et al[4] studied a cohort of 361 heterogeneous shoulder patients treated by physiotherapists, using a comparable anchor-based mean change method; they found an SEM 3.9, an SDC of 10.7, and a MIC of 11.5. Gummeson[14] found a MIC of 10 in a comparable study in 109 upper extremity patients.

The results of the *Quick*DASH were also comparable with those in the current literature. Mintken et al[20] analyzed 101 shoulder patients. Using a comparable anchor-based technique, they found a MIC of 8.2. They calculated SDC using the "unchanged group" at follow-up,

which is a suboptimal technique for determining the measurement error because of the risk of bias due to the lack of validity of the anchor.[24] They also used a 90% interval for the SDC calculations; we recalculated the SDC to a 95% interval, resulting in an SDC of 13.3. Polson et al[26] analyzed 35 upper extremity patient with an anchor-based mean change technique. They found a higher MIC of 19 points, most likely because they used the "much improved" group for the MIC calculations instead of the "slightly improved" group as we did in this study. Polson et al[26] also reported the change score of the "slightly improved" group to be 13.1; this information is used in Table V to improve the comparability of our results.

To the best of our knowledge, there is no previous data on SDC and MIC for the Oxford Shoulder Score.[1] One-third of the questions in the OSS are pain related so we used both anchors. We found an SDC of 6.0 points on a scale from 12 to 60. The MIC was 6.0 corresponding to the functional anchor, and 4.7 to the pain anchor.

Strengths of this study are that there were almost no missing data and we had very high response rates at all time-points. This is a clear advantage of web-based questionnaire administration. Furthermore, we included twice the recommended minimal number of patients.

There are several limitations to our study. First, we used a heterogeneous population for calculation of the MIC. There is no evidence in the literature that the MIC differs among (sub) populations, but it has been suggested that this should be evaluated.[18,27] This was not possible in our study because the subgroups would be too small. The advantage of using a heterogenic cohort is that it provides a MIC estimation that can be used in all kind of shoulder disorders. Future studies should examine if and how the MIC varies among subgroups. Second, our patients had to complete three different PROMs at the same time. This could be a response burden to the patient, which might lead to loss of interest during completion. Theoretically, this could result in increased measurement error and a higher SDC. Third, we computed the *Quick*DASH from the full DASH questionnaire. This is not the same as completing the *Quick*DASH questionnaire independently. Fourth, although anchor-based techniques are considered the best method for assessing the MIC,[27] there is debate in the literature about the validity of anchors and the best statistical approach for calculating the MIC.[32] For example, a disadvantage of the mean change method is that it uses only the average change score of one patient subgroup for the MIC calculation, meaning that only 23 patients determined the MIC value in this study. For these methodological reasons, it has been recommended that the MIC of PROMs should be determined in multiple studies.[36] Our study therefore contributes to a better understanding of the change scores of PROMs in shoulder patients.

CONCLUSION

This study shows that on an individual patient-based level, when taking into account the SDC and MIC, the change score should be above 2.6 points for the SST, above 16.0 points for the DASH, above 16.4 points for the *Quick*DASH , and above 6.0 points for the OSS to show a clinical relevant change that is not due to measurement error.

REFERENCES

1. Angst F, Schwyzer H-K, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and Its Short Version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society Standardized Shoulder. Arthritis Care Res. 2011 Nov 7;63(S11):S174–S188. doi:10.1002/acr.20630
2. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg. 1998 Nov;7(6):565–572.
3. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. J Hand Ther. 2001 Apr;14(2):128–146.
4. Beaton DE, van Eerd D, Smith P, van der Velde G, Cullen K, Kennedy CA, et al. Minimal change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. J Clin Epidemiol. 2011 May;64(5):487–496. doi:10.1016/j.jclinepi.2010.07.012
5. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005 May;87(5):1038–1046. doi:10.2106/JBJS.D.02060
6. Berendes T, Pilot P, Willems WJ, Verburg H, Slaa te R. Validation of the Dutch version of the Oxford Shoulder Score. J Shoulder Elbow Surg. 2010 Sep;19(6):829–836. doi:10.1016/j.jse.2010.01.017
7. Cloke DJ, Lynn SE, Watson H, Steen IN, Purdy S, Williams JR. A comparison of functional, patient-based scores in subacromial impingement. J Shoulder Elbow Surg. 2005 Jul;14(4):380–384. doi:10.1016/j.jse.2004.08.008
8. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003 May;56(5):395–407.
9. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br. 1996 Jul;78(4):593–600.
10. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006;4:54. doi:10.1186/1477-7525-4-54
11. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006 Oct;59(10):1033–1039. doi:10.1016/j.jclinepi.2005.10.015
12. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. 2011.
13. Desai AS, Dramis A, Hearnden AJ. Critical appraisal of subjective outcome measures used in the assessment of shoulder disability. Ann R Coll Surg Engl. 2010 Jan;92(1):9–13. doi:10.1308/003588410X12518836440522
14. Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. BMC Musculoskelet Disord. 2003 Jun 16;4:11. doi:10.1186/1471-2474-4-11
15. Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. BMC Musculoskelet Disord. 2006;7:44. doi:10.1186/1471-2474-7-44
16. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. COPD. 2005 Mar;2(1):63–67. doi:10.1081/COPD-200050663

17. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). The Upper Extremity Collaborative Group (UECG). Am. J. Ind. Med. 1996 Jun;29(6):602–608.
18. King M. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res. 2011 Apr;11(2):171–184. doi:10.1586/erp.11.9
19. Lippitt SB, Matsen FA. A practical tool for evaluating function: The simple shoulder test. In The Shoulder: A Balance of Mobility and Stability. American Academy of Orthopaedic Surgeons. 1993.
20. Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. J Shoulder Elbow Surg. 2009 Nov;18(6):920–926. doi:10.1016/j.jse.2008.12.015
21. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol. 2006;6:2. doi:10.1186/1471-2288-6-2
22. 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. Qual Life Res. 2010 May;19(4):539–549. doi:10.1007/s11136-010-9606-8
23. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul;63(7):737–745. doi:10.1016/j.jclinepi.2010.02.006
24. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol. 1997 Aug;50(8):869–879.
25. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health. 2007 Nov;10 Suppl 2:S125–37. doi:10.1111/j.1524-4733.2007.00275.x
26. Polson K, Reid D, McNair PJ, Larmer P. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. Man Ther. 2010 Aug;15(4):404–407. doi:10.1016/j.math.2010.03.008
27. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008 Feb;61(2):102–109. doi:10.1016/j.jclinepi.2007.03.012
28. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004 Oct;57(10):1008–1018. doi:10.1016/j.jclinepi.2004.02.007
29. Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. Health Qual Life Outcomes. 2006;4:62. doi:10.1186/1477-7525-4-62
30. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. The Journal of Bone and Joint Surgery. 2010 Feb 1;92(2):296–303. doi:10.2106/JBJS.H.01296
31. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012 May;21(4):651–657. doi:10.1007/s11136-011-9960-1

32. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010 May;63(5):524–534. doi:10.1016/j.jclinepi.2009.08.010
33. Terwee CB, Roorda LD, Knol DL, De Boer MR, de Vet HCW. Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol. 2009 Oct;62(10):1062–1067. doi:10.1016/j.jclinepi.2008.10.011
34. van Kampen DA, van Beers LWAH, Scholtes VAB, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg. 2011 Dec 23;doi:10.1016/j.jse.2011.09.026
35. Veehof MM, Sleegers EJA, van Veldhoven NHMJ, Schuurman AH, van Meeteren NLU. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). J Hand Ther. 2002 Oct;15(4):347–354. doi:10.1016/S0894-1130(02)80006-0
36. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011 May;64(5):507–516. doi:10.1016/j.jclinepi.2010.11.018
37. National Ambulatory Medical Care Survey 1998-2006. Data obtained from: U.S. Department of Health and Human Services; Centers for Disease Control and Prevention; National Center for Health Statistics.

Chapter 4

The diagnostic value of the combination of patient characteristics, history, and clinical tests for the diagnosis rotator cuff tear

DA van Kampen
T van den Berg
HJ van der Woude
RM Castelein
CB Terwee
WJ Willems

Submitted for journal publication

ABSTRACT

Purpose: The purpose of this study was to improve the diagnostic value for rotator cuff tears, by combing patient characteristics, history, and clinical test results.

Methods: This prospective cohort included 169 patients with shoulder complaints that visited an orthopaedic outpatient clinic. One experienced clinician conducted 26 clinical tests of which nine are specifically for rotator cuff pathology (empty can, Neer, Hawkins-Kenney, drop arm, lift off test, pain full arc, external rotation lag sign, dropsign, infraspinatus muscle strength test). The final diagnosis, based on magnetic resonance arthrography (MRA), was determined by consensus between the clinician and a radiologist, who were blinded to patient information. A prediction model was developed by logistic regression analysis, where eight test scores were considered potential predictors. The model was internally validated and shrinkage was applied to correct for overoptimism.

Results: Out of 169 patients evaluated, 43 had rotator cuff tears. The individual overall accuracy of the rotator cuff clinical tests was 68-82%. After backward selection, the model determined that the most important predictors of rotator cuff tears were higher age, no previous shoulder dislocation and a positive Neer test. This internally-validated prediction model had good discriminative ability (area under the curve, 0.81).

Conclusion: Our results showed that individual clinical shoulder tests had moderate diagnostic value for patients with shoulder complaints. Our prediction model showed improved diagnostic value for diagnosing rotator cuff tears.

INTRODUCTION

Shoulder disorders rank among the most prevalent musculoskeletal disorders, together with back and knee pain. These disorders can be caused by many different pathologies, each requiring their own specific surgical or non-surgical treatment. Diagnostically, the shoulder is one of the most complex joints, due to its multiple directional movements. Furthermore, direct observation of shoulder motion is obscured by the muscles.[19] Rotator cuff (RC)-related disorders are among the most important causes for visiting the orthopaedic outpatient clinic. The correct diagnosis is essential for selecting the appropriate treatment plan. Several shoulder specific clinical tests have been developed for diagnosing RC tears. However, a recent meta-analysis showed that data was lacking to support most clinical tests used for diagnosing RC tears; moreover, there is a need for high quality studies to test the diagnostic performance of parameters from patient history and physical examinations.[12,13] It is difficult to diagnose RC tears based purely on patient history and physical examination; therefore, the use of other techniques for establishing the diagnosis has increased, including magnetic resonance (MR) imaging, ultrasound scans, and diagnostic arthroscopy. However, these tests are time-consuming, expensive, and/or invasive; thus, they should be restricted as much as possible.

Murrel et al and Park et al reported that the combination of individual clinical test results and patient age could lead to improved diagnostic value for diagnosing a RC tear.[21,23] However, in those studies, the reference standard (arthroscopy) was not performed in all patients, which could lead to a verification bias.[5] In the ideal research world, an arthroscopy could be used as a reference standard to determine the diagnosis for every new patient with a shoulder complaint in the outpatient clinic. However, because it is invasive, this approach is not ethically justified. Therefore, in our view, the reference standard should be a MR arthrography (MRA).

It would be very useful to have a prediction model, which combined patient characteristics, history, and results from a few clinical tests, for predicting the probability of a RC tear in individual patients. For example, the Ottawa Ankle Rules comprise one of the most famous prediction models presently used in orthopaedic surgery.[32]

The first purpose of the present study was to estimate the diagnostic performance of individual clinical tests for diagnosing rotor cuff tears; thus, we compared the performance of each test to that of the MRA, as a reference standard. The second purpose was to develop a prediction model, which combined patient characteristics, history, and clinical test scores, for predicting the diagnosis of a RC tear. We hypothesized that the combined use of patient characteristics, history, and clinical tests will improve the diagnostic value for RC tears.

METHODS

Patients

We performed a prospective cohort study of new patients with acute and chronic shoulder complaints, recruited consecutively between February 2009 and June 2012 at the orthopaedic outpatient clinic of the Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands. Exclusion criteria were: previously diagnosed shoulder disorders, fractures, frozen shoulder, arthritis, and deficiencies in reading and understanding the Dutch language.

Institutional approval was obtained by our local ethics committee, and written, signed, informed consent was obtained from all participants.

Data collection

One experienced orthopaedic surgeon (WJW) performed 26 shoulder-specific clinical tests on all patients, according to a standardised diagnostic protocol based on the original descriptions of the clinical tests. Nine of these clinical tests are regarded as specific for the rotator cuff and were selected for this study. The examiner was blinded to the imaging analyses. Within a few days after the initial visit, all patients were asked to complete an online (web-based) questionnaire to evaluate patient history. Patients were asked about previous shoulder dislocations and we used validated Patient Reported Outcome Measures (PROMs) to standardise the history questions.[24,28] Two PROMS were administered: the Simple Shoulder Test and the Oxford Shoulder Score. All patients underwent MRA to obtain a reference standard.

Clinical examination tests for rotator cuff tears

The empty can test, also known as the Jobe test, was performed with the patient standing, the shoulder in 90° abduction in the scapular plane, and with full internal rotation.[16] The thumbs were pointing toward the floor. The patient maintained this position against downward resistance applied by the examiner. The test was considered positive when the patient demonstrated weakness or pain during the applied resistance.

The Neer test was performed with the patient sitting or standing.[22] The ipsilateral scapula was stabilised with the examiner's hand, and the patient's arm was passively elevated forward. The test was considered positive when the patient experienced pain. In the original description, Neer advised giving an injection of lidocaine in the subacromial space to relieve pain. Due to time limitations in the orthopaedic outpatient setting, we decided not to give patients a lidocaine injection. This was comparable with common practice, and it was consistent with the study by Park et al.[23]

The Hawkins Kennedy test was performed with the examiner facing the seated or standing patient.[11] The patient's arm was elevated forward at 90°, and the elbow was flexed at 90°. The test was considered positive when pain occurred with passive internal rotation.

The drop arm test, also known as Codman's sign, was performed with the patient standing.[6] The patient was asked to abduct the arm fully, and then, to reverse the motion slowly, in the same arc. When the arm dropped suddenly, the test was considered positive.

The lift off test, also known as the Gerber test, was performed with the patient standing.[9] The patient was asked to place their hand on their back for maximum internal rotation, and then, to lift their hand off their back. The test was considered positive when the patient was not able perform this.

The painful arc test was performed with the patient standing.[17] The patient was asked to elevate the arm actively in the scapular plane, until the arm was fully elevated, and then, to let the arm down in the same arc. The test was considered positive when the patient demonstrated pain, or reported a painful catching between 60° and 120° elevation.

The external rotation lag sign was performed with the patient seated.[15] The elbow was passively flexed to 90°, and the examiner held the shoulder at 20° elevation (in the scapular plane), near maximal external rotation (i.e., maximum external rotation minus 5, to avoid elastic recoil in the shoulder). The patient was then asked to maintain the external rotation in elevation as the examiner released the wrist, but maintained support of the limb at the elbow. The sign was considered positive when a lag, or angular drop occurred.

The dropsign, also known as the infraspinatus drop sign, was similar to the ERLS, but the arm was held at 90° elevation (in the scapular plane) by the examiner, instead of the 20° elevation.[15]

The infraspinatus muscle strength test was performed with the patient standing or sitting.[23] The elbow was flexed at 90°, and the arm adducted to the trunk in neutral rotation. The examiner applied an internal rotation force to the arm while the patient resisted. The test was considered positive when the patient demonstrated weakness compared to the other side.

Imaging technique, MRA

MRA was performed after intra-articular administration of 10 mL omnipaque 300 (300 mg I/ml iohexol; GE Health- care BV, Eindhoven) and a 10 mL mixture of 0.5 mL omniscan (0.5 mmol/ml Gd-DTPA-BMA; GE Healthcare BV) added to 100 mL saline (0.9%). This solution was administered (12– 15 cc) by inserting an 18-gauge needle in the glenohumeral joint, with fluoroscopic guidance, from either an anterior or a posterior approach. MR images were acquired within 30 min after injection; patients were instructed to immobilise the shoulder of interest between the injection and the MR examination. Imaging was performed with either a 1.0 T unit (MR Systems NT Release 4.5; Philips Medical Systems, Best, The Netherlands) or a 1.5 T unit (MR Systems Intera, Release 9.0, Philips Medical Systems). The following sequences were performed: T1-weighted fast spin echo (FSE) with fat-selective presaturation in an axial

plane, oblique coronal, and oblique sagittal plane; oblique coronal proton-density (PD); and T2-weighted FSE and T1-weighted FSE with fat-selective presaturation with the shoulder in the abduction-external rotation (ABER) position.

Reference standard

The diagnosis based on the MRA was defined as the reference standard. The MRAs were reviewed in random order, and the evaluators were blinded to the patient's personal details, clinical history, and symptoms. The final diagnosis was made in consensus by the orthopaedic surgeon (W.J.W.) and a musculoskeletal radiologist (H.J.W.). Both had more than 15 years experience in evaluating shoulder MRAs. In the case of no consensus, a second experienced musculoskeletal radiologist was available to make the final diagnosis. We chose a consensus diagnosis for the MRA, because previous studies have shown inter-observer variability for detecting full thickness and partial tears in the RC.[29,30] All potential diagnoses for shoulder complaints were made in accordance with standard radiologic criteria.[26]

Specifically, for RC tears, we used the following criteria. A complete (full-thickness) tear, with or without retraction of tendon edges, was identified as a gap, with hyperintense fluid signal intensity equal to water on a T2 FSE, with or without fat suppression, that extended from the articular space to the subacromial space and/or a hyperintense signal intensity on T1-weighted, fat-suppressed, MRA images in the various planes. An incomplete, or partial, tear was identified as an incomplete tendon defect, either on the bursal side or the articular side, with a hyperintense fluid signal intensity that extended within, but did not traverse, the tendon. Both partial and full thickness tears were considered RC tears.

PROMs

The Simple Shoulder test

The Simple Shoulder test (SST) measures functional limitations in patients with shoulder complaints.[18] It consists of 12 questions with dichotomous response options. Scores were summarised to a total score, which ranged from 0 (worst) to 12 (excellent). The SST has been validated in Dutch patients with shoulder complaints.[33]

Oxford Shoulder Score

The Oxford Shoulder Score (OSS) measures functional limitations in patients with shoulder complaints.[7] It contains 12 items with 5 response options. Scores were summarised to a total score, which ranged from 12 (excellent) to 60 (worst). The OSS has been validated in Dutch patients with shoulder complaints.[4]

Statistical analysis

All patients with any type of RC tear were allocated to the RC tear group. Patients that had other diagnoses, in addition to the RC tear remained in the RC tear group. All patients without a RC tear, based on the MRA, were allocated to the No-RC tear group, independent of their final diagnosis. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio, negative likelihood ratio, and overall accuracy were calculated with a 2x2 table. We defined overall accuracy as the percentage of true results (both true positives and true negatives).

We used logistic regression to develop the prediction model. Based on clinical experience and Hegedus et al,[13] we chose eight candidate predictors prior to the data analysis, including age, previous shoulder dislocation, night pain (OSS-question 4 was dichotomised to no night pain versus any night pain), weakness (SST-question 6), and four clinical tests (Empty can, Neer, ERLS, and the Hawkins-Kennedy test).[1] Strong correlations between predictors were investigated during the modelling process to make sure multicollinearity was not troublesome. A logistic regression model was used to select relevant predictors and to estimate the regression coefficients. This was performed with a backward selection strategy in the full, eight-predictor model. Predictors were deleted step by step from the model based on the highest p-value, until a stopping rule was reached, based on Akaike's Information Criterion.[1] The final model consisted only of predictors with a p-value below 0.157.

We assessed the diagnostic performance of our model by determining calibration and discrimination.[27] Calibration referred to the agreement between observed and predicted outcomes. The Hosmer and Lemeshow "goodness-of-fit" test indicated whether the model was a good fit to the data. The discriminative ability of the prediction model was assessed by the area under the receiver operating characteristic curve (AUC) or the equivalent c (concordance) index.[10] The receiver operator curve (ROC) is a graphical presentation of the true positives (sensitivity) vs. false positives (1 – specificity) performance of the prediction model. The AUC of ROC curve is a way to reduce ROC performance to a single value representing expected performance. The AUC has a value between 0.5 (no discriminative ability) and 1.0 (perfect discriminative ability).

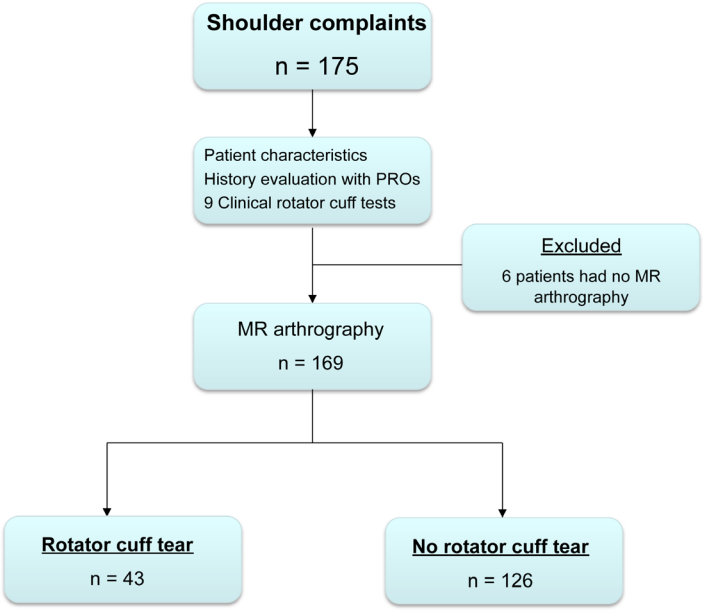
A prediction model was fitted to the dataset at hand, and therefore, it was prone to overoptimism in new patients. Overoptimism is particularly common in small datasets, where the number of (starting) predictors is large compared to the smallest outcome group. Internal validation can correct for some of this overfitting with the bootstrapping method.[2,27] In this case, 500 bootstraps were performed and a shrinkage factor was calculated to penalise the regression coefficients. For guidance on this protocol, see Steyerberg et al.[31] Analyses were performed with R version 2.14.2 (http://www.R-project.org).

RESULTS

Patients

The inclusion protocol (Figure I) shows that 174 new patients were included. One patient had complaints in both shoulders; thus, 175 shoulders were included. All patients completed the 26 clinical tests of which nine, specifically for the rotator cuff, were evaluated in this study. Six patients were lost to follow up: one patient went to another hospital, due to the long waiting list; five patients declined the MRA. Thus, 169 (97%) shoulders were analysed. Twelve (7%) patients did not complete the online questionnaire, but they were included, because previous shoulder dislocation data had been documented in the patient files. The patient demographic characteristics are presented in Table I. The average time between the clinical tests and the MRA was 38 days. In two cases, there was no consensus between the two reviewers, and the second radiologist made the final decision on the diagnosis. There were no adverse events related to the MRA. Forty-three patients were diagnosed with a RC tear (Table II). Among patients in the RC tear group, some were also diagnosed with traumatic anterior shoulder instability (N=2), biceps pathology (N=7), and superior labrum tears from anterior to posterior (SLAP; N=4). In the no-RC tear group (N=126), patients were diagnosed with traumatic anterior shoulder instability (N=58), Biceps pathology (N=3), SLAP (N=12), and impingement syndromes (N=9). Thus, some patients had multiple diagnoses. In 55 patients, we could not find an explanation for the shoulder complaints based on the MRA.

Figure I. Flow chart of the study population



n: the number of shoulders evaluated

Table I. Demographic characteristics of patients with shoulder complaints

		Baseline
Number of shoulders		169
Mean age, years (SD)		38 (14)
Gender	M	116 (69%)
	F	53 (31%)
Side	R	103 (61%)
	L	64 (38%)
	B	1 (1%)
Simple Shoulder Test	[0-12], mean (SD)*	8.5 (2.8)
Oxford Shoulder score	[12-60], mean (SD)+	24.1 (7.3)

*Abbreviations: M=male, F= female, L=left, R=right, B= both. * Higher score is considered a better functioning of the shoulder. + Lower score is considered a better functioning of the shoulder.*

Table II. The different types of rotator cuff tears

Tendon		N
Supraspinatus	Partial tear	22
	Full thickness	20
Infraspinatus	Partial tear	2*
	Full thickness tear	7*
Subscapularis	Partial tear	3*
	Full thickness tear	3**

**all these patients also had a supraspinatus tear,*

***2 patients also had a supraspinatus tear, 1 patient had isolated tears (traumatic)*

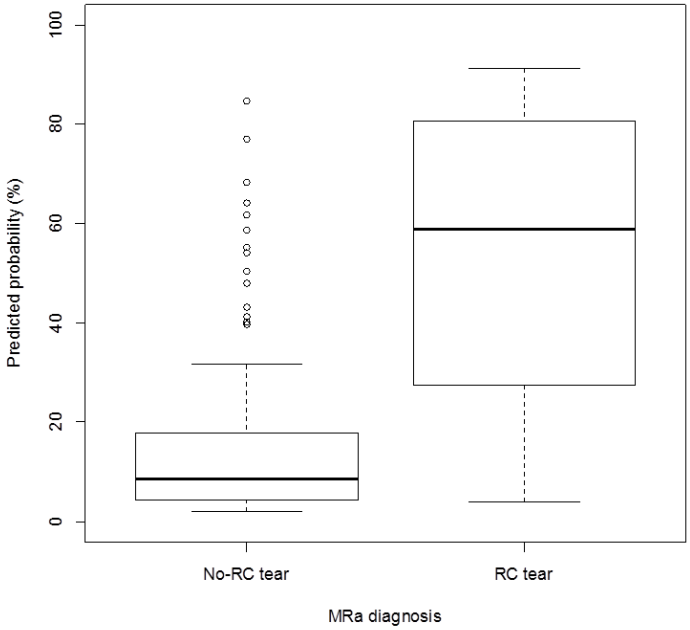
Individual clinical shoulder tests

The diagnostic results of the eight individual clinical tests are presented in Table III. The Empty can test was the most sensitive (65.1%), the Drop arm test had the highest specificity (100%), and the Neer test had the best overall accuracy (82%).

Prediction model

Of the eight preselected candidate predictors, age, previous shoulder dislocation, and the Neer test remained in the model as independent predictors. The combination of clinical tests did not provide additional diagnostic value. Table IV shows a simplified score chart that illustrates predictions from the final model after internal validation. The discriminative ability (AUC) of the model was 0.81. The 'Hosmer-Lemeshow goodness of fit' test was not significant, which indicated a good fit of the model to the data. In Figure II, the patients were grouped according to the reference standard results into the RC tear group or the No-RC tear group; then, we plotted the estimated probabilities of a RC tear, according to the prediction model. According to our prediction model, the median probability of a RC tear was 60% in the RC tear group and 9% in the No-RC tear group.

Figure II. Rotator cuff (RC) tears or no-RC tears were diagnosed according to magnetic resonance arthrography and compared to the predicted probabilities of a RC tear, based on the prediction model.



The X-axis represents the patients diagnosed with a RC tear or with no RC tear, according to magnetic resonance arthrography. The Y-axis represents the predicted probabilities of a RC tear according to the prediction model. The box represents graphically 50% of the patients; the heavy line inside the box is the median.

Table III. Evaluations of diagnostic clinical tests; values represent percentages (%)

	Sensitivity	Specificity	PPV	NPV	LR (+)	LR (-)	Overall accuracy
Empty can	65.1	73.0	45.2	86.0	2.41	0.48	71.0
Neer	58.1	90.5	67.6	86.4	6.10	0.46	82.2
Hawkins Kennedy	48.8	85.7	53.8	83.1	3.42	0.60	76.3
Drop arm	4.7	100.0	100.0	75.4	∞	0.95	75.7
Lift off test	11.6	99.2	83.3	76.7	14.6	0.89	76.9
Painful arc	39.5	92.1	63.0	81.7	4.98	0.66	78.7
ERLS	11.6	98.4	71.4	76.5	7.32	0.90	76.3
Dropsign	11.6	99.2	83.3	76.7	16.7	0.89	76.9
ISMST	14.0	98.4	75.0	77.2	8.86	0.87	77.1

PPV: positive predictive value, NPV: negative predictive value, LR+: likelihood ratio (for positive test), LR-: likelihood ratio (for negative test), ERLS: external rotation lag sign, ISMST: infraspinatus muscle strength test

Table IV. Estimations of the probability of a RC tear, based on patient age (years), previous shoulder dislocation, and the Neer test result, according to the validated prediction model.

	No dislocation		Dislocation	
Age group	- Neer	+ Neer	- Neer	+ Neer
20	9%	26%	5%	14%
30	14%	36%	7%	22%
40	21%	48%	11%	31%
50	31%	61%	18%	43%
60	42%	72%	27%	55%
70	67%	81%	37%	67%

The diagnostic odds ratios of the model were: older age (per 10 years): 1.63; no previous shoulder dislocation: 2.05; positive Neer test: 3.43

DISCUSSION

We assessed the diagnostic value of individual clinical shoulder tests for RC tears, and we developed a prediction model that combined patient characteristics, history, and clinical test scores to improve the overall diagnostic value for RC tears. We found that the individual clinical tests had moderate sensitivity and specificity, and overall accuracies that ranged from 68 to 82%. No single test had a good discriminative value. Our hypothesis was confirmed, the prediction model, which included higher age, no previous shoulder dislocation, and a positive Neer test, clearly improved the diagnostic value to detect a RC tear.

A few clinical tests have a very high specificity (Table III); the drop arm, lift off test, external rotation lag sign, and infraspinatus muscle strength test, which was also found by Bak et al.[3] This could suggest that these are very useful clinical tests. However, because of the low incidence of a positive test result [2-7 times], they are less useful as a general screening tool for RC tears.

Our prediction model performed well and had a good discriminative ability (AUC 0.82). Table IV aims to serve as a simplified score chart that illustrates estimations of the probability that patient will have a RC tear. After external validation, such a score could potentially be useful in deciding whether a MRA would provide added value for diagnosing rotator cuff tear.

Our prediction model has potential for being implemented in clinical practice, because it contains only three clear prediction factors. The patient's age and previous shoulder dislocation are simply determined, and the Neer test is one of the easiest clinical tests to perform. Even without the use of the lidocaine injection in the subacromial space, as originally described by Neer, the test performed very well in our prediction model. Henkus et al showed that it was difficult to place the injection exactly in the subacromial space; thus, they considered the Neer test in combination with the injection a poor diagnostic tool.[14]

For the development of a prediction model, it is recommended that, for each potential important predictor studied, at least 5 events (in this study: patients with a RC tear) are required to avoid the risk that overestimation might become problematic.[20,34] Therefore, we assessed the prognostic value of eight pre-determined potentially important predictors. The fact that our prediction model showed a strong effect of age on the probability of a RC tear was consistent with findings in the literature. RC tears have been described as a degenerative condition that increases linearly with age. [3,25] This prediction model was developed to be used for every patient that presents at the orthopaedic outpatient clinic with shoulder complaints. This explains why a previous shoulder dislocation is a strong predictor against a RC tear, as most of these patients are diagnosed with anterior shoulder instability instead of a RC tear.

Most previous studies that evaluated clinical tests for RC tears used arthroscopy as the reference standard. That type of study design can induce verification bias, because typically, only patients with a surgical indication were tested with the reference standard.[5] A recent meta analysis showed that MRA was the most sensitive and specific technique for diagnosing both full- and partial-thickness RC tears, compared to native MR imaging or ultrasound scans.[8] Therefore, we chose the MRA as our reference standard and performed MRAs on all patients.

Consistent with the studies of Murrel and Park, we found that the combination of age and a clinical test improved the diagnostic value for rotator cuff tears.[21,23] In contrast to their results, and in agreement with the results of Bak et al, we did not find that the combination of multiple clinical tests improved the diagnostic value.[3] Our study provided additional value compared to the mentioned studies for several reasons. First, we used a rigorous study design; we attempted to replicate clinical practice by combining data on patient characteristics, history, and clinical tests in our prediction model. Second, we included every patient with a shoulder complaint that visited the outpatient clinic, and we confirmed the diagnosis with MRA as the reference standard; this strategy prevented a verification bias.[5] Third, the facts that diagnoses were made by individuals blinded to patient information and decisions were made by consensus ensured that this study was reproducible. Finally, we used state of the art methodology to develop our prediction model, and we internally validated it in our dataset.

Our study also had limitations. First, we did not investigate the inter-examiner reliability of the physical examinations. However, a previous study by Young et al reported good inter-observer agreement for clinical tests applied to diagnose RC tears.[35] Second, the examiner was not blinded to the patient information and history during the clinical tests; this could have been a source of bias. However, it is questionable whether the diagnostic value of the clinical tests would be valid in conditions where the examiner lacked prior information, because this scenario never occurs in clinical practice. Third, we did not evaluate every clinical rotator cuff test published; therefore, it is possible that other clinical tests, which we did not include, might also be good predictors. Fourth, although we included 175 patients, only 43 had RC

tears. This small sample limited our analyses by preventing the inclusion of more potential important predictors. Moreover, we may have included a noise variable in our final prediction model. However, the internal validation procedure showed that the predictors used in our study were robust, and the shrinkage factor was fairly high. In a larger study sample, it would be interesting to do subgroup analyses to differentiate among the different tendons of the RC.

Before implementing our prediction model, it must first be validated with a new cohort of patients (external validation). It is also important to stress that our prediction model was developed for orthopaedic outpatient clinic patients; therefore, it may not be generalizable to primary care. The incidence of anatomical abnormalities is much higher for patients with orthopaedic complaints than for patients examined in primary care; therefore, the probability of finding a RC tear is much higher in an orthopaedic outpatient clinic.

Conclusion

This prospective cohort study showed that individual clinical shoulder tests had moderate diagnostic value for the diagnosis of RC tears. Our prediction model, which combined age, previous shoulder dislocation, and the Neer test, improved the diagnostic value for diagnosing rotator cuff tears.

REFERENCES

1. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. Stat Med. 2002;21(24):3803–3822. doi:10.1002/sim.1422
2. Austin PC, Tu JV. Bootstrap Methods for Developing Predictive Models. The American Statistician. 2004 May;58(2):131–137. doi:10.1198/0003130043277
3. Bak K, Sørensen AKB, Jørgensen U, Nygaard M, Krarup AL, Thune C, et al. The Value of Clinical Tests in Acute Full-Thickness Tears of the Supraspinatus Tendon: Does a Subacromial Lidocaine Injection Help in the Clinical Diagnosis? A Prospective Study. YJARS. 2010 Jun 1;26(6):734–742. doi:10.1016/j.arthro.2009.11.005
4. Berendes T, Pilot P, Willems WJ, Verburg H, Slaa te R. Validation of the Dutch version of the Oxford Shoulder Score. J Shoulder Elbow Surg. 2010 Sep;19(6):829–836. doi:10.1016/j.jse.2010.01.017
5. Bhandari M, Montori VM, Swiontkowski MF, Guyatt GH. User's guide to the surgical literature: how to use an article about a diagnostic test. J Bone Joint Surg Am. 2003;85(6):1133–1140.
6. Codman EA. The shoulder. Rupture of the supraspinatus tendon and other lesions in or about the subacromial bursa. Brooklyn, New York, USA: G Miller & Co Medical Publishers; 1934.
7. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br. 1996 Jul;78(4):593–600.
8. de Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. American Journal of Roentgenology. 2009 Jun; 192(6):1701–1707. doi:10.2214/AJR.08.1241
9. Gerber C, Krushell RJ. Isolated rupture of the tendon of the subscapularis muscle. Clinical features in 16 cases. J Bone Joint Surg Br. 1991 May;73(3):389–394.
10. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996 Feb 28;15(4):361–387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
11. Hawkins RJ, Kennedy JC. Impingement syndrome in athletes. Am J Sports Med. 1980 May;8(3):151–158.
12. Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, et al. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. 2007 Jun 4;42(2):80–92. doi:10.1136/bjsm.2007.038406
13. Hegedus EJ. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med. 2012 Jul 7;doi:10.1136/bjsports-2012-091066
14. Henkus H-E, Cobben LPJ, Coerkamp EG, Nelissen RGHH, van Arkel ERA. The accuracy of subacromial injections: a prospective randomized magnetic resonance imaging study. Arthroscopy. 2006 Mar; 22(3):277–282. doi:10.1016/j.arthro.2005.12.019
15. Hertel R, Ballmer FT, Lombert SM, Gerber C. Lag signs in the diagnosis of rotator cuff rupture. J Shoulder Elbow Surg. 1996 Jul;5(4):307–313.
16. Jobe FW, Jobe CM. Painful athletic injuries of the shoulder. Clin. Orthop. Relat. Res. 1983 Mar;(173): 117–124.
17. Kessel L, Watson M. The painful arc syndrome. Clinical classification as a guide to management. J Bone Joint Surg Br. 1977 May;59(2):166–172.
18. Lippitt SB, Matsen FA. A practical tool for evaluating function: The simple shoulder test. In The Shoulder: A Balance of Mobility and Stability. American Academy of Orthopaedic Surgeons. 1993.
19. McFarland EG, Garzon-Muvdi J, Jia X, Desai P, Petersen SA. Clinical and diagnostic tests for shoulder disorders: a critical review. Br J Sports Med. 2010 Apr;44(5):328–332. doi:10.1136/bjsm.2009.067314

20. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009 Feb 23;338(feb23 1):b375–b375. doi:10.1136/bmj.b375

21. Murrell GA, Walton JR. Diagnosis of rotator cuff tears. Lancet. 2001 Mar 10;357(9258):769–770. doi: 10.1016/S0140-6736(00)04161-1

22. Neer CS. Impingement lesions. Clin. Orthop. Relat. Res. 1983 Mar;(173):70–77.

23. Park HB, Yokota A, Gill HS, Rassi El G, McFarland EG. Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. J Bone Joint Surg Am. 2005 Jul;87(7):1446–1455. doi:10.2106/JBJS.D.02335

24. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health. 2007 Nov;10 Suppl 2:S125–37. doi:10.1111/j.1524-4733.2007.00275.x

25. Reilly P, Macleod I, Macfarlane R, Windley J, Emery RJH. Dead men and radiologists don't lie: a review of cadaveric and radiological studies of rotator cuff tear prevalence. Ann R Coll Surg Engl. 2006 Mar; 88(2):116–121. doi:10.1308/003588406X94968

26. Resnick D. Diagnosis of bone and joint disorders. W B Saunders Co; 2002.

27. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009 Mar 31;338:b604. doi:10.1136/bmj.b604

28. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011 Mar;42(3):236–240. doi:10.1016/j.injury.2010.11.042

29. Schreinemachers SA, Hulst VPM, Willems WJ, Bipat S, Woude H-J. Detection of partial-thickness supraspinatus tendon tears: is a single direct MR arthrography series in ABER position as accurate as conventional MR arthrography? Skeletal Radiol. 2009 Mar 18;38(10):967–975. doi:10.1007/s00256-009-0680-3

30. Slabaugh MA, Friel NA, Karas V, Romeo AA, Verma NN, Cole BJ. Interobserver and intraobserver reliability of the goutallier classification using magnetic resonance imaging: proposal of a simplified classification system to increase reliability. Am J Sports Med. 2012 Aug;40(8):1728–1734. doi: 10.1177/0363546512452714

31. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999 Oct;52(10):935–942.

32. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med. 1992 Apr; 21(4):384–390.

33. van Kampen DA, van Beers LWAH, Scholtes VAB, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg. 2012 Jun;21(6):808–814. doi:10.1016/j.jse.2011.09.026

34. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am. J. Epidemiol. 2007 Mar 15;165(6):710–718. doi:10.1093/aje/kwk052

35. Young C, Nanda R, Liow R, Rangan A. Diagnostic accuracy of clinical test tests for the different degrees of subacromial impigment syndrome in rotator cuff disease. J Bone Joint Surg Br. 2003 Jan 27;85 Suppl 1:69.

Chapter 5

The diagnostic value of the combination of patient characteristics, history and clinical tests for traumatic anterior shoulder instability

DA van Kampen
T van den Berg
HJ van der Woude
RM Castelein
CB Terwee
WJ Willems

Accepted in the Journal of Shoulder and Elbow Surgery

ABSTRACT

Background: It is unknown what combination of patient information and clinical tests might be optimal for the diagnosis of traumatic anterior shoulder instability. This study aimed to determine the diagnostic value of individual clinical tests and to develop a prediction model that combined patient characteristics, history, and clinical tests for diagnosing traumatic anterior shoulder instability.

Material and methods: This prospective cohort study included 169 consecutive patients with shoulder complaints that were examined at an orthopaedic outpatient clinic. One experienced clinician conducted 26 clinical tests; of these, six were specific for testing traumatic anterior shoulder instability (apprehension, relocation, release, anterior drawer, load and shift, and hyperabduction tests). Magnetic resonance arthrography (MRA) was used to determine the final diagnosis. A prediction model was developed by logistic regression analysis.

Results: In this cohort, 60 patients (36%) were diagnosed with anterior shoulder instability based on the MRA. The overall accuracy of individual clinical tests was 80.5-86.4%. Age, previous shoulder dislocation, sudden onset of complaints, and the release test were important predictors for the diagnosis of traumatic anterior shoulder instability. The prediction model demonstrated high discriminative ability (AUC 0.95).

Conclusion: Individual clinical shoulder tests provide good diagnostic accuracy. Young age, a history of shoulder dislocation, sudden onset of complaints, and a positive release test were the most important predictors for traumatic anterior shoulder instability.

INTRODUCTION

Shoulder dislocation is a common problem. The incidence in the Netherlands is estimated to be approximately 38 cases per 100 000 persons per year.[33] In most cases, the shoulder dislocates anteriorly.[16,19,33] After an initial dislocation, the labral structures often tear loose from the bony glenoid, and reinsert more inferiomedially. The capsular ligaments stretch, which widens the joint capsule, and this often results in recurrent dislocations. Initial treatment is controversial; there is no consensus on whether a primary shoulder dislocation should be treated surgically or conservatively. Particularly in young patients, an acute dislocation may ultimately lead to chronic impairment.[28] Symptoms may range from a vague uncomfortable sensation to multiple shoulder dislocations. Due to these complaints, patients may be limited in sports and daily activities.

A careful history and physical examination of the shoulder is the cornerstone of diagnosing traumatic anterior shoulder instability. Sometimes the diagnosis is not clear, due to a non-specific medical history and physical examination. Several shoulder complaints may arise from different, shoulder-related disorders. Depending on the specific diagnosis, there are different surgical or non-surgical treatment options.

There are numerous clinical tests for traumatic anterior shoulder instability. The most commonly used clinical tests are the apprehension, relocation, and release tests. They all have fair to good diagnostic accuracy.[11,15,18,26] Previous studies only focused on the diagnostic value of individual clinical tests; however, in clinical practice, patient characteristics and history are also very important factors.[7,17,22,34] To our knowledge, no previous studies have considered history or patient characteristics as factors for diagnosing traumatic anterior shoulder instability.[23]

We hypothesised that clinical evaluations might be improved with a prediction model that combines patient characteristics, history, and results from a few clinical tests to predict traumatic anterior shoulder instability in patients with shoulder complaints that visit the orthopaedic outpatient clinic. This prediction model could guide clinicians in diagnosing and determining which patients require additional imaging, like magnetic resonance arthrography (MRA), to establish the diagnosis. This type of model may potentially reduce unnecessary health care expenses.

The first objective of this study was to estimate the diagnostic value of different individual clinical tests for traumatic anterior shoulder instability. The MRA was taken as the reference standard. The second objective was to develop a prediction model, which combined patient characteristics, history, and clinical tests to predict the presence of traumatic anterior shoulder instability.

MATERIALS AND METHODS

Patients

This prospective cohort study included new patients with shoulder complaints, recruited consecutively between February 2009 and June 2012 at the orthopaedic outpatient clinic. Exclusion criteria were fractures, frozen shoulder, arthritis, and problems with reading and understanding the Dutch language. Institutional approval was obtained by our local Ethics Committee, and written, signed, informed consent was obtained from all participants.

Data collection

One experienced orthopaedic surgeon (WJW) performed 26 shoulder-specific clinical tests on all patients, according to a standardised diagnostic protocol, based on the original descriptions of the clinical tests. Six of these clinical tests were considered specific for traumatic anterior shoulder instability, and these were selected for evaluation in this study. The examiner was blinded to the imaging analyses. In addition to performing the typical clinical patient history, all patients were asked to complete an online (web-based) questionnaire, which allowed a standardised evaluation of patient history. Patients were asked about previous shoulder dislocations and the onset of the current shoulder complaint. We used validated, patient reported outcome measures (PROMs), to ensure the questions were standardised. Three PROMS were administered: the Simple Shoulder Test, the Oxford Shoulder Score, and the Western Ontario Shoulder Instability Index. Subsequently, all patients underwent MRA of the involved shoulder as a reference standard for the final diagnosis.

Clinical examination tests for assessing traumatic anterior shoulder instability.

The apprehension test was carried out with the patient supine with the arm in 90° abduction, the elbow in 90° flexion, and maximum external rotation.[29] The examiner applied an anterior, external, rotatory force. The test was considered positive when the patient demonstrated an apprehensive feeling; when only pain was experienced, the result was not considered positive.

The relocation test was carried out immediately after the apprehension test.[18] The patient remained in the position that evoked symptoms, and the humeral head was depressed in the posterior direction with direct force to the humeral head. The test was considered positive when the applied posterior-directed force provided relief of the apprehensive feeling, and the patient was able to tolerate maximal external rotation.

The anterior release test, also known as the surprise test, was carried out immediately after the relocation test.[11,32] The examiner suddenly released the pressure on the humeral head (while maintaining the patient's arm in the position of apprehension). The test was considered positive when the patient experienced a sudden apprehensive feeling.

The anterior drawer test was carried out with the patient in the supine position.[10] The patient's hand was positioned on the examiner's axilla. The patient's arm was abducted to 80° to 120°, 0° to 20° forward flexion, and 0° to 30° external rotation. With one hand, the examiner

stabilised the scapula by applying pressure on the coracoid process. With the other hand, the examiner grasped the humerus and drew it out anteriorly. The test was considered positive when there was an increased translation of the humeral head compared to the other shoulder, or when the patient became apprehensive.

The load and shift test, also known as the push-pull test, was carried out with the patient in a sitting position.[24] The examiner grasped the patient's elbow with the corresponding hand. The examiner's other hand grasped the patient's upper arm. The examiner then positioned the patient's arm in 90° abduction in the scapular plane, in neutral rotation, and centred the patient's humeral head on the glenoid by applying a load along the axis of the humerus with the hand that was grasping the patient's elbow. The examiner then attempted to shift the patient's humeral head off the glenoid in the anterior direction. The test was considered positive when the humeral head could be shifted anteriorly off the glenoid (grade II and III) or when the patients displayed apprehension.[24]

The hyperabduction test was carried out with the patient standing,[9] and the examiner standing behind the patient. With the examiner's forearm, the shoulder girdle was pushed down firmly, while the examiner's other hand lifted the patient's upper limb, which was relaxed in abduction. During the test, the elbow was flexed at 90°, and the forearm was horizontal. The test was considered positive when the arm could be hyperabducted above 105°, or when the patient displayed apprehension.

Imaging technique: MRA

Each patient first received an intra-articular administration of 10 mL omnipaque 300 (300 mg I/ml iohexol; GE Health- care BV, Eindhoven, The Netherlands). A 10 mL mixture of 0.5 mL omniscan (0.5 mmol/mL Gd-DTPA-BMA; GE Healthcare BV) was added to 100 mL of 0.9% saline. Then, the patient received 12– 15 cc of this solution, delivered with an 18-gauge needle inserted into the glenohumeral joint under fluoroscopic guidance, with either an anterior or a posterior approach. MRA images were acquired within 30 min after injection. Patients were instructed to immobilise the shoulder of interest after the injection and during the MRA. Imaging was performed with either a 1.0 T unit (MR Systems NT Release 4.5; Philips Medical Systems, Best, The Netherlands) or a 1.5 T unit (MR Systems Intera, Release 9.0, Philips Medical Systems). The following sequences were performed: T1-weighted fast spin echo (FSE) with fat-selective presaturation in an axial plane, oblique coronal plane, and oblique sagittal plane; oblique coronal proton-density (PD); and T2-weighted FSE and T1-weighted FSE with fat-selective presaturation, with the shoulder in an abduction-external rotation (ABER) position.

Reference standard

The diagnosis based on the MRA was defined as the reference standard. The MRAs were reviewed in random order, blinded of the patient's personal details, clinical history, and symptoms. The final diagnosis was determined in consensus between the orthopaedic surgeon (W.J.W.) and a musculoskeletal radiologist (H.J.W.). Both had more than 15 years experience in evaluating shoulder MRAs. In cases that no consensus could be reached, a second experienced

musculoskeletal radiologist was available to make the final diagnosis. We chose a consensus diagnosis for the MRA, because previous studies have shown inter-observer variability for traumatic anterior shoulder instability.[12,31] All different possible diagnoses for shoulder complaints were determined in accordance with standard radiologic criteria.[27]

Specifically, for traumatic anterior shoulder instability, we used the following criteria: the presence of contrast medium between the anterior-inferior glenoid, and a detached labroligamentous complex on a combination of MRA acquisitions. Detachments included a complete detachment, referred to as Bankart lesion, or an avulsed anterior-inferior labral structure, which could be displaced medially on the glenoid neck, with absence of the labrum on the glenoid rim, or non-displaced, with residual attachment of the labrum through an intact periosteum.[40] Furthermore, the presence or absence of a (cartilaginous or bony) Hill-Sachs lesion on the axial images was assessed. The presence of either a labrum lesion and/or a Hill-Sachs lesion was defined as the presence of anterior shoulder instability.

PROMs

We selected three shoulder-related PROMs to standardise the patient history records. We were not able to use the Oxford Shoulder Instability Score, because this PROM was not validated for the Dutch language.[5] Based on clinical experience, we selected four questions out of these PROMs that we expected to facilitate differentiation between several potential diagnoses.

The Simple Shoulder test (SST)

The SST measured functional limitations in patients with shoulder complaints.[21] The SST consisted of 12 questions with dichotomous response options, which were scored (0, 1). The sum of the scores gave the total score, which ranged from 0 (worst) to 12 (excellent). The SST had been validated in Dutch patients with shoulder complaints.[38] Question 8 (weakness) was used as a potential predictor in our prediction model.

Oxford Shoulder Score (OSS)

The OSS also measured functional limitations in patients with shoulder complaints.[4] It contained 12 items with 5 response options, scored from 1 to 5. The total score ranged from 12 (excellent) to 60 (worst). The OSS had been validated in Dutch patients with shoulder complaints.[2] Question 4 (night pain) was used as a potential predictor in our prediction model.

Western Ontario Shoulder Instability Index (WOSI)

The WOSI was specifically designed for shoulder instability evaluations.[20] It contained 21 items on a visual analogue scale (VAS). The total score ranged from 0 (excellent) to 100 (worst). This instrument was validated for patients with shoulder instability, and it is currently being validated in our institution for Dutch patients with shoulder complaints. Question 5 (clicking, cracking, or snapping) and question 8 (feeling of instability or looseness) were used as potential predictors in our prediction model.

Statistical analysis

All patients with any type of traumatic anterior shoulder instability based on the MRA were allocated to the instability group. If patients had other diagnosis, in addition to the anterior shoulder instability, they remained in the instability group. All patients without anterior shoulder instability were, independent of their final diagnosis, allocated to the no anterior instability group. We calculated sensitivity, specificity, positive (PPV) and negative predictive values (PPV and (NPV, respectively), positive likelihood ratio, negative likelihood ratio, and overall accuracy were calculated for all clinical tests using with a 2×2 table. We defined overall accuracy as the percentage of true results (true positives and true negatives).

We used logistic regression to develop the prediction model. Based on clinical experience and a meta-analysis by Hegedus et al.[14], we chose 12 candidate predictors prior to the data analysis. These included age, previous shoulder dislocation, acuteness of the onset of complaints, OSS question 4 (the five possible answers were dichotomised to night pain and no night pain), SST question 8, WOSI questions 5 and 8, and the five clinical tests (apprehension, relocation, release, load and shift, and hyperabduction tests).[1]

Correlations between predictors were investigated to avoid multi-collinearity. A logistic regression prediction model was built with a backward elimination strategy. Thus, predictors were deleted step by step from the model, based on the highest p-value, until a stopping rule was reached, based on Akaike's Information Criterion.[1] Therefore, the final model consisted only of predictors with a p-value below 0.157. When the selected predictor was treated as a continuous variable, like age and the WOSI questions, it was checked to determine whether there was a linear relation between the predictor and the outcome.

The diagnostic value of the model was assessed by determining its calibration and discrimination accuracies.[30] Calibration refers to the agreement between observed and predicted outcomes. This was tested with the Hosmer and Lemeshow "goodness-of-fit" method. The discriminative ability of the prediction model was assessed by the area under the receiver operating characteristic curve (AUC) or the equivalent c (concordance) index.[13]

The receiver operator curve (ROC) is a graphical presentation of the true positives (sensitivity) vs. false positives (1 – specificity) performance of the prediction model. The AUC of ROC curve is a way to reduce ROC performance to a single value representing expected performance. The AUC value ranged between 0.5 (no discriminative value) and 1.0 (perfect discriminative value).

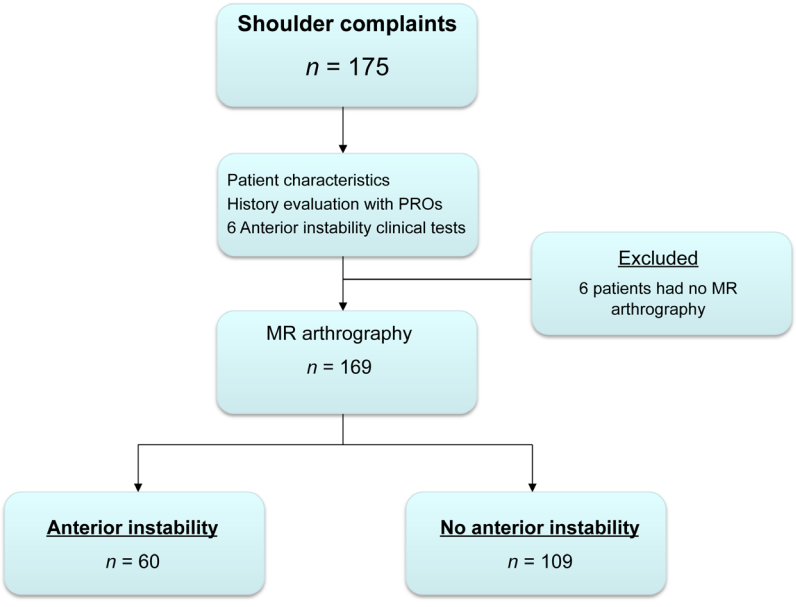
A prediction model was fitted to the dataset at hand; therefore, it was prone to overoptimism in new patients. Overoptimism is particularly common in small datasets, where the number of (starting) predictors is large compared to the smallest outcome group. Internal validation can correct for some of this overfitting with the bootstrapping method. In this case, 500 bootstraps were performed and a shrinkage factor was calculated to penalise the regression coefficients. For guidance on this protocol, see Steyerberg et al.[35] Analyses were performed with R version 2.14.2 (http://www.R-project.org).

RESULTS

Patients

The flowchart of the selection process for the study population is presented in Figure I. The study included 174 patients with shoulder complaints. One patient had bilateral complaints; thus, 175 shoulders were included. Each patient underwent the standardised testing protocol. Six patients were subsequently excluded: one patient went to another hospital due to our waiting list; five patients refused to undergo the MRA. Thus, 169 (97%) shoulders were analysed. Twelve (7%) patients did not complete the online questionnaire, but previous shoulder dislocations had been documented in their files; therefore, they were not excluded. The demographic characteristics are presented in Table I. The average time between the clinical tests and the MRA was 38 days. In two cases, the two reviewers could not reach a consensus on interpretation of the images, and the second radiologist established the final diagnosis. There were no adverse events related to the MRA. Sixty patients (36%) were diagnosed with traumatic anterior shoulder instability based on the MRA. In this group, concomitant diagnoses included superior labral tears from anterior to posterior (SLAP lesions; N=7), rotator cuff (RC) tears (N=2), and biceps pathology (N=1). In the group without anterior instability (N=109), patients were diagnosed with RC tears (N=41), SLAP lesions (N=9), impingement syndrome (N=9), and biceps pathology (N=9). Thus, some patients had multiple diagnoses. In 55 patients, we could not find an explanation for the shoulder complaints based on the MRA.

Figure I. Flow chart of the study population



n: the number of shoulders evaluated

Table I. Demographic characteristics of patients with shoulder complaints

		Baseline
Number of shoulders		169
Mean age, years (SD)		38 (14)
Gender	M	116 (69%)
	F	53 (31%)
Side	R	103 (61%)
	L	64 (38%)
	B	1 (1%)
Simple Shoulder Test	[0-12], mean (SD)*	8.5 (2.8)
Oxford Shoulder score	[12-60], mean (SD)+	24.1 (7.3)
WOSI	[0-100], mean (SD)+	53.1 (19)

M=male, F= female, L=left, R=right, B= both, WOSI: Western Ontario Shoulder Instability Index
**High scores indicate high levels of shoulder function. +Low scores indicate high levels of shoulder function.*

Individual clinical shoulder tests

The estimated diagnostic values of the six individual clinical tests for traumatic anterior shoulder instability are presented in Table II. The overall accuracy, compared to the reference standard (MRA), ranged from 80.5 to 86.4%. The release test showed the best overall accuracy (86.4%). The relocation test was the most sensitive (96.7%), and the anterior drawer test was the most specific (92.7%).

Table II. Diagnostic value of 6 clinical tests for traumatic anterior shoulder instability

Clinical test	Sensitivity	Specificity	PPV	NPV	LR+	LR -	Overall accuracy (%)
Apprehension	98.3	71.6	65.9	98.7	3.46	0.02	81.7
Relocation	96.7	78.0	71.1	97.7	4.39	0.04	85.2
Release test	91.7	83.5	75.3	94.8	5.55	0.10	86.4
Anterior drawer	58.3	92.7	81.4	80.2	7.95	0.45	80.5
Load and shift	71.7	89.9	79.6	85.2	7.10	0.32	83.4
Hyperabduction	66.7	89.0	76.9	82.9	6.06	0.37	81.1

Abbreviations: PPV: positive predictive value, NPV: negative predictive value, LR+: likelihood ratio (for positive test), LR-: likelihood ratio (for negative test)

Prediction model

Of the twelve, pre-selected, candidate predictors, six remained in the prediction model as independent predictors of traumatic anterior shoulder instability: young age, a history of previous shoulder dislocation, sudden onset of shoulder complaints, high scores on WOSI question 5 (clicking, cracking, or snapping) and WOSI question 8 (feeling of instability or looseness), and a positive release test. The WOSI questions were not linearly related to the outcome. Therefore, they were not used in the final prediction model. The other clinical tests

or any combination of clinical tests did not provide additional diagnostic value in our prediction model. Table III shows the final model with the diagnostic odds ratios and a simplified score chart. The discriminative ability, after internal validation (AUC), was 0.95. The 'Hosmer-Lemeshow goodness of fit' test was not significant, which indicated that the model provided a good fit to the data. In Figure II, we plotted the predicted probabilities of traumatic anterior instability for patients in the instability and no instability groups. Based on our prediction model, the median predicted probability of anterior shoulder instability was 91% for patients in the anterior instability group, and only 1% for patients in the no anterior instability group.

Table III. Estimation of the probability of traumatic anterior shoulder instability, according to the validated prediction model. The model is based on previous shoulder dislocation, gradual or sudden onset of the shoulder complaint, the release test, and patient age.

History of Previous dislocation	Onset of complaints	Release test	Patient Age					
			20 y	30 y	40 y	50 y	60 y	70 y
No dislocation	Gradual	-	4%	2%	1%	1%	1%	1%
		+	28%	19%	12%	7%	6%	3%
	Sudden	-	10%	6%	4%	2%	1%	1%
		+	52%	39%	27%	18%	11%	7%
Dislocation	Gradual	-	38%	27%	18%	11%	7%	4%
		+	85%	77%	67%	55%	42%	30%
	Sudden	-	63%	50%	37%	26%	17%	11%
		+	94%	90%	85%	77%	66%	53%

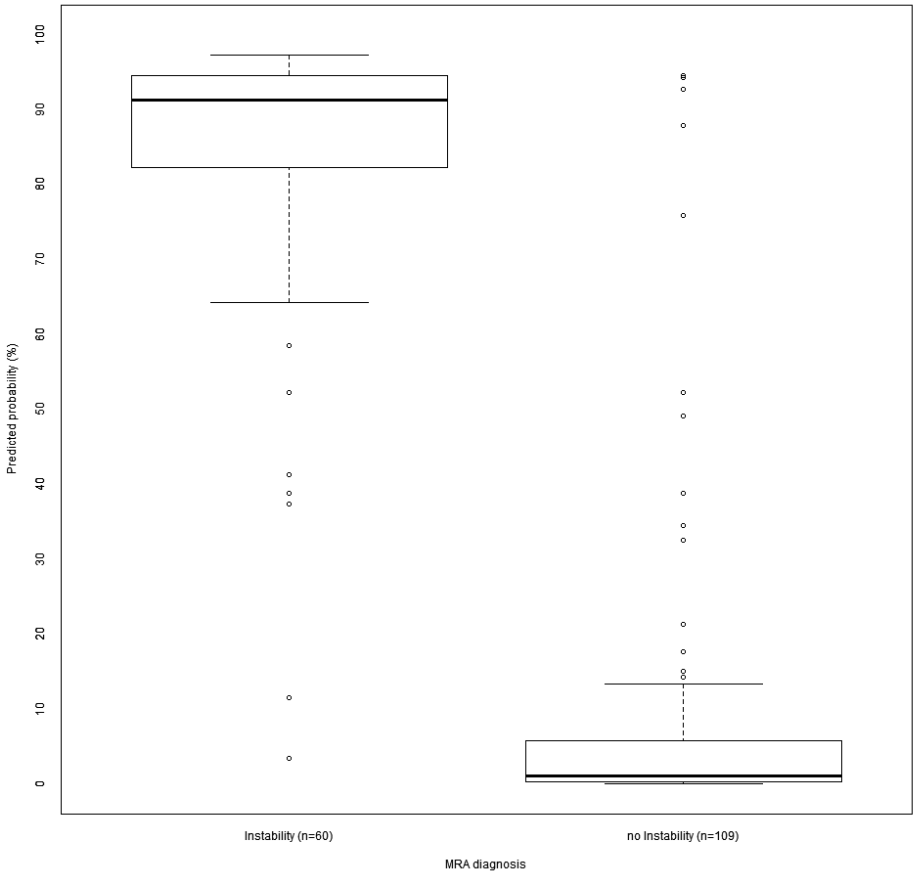
The diagnostic odds ratios of the model are: increasing age (per 10 years): 0.60; previous shoulder dislocation: 14.89; sudden onset of complaints: 9.47; positive release test: 2.73

DISCUSSION

This study evaluated the value of individual clinical tests for diagnosing traumatic anterior shoulder instability. In addition, we developed a prediction model, which combined patient characteristics, history, and clinical tests. We found that the patient information was of paramount importance in establishing the diagnosis of traumatic anterior shoulder instability. The individual clinical tests had good diagnostic value, with an overall accuracy that varied between 80.5 and 86.4%. In constructing a prediction model, we found that the most important predictors were young age, previous shoulder dislocation, sudden unset of complaints, and the release test.

Our results clearly showed that patient characteristics and an adequate history added diagnostic value to the clinical tests for evaluating patients with traumatic anterior shoulder instability. Our prediction model had high discriminative ability (AUC 0.95). Table III serves as a simplified score chart to illustrate model estimations of the probability that a patient will

Figure II. Boxplot



Anterior instability and no anterior instability were diagnosed according to the magnetic resonance arthrography and compared to the predicted probabilities of traumatic anterior instability based on the prediction model. The X-axis represents the patients diagnosed with anterior instability or no anterior instability, according to magnetic resonance arthrography. The Y-axis represents the predicted probabilities of anterior shoulder instability according to the prediction model. The box represents graphically 50% of the patients; the heavy line inside the box is the median.

have traumatic anterior shoulder instability, based on the four important predictors shown. For example, a 20-year old patient, with a history of a previous shoulder dislocation, with sudden onset of complaints, and with a positive release test, had a 94% chance of traumatic anterior instability. In contrast, a 40-year old patient with no history of shoulder dislocation, with gradual development of complaints, and with a negative release test had a only a 4% chance of displaying abnormalities on the MRA that would indicate traumatic anterior shoulder instability.

The use of this prediction model in the clinic could substantially reduce the need for a MRA in establishing the diagnosis of traumatic anterior shoulder instability. However, the MRA can still be useful for visualising other structures in the shoulder, for example the bony defect of the glenoid in order to make an adapted treatment plan.

Most previous studies evaluated these clinical tests with arthroscopy as the reference standard. That type of study design induced a verification bias, because only patients with a surgical indication would be tested with the reference standard, and conservatively treated patients would lack a reference.[3,25] The ideal study design would be to perform an arthroscopy as a reference standard in every new patient with shoulder complaints. This would allow an accurate diagnosis for every patient; however, this approach is not ethically feasible, because arthroscopy is an invasive procedure. Alternatively, MRA is considered to be a highly accurate imaging technique for traumatic anterior shoulder instability; in our view, it is the best reference standard for this type of study.[31,40]

In accordance with previous studies by Tzannes et al., we considered the apprehension, relocation, and release test positive only when patients became apprehensive, not when they experienced pain without apprehension.[36,37] Consistent with the meta-analysis of Hegedus et al., we found high sensitivity for all the individual clinical tests.[15] However, the specificities we observed in our clinical tests were lower than those reported in previous publications.[15] This could be due to the fact that we studied a group of orthopaedic patients with general shoulder complaints, rather than patients with an established surgical indication. Moreover, the results described by Hegedus et al. were all marked as having a high risk of bias.[8,15,17,22]

No previous study has described a prediction model for traumatic anterior shoulder instability. One study with a similar aim, by Farber et al., found that diagnostic value increased when multiple clinical tests were combined.[7] However, there were some limitations to that study. First, the clinical tests were performed only in patients that were scheduled for surgery. Therefore, the clinical tests were not blinded to the diagnosis; this led to a test review bias.[41] Second, conservatively treated patients were not evaluated with the reference standard; this potentially led to a verification bias.[3,23] Finally, the relevance of that study was low, because the greatest value in the clinical tests would be to achieve a diagnosis of traumatic anterior shoulder instability at the first visit to the outpatient clinic, not after the patient has been scheduled for an operation. In our prediction model, no combination of clinical tests increased the diagnostic value.

In developing a prediction model, it is recommended that, for each potential important predictor studied, at least 5 patients must have a certain diagnosis (in this study: 60 patients were diagnosed with traumatic anterior shoulder); otherwise, the risk of overestimating the model would become a problem.[25,39] Therefore, we assessed the prognostic value of 12 pre-determined, potentially important predictors. In addition to the four predictors that we used in the prediction model, we also found two important WOSI questions (5 and 8). These questions focused on 'clicking, cracking, or snapping' and 'an unstable or loose feeling in the shoulder'. However, the increasing steps on the VAS scale of the WOSI were not linearly

related to the outcome. Because we could not clinically explain the meandering values of the WOSI scores in relation to the instability outcome, these two questions were not used in the final prediction model. The absence of a linear relation might have been an artefact due to the small sample size. We selected the WOSI for our study, because it was one of the best-validated PROMs for anterior shoulder instability. In future research, a sensible strategy might be to dichotomise the responses to questions on the WOSI, rather than use the scale of 0-100; this might facilitate the examination of added value of this questionnaire in a prediction model.

Our study is of additional value to the previous studies because of several reasons. First, we came closer to replicating clinical practice by including patient characteristics, patient history, and clinical tests in the diagnostic process. Also, we included every patient with shoulder complaints that visited the orthopaedic outpatient clinic. This prevented a verification bias, as recommended by Luime et al.[23] Furthermore, the reference standard (MRA) was evaluated by individuals that were blinded from patient information. This increased the reproducibility. Finally, we used the same reference standard for every included patient with shoulder complaints.[3]

Our study also had some limitations. First, we did not investigate the inter-observer reliability of the physical examination. However, multiple previous studies showed no significant differences among examiners in performing clinical tests for traumatic anterior shoulder instability.[6,22,24] Second, the examiner was not blinded to the patient characteristics and history information, when the clinical tests were performed. This could lead to a bias. This might explain the high diagnostic value of the individual tests in our study, because the clinical tests might have been influenced by the history information. However, our goal was to develop a prediction model that mimicked clinical practice; therefore, this study design was justified, because it reflected clinical circumstances.

Before implementing our prediction model in clinical practice, it should first be confirmed in a different cohort of patients (external validation). It is also important to stress that our prediction model was developed for patients that visited an orthopaedic outpatient clinic, and thus, it may not be generalizable to primary care.

Conclusion

We hypothesised that clinical evaluation might improve with a prediction model that combine patients characteristics, history and clinical tests for the diagnosis traumatic anterior shoulder instability. In this cohort study, we found that the individual clinical tests for traumatic anterior shoulder instability provided good diagnostic value. Moreover, we developed a prediction model that provided improved diagnostic value, with an AUC of 0.95. Young age, a history of shoulder dislocation, sudden onset of complaints, and a positive release test were the most important predictors for traumatic anterior shoulder instability.

REFERENCES:

1. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. Stat Med. 2002;21(24):3803–3822. doi:10.1002/sim.1422
2. Berendes T, Pilot P, Willems WJ, Verburg H, Slaa te R. Validation of the Dutch version of the Oxford Shoulder Score. J Shoulder Elbow Surg. 2010 Sep;19(6):829–836. doi:10.1016/j.jse.2010.01.017
3. Bhandari M, Montori VM, Swiontkowski MF, Guyatt GH. User's guide to the surgical literature: how to use an article about a diagnostic test. J Bone Joint Surg Am. 2003;85(6):1133–1140.
4. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br. 1996 Jul;78(4):593–600.
5. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. J Bone Joint Surg Br. 1999 May;81(3):420–426.
6. Ellenbecker TS, Bailie DS, Mattalino AJ, Carfagno DG, Wolff MW, Brown SW, et al. Intrarater and interrater reliability of a manual technique to assess anterior humeral head translation of the glenohumeral joint. J Shoulder Elbow Surg. 2002 Sep;11(5):470–475.
7. Farber AJ, Castillo R, Clough M, Bahk M, McFarland EG. Clinical assessment of three common tests for traumatic anterior shoulder instability. J Bone Joint Surg Am. 2006 Jul;88(7):1467–1474. doi:10.2106/JBJS.E.00594
8. Fowler EM, Horsley IG, Rolf CG. Clinical and arthroscopic findings in recreationally active patients. Sports Med Arthrosc Rehabil Ther Technol. 2010;2:2. doi:10.1186/1758-2555-2-2
9. Gagey OJ, Gagey N. The hyperabduction test. J Bone Joint Surg Br. 2001 Jan;83(1):69–74.
10. Gerber C, Ganz R. Clinical assessment of instability of the shoulder. With special reference to anterior and posterior drawer tests. J Bone Joint Surg Br. 1984 Aug;66(4):551–556.
11. Gross ML, Distefano MC. Anterior release test. A new test for occult shoulder instability. Clin. Orthop. Relat. Res. 1997 Jun;(339):105–108.
12. Halma JJ, Eshuis R, Krebbers YMJ, Weits T, de Gast A. Interdisciplinary inter-observer agreement and accuracy of MR imaging of the shoulder with arthroscopic correlation. Arch Orthop Trauma Surg. 2012 Mar;132(3):311–320. doi:10.1007/s00402-011-1370-x
13. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996 Feb 28;15(4):361–387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
14. Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, et al. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. 2007 Jun 4;42(2):80–92. doi:10.1136/bjsm.2007.038406
15. Hegedus EJ. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med. 2012 Jul 7;doi:10.1136/bjsports-2012-091066
16. Hovelius L, Olofsson A, Sandstrom B, Augustini BG, Krantz L, Fredin H, et al. Nonoperative Treatment of Primary Anterior Shoulder Dislocation in Patients Forty Years of Age and Younger. A Prospective Twenty-five-Year Follow-up. The Journal of Bone and Joint Surgery. 2008 May 1;90(5):945–952. doi:10.2106/JBJS.G.00070
17. Jia X, Petersen SA, Khosravi AH, Almareddi V, Pannirselvam V, McFarland EG. Examination of the Shoulder: The Past, the Present, and the Future. The Journal of Bone and Joint Surgery. 2009 Nov 2;91(Supplement 6):10–18. doi:10.2106/JBJS.I.00534
18. Jobe FW, Kvitne RS, Giangarra CE. Shoulder pain in the overhand or throwing athlete. The relationship of anterior instability and rotator cuff impingement. Orthop Rev. 1989 Sep;18(9):963–975.

19. Kazár B, Relovszky E. Prognosis of primary dislocation of the shoulder. Acta Orthop Scand. 1969;40(2): 216–224.
20. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med. 1998 Nov;26(6):764–772.
21. Lippitt SB, Matsen FA. A practical tool for evaluating function: The simple shoulder test. In The Shoulder: A Balance of Mobility and Stability. American Academy of Orthopaedic Surgeons. 1993.
22. Lo IKY, Nonweiler B, Woolfrey M, Litchfield R, Kirkley A. An evaluation of the apprehension, relocation, and surprise tests for anterior shoulder instability. Am J Sports Med. 2004 Mar;32(2):301–307.
23. Luime JJ, Verhagen AP, Miedema HS, Kuiper JI, Burdorf A, Verhaar JAN, et al. Does this patient have an instability of the shoulder or a labrum lesion? JAMA. 2004 Oct 27;292(16):1989–1999. doi:10.1001/jama.292.16.1989
24. McFarland EG, Torpey BM, Curl LA. Evaluation of shoulder laxity. Sports Med. 1996;22:264–272.
25. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009 Feb 23;338(feb23 1):b375–b375. doi:10.1136/bmj.b375
26. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health. 2007 Nov;10 Suppl 2:S125–37. doi:10.1111/j.1524-4733.2007.00275.x
27. Resnick D. Diagnosis of bone and joint disorders. W B Saunders Co; 2002.
28. Robinson CM, Howes J, Murdoch H, Will E, Graham C. Functional outcome and risk of recurrent instability after primary traumatic anterior shoulder dislocation in young patients. J Bone Joint Surg Am. 2006 Nov;88(11):2326–2336. doi:10.2106/JBJS.E.01327
29. Rowe CR, Zarins B. Recurrent transient subluxation of the shoulder. J Bone Joint Surg Am. 1981 Jul; 63(6):863–872.
30. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009 Mar 31;338:b604. doi:10.1136/bmj.b604
31. Schreinemachers SA, van der Hulst VPM, Jaap Willems W, Bipat S, van der Woude H-J. Is a single direct MR arthrography series in ABER position as accurate in detecting anteroinferior labroligamentous lesions as conventional MR arthography? Skeletal Radiol. 2009 Jul;38(7):675–683. doi:10.1007/s00256-009-0692-z
32. Silliman JF, Hawkins RJ. Classification and physical diagnosis of instability of the shoulder. Clin. Orthop. Relat. Res. 1993 Jun;(291):7–19.
33. Slaa te RL, Wijffels MPJM, Brand R, Marti RK. The prognosis following acute primary glenohumeral dislocation. J Bone Joint Surg Br. 2004 Jan;86(1):58–64. doi:10.1302/0301-620X.86B1.13695
34. Speer KP, Hannafin JA, Altchek DW, Warren RF. An evaluation of the shoulder relocation test. Am J Sports Med. 1994 Mar;22(2):177–183.
35. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999 Oct;52(10):935–942.
36. Tzannes A, Murrell GAC. Clinical examination of the unstable shoulder. Sports Med. 2002;32(7): 447–457.
37. Tzannes A, Paxinos A, Callanan M, Murrell GAC. An assessment of the interexaminer reliability of tests for shoulder instability. J Shoulder Elbow Surg. 2004 Jan;13(1):18–23. doi:10.1016/S1058274603002453
38. van Kampen DA, van Beers LWAH, Scholtes VAB, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg. 2012 Jun;21(6):808–814. doi:10.1016/j.jse.2011.09.026

39. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am. J. Epidemiol. 2007 Mar 15;165(6):710–718. doi:10.1093/aje/kwk052

40. Waldt S, Burkart A, Imhoff AB, Bruegel M, Rummeny EJ, Woertler K. Anterior shoulder instability: accuracy of MR arthrography in the classification of anteroinferior labroligamentous injuries. Radiology. 2005 Nov;237(2):578–583. doi:10.1148/radiol.2372041429

41. Whiting P, Rutjes AWS, Dinnes J, Reitsma J, Bossuyt PMM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess. 2004 Jun;8(25): iii–1–234.

Chapter 6

Long-term results after arthroscopic shoulder stabilization using suture anchors; a 8-10 year follow up.

JA v/d Linde
DA Van Kampen
CB Terwee
LM Dijksman
G KleinJan
WJ Willems

The American Journal of Sports Medicine 2011 vol. 39 no. (11): 2396-2403

ABSTRACT

Background: Arthroscopic stabilization using suture anchors is widely used to restore stability after anterior shoulder dislocations and is associated with low recurrence rates in short-term follow-up studies.

Purpose: To evaluate the long-term follow up after arthroscopic stabilization for traumatic recurrent anterior instability using suture anchors with emphasis on both re-dislocations and subjective shoulder function.

Methods: We included 67 consecutive patients with 70 affected shoulders. After 8-10 years, patients were asked to report the presence and course of their redislocations. Subjective shoulder function was addressed using the Oxford Shoulder Instability Score (OSIS), the Western Ontario Shoulder Instability Index (WOSI) and the Simple Shoulder Test (SST). Patients rated their health status using the Short Form-36 (SF-36).

Results: Sixty-five patients with 68 affected shoulders (97%) were evaluated for follow-up; 35% reported a redislocation. Median shoulder function scores were 16 of 12 to 60, 22 of 0 to 210, and 12 of 0 to 12 for the OSIS, WOSI, and SST, respectively. There was a significant difference in subjective function between patients with and without recurrent instability, respectively, 16 versus 24 for the OSIS (P = .004), and 16 versus 47 for the WOSI (P = .05). We found a trend for an inverse relationship between the number of suture anchors and recurrent instability, with 2 having a higher recurrence rate than 3 or more (P = .06). Another trend was found with the presence of a Hill-Sachs defect slightly increasing the risk of a redislocation (P = .07).

Conclusion: With a follow-up of 97%, about one third of the stabilized shoulders experienced at least one redislocation after 8 to 10 years. The presence of a Hill-Sachs defect and the use of less than 3 suture anchors might increase the chance of a redislocation. Patients without a redislocation have a significantly better shoulder function compared with patients with a redislocation.

INTRODUCTION

Shoulder dislocations have an incidence of 17 to 24 per 100 000 per year[20,21] and are more than 90% directed anteriorly.[11] After an initial dislocation, the labral structures often tear from the bony glenoid to reinsert more inferiomedially. The capsular ligaments stretch, thus widening the joint, often resulting in recurrent dislocations.

Several risk factors for primary dislocation have been identified. Being adolescent or a young adult at the time of initial dislocation is clearly associated with a higher risk of subsequent instability, just as is occupational use of the arm at or above chest level, hyperlaxity, and participation in collision sports.[12,36] Although nonoperative therapy, that is, shoulder immobilization and physical therapy, can lead to satisfactory results,[14,17] it is also reported to be associated with a recurrence rate as high as 57%, which is inversely related to the age of the initial dislocation.[12,15] Surgical stabilization can be achieved using several techniques such as the open and arthroscopic Bankart procedure at which the labrum is reattached to the glenoid.2,3 The presence of recurrent instability after the open Bankart procedure varies from 3.5% to 23% after 4 to 6 years postoperatively and 10% to 22.6% after the modified Bankart procedure 11 to 29 years postoperatively.[2,21,29,35,38]

Open stabilization has long been considered the gold standard for surgical stabilization and is still reported in several reviews to be superior to arthroscopic stabilization.[10,23,27] However, arthroscopic stabilization evolved and shows to have good results too. Arthroscopic advantages include less chance of loss of motion, especially external rotation that could limit the shoulder function, shorter surgery time, improved cosmetics, and less postoperative pain.14 The most recent arthroscopic technique involves the use of suture anchors and decreases the failure rate when compared with the previous arthroscopic techniques of capsulorrhaphy, transglenoid sutures, and bioabsorbable tacks.[13]

Most studies with this technique show recurrence rates around 10% up to 3.6 years postoperatively.[4,7,13,41] With more extended follow-up, one study in rugby players shows a success rate of .90% 5.9 years postoperatively.[22] Recently, Castagna et al[8] showed a recurrence rate of 23% in 10.9 years after arthroscopic stabilization using suture anchors.

Predisposing factors identified for recurrent dislocation after arthroscopic stabilization include a young age, being male, and having an interval of more than 6 months between the first dislocation and time of surgery. [18,32] Also, both humeral head and glenoidal defects are described as risk factors as well as the number of suture anchors that have been used.[4,41]

Increasing evidence is emerging that patients and doc- tors do not always agree on functional improvements after therapeutic interventions.[16] The patient's subjective well- being is increasingly considered to be important, in addition to objective shoulder scores. This implies the start of using patient-based questionnaires more regularly. To investigate the outcome of a surgical intervention ideally, both system-specific (shoulder) and condition-specific (instability) instruments should be used.[30,31]

The purpose of the present study was to prospectively evaluate the long-term surgical outcomes after arthroscopic shoulder stabilization in patients with traumatic recurrent anterior shoulder instability using absorbable suture anchors with emphasis on both redislocations and subjective shoulder function.

MATERIAL AND METHODS

Design

Institutional approval was obtained from our local ethics committee. Patients signed informed consent forms. We performed an observational prospective case series of 67 consecutive patients with 70 affected shoulders who under- went an arthroscopic stabilization using suture anchors from January 1999 to December 2001, with a mean follow-up of 9 years (range, 8-10 years). This technique was introduced in our hospital in 1997. All patients were operated on by one single senior surgeon.

These patients met the following inclusion criteria: (1) all patients were 18 years or older and had repeated involuntary anterior instability after an initial episode caused by a traumatic event, and patients with atraumatic or multidirectional instability were excluded; (2) arthroscopic repair was performed using absorbable suture anchors; and (3) a consistent postoperative treatment program was followed, which included 6 weeks of immobilization in a sling after which a period of active exercises was started. Sport activities were allowed 4 months postoperatively.

Preoperative evaluation consisted of a detailed history including their level of sports activity and physical work- load, and physical examination included Rowe scores (version 1978; 0-25 = poor, 25-50 = fair, 50-75 = good, and 75-100 = excellent) to enable prospective evaluation.[35] We have a heterogeneous population including professional athletes, recreational athletes, and sedentary patients.

Surgical Procedure

Patients received an interscalene block before induction of general anesthesia to diminish postoperative pain. Patients were operated in the lateral decubitus position with traction in 2 directions. For the procedure, we used Rodtag (Smith & Nephew, Andover, Massachusetts) made of polyglycolide/polylactate, 3.7 mm in size with a PDS 2.0. The anchors were applied using a push-in technique and were single loaded.

Three portals were used: the standard posterior portal, the standard anterior portal just above the subscapularis tendon through which the anchors were placed, and one superior portal just anterolateral to the acromion. The first anchor was placed at the 5-o'clock position (right shoulder) or 7-o'clock position (left shoulder), which was sufficient to fix the advanced anteroinferior capsula/labrum complex anterosuperior.

With a suture passer through the anterior portal, a shuttle relay was passed through the labrum and advanced through the superior portal, where one of the strands of the anchor suture was brought as well. With the shuttle relay, the first strand was led through the labrum. The same was repeated with the second strand of the anchor suture. After having led both sutures through the labrum, they were tied, thus achieving a mattress configuration. Capsular plication was performed when the capsule was stretched out.

Postoperative Evaluation

Patients were contacted by telephone and asked for writ- ten follow-up using Web-based questionnaires. Our primary outcome is recurrent instability, defined as either a subluxation or a full dislocation. Our secondary outcome is the subjective improvement of shoulder function. The shoulder function was not objectified by physical or radio- graphic examination.

We conducted a subanalysis addressing the influence of several possible risk factors on redislocation. These risk factors include age, dominance, gender, preoperative shoulder function (Rowe score), the number of dislocations preoperatively, the time to surgery in months, and the number of suture anchors that were used. Also, possible influences of both preoperative overhead and contact sports are assessed, as well as the presence of bony defects on both sides of the shoulder joint.

Because both glenoidal bone loss as well as Hill-Sachs lesions have proven to significantly influence the rate of a redislocation after soft tissue shoulder stabilization, the extent of osseous defects was measured on magnetic resonance imaging (MRI) scans.[5,33] Glenoid defects (compression fracture) were assessed using the Bigliani classification, ranging from an ununited glenoidal fragment attached to the separated labrum (type 1) to a malunited fragment detached from the labrum (type 2) to an anterior glenoidal erosion <25% (type 3A) or >25% (type 3B).[3] Defects on the humeral head (Hill-Sachs lesions) were calculated by taking the MRI slice with the largest observed defect and measuring the size of the defect in relation to the total joint circumference in the same slice, expressed as a percentage.

We used the Oxford Shoulder Instability Score (OSIS),[9] the Western Ontario Shoulder Instability Index (WOSI),[19] and the Dutch version of the Simple Shoulder Test (SST) as validated patient-based questionnaires to evaluate the shoulder function at follow-up.[24,39] We used the Short Form-36 (SF-36) to assess perceived general health status from our patients compared with the normal Dutch population.[1]

Statistical Analysis

Analyses were performed using SPSS version 17.0 (Chicago, Illinois). Characteristics of our study population were described using the median and standard deviation (SD) when normally distributed or the interquartile range when nonnormally distributed. The number of redislocations was calculated as a percentage of the examined shoulders. To evaluate the influence of gender, dominance, number of suture anchors, preoperative sports participation, and degree of bone defect on the risk of redislocation (yes/no), we calculated relative risks

with 95% confidence intervals. For the categorical and continuous variables of age, Rowe score, months to surgery, and number of preoperative dislocations, we used logistic regression analysis, with redislocation (yes/no) as the dependent variable, and calculated odds ratios with 95% confidence intervals. For the number of preoperative dislocations, we divided our patients into 3 groups: having experienced up to 2, up to 5, and more than 10 preoperative dislocations. The differences in OSIS, WOSI, and SST between patients with and without redislocations were evaluated using the Mann-Whitney test. Scores on the SF-36 were compared with reference scores from the Dutch general population (version SF-36-1), adjusted for age and sex.1 A value of $P < .05$ was considered statistically significant.

RESULTS

Of 67 patients in total, 65 patients with 68 affected shoulders (97%) could be reached for follow-up to evaluate redislocation; 2 patients were lost to follow-up. One patient died 5 years postoperatively without ever having experienced a redislocation. There were 43 men and 22 women; in 47%, their dominant side was involved. The mean age at time of surgery was 31 years (range, 19-56 years). One patient who reported 2 severe subluxations was included because of the presence of a labral lesion on his MRI arthrography. The median number of dislocations before surgery was 5 (range, 2-40). The median number of months between the first dislocation and surgery was 51 months (range, 8-479 months). Sixty-five percent of the shoulders were stabilized using 2 suture anchors and 35% using 3 or more suture anchors. Before stabilization, 17 patients (19 shoulders) participated in contact sports, 22 patients (23 shoulders) participated in overhead sports, and 8 patients (9 should- ers) participated in a combination of both.

Recurrent Instability

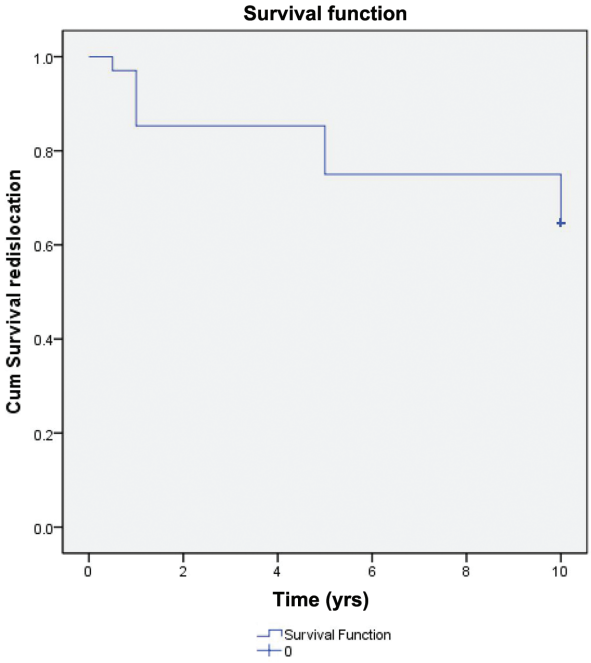
At follow-up, 8 to 10 years postoperatively, a total of 24 (35%) shoulders had experienced a redislocation. In 10 (15%) shoulders, redislocation took place within the first 2 years postoperatively, another 7 (10%) shoulders redislocated 2 to 5 years postoperatively, and another 7 (10%) experienced a redislocation after more than 5 years (Figure I).

Of all patients who experienced a redislocation, 18% (12 shoulders) experienced 1 to 2 recurrent dislocations at most, 7% (5 shoulders) experienced 3 to 4 recurrent dislocations, another 7% (5 shoulders) experienced 5 to 10, and 3% (2 shoulders) experienced more than 10 recurrent dislocations. Nine patients (13%) in whom primary stabilization failed underwent a new operation; others were treated nonoperatively or refused a reoperation.

Risk Factors

The results are shown in Table I. Although not significant, shoulders stabilized with 3 or 4 anchors tended to have a less chance to redislocate than ones stabilized with 2 anchors ($P = .06$).

Figure I. Kaplan-Meier curve of the redislocation rate over time



No relationship was found between the number of suture anchors and the postoperative period in which a redislocation took place (P = .48).

No other possible risk factors could be confirmed; being operated on the dominant side was not associated with having an increased risk for experiencing a redislocation (P = .60), nor was being male (P = .79). Age (P = .43) and preoperative Rowe score (P = .84) were neither associated to have an increased risk, and also time span to surgery (P = .09) and the number of preoperative dislocations in general(P=.23) or up to 2, up to 5, or more than 10 were not associated with an increased risk of experiencing a redislocation (P = 1.00, P = .60, and P = .34, respectively).

Preoperative Sports Participation

In total, 40 patients (42 shoulders, 62%) participated in either contact sports or overhead sports or both preoperatively. Twenty-five patients (26 shoulders, 38%) either participated in other sports or no sport at all. Preoperative participation in contact sports alone or in both contact and overhead sports simultaneously did not increase the chance of experiencing a re-dislocation (P = .57 and P = .48, respectively). Participation in overhead sports alone, however, seemed to significantly decrease the chance of experiencing a redislocation (P = .03) (Table II).

Table I. Sub analysis for possible risk factors for re-dislocation.

Variable		N	Re-dislocation %	RR	95% CI	P-value
Gender	Male	43	37	1.16	0.57 – 2.38	0.79
	Female	22	32			
Surgery						
	Dominant	36	33	0.82	0.43 – 1.57	0.60
	Non dominant	32	37			
Anchors						
	2	43	44	2.12	0.91 – 4.96	0.06
	3 or more	25	20			
Pre-op dislocations						
	> 2	43	37	1.06	0.52 – 2.17	1.00
	≤ 2	20	35			
	> 5	28	32	0.80	0.41 – 1.58	0.60
	≥ 5	35	40			
	> 10	13	23	0.58	0.20 – 1.65	0.34
	≤ 10	50	40			

Variable	n	Median (range)	OR	95% CI	P-value
Age (per 10 years older)	68	31 (19-56)	0.77	0.41-1.46	0.43
Rowe-score (per point higher)	64	45 (15-75)	1.00	0.95-1.05	0.84
Months to surgery (per 6 months longer)	61	51 (8-479)	0.96	0.91-1.00	0.09
Number of pre-op dislocations	61	5 (0-40)	0.95	0.88-1.0	0.23

RR: relative risk, CI: confidence interval, OR: odds ratio

Table II. Sub analysis for sport activities as risk factors for re-dislocation.

Type of sport		N	Re-dislocation %	RR	95% CI	P-value
Contact			42	1.29	0.66 – 2.50	0.57
	Yes	19				
	No	49	33			
Overhead				0.39	0.15 – 1.01	0.03
	Yes	23	17			
	No	45	44			
Both				0.60	0.17 – 2.11	0.48
	Yes	9	22			
	No	59	37			

RR: relative risk, CI: confidence interval

Bony Defects

To assess the effect of bony defects on redislocations, we conducted a subanalysis using 54 MRI scans in which we evaluated both the glenoid and humeral head (Hill-Sachs) (Table III).

Table III. Sub analysis for bony defects as risk factors for re-dislocation.

Degree bone defect			N	Re-dislocation %	RR	95% CI	P-value
Glenoid	Type 2	Yes	6	50	1.32	0.55 – 3.20	0.63
		No	48	38			
	Type 3A	Yes	33	39	1.03	0.52 – 2.06	1.0
		No	21	38			
	Any degree	Yes	39	41	1.23	0.55 – 2.76	0.60
		No	15	33			
Hill Sachs	<25%	Yes	43	42	1.56	1.08 – 2.21	0.13
		No	11	27			
	>25%	Yes	3	67	1.79	0.75 – 4.30	0.55
		No	51	37			
	Any degree	Yes	45	44	4.00	0.61 – 26.1	0.07
		No	9	11			

RR: relative risk, CI: confidence interval

In 15 shoulders (28%), the glenoid was intact; in 39 shoulders (72%), the glenoid had some degree of damage, varying from Bigliani type 2 (6 cases) to type 3A (33 cases). When compared with shoulders without a glenoid defect, neither type 2 nor type 3A nor both types combined increased the risk of redislocation (P = .63, P = 1.00, and P = .60, respectively).

Addressing the humeral head, in 9 shoulders (17%), no Hill-Sachs lesion was identified; in 42 shoulders (78%), a Hill-Sachs lesion comprised less than 25% of the humeral head circumference, and in only 3 shoulders (6%), the Hill- Sachs lesion comprised more than 25%. Although not significant, there might be a relationship between the presence of any degree of Hill-Sachs lesion and the presentation of a redislocation (P = .07). The lesion's size, however, was of no significant influence, with P = .13 for lesions <25% and P = .55 for lesions >25%.

Postoperative Shoulder Function

Our secondary outcome was the subjective shoulder function, evaluated with the above-mentioned questionnaires. This function was not objectified. Although 57 patients (59 shoulders, 84%) were willing to answer the questionnaires, the postoperative function was conducted

Table IV. Subjective shoulder scores and SF-36 scores (median and IQR).[a]

					Re-dislocation		
Score	No. items	Range	Best	Total	No (n=35)	Yes (n=15)	P-value
OIS	12	12-60	Lowest	16 (13-24)	16 (13-17)	24 (15-28)	0.004
WOSI	21	0-210	Lowest	22 (11-70)	16 (9-45)	47 (19-75)	0.05
SST	12	0-12	Highest	12 (10-12)	12 (10-12)	12 (11-12)	0.48

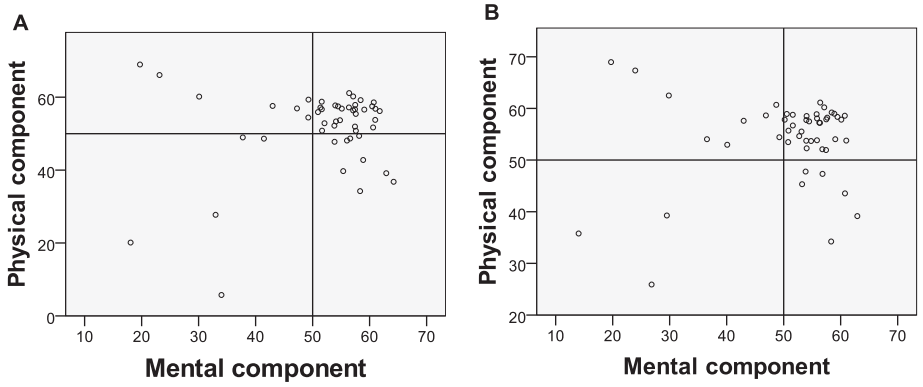
[a] *Patients with recurrent surgery are excluded. IQR: interquartile range, OIS: Oxford Instability Score, WOSI: Western Ontario Shoulder Instability Index, SST: Simple Shoulder Test.*

only from those 48 patients (50 shoulders) without a reoperation during follow-up (71%). The characteristics of these scores are summarized in Table IV.

The median scores for both the OSIS and the WOSI were low (meaning good function), with 16 of 12 to 60 for the OSIS and 22 of 0 to 210 for the WOSI. The scores of patients without a redislocation were significantly better compared with patients with a redislocation. The median SST score was high (meaning good function), with 12 of 0 to 12, and there was no difference between SST scores of patients with or without a redislocation (Table IV).

In the SF-36 score, when corrected for age (16-40 years), only 5 participants scored less in both the physical and mental component compared with the normal Dutch population. Because not all patients were athletes, the SF-36 score was not corrected for this entity, even though athletes might have a higher baseline SF-36 score. When corrected for sex, there were only 3 patients to score less than the norm (Figure II).

Figure II. The SF-36 scores corrected for age (a) (16-40 yrs) and sex (b) compared to the normal Dutch population. The physical and mental scores above 50 correspond to better physical or mental function than the comparison group.



Complications

No infections or other complications occurred in this series.

DISCUSSION

The optimal surgical technique to treat recurrent anterior glenohumeral instability remains a controversial topic, as good results have been shown by both open and arthroscopic stabilization. Although often described to show a lower recurrence rate, the open Bankart procedure can result in a limited range of motion.[17,23]

In our study on the long-term follow-up after arthroscopic stabilizations for traumatic onset of recurrent anterior shoulder instability, redislocation was defined as our primary

outcome. Because the subjective questionnaires reflect the experience of the patient, we used strict criteria for redislocation, including both a subluxation and a full dislocation, to avoid possible confusion for the patient. Our results up to 2 years postoperatively resemble previous studies with a 15% failure rate.[4,13,41] However, another 20% of the patients experienced their first redislocation after 2 years, leading to a total 35% of our patients having experienced at least one redislocation 8 to 10 years postoperatively. This is slightly more compared with long-term follow-up studies in open stabilization and one previous long-term follow-up study in arthroscopic stabilization, which only reached 71% for final follow-up.[2,8,21,29]

No previously identified risk factors could be confirmed in our study. Contrary to previous studies, no significance was found for age as a risk factor.[32,34,40] Because young age has been identified as a risk factor to increase the chance of a redislocation, the relatively high age of our population might positively influence our results. The interval between first dislocation and surgery did not increase the chance of a redislocation. Nor did the number of preoperative dislocations, up to 2, up to 5, or more than 10, influence the chance of having a redislocation. Dominance, gender, and preoperative shoulder function did not influence the chance of having a redislocation either.

As Boileau et al[4] previously described, we also did find, although not significantly, a relationship between the number of suture anchors that were used and the chance of a redislocation. This trend shows that patients stabilized with 3 or more suture anchors have less chance to experience a redislocation than patients stabilized with 2 suture anchors. The fact that 43 shoulders (63%) were stabilized using 2 suture anchors might partly explain the higher recurrence rate in our study compared with previous studies. We did not find any confounding factors between the different risk factors.

We also found, as previously described by Voos et al[41] and Boileau et al,[4] that the presence of a Hill-Sachs defect, although not significantly, increased the redislocation risk. We did not find a relationship between glenoid lesions and redislocations.

Preoperative participation in contact sports, as Ide et al[15] described before, did not influence the redislocation risk. We observed that preoperative participation in over- head sports, however, decreased this risk significantly, although no specific measures were taken in overhead athletes. This is contrary to the findings of Ide et al[15] and Pagnani et al,[28] who previously described no relationship, and to Sachs et al[36] and Calvo et al,[6]who described a significant increased redislocation risk in patients participating in overhead sport activities.

This finding might be because of higher awareness, better muscular control, and better proprioceptive abilities in patients participating in overhead sports compared with their peers who do not share their experiences. Another plausible explanation is that these patients might have diminished their sports intensity or did not return to their preinjury level of shoulder function compared with patients who were not engaged in overhead activity.

To investigate the subjective functional improvement of a surgical intervention ideally, both system (shoulder)– specific and condition (instability)–specific instruments should be used.[30,31] The OSIS, WOSI, and SST are designed to do so, including questions addressing

sport, work, and daily activities. To objectify the results of our initial procedure, in this subanalysis, we only included patients without a reoperation during follow-up. Based on the good average scores, we conclude that our patients in general are satisfied with their shoulder function.

We furthermore found that only the instability-specific questionnaires (OSIS and WOSI) distinguished significantly between patients with and without redislocation, whereas the more general shoulder score (SST) did not. In our opinion, only instability-specific questionnaires should therefore be used in long-term follow-up studies for instability management.

The SF-36 showed that the majority of the patients score above average on both the mental and physical component. With our heterogeneous patient population, we did not correct for the fact that athletes might have a higher baseline SF-36 score.

Because the OSIS, WOSI, and SST have only recently been developed and had not yet been validated thoroughly at the time of surgery, we conducted the Rowe score before surgery. The Rowe score is a very commonly used scoring system largely based on the objective shoulder function.

An important advantage of using validated patient- based questionnaires is that patients can be included for final follow-up without visiting the hospital. Especially in this relatively young and highly mobile patient population, it is a very practical way to follow up on patients.

Although long-term follow-up studies on previous arthroscopic techniques are available,[26,37] to our knowledge, only one previous study has been published on extended follow- up after arthroscopic stabilization using suture anchors in a community-based patient population. This study, however, included only 71% for their final follow-up.[8]

One study with an extended follow-up included only male contact athletes who are likely to be in excellent condition and might not be as representative as our population.[26] Another previous study described the results after open stabilization with suture anchors using the OSIS. Eleven years after their initial operation, they have a response rate of 64% and have a mean score of 21.7, with 12% experiencing further dislocations or ongoing symptoms of instability.[25]

Strong points of this study are the 97% long-term follow-up and the presence of an independent observer using validated patient-based questionnaires. A weak point is the fact that 10 years ago, generally 2 to 3 anchors were used, while presently, a large consensus exists to use 3 or more anchors. Another weak point is the fact that, although the used subjective scores are highly valuable, patients were not physically or radiographically evaluated, and instability, shoulder range of motion, or strength is not objectified. The authors, however, argue that a stiff shoulder or an impaired shoulder function will translate in the outcome of the questionnaires. Instability-specific parameters like passive apprehension could therefore not be included in our results.

Conclusion

Eight to 10 years postoperatively, about one third of the patients undergoing this type of procedure reported a redislocation. More than half of these occurred for the first time more than 2 years after surgery.

We found that the number of anchors used, as well as the presence of a Hill-Sachs defect, tended to be predisposing factors for experiencing a redislocation. Other previously identified predisposing factors for redislocation could not be confirmed. Although all patients in general reported few functional problems, patients without a redislocation have a significantly better subjective shoulder function compared with their peers with a redislocation.

REFERENCES

1. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol. 1998 Nov;51(11):1055–1068.
2. Berendes TD, Wolterbeek R, Pilot P, Verburg H, Slaa te RL. The open modified Bankart procedure: outcome at follow-up of 10 to 15 years. J Bone Joint Surg Br. 2007 Aug;89(8):1064–1068. doi: 10.1302/0301-620X.89B8.19280
3. Bigliani LU, Newton PM, Steinmann SP, Connor PM, Mcllveen SJ. Glenoid rim lesions associated with recurrent anterior dislocation of the shoulder. Am J Sports Med. 1998 Jan;26(1):41–45.
4. Boileau P, Villalba M, Héry J-Y, Balg F, Ahrens P, Neyton L. Risk factors for recurrence of shoulder instability after arthroscopic Bankart repair. J Bone Joint Surg Am. 2006 Aug;88(8):1755–1763. doi: 10.2106/JBJS.E.00817
5. Burkhart SS, De Beer JF, Barth JRH, Cresswell T, Criswell T, Roberts C, et al. Results of modified Latarjet reconstruction in patients with anteroinferior instability and significant bone loss. Arthroscopy. 2007 Oct;23(10):1033–1041. doi:10.1016/j.arthro.2007.08.009
6. Calvo E, Granizo JJ, Fernández-Yruegas D. Criteria for arthroscopic treatment of anterior instability of the shoulder: a prospective study. J Bone Joint Surg Br. 2005 May;87(5):677–683. doi: 10.1302/0301-620X.87B5.15794
7. Carreira DS, Mazzocca AD, Oryhon J, Brown FM, Hayden JK, Romeo AA. A prospective outcome evaluation of arthroscopic Bankart repairs: minimum 2-year follow-up. Am J Sports Med. 2006 May; 34(5):771–777. doi:10.1177/0363546505283259
8. Castagna A, Markopoulos N, Conti M, Delle Rose G, Papadakou E, Garofalo R. Arthroscopic bankart suture-anchor repair: radiological and clinical outcome at minimum 10 years of follow-up. Am J Sports Med. 2010 Oct;38(10):2012–2016. doi:10.1177/0363546510372614
9. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. J Bone Joint Surg Br. 1999 May;81(3):420–426.
10. Freedman KB, Smith AP, Romeo AA, Cole BJ, Bach BR. Open Bankart repair versus arthroscopic repair with transglenoid sutures or bioabsorbable tacks for Recurrent Anterior instability of the shoulder: a meta-analysis. Am J Sports Med. 2004 Sep;32(6):1520–1527. doi:10.1177/0363546504265188
11. Goss TP. Anterior glenohumeral instability. Orthopedics. 1988 Jan;11(1):87–95.
12. Handoll HHG, Almaiyah MA, Rangan A. Surgical versus non-surgical treatment for acute anterior shoulder dislocation. Cochrane Database Syst Rev. 2004;(1):CD004325. doi:10.1002/14651858.CD004325.pub2
13. Hobby J, Griffin D, Dunbar M, Boileau P. Is arthroscopic surgery for stabilisation of chronic shoulder instability as effective as open surgery? A systematic review and meta-analysis of 62 studies including 3044 arthroscopic operations. J Bone Joint Surg Br. 2007 Sep;89(9):1188–1196. doi: 10.1302/0301-620X.89B9.18467
14. Hovelius L, Augustini BG, Fredin H, Johansson O, Norlin R, Thorling J. Primary anterior dislocation of the shoulder in young patients. A ten-year prospective study. J Bone Joint Surg Am. 1996 Nov;78(11): 1677–1684.
15. Ide J, Maeda S, Takagi K. Arthroscopic Bankart repair using suture anchors in athletes: patient selection and postoperative sports activity. Am J Sports Med. 2004 Dec;32(8):1899–1905.
16. Janse AJ, Gemke RJBJ, Uiterwaal CSPM, van der Tweel I, Kimpen JLL, Sinnema G. Quality of life: patients and doctors don't always agree: a meta-analysis. J Clin Epidemiol. 2004 Jul;57(7):653–661. doi: 10.1016/j.jclinepi.2003.11.013

17. Karlsson J, Magnusson L, Ejerhed L, Hultenheim I, Lundin O, Kartus J. Comparison of open and arthroscopic stabilization for recurrent shoulder dislocation in patients with a Bankart lesion. Am J Sports Med. 2001 Sep;29(5):538–542.
18. Kim S-H, Ha K-I, Cho Y-B, Ryu B-D, Oh I. Arthroscopic anterior stabilization of the shoulder: two to six-year follow-up. J Bone Joint Surg Am. 2003 Aug;85-A(8):1511–1518.
19. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med. 1998 Nov;26(6):764–772.
20. Krøner K, Lind T, Jensen J. The epidemiology of shoulder dislocations. Arch Orthop Trauma Surg. 1989; 108(5):288–290.
21. Langford J, Bishop J, Lee E, Flatow E. Outcomes following open repair of Bankart lesions for recurrent, traumatic anterior glenohumeral dislocations. Orthopedics. 2006 Nov;29(11):1008–1013.
22. Larrain MV, Montenegro HJ, Mauas DM, Collazo CC, Pavón F. Arthroscopic management of traumatic anterior shoulder instability in collision athletes: analysis of 204 cases with a 4- to 9-year follow-up and results with the suture anchor technique. Arthroscopy. 2006 Dec;22(12):1283–1289. doi:10.1016/j.arthro.2006.07.052
23. Lenters TR, Franta AK, Wolf FM, Leopold SS, Matsen FA. Arthroscopic compared with open repairs for recurrent anterior shoulder instability. A systematic review and meta-analysis of the literature. J Bone Joint Surg Am. 2007 Feb;89(2):244–254. doi:10.2106/JBJS.E.01139
24. Lippitt SB, Matsen FA. A practical tool for evaluating function: The simple shoulder test. In The Shoulder: A Balance of Mobility and Stability. American Academy of Orthopaedic Surgeons. 1993.
25. Loughead JM, Williams JR. Outcome following Bankart repair for shoulder instability 1998-2002 using a subjective patient based shoulder questionnaire (Oxford). Injury. 2005 Apr;36(4):539–543. doi:10.1016/j.injury.2004.09.031
26. Marquardt B, Witt K-A, Götze C, Liem D, Steinbeck J, Pötzl W. Long-term results of arthroscopic Bankart repair with a bioabsorbable tack. Am J Sports Med. 2006 Dec;34(12):1906–1910. doi:10.1177/0363546506290404
27. Mohtadi NGH, Bitar IJ, Sasyniuk TM, Hollinshead RM, Harper WP. Arthroscopic versus open repair for traumatic anterior shoulder instability: a meta-analysis. Arthroscopy. 2005 Jun;21(6):652–658. doi:10.1016/j.arthro.2005.02.021
28. Pagnani MJ, Warren RF, Altchek DW, Wickiewicz TL, Anderson AF. Arthroscopic shoulder stabilization using transglenoid sutures. A four-year minimum followup. Am J Sports Med. 1996 Jul;24(4):459–467.
29. Pelet S, Jolles BM, Farron A. Bankart repair for recurrent anterior glenohumeral instability: results at twenty-nine years' follow-up. J Shoulder Elbow Surg. 2006 Mar;15(2):203–207. doi:10.1016/j.jse.2005.06.011
30. Plancher KD, Lipnick SL. Analysis of evidence-based medicine for shoulder instability. Arthroscopy. 2009 Aug;25(8):897–908. doi:10.1016/j.arthro.2009.03.017
31. Poolman RW, Swiontkowski MF, Fairbank JCT, Schemitsch EH, Sprague S, de Vet HCW. Outcome instruments: rationale for their use. The Journal of Bone and Joint Surgery. 2009 May;91 Suppl 3:41–49. doi:10.2106/JBJS.H.01551
32. Porcellini G, Campi F, Pegreffi F, Castagna A, Paladini P. Predisposing factors for recurrent shoulder dislocation after arthroscopic treatment. The Journal of Bone and Joint Surgery. 2009 Nov;91(11):2537–2542. doi:10.2106/JBJS.H.01126

33. Purchase RJ, Wolf EM, Hobgood ER, Pollock ME, Smalley CC. Hill-sachs "remplissage": an arthroscopic solution for the engaging hill-sachs lesion. Arthroscopy. 2008 Jun;24(6):723–726. doi:10.1016/j.arthro.2008.03.015
34. Robinson CM, Howes J, Murdoch H, Will E, Graham C. Functional outcome and risk of recurrent instability after primary traumatic anterior shoulder dislocation in young patients. J Bone Joint Surg Am. 2006 Nov;88(11):2326–2336. doi:10.2106/JBJS.E.01327
35. Rowe CR, Patel D, Southmayd WW. The Bankart procedure: a long-term end-result study. J Bone Joint Surg Am. 1978 Jan;60(1):1–16.
36. Sachs RA, Lin D, Stone ML, Paxton E, Kuney M. Can the need for future surgery for acute traumatic anterior shoulder dislocation be predicted? J Bone Joint Surg Am. 2007 Aug;89(8):1665–1674. doi:10.2106/JBJS.F.00261
37. Söderlund T, Mattila VM, Visuri TI, Pihlajamäki HK. Long-term outcome of a transglenoid suture technique for anterior shoulder instability in young adults. J Bone Joint Surg Br. 2008 Feb;90(2):189–193. doi:10.1302/0301-620X.90B2.20040
38. Uhorchak JM, Arciero RA, Huggard D, Taylor DC. Recurrent shoulder instability after open reconstruction in athletes involved in collision and contact sports. Am J Sports Med. 2000 Nov;28(6):794–799.
39. van Kampen DA, van Beers LWAH, Scholtes VAB, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg. 2012 Jun;21(6):808–814. doi:10.1016/j.jse.2011.09.026
40. Vermeiren J, Handelberg F, Casteleyn PP, Opdecam P. The rate of recurrence of traumatic anterior dislocation of the shoulder. A study of 154 cases and a review of the literature. Int Orthop. 1993 Dec;17(6):337–341.
41. Voos JE, Livermore RW, Feeley BT, Altchek DW, Williams RJ, Warren RF, et al. Prospective evaluation of arthroscopic bankart repairs for anterior instability. Am J Sports Med. 2010 Feb;38(2):302–307. doi:10.1177/0363546509348049

Chapter 7

The "value added" of neurocognitive testing following sport-related concussion

Derk A. Van Kampen
Mark R. Lovell
Michael W. Collins
Jamie Pardini
Freddie H. Fu

The American Journal of Sports Medicine. 2006 Oct;34(10):1630

ABSTRACT

Background: Neurocognitive testing has been endorsed as a "cornerstone" of concussion management by recent Vienna and Prague meetings of the Concussion in Sport Group. Neurocognitive testing is important given the potential unreliability of athlete self-report after injury. Relying only on athletes' reports of symptoms may result in premature return of athletes to contact sport, potentially exposing them to additional injury.

Hypothesis: Use of computer-based neurocognitive testing results in an increased capacity to detect postconcussive abnormalities after injury.

Methods: High school and college athletes with a diagnosed concussion were tested 2 days after injury. Postinjury neurocognitive performance (Immediate Postconcussion Assessment and Cognitive Testing) and symptom (postconcussion symptom) scores were compared with preinjury (baseline) scores and with those of an age- and education-matched noninjured athlete control group. "Abnormal" test performance was determined statistically with Reliable Change Index scores.

Results: Sixty-four percent of concussed athletes reported a significant increase in symptoms, as judged by postconcussion symptom scores, compared with preinjury baseline at 2 days after injury. Eighty-three percent of the concussed sample demonstrated significantly poorer neurocognitive test results relative to their own baseline performance. The addition of neurocognitive testing resulted in a net increase in sensitivity of 19%. Ninety-three percent of the sample had either abnormal neurocognitive test results or a significant increase in symptoms, relative to their own baseline; 30% of a control group demonstrated either abnormalities in neurocognitive testing or elevated symptoms, as judged by postconcussion symptom scores. For the concussed group, use of symptom and neurocognitive test results resulted in an increased yield of 29% overreliance on symptoms alone. In contrast, 0% of the control group had both symptoms and abnormal neurocognitive testing.

Conclusion: Reliance on patients' self-reported symptoms after concussion is likely to result in under diagnosis of concussion and may result in premature return to play. Neurocognitive testing increases diagnostic accuracy when used in conjunction with self-reported symptoms.

INTRODUCTION

Sports-related concussion is a transient neurologic condition that occurs as a result of traumatic biomechanical force.[1] Symptoms may include confusion, disorientation, memory loss, motor unsteadiness, dizziness, headache, or visual disturbances. These symptoms usually occur with no detectable pathologic changes, and traditional neurodiagnostic tests such as CT, MRI, and electroencephalogram are generally insensitive in measuring the subtle neurologic changes after injury.[17] Recent research has indicated that sports-related concussion is a very common injury and that a minimum of 1.5 million concussion injuries occur in American football in the United States alone.[2]

The diagnosis and management of sports-related concussion have traditionally relied heavily on an athlete's self- report of symptoms, but these symptoms may not always be accurately reported to team medical personnel. However, as many clinicians have recognized, and recent research has suggested, an exclusive reliance on the athlete's report of symptoms may result in potential exposure to additional injury.[16,20]

Recent research has demonstrated that even in mildly concussed athletes, there can be a pronounced memory decline, lasting for at least 7 days after injury.[10,18] These data have led to a reexamination of previous return-to-play guidelines and a reconsideration of return-to-play standards that were heavily symptom based. More recently, neurocognitive testing has been endorsed as a "cornerstone" of concussion management by the Vienna Concussion in Sports Group. Specifically, neurocognitive testing has been identified as a helpful piece of additional information to assist in diagnosing and managing concussions.[1] This position was reaffirmed by a second international conference held in Prague in 2004.[23] The role of neurocognitive testing in the diagnosis and management of concussion has been emphasized because of the potential unreliability of athlete self- report of symptoms. The minimization of postconcussion symptoms (PCS) is a well-known phenomenon at all levels of competition.[7,20,21] An athlete's apparent fear of removal from a game or of losing his or her position on the team may tempt some athletes to deny or underreport postconcussive symptoms. Furthermore, prior research has suggested that premature return to play may be a particularly dangerous practice in children given a likely heightened degree of vulnerability in this group.[4,5,9]

Despite the widespread acceptance of neurocognitive testing in professional, collegiate, and high school sports, few studies have been completed regarding the clinical utility of neurocognitive testing relative to player report of symptoms.[6,8,18,19,21] In addition, although most concussion protocols espouse "return to baseline" on neurocognitive testing before return to sport activity,this fails to take into account test error or "practice effects" as a result of multiple exposures to the test or test battery.[2,15,20]

Immediate Postconcussion Assessment and Cognitive Testing (ImPACT) is a computer-based neurocognitive test battery designed specifically for sports-related concussion. This is a widely used program, allowing completion of neurocognitive testing in an expeditious and

standardized manner. The ImPACT test battery has undergone extensive validation through multiple studies and is currently used throughout professional and amateur sports.[11-14,18,19,28]

This study was designed to evaluate the individual and combined sensitivity and specificity of player symptom reporting, as judged by PCS score and neurocognitive testing in a group of high school and college athletes. Athletes were evaluated 2 days after concussion, and their test results were compared with the on-field diagnoses by a medical doctor or certified athletic trainer. The on-field diagnosis by medical staff has traditionally represented the "gold standard" for concussion diagnosis. We hypothesized that the use of computer-based neurocognitive testing (ImPACT) would result in an increased capacity to detect postconcussive abnormalities, compared with PCS alone, in a large group of athletes with diagnosed concussions.

MATERIAL AND METHODS

This study received approval from the University of Pittsburgh Institutional Review Board. All concussed athletes (N = 122) had undergone preseason baseline testing with ImPACT and had completed at least 1 follow-up evaluation within 2 days of injury. Athletes within the concussed sample were included from high schools and colleges within the states of Pennsylvania, Michigan, Illinois, Oregon, Maine, and California. This ongoing clinical program implements the use of baseline and postinjury neurocognitive testing to assist team medical staff in making return-to-play decisions after the occurrence of sports-related concussions.

All athletes in our clinical sample were included with the exception of athletes with a history of attention deficit disorder or a psychiatric disorder for which they were receiving medication. No athletes were included with a history of seizures or any other known neurologic disorder. To take into account the possible impact of prior concussions in the injured sample, a series of analyses was conducted to evaluate group differences between athletes with and without a history of concussion. There were no statistically significant differences in ImPACT test performance or in symptom reporting at either baseline or postconcussion with the exception of differences between the 2 groups at baseline with regard to the verbal memory composite score ($t = -2.72$, $p < 0.007$). However, the group with a history of concussion actually performed better than the no concussion group.

For the purpose of comparison, a sample of 70 non concussed athletes composed a control group. This group underwent baseline testing followed by a second evaluation within 1 week of baseline testing to determine test- retest fluctuations. This group was employed in this study to allow a direct comparison of changes in ImPACT and PCS scores during 2 successive testing periods (as was the case with the concussed group). In addition, this control group made possible the completion of statistical analyses to evaluate the specificity of ImPACT testing and PCS scores.

For this study, concussion was defined as a "traumatically induced alteration in mental status with or without a loss of consciousness," based on the standard American Academy

of Neurology nomenclature.[16] In addition to alteration of consciousness, athletes were diagnosed with concussion if they reported other typical symptoms of injury, such as headache, dizziness, balance dysfunction, or nausea, after a blow to the head or body. All injuries were diagnosed by a physician or certified athletic trainer who was present at the time of the injury.

The test battery used in this study was ImPACT.[22] The computer-based neurocognitive assessment tool includes a demographic questionnaire, symptom inventory, injury evaluation form, and a 20-minute neurocognitive test battery. The standardized demographic questionnaire requires the athlete to document relevant educational, sports participation, and personal medical history. This section also requires the athlete to report each prior concussion that had been formally diagnosed by a team physician or a certified athletic trainer. Also, ImPACT contains the 22-item PCS scale, which is also administered along with the test battery. The PCS scale evaluates common postconcussive symptoms (such as headache, nausea, dizziness, and trouble sleeping) as rated by the athlete on a Likert scale from 0 (asymptomatic) to 6 (symptomatic) according to his or her condition at the moment of testing.

The ImPACT test battery evaluates multiple aspects of cognitive functioning and is relatively brief. The entire battery, including the demographic information and PCS scale, takes less than 25 minutes to administer, is automatically scored, and produces a 6-page report that is complete with age- referenced percentile scores for select indices. The ImPACT test battery is heavily oriented toward the evaluation of attention, visual scanning, and information processing, although it also evaluates visual memory, verbal memory, and visual motor speed. Multiple studies using the ImPACT test battery have indicated that it is both reliable and valid. For example, Iverson et al found no significant practice effects in a sample of noninjured high school athletes tested twice within several days.[13] With regard to validity studies, the ImPACT test battery has been found to correlate highly with the Symbol Digit Modalities Test, an often-used test of cognitive speed in research with athletes.[14] This test battery has also demonstrated good sensitivity and specificity in prior studies of young athletes, and ImPACT has the capability to discriminate even mildly concussed high school athletes.[24,25,28] It has also been found to correlate with athlete self-report of neurocognitive decline and "fogginess."[11]

Table I provides a listing of the individual ImPACT tests and a description of neurocognitive abilities assessed. From these 6 tests, 4 separate composite scores are generated: verbal memory, visual memory, visual motor speed, and reaction time. In addition, an impulse control composite score is calculated that serves as 1 indicator of test validity. These composite scores were constructed to measure the broad neurocognitive domains that their names suggest, and recent validity studies have indicated good convergence with more traditional neuropsychological tests.[14] Multiple composite scores were constructed to reflect the reality that athletes who have suffered a concussion may present with different neurocognitive deficits depending on the biomechanics of their injuries, their ages, and a variety of other factors.[9,26,27] Therefore, no one score can be used to assess severity of injury. The administration of the ImPACT test battery was supervised by a team of clinical neuropsychologists, athletic trainers, and/or physicians who were trained and supervised in the administration of the standardized inventory.

Table I. ImPACT Neurocognitive Test Battery

Test Name	Neurocognitive Domain Measured
Word Memory	Verbal recognition memory (learning and retention)
Design Memory	Spatial recognition memory (learning and retention)
X's and O's	Visual working memory and cognitive speed
Symbol Match	Memory and visual-motor speed
Color Match	Impulse inhibition and visual-motor speed
Three letter memory	Verbal working memory and cognitive speed
Symptom Scale	Rating of individual self-reported symptoms
Composite Score	Contributing Score
Verbal Memory	Averaged percentage correct scores for the Word Memory (learning and delayed), Symbol Match memory test, and Three Letters Memory test
Visual Memory	Average percentage correct scores for the Design Memory (learning and delayed) X's and O's test
Reaction Time	Mean time in miliseconds for the X's and O's (mean correct counted reaction time), Symbol Match (mean weighted reaction time for correct responses), Color Match (mean reaction time for correct rersponse)
Visual motor processing speed	X's and O's (mean correct distracters), Symbol Match (mean correct responses), and Three Letters Memory (number of correct numbers correctly counted)

The ImPACT test battery, including the PCS scale, was administered within 2 days of injury. All of the data obtained from the administration process were automatically generated within the ImPACT clinical report and used in the current analysis.

Significant declines in test scores after concussion and significant increases in symptom scores were determined by the application of Reliable Change Index (RCI) (this is also known as the smallest detectable change) scores as described by Iverson et al and presented in Figure I.[13] The use of RCI scores is an increasingly popular method to account for practice effect and other factors that can influence test scores over repeated testing.[3,10,25] The RCI scores allow a clinician to account for measurement error surrounding test-retest difference score and therefore adjust each score for practice effects secondary to multiple exposures to the particular test.

For this study, RCI indices were established for the verbal memory, visual memory, reaction time, visual motor processing speed, and PCS composite scores produced in the ImPACT report. An athlete's test performance was deemed to be reliably different relative to his or her own baseline if the difference between postconcussion and baseline scores on a given composite index of ImPACT was larger than the established RCI scores, as determined in previously published research by Iverson et al Iverson et al have used these RCI scores in researching the ImPACT test battery by testing 56 healthy "not concussed" athletes twice within a few days to examine their test- retest reliability, practice effect, and reliable change parameters and to ultimately determine the normal variability of testing.[13] Whenever an athlete exceeds these normal variations, he or she is judged as abnormal on the test score in question. For example,

Figure I. *Reliable Change Index Score Formula (also known as the Smallest Detectable Change).*

$SEM_1 = SD\sqrt{1 - r_{12}}$ *Standard deviation from time 1 multiplied by the square root of 1 minus the test-retest coefficient.*

$SEM_2 = SD\sqrt{1 - r_{12}}$ *Standard deviation from time 2 multiplied by the square root of 1 minus the test-retest coefficient.*

$S_{diff} = \sqrt{SEM_1^2 + SEM_2^2}$ *Square root of the sum of the squared SEMs for each testing occasion.*

SEM, standard error of the mean

because the established RCI value for verbal memory is 8.75, any decline on this index (relative to baseline) that exceeds this value is rated as significantly different. Because the verbal memory composite scores are expressed as integers, a score that has decreased by 9 points or more would be categorized as abnormal. Additional RCI values are provided in Table II.

Tabel II. Group Means and RCI values for ImPACT Composite Scores

ImPACT Composite Score	Concussed Group Baseline	Concussed Groep at Follow-Up	RCI Value (.80) Confidence Interval
Verbal Memory	85.7 (8.9)	76.0 (14.4)	8.75
Visual Memory	74.0 (12.8)	64.3 (13.8)	13.55
Reaction Time	0.57 (0.08)	.64 (.13)	0.06
Processing Speed	36.0 (6.8)	32.7 (8.6)	4.98
Symptom Report	6.8 (9.6)	25.6 (19.9)	9.18

RCI: Reliable Change Index

Postconcussion Evaluation

All the athletes in our study had taken a preinjury (baseline) test from which difference scores could be calculated after injury. Whenever an athlete experienced a concussion during the period 2001 to 2004, he or she was referred for evaluation, which involved completion of the ImPACT test and PCS score. Concussions were diagnosed on the basis of the criteria described earlier.

Statistical Analysis

Abnormal performance was determined by comparing the athlete's postinjury scores to his or her baseline performance. Deviations from baseline performance larger than the established RCI scores for the particular composite score were deemed to be abnormal. Statistical differences in concussion classification using symptoms and ImPACT test results were determined via X^2 analyses. All statistical calculations were performed with the Statistical Package for the Social Sciences (SPSS Science Inc, Chicago, Ill).

RESULTS

Sample characteristics are displayed in Table III for both concussed and control groups. Ninety-seven of the 122 concussed athletes (80%) were high school students, and 25 (20%) were college students. The control group was composed of 50 (71%) high school and 20 college (29%) athletes. Mean education level for the collective sample was 10.2 years (range, 8-15 years). The concussed sample was largely male athletes (82%), whereas the control group consisted of more female than male athletes (54%). American football athletes composed a majority of the concussed sample (68%). Eleven percent were soccer athletes, 8% were basketball athletes, and the remaining 13% competed in ice hockey, wrestling, or lacrosse. For the control group, 50% were swimmers, 24% were soccer players, 17% were track athletes, and the remaining athletes participated in wrestling and lacrosse. With regard to concussion history, 76% of the

Table III. Demographic data of the 170 concussed athletes

Variable	Concussed Subjects (N=122)	Controle Subjects (N=70)
Mean (SD) age (yrs)	16.7 (12-27)	17.3 (14-22)
Mean (SD) education (yrs)	10.3 (8-16)	10.9 (8-16)
Highschool, %	80	71
College, %	20	29
Previous concussions, %		
0	76	90
1	14	10
2	8	0
3	2	0
Gender: male, %	82	47
Sport, %		
American Football	68.0	0
Soccer	11.0	24
Basketball	7.6	0
Swimmers	0	50
Track	0	17
Other	14.4	26
Time between injury to testing (days)	2	3
On-field markers[a]		
Positive LOC[b]	12.3	NA
Retrograde Amnesia	53,5	NA
Anterograde Amnesia	1.8	NA
Confusion	17.8	NA

[A] *Because of the natural difficulty of collecting on-field markers, some data were missing.*
[B] *LOC=Loss of conscious*

concussed sample had no prior concussion, and 24% had a history of concussion. Fourteen per- cent of the concussed sample had a history of 1 prior injury. Eight percent of the sample had experienced 2 prior concussions, and only 2% had experienced 3 or more concussions. The control group had a slightly lower rate of concussion, with only 10% of the group having experienced a past concussion and none of the group having more than 1 concussion.

Based on their total PCS scores only, 64% of the athletes reported an increase in symptoms from their baseline that exceeded the RCI score. In contrast, only 9% of the control group reported a subjective increase in symptoms from baseline to their second evaluation ($x^2 = 55.4$, $P < 0.000$). Eighty-three percent of the concussed sample demonstrated at least 1 ImPACT score that exceeded the RCI for that score, whereas 30% of the control group had 1 abnormal ImPACT score. Therefore, the addition of neurocognitive testing resulted in an increase in sensitivity from 64% to 83%, a net increase of 19% for the concussed group.

When either the symptom score or at least 1 neurocognitive test result was abnormal, 93% of concussed athletes were correctly identified as concussed when compared with the gold standard of on-field diagnosis. Whereas 30% of the control group did have 1 abnormal ImPACT score, no one (0%) in the control group had both abnormal neurocognitive performance and an increase in symptoms. Overall, the predictive value of having an abnormal PCS score was 93%, but the predictive value of not having an elevated symptom score was only 59%. If ImPACT was used in the absence of symptom data, the predictive value of having at least 1 abnormal neurocognitive test score was 83%, and the predictive value of a negative test result was 70%. However, when criteria for concussion classification were changed to requiring at least 1 abnormal ImPACT test and an abnormal PCS score, the predictive value of neurocognitive testing was 81%, and the predictive value of having a negative score was 83%.

DISCUSSION

Concussion has become a major public health issue because of the risk of both short- and long-term morbidity. Historically, return-to-play guidelines have relied heavily on the athletes' self-reports of symptoms. However, overreliance on athlete symptoms has recently been criticized based on t he tendency of some athletes to underreport symptoms, presumably in an attempt to speed their return to the playing field.[20] We present data in this study that suggest reliance on symptoms alone is inadequate and is likely to lead to missed diagnosis of the injury in a significant number of athletes. We found that only 64% of our recently con- cussed sample reported a significant increase in symptoms on the PCS scale within 2 days of evaluation. Adding neurocognitive testing increased the number of athletes who were identified as being abnormal to 83%. However, if a significant increase in symptom self-report and a decline on neurocognitive testing were used as classificatory criteria, the "diagnostic yield" increased to 93% compared with the gold standard of on-field diagnosis. Furthermore, we found that although 93% of our concussed sample had either ImPACT or symptom scores that fell within

the abnormal range compared to baseline level, none of the nonconcussed sample of athletes had both abnormal symptoms and abnormal ImPACT performance. These findings support previous studies that have indicated an imperfect agreement between self-reported symptoms and decreased neurocognitive test scores after concussion. [6,10,18]

This is the first study to formally evaluate the sensitivity and specificity of the ImPACT test when used in combination with athlete report of symptoms. Given these results, it is of concern that most return-to-play decisions after concussion have relied heavily on the athlete's self- report of symptoms. In fact, in many sports settings, return-to-play decisions have been based almost exclusively on the self-reported symptoms.[4,16] This study demonstrates that even athletes who report being symptom free may continue to exhibit neurocognitive deficits that they are either unaware of or are failing to report.

Recently, the Concussion in Sports Group recommended the use of neurocognitive testing in conjunction with other diagnostic information such as symptoms.[1,23] This current study provides support for this recommendation. Furthermore, our data suggest that if neurocognitive testing is unavailable, the treating physician should be cautious in returning athletes to play based on their self-report of symptoms. This study also provides preliminary support for the use of the ImPACT composite scores as diagnostic indicators, with a higher number of abnormal composite scores suggesting a more severe concussion. In this study, 2 abnormal ImPACT scores did not occur in any of the non- concussed athletes and may provide a clear marker of injury. However, this is not to suggest that athletes with 1 abnormal ImPACT score are presumed to be normal. Clearly, further study of the individual and aggregate use of ImPACT scores to evaluate the recovery process is needed.

We recognize several limitations with this study. First, our approach used a rigorous statistical method for determination of significant change after concussion, rather than a clinical approach. Therefore, given the relatively conservative nature of RCI scores, it is possible that we may have failed to correctly classify milder concussions in the sample whose scores did not fall outside of the RCI scores. Second, our sample primarily consisted of male high school and collegiate foot- ball players, which limits generalizability to other groups. In contrast, our control group consisted of athletes from more traditionally noncontact sports such as swimming and track and field. Therefore, our concussed and control groups were not identical. However, it is important to note that our assessment of significant change after injury was based on whether the athletes differed relative to their own baseline scores rather than comparison with the control group. Therefore, differences between the concussed and control group with regard to the sport participated in and concussion history did not affect the classification of athletes with regard to whether their test performance was normal or abnormal. In the future, we hope to continue to investigate the relationship of neurocognitive performance and athletes' report of symptoms in other sport groups outside of football. In addition, because the study was conducted exclusively with nonprofessional athletes, our findings should not be generalized beyond those sports levels. Recent published studies of professional football athletes in the United States have suggested a quicker recovery rate and no significant effect

of multiple injuries in this group when compared with younger nonprofessional athletes.[26,27] Therefore, the development of different RCI criteria based on age and level of competition may be useful, as recommended by the recent Prague conference.[23]

Based on the current study, we conclude that the use of neurocognitive testing (ImPACT) results in an increased sensitivity to detect postconcussion abnormalities. Therefore, we believe that neurocognitive assessment tools such as ImPACT provide "added value" to the more traditional assessment of symptoms.

REFERENCES:

1. Aubry M, Cantu R, Dvorak J, Graf-Baumann T, Johnston KM, Kelly J, et al. Summary and agreement statement of the 1st International Symposium on Concussion in Sport, Vienna 2001. 2002. p. 6–11.
2. Bailes JE, Cantu RC. Head injury in athletes. Neurosurgery. 2001 Jan;48(1):26–45– discussion 45–6.
3. Barr WB, McCrea M. Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. J Int Neuropsychol Soc. 2001 Sep;7(6):693–702.
4. Cantu RC. Cerebral concussion in sport. Management and prevention. Sports Med. 1992 Jul;14(1): 64–74.
5. Cantu RC. Posttraumatic Retrograde and Anterograde Amnesia: Pathophysiology and Implications in Grading and Safe Return to Play. J Athl Train. 2001 Sep;36(3):244–248.
6. Collins MW, Grindel SH, Lovell MR, Dede DE, Moser DJ, Phalin BR, et al. Relationship between concussion and neuropsychological performance in college football players. JAMA. 1999 Sep 8;282(10): 964–970.
7. Delaney JS, Lacroix VJ, Leclerc S, Johnston KM. Concussions during the 1997 Canadian Football League season. Clin J Sport Med. 2000 Jan;10(1):9–14.
8. Echemendia RJ, Putukian M, Mackin RS, Julian L, Shoss N. Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. Clin J Sport Med. 2001 Jan;11(1):23–31.
9. Field M, Collins MW, lovell MR, Maroon J. Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. J. Pediatr. 2003 May;142(5):546–553. doi: 10.1067/mpd.2003.190
10. Hinton-Bayre AD, Geffen GM, Geffen LB, McFarland KA, Friis P. Concussion in contact sports: reliable change indices of impairment and recovery. J Clin Exp Neuropsychol. 1999 Feb;21(1):70–86. doi: 10.1076/jcen.21.1.70.945
11. Iverson GL, Gaetz M, lovell MR, Collins MW. Cumulative effects of concussion in amateur athletes. Brain Inj. 2004 May;18(5):433–443. doi:10.1080/02699050310001617352
12. Iverson GL, Gaetz M, lovell MR, Collins MW. Relation between subjective fogginess and neuropsychological testing following concussion. J Int Neuropsychol Soc. 2004 Oct;10(6):904–906.
13. Iverson GL, lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. Clin Neuropsychol. 2003 Nov;17(4):460–467. doi:10.1076/clin.17.4.460.27934
14. Iverson GL, lovell MR, Collins MW. Validity of ImPACT for measuring processing speed following sports-related concussion. J Clin Exp Neuropsychol. 2005 Aug;27(6):683–689. doi:10.1081/13803390490918435
15. Jia X, Petersen SA, Khosravi AH, Almareddi V, Pannirselvam V, McFarland EG. Examination of the Shoulder: The Past, the Present, and the Future. The Journal of Bone and Joint Surgery. 2009 Nov 2; 91(Supplement 6):10–18. doi:10.2106/JBJS.I.00534
16. Kelly JP, Rosenberg JH. Diagnosis and management of concussion in sports. Neurology. 1997 Mar;48(3): 575–580.
17. Lovell M, Collins M, Bradley J. Return to play following sports-related concussion. Clin Sports Med. 2004 Jul;23(3):421–41– ix. doi:10.1016/j.csm.2004.04.001
18. Lovell MR, Collins MW, Iverson GL, Field M, Maroon JC, Cantu R, et al. Recovery from mild concussion in high school athletes. J. Neurosurg. 2003 Feb;98(2):296–301. doi:10.3171/jns.2003.98.2.0296
19. Lovell MR, Collins MW, Iverson GL, Johnston KM, Bradley JP. Grade 1 or "ding" concussions in high school athletes. Am J Sports Med. 2004 Jan;32(1):47–54.
20. Lovell MR, Collins MW, Maroon J. Inaccuracy of symptom reporting following concussion in athletes. Medicine and Science in Sports Exercise. 2002 Nov 19;(34):S298.

21. Lovell MR, Collins MW. Neuropsychological assessment of the college football player. J Head Trauma Rehabil. 1998 Apr;13(2):9–26.

22. Maroon JC, Lovell MR, Norwig J, Podell K, Powell JW, Hartl R. Cerebral concussion in athletes: evaluation and neuropsychological testing. Neurosurgery. 2000 Sep;47(3):659–69– discussion 669–72.

23. McCrory P, Johnston K, Meeuwisse W, Aubry M, Cantu R, Dvorak J, et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. 2005. p. 48–55.

24. Moser RS, Schatz P. Enduring effects of concussion in youth athletes. Arch Clin Neuropsychol. 2002 Jan;17(1):91–100.

25. Naugle RI, Chelune GJ, Cheek R, Lüders H, Awad IA. Detection of changes in material-specific memory following temporal lobectomy using the Wechsler Memory Scale-Revised. Arch Clin Neuropsychol. 1993 Oct;8(5):381–395.

26. Pellman EJ, lovell MR, Viano DC, Casson IR, Tucker AM. Concussion in professional football: neuropsychological testing—part 6. Neurosurgery. 2004 Dec;55(6):1290–303– discussion 1303–5.

27. Pellman EJ, lovell MR, Viano DC, Casson IR. Concussion in professional football: recovery of NFL and high school athletes assessed by computerized neuropsychological testing—Part 12. Neurosurgery. 2006 Feb;58(2):263–74– discussion 263–74. doi:10.1227/01.NEU.0000200272.56192.62

28. Schatz P, Pardini JE, lovell MR, Collins MW, Podell K. Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. Arch Clin Neuropsychol. 2006 Jan;21(1):91–99. doi:10.1016/j.acn.2005.08.001

Chapter 8

Summary and answers to the questions

SUMMARY AND ANSWERS TO THE QUESTIONS

This chapter summarizes the results of the studies that were performed for this thesis by answering the questions that were formulated in **Chapter 1** (Introduction and aims).

PART I: PATIENT REPORTED OUTCOME MEASURES (PROMS)

Question 1: Is the Dutch version of the Simple Shoulder Test a valid and reliable tool that can be used in Dutch shoulder patients?

The Simple Shoulder Test (SST) is an international common used PROM. Until now, there has been no Dutch equivalent of the SST, making international comparison of study results impossible. In **Chapter 2,** we created a Dutch translation of the SST, and we assessed the reliability and validity of this translation in a cohort of 110 patients with shoulder complaints.

Our study results showed that the Dutch version of the SST is a valid and reliable instrument for the assessment of patients with shoulder complaints. Thus, the SST can be used for clinical trials and for comparisons of study results from different countries. We recommend the use of the Dutch translation of the SST for evaluating patients with shoulder complaints.

Question 2: How many points should an individual patient improve on a PROM to experience a clinically relevant change?

In the current healthcare environment, it is increasingly important to evaluate the outcome of treatment. Therefore it is important to find out, from the patient's perspective, how many points a patient should improve on a PROM for the improvement to be considered a clinically relevant change. The Simple Shoulder Test (SST), the Disabilities of the Arm, Shoulder and Hand (DASH), the Shortened Disabilities of the Arm, Shoulder, and Hand Questionnaire (*Quick*DASH) and the Oxford Shoulder Score (OSS) are among the most commonly used PROMs for the shoulder. In **Chapter 3** we defined the measurement error as expressed as the Smallest Detectable Change (SDC) and the Minimal Clinically Important Change (MIC) for each of these four commonly used shoulder-related PROMs in a cohort of 164 patients with shoulder complaints.

Our study results showed that a patient with shoulder complaints should improve an average of 12–20% on a PROM for it to be considered a clinically relevant change that is not due to a measurement error. For the SST (range 0–12), this was 2.6 points. For the DASH and the *Quick*DASH (range 0–100), this was 16 points. For the OSS (range 12–60), this was 6 points.

PART II: THE DIAGNOSTIC VALUE OF COMBINING PATIENT INFORMATION AND CLINICAL TESTS

Question 3: Can the diagnostic value for detecting a rotator cuff tear be improved by combining patient characteristics, history and multiple clinical tests?

Rotator cuff tears are difficult to diagnose clinically. The aims of the study in **Chapter 4** were to determine the diagnostic value of individual clinical tests and to develop a prediction model that combined patient characteristics, history and clinical test results to improve the diagnosis of rotator cuff tears using magnetic resonance arthrography (MRA) as the reference standard. This study included 169 patients with shoulder complaints. A single experienced clinician conducted 25 clinical tests, 9 of which were specific for rotator cuff tears. All patients underwent MRA to determine the final diagnosis. In this cohort, 43 patients (26%) were diagnosed with rotator cuff tear based on the MRA. We determined the diagnostic value of individual tests based on each one's sensitivity, specificity and overall accuracy. We defined the diagnostic value of our prediction model for detecting rotator cuff tear as the area under the receiver operating curve (AUC) as explained in Chapter 4.

Our study showed that the individual clinical tests for rotator cuff tears had limited diagnostic value, with the accuracy ranging from 68–82%. Combining patient information and clinical tests improved the diagnostic value for rotator cuff tear compared to individual clinical tests with an AUC of 0.81. In our population of patients, rotator cuff tears were associated with higher age, no previous shoulder dislocation and a positive Neer test.

Question 4: Can the diagnostic value for detecting traumatic anterior shoulder instability be improved by combining patient characteristics, history and multiple clinical tests?

There is a lack of evidence regarding the diagnostic value of using a combination of patient characteristics and history to diagnose traumatic anterior shoulder instability. The aims of the study in **Chapter 5** were to determine the diagnostic values of individual clinical tests and to develop a prediction model that combined patient characteristics, history and clinical tests to improve the diagnostic value for detecting traumatic anterior shoulder instability with MRA as the reference standard. We used the same study population as in the study in Chapter 4. One experienced clinician conducted 25 clinical tests, 6 of which were specific for traumatic anterior shoulder instability. All patients underwent MRA to determine the final diagnosis. In this cohort, 60 patients (36%) were diagnosed with traumatic anterior shoulder instability based on the MRA. We defined the diagnostic values of individual tests based on sensitivity, specificity and overall accuracy. We defined the diagnostic value of our prediction model to detect traumatic anterior shoulder instability using the AUC as explained in Chapter 5.

Our study showed that individual clinical tests have good diagnostic value for traumatic anterior shoulder instability, with an overall accuracy of 81–86%. The prediction model, which combined patient information and clinical tests, clearly improved the diagnostic value of traumatic anterior shoulder instability compared to individual clinical tests. Our prediction model has high diagnostic value for traumatic anterior shoulder instability, with an AUC of 0.95. Within our population of patients, traumatic anterior shoulder instability was associated with young age, a history of shoulder dislocations, sudden onset of complaints and a positive release test. Using our prediction model in clinical practice could potentially decrease the need to perform MRA for the diagnosis of traumatic anterior shoulder instability. This study underscores the value of patient history and clinical examination for diagnosing traumatic anterior shoulder instability.

PART III: APPLICATION OF MEASUREMENT INSTRUMENTS IN CLINICAL PRACTICE

Question 5: What are the long-term results of arthroscopic shoulder stabilization?

Traumatic anterior shoulder instability is a common problem, and many patients are treated with arthroscopic shoulder stabilisation. This treatment seems to be successful in general, but long-term follow-up is lacking. In **Chapter 6** we conducted a prospective case series of 67 patients (70 shoulders) who were evaluated 8- to 10-years post-operatively after arthroscopic shoulder stabilisation with suture anchors.

Our study results showed a high recurrence rate of up to 35% after 8–10 years. This is higher than we expected and highlights the importance of long-term follow-up. New failures are seen after 2 years of follow-up, which is the follow-up time used in most clinical studies. In general, all patients had good shoulder function based on their scores on the PROMs. There was a clear difference in scores between the failed and successful shoulder stabilization groups when we looked at the instability-specific PROMs.

Question 6: Are neurocognitive testing results a valuable adjunct to a patient's self-reported symptoms for detecting post-concussive abnormalities after sports-related brain concussions?

In the field of sports medicine, there is a need for more standardised measurements for diagnosing sports-related concussions in athletes. Historically, the focus has been on self-reported concussion symptoms. However, there is some doubt about whether athletes accurately report symptoms to their team's medical personnel. Thus, there is a need for more objective data to determine better return-to-play recommendations. Neurocognitive testing has potential as a useful extra tool for diagnosing concussions. In **Chapter 7** we evaluated the diagnostic value of the players' self-reported symptoms in combination with neurocognitive chance scores

(baseline and post-concussion scores) for detecting concussions in a group of high school and college athletes (N=122).

Our study results showed that relying purely on patient-reported symptoms leads to an underdiagnosis of 36% of the concussed patients. Using a combination of significant changes in patient symptoms and/or significant decreases of neurocognitive test scores leads to improved sensitivity of 93%. We conclude that neurocognitive testing has added value for diagnosing post-concussion abnormalities. This study highlights the importance of using both clinician- and patient-based outcomes.

CONCLUSION

This thesis investigated the use of clinical information and PROMs for patient evaluation in orthopaedic surgery and sports medicine. We have shown that both are relevant measurement tools for the diagnosis of shoulder diseases and for the assessment of patient outcomes. PROMs can therefore be considered valuable tools that can help physicians better understand the patient's perspective in clinical orthopaedic practice and research.

Chapter 9

Discussion and future perspectives

DISCUSSION AND FUTURE PERSPECTIVES

The overall objective of this thesis was to study the added value of combining clinical information and Patient Reported Outcome Measures scores (PROMs) for evaluating patients in orthopaedic surgery and sports medicine. The most important findings are summarised in **Chapter 8**. In the current chapter, the findings of this thesis are placed in a broader perspective.

PART I: PATIENT REPORTED OUTCOME MEASURES (PROMS)

In **Chapter 2**, we validated the Dutch version of the Simple Shoulder Test (SST). This Dutch translation of the SST showed good validity and reliability and can now be used in the Netherlands in clinical studies. In 1993, Lippitt and Matsen suggested using specific questions from the SST to help differentiate among possible shoulder-related diagnoses.[8] In Part II of this thesis we used a few specific questions from the validated Dutch version of the SST to help diagnose rotator cuff tears and traumatic anterior shoulder instability in patients with shoulder complaints.

In **Chapter 3**, we determined the improvements (in points) in four different shoulder-related PROMs that were needed to show that a patient with shoulder complaints had a minimal clinically important change (MIC). In a research setting, this information can be used to calculate the percentage of patients who report a change greater than the MIC (i.e. the responders) in each arm of a trial, and then the percentages of responders can be compared.[13] This information can also be used to calculate required sample sizes for clinical studies.[15]

The use of PROMs to evaluate treatment effects in clinical practice is an important shift in paradigm in terms of evaluating orthopaedic treatment: It is useful for the patient to look at changes in his or her condition over time, it is important for the clinician to gain personal insight into the patient's performance and the results of PROMs can potentially be used to compare treatment results between different doctors or hospitals. At the same time, the PROM scores obtained in daily clinical practice could be used for research purposes. Standardised evaluations using PROMs make it possible to detect small differences between patients using different treatment protocols.[13]

For clinical practice and research purposes, it is important to have international consensus about which PROM or a combination of PROMs are best for use in a particular patient population, thus increasing the comparability of study results. International collaborations of orthopaedic surgeons will play an important role in selecting appropriate PROMs for different conditions.

One of the limitations of PROMs is that they contain many general questions that are not relevant to every patient. For example, "Can you walk 100 meters?" is not relevant to a patient who is very athletic. In such a patient, it would be preferable to evaluate the patient's capacity to run 5 or 25 kilometres. To address this problem, the Patient-Reported Outcomes Measurement Information System (PROMIS®) was developed in 2004 with the financial support of the US National Institutes of Health (NIH).[4,16] The PROMIS® group created a huge database with ranked questions based on commonly used PROMs. They developed a separate database for: physical-, mental- and social health. These questions can be administered computer based with a computer adaptive test (CAT) system. CAT is a dynamic process of test administration in which items are selected on the basis of the patients' responses to previously administered items.[7] This will lead to a more precise estimate of one's score with a minimal number of questions, limiting the burden for the patient. All PROMIS® instruments are scored on a same scale making it very easy to compare scores between patient populations. In the future this seems to be a potentially useful tool for patient based evaluation. However PROMIS® should first be validated in the general Dutch population and in different orthopaedic patients groups before it can be implemented in the Netherlands.

PART II: THE DIAGNOSTIC VALUE OF COMBINING PATIENT INFORMATION AND CLINICAL TESTS

Most previous clinical studies have focussed on the diagnostic value of individual clinical tests for a particular shoulder-related diagnosis. However, in practice, a single clinical test is often insufficient for diagnosing a specific problem. In addition, specific patient characteristics and information from the patient's history often influence the probability of a particular diagnosis. In Part II of this thesis, the aim was to find the best combination of medical signs, symptoms and other findings for predicting the probability of a specific shoulder-related disease.

In **Chapter 4**, in which we evaluated the diagnostic problems associated with diagnosing rotator cuff tears, it was shown that individual clinical shoulder tests had moderate diagnostic value for this clinical entity. The most important predictors for rotator cuff tears were higher age, no previous shoulder dislocation and a positive Neer test. For example, according to our internally-validated prediction model, a 30-year-old patient with a history of previous shoulder dislocation and a negative Neer test had only a 7% chance of rotator cuff tear. In contrast, a 60-year-old patient with no history of shoulder dislocation with a positive Neer test had a 72% chance of displaying abnormalities on the MRA that would indicate a rotator cuff tear. Consistent with the studies of Murrel and Park, we found that using a combination of age and results on a clinical test improved the diagnostic value for rotator cuff tears.[9,10] In contrast to their results, and in agreement with the results of Bak et al., we did not find that using a combination of multiple clinical tests improved the diagnostic value.[1] Our study results showed that many of the clinical tests for rotator cuff tear were interchangeable; therefore,

using a combination of clinical tests does not appear to have added diagnostic value over important factors such as patient age and history.

In **Chapter 5**, which evaluated traumatic anterior shoulder instability, it was shown that individual clinical shoulder tests provided good diagnostic value. Young age, a history of shoulder dislocation, sudden onset of complaints and a positive release test were the most important predictors of traumatic anterior shoulder instability. For example, according to our internally-validated prediction model, a 20-year-old patient with a previous shoulder dislocation, sudden onset of complaints and a positive release test had a 94% chance of traumatic anterior instability. In contrast, a 40-year-old patient with no history of shoulder dislocation, gradual development of complaints and a negative release test had only a 4% chance of displaying abnormalities on the MRA that would indicate traumatic anterior shoulder instability.

Our results are in accordance with a recent review by Hegedus et al., which also showed that the release test is the best clinical test for determining traumatic anterior shoulder instability.[5] The anamnesis of a previous shoulder dislocation is of course the most important history information to diagnose traumatic anterior shoulder instability. However, sometimes it is unclear whether there was a previous shoulder dislocation if the presentation was vague or if the patient does not have a good recollection of the traumatic event. Our study showed that even without a clear anamnestic shoulder dislocation, other parameters, such as sudden onset of complaints, young age and positive release tests can still guide the clinician to the diagnosis of traumatic anterior shoulder instability.

In this thesis, we developed individual prediction models for rotator cuff tear and traumatic shoulder instability. However, there are multiple other possible diagnoses related to shoulder complaints, such as SLAP, impingement, biceps pathology, glenohumeral joint arthritis and acromioclavicular joint arthritis. For clinicians, it would be useful to have one prediction model or decision tree that would be appropriate for use for any possible shoulder diagnosis to help determine which measurements should be performed, in which order and, if needed, what additional imaging should be performed in order to determine the diagnosis. It was not possible to develop this in our cohort, as the cohort sample size was not large enough. In the future, using a larger cohort of patients, it should be possible to develop one prediction rule or decision tree for use by any patient with shoulder complaints. Based on age, a few simple history questions and two or three specific clinical tests may be able to predict the probability of a certain diagnosis and provide recommendations for additional imaging tests. Finally, and most importantly, the model should provide the best evidence-based treatment plan. It can be assumed that this would lead to more efficient care and lower costs.

It is important to stress that our prediction model was developed for patients being treated at orthopaedic outpatient clinics; therefore, it may not be generalizable to primary care. The incidence of anatomical abnormalities is much higher for patients presenting to orthopaedic

clinics than for patients examined in primary care settings; therefore, the probability of finding anatomical abnormalities is much higher in an orthopaedic outpatient clinic.

However, primary care has a very important screening function for orthopaedic surgery. For example, previous studies have shown that for rotator cuff tears, early repair (within 3 months) gives better results; therefore, early detection is important.[2,11,14] Notably, primary care doctors generally have less access to MRI scanner or ultrasound machines and thus rely heavily on the patient's history and on the clinical examination. It will be very useful to perform a comparable study in the primary care setting.

PART III: APPLICATION OF MEASUREMENT INSTRUMENTS IN CLINICAL PRACTICE

In Part III of this thesis, a combination of clinician-based and patient-based information (PROMs) was used to evaluate patients in clinical practice. In **Chapter 6**, we analysed the long-term outcome of arthroscopic labrum repair for traumatic anterior shoulder instability, using PROMs to look at both the re-dislocation rate and subjective shoulder functioning. We found a found a high failure rate (re-dislocation) of up to 35% 8- to 10-years post-operatively. Our study showed that only the instability-specific questionnaires distinguished between patients with and without re-dislocation. This suggests that the use of a general shoulder PROM is not sufficient for evaluating traumatic anterior instability and will be part of future research in our institution. Specifically, in future studies it would be interesting to perform a randomized controlled trial that compared different surgical techniques for traumatic anterior shoulder instability by assessing re-dislocation rate and functional outcome using instability-specific PROMs.[3,6,13] The recent review by Rouleau et al. showed that the Western Ontario Shoulder Instability Index (WOSI) questionnaire has the best validity and reliability compared to other PROMs and that the WOSI is the most widely used PROM for shoulder instability outcome evaluation in clinical articles.[8,12]

One limitation of the WOSI is the use of a 100-mm VAS scale for all of the questions, which makes it necessary to measure each of the 21 items and then calculate subscales and total scores. This is time consuming, making it less practical for use in a busy clinical practice. A methodologically good alternative is the Oxford Shoulder Instability Score (OSIS), which contains only 12 questions and 5 answer options. We recommend the use of either the WOSI or the OSIS for traumatic anterior shoulder instability treatment evaluation. In the future, we intend to compare the measurement properties of these two PROMs in a single clinical study to help to determine the best option.

In **Chapter 7**, we compared the value of patient self-reported symptoms and neurocognitive testing for diagnosing sports-related concussion. We found that only 64% of patients with a recent concussion reported a significant increase in symptoms within 2 days of evaluation. However 83% of the concussed athletes showed significant neurocognitive impairment. Our study demonstrated that if a significant increase in self-reported symptoms and/or a decline in

neurocognitive testing scores were used as classifying criteria, the "diagnostic yield" increased to 93% compared with the gold standard of on-field diagnosis. In this study we showed that the combination of patient information and neurocognitive testing provided the best sensitivity for diagnosing brain concussion. We therefore recommend the use of neurocognitive assessment tools in addition to the more traditional assessment of symptoms in post-concussion evaluation.

OVERALL CONCLUSION

This thesis investigated the use of clinical information and PROMs for patient evaluation in orthopaedic surgery and sports medicine. We have shown that both are important for diagnosing and evaluating patients' outcome. This work should help clinicians and researchers to better understand and interpret PROMs and will hopefully motivate more clinicians to use PROMs as part of their standard care.

RECOMMENDATIONS FOR CLINICAL PRACTICE

- Start using PROMs in standard orthopaedic outpatient clinics for treatment evaluation.
- Develop international recommendations that indicate which PROMs should be used for which orthopaedic conditions.

RECOMMENDATIONS FOR FUTURE RESEARCH

- Determine the SDC and MIC of other PROMs.
- Validate PROMIS® for different orthopaedic conditions.
- Determine the best PROM for traumatic anterior shoulder instability treatment evaluation by comparing the WOSI and the OSIS in a clinical study.
- Use PROMs as outcome measures in addition to clinician-based outcomes in research.
- Perform external validation of the two prediction models developed in this thesis.
- Develop a general decision tree for patients with shoulder complaints.

REFERENCES

1. Bak K, Sørensen AKB, Jørgensen U, Nygaard M, Krarup AL, Thune C, et al. The Value of Clinical Tests in Acute Full-Thickness Tears of the Supraspinatus Tendon: Does a Subacromial Lidocaine Injection Help in the Clinical Diagnosis? A Prospective Study. YJARS. 2010 Jun 1;26(6):734–742. doi:10.1016/j.arthro.2009.11.005
2. Bassett RW, Cofield RH. Acute tears of the rotator cuff. The timing of surgical repair. Clin Ortho Relat Res. 1983 May;(175):18–24.
3. Freedman KB. Open Bankart Repair Versus Arthroscopic Repair With Transglenoid Sutures or Bioabsorbable Tacks for Recurrent Anterior Instability of the Shoulder: A Meta-analysis. Am J Sports Med. 2004 Sep 1;32(6):1520–1527. doi:10.1177/0363546504265188
4. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. J Rheumatol. 2009 Sep;36(9):2061–2066. doi:10.3899/jrheum.090358
5. Hegedus EJ. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med. 2012 Jul 7;doi:10.1136/bjsports-2012-091066
6. Hiemstra LA, Sasyniuk TM, Mohtadi NGH, Fick GH. Shoulder strength after open versus arthroscopic stabilization. Am J Sports Med. 2008 May;36(5):861–867. doi:10.1177/0363546508314429
7. Khanna D, Krishnan E, Dewitt EM, Khanna PP, Spiegel B, Hays RD. Patient-Reported Outcomes Measurement Information System (PROMIS®) — The future of measuring patient reported outcomes in rheumatology. Arthritis Care Res. 2011 Nov 7;63 Suppl 11(S11):S486–S490. doi:10.1002/acr.20581
8. Lippitt SB, Harryman DT, matsen FA. The Shoulder: a balance of mobility and stability. American Academy of Orthopedic Surgeons Rosemont IL; 1993.
9. Murrell GA, Walton JR. Diagnosis of rotator cuff tears. Lancet. 2001 Mar 10;357(9258):769–770. doi:10.1016/S0140-6736(00)04161-1
10. Park HB, Yokota A, Gill HS, Rassi El G, McFarland EG. Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. J Bone Joint Surg Am. 2005 Jul;87(7):1446–1455. doi:10.2106/JBJS.D.02335
11. Petersen SA, Murphy TP. The timing of rotator cuff repair for the restoration of function. J Shoulder Elbow Surg. 2011 Jan;20(1):62–68. doi:10.1016/j.jse.2010.04.045
12. Rouleau DM, Farber K, MacDermid JC. Systematic review of patient-administered shoulder functional scores on instability. J Shoulder Elbow Surg. 2010 Dec 1;19(8):1121–1128. doi:10.1016/j.jse.2010.07.003
13. Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. Health Qual Life Outcomes. 2006;4:62. doi:10.1186/1477-7525-4-62
14. Tashjian RZ. Epidemiology, Natural History, and Indications for Treatment of Rotator Cuff Tears. Clin Sports Med. 2012 Oct 1;31(4):589–604. doi:10.1016/j.csm.2012.07.001
15. Terwee CB, Roorda LD, Knol DL, De Boer MR, de Vet HCW. Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol. 2009 Oct;62(10):1062–1067. doi:10.1016/j.jclinepi.2008.10.011
16. National Institutes of Health [Internet]. [cited 2012 Dec 17]. Available from: http://www.nihpromis.org/

Chapter 10

Nederlandse samenvatting en de beantwoording van de onderzoeksvragen

NEDERLANDSE SAMENVATTING EN DE BEANTWOORDING VAN DE ONDERZOEKSVRAGEN

Dit hoofdstuk vat de resultaten van dit proefschrift samen aan de hand van de in **Hoofdstuk 1** (Introduction and aims) geformuleerde onderzoeksvragen.

DEEL I: PATIËNT-GERAPPORTEERDE UITKOMSTMATEN (PROMS)

Vraag 1: Is de Nederlandse versie van de Simple Shoulder Test een valide en betrouwbaar instrument voor het evalueren van Nederlandse patiënten met schouder klachten?

De Simple Shoulder Test (SST) is een internationaal veel gebruikte patiënt-gerapporteerde uitkomstmaat (PROM). Tot nu toe was er geen officiële Nederlandse versie van deze vragenlijst, waardoor een internationale vergelijking van studie resultaten niet mogelijk was.

In **Hoofdstuk 2**, hebben we de SST vertaald en vervolgens gevalideerd in een cohort van 110 patiënten met schouder klachten. We vonden een hoge reliability (ICC, 0.92) en een hoge interne consistentie (Cronbachs alfa, 0.78). Wij adviseren het gebruik van de SST voor het evalueren van patiënten met schouder klachten.

Vraag 2: Hoeveel punten moet een patiënt verbeteren op een PROM voor een klinisch relevant verschil?

In de moderne gezondheidzorg is het belangrijk om het effect van een behandeling te evalueren. Daarom is het van belang om, vanuit het perspectief van de patiënt, te bepalen hoeveel punten een patiënt moet verbeteren op een PROM om een klinisch relevante verbetering te verwachten.

De SST, de Disabilities of the Arm, Shoulder and Hand outcome measure (DASH), de Shortened Disabilities of the Arm, Shoulder, and Hand Questionnaire (*Quick*DASH) en de Oxford Shoulder Score (OSS) zijn veel gebruikte PROMs voor de schouder.

In **Hoofdstuk 3**, hebben we de measurement error uitgedrukt als Smallest Detectable Change (SDC) en de Minimal Clinically Important Change (MIC) bepaald voor deze vier veel gebruikte schouder specifieke PROMs in een cohort van 164 patiënten met schouder klachten.

Onze studie toonde aan dat een patiënt met schouder klachten gemiddeld 12-20% moet verbeteren op een PROM voordat de verandering als klinisch relevant kan worden beschouwd en niet onderdeel is van de meetfout van het instrument.

Voor de SST (range 0–12), was dit 2.6 punten. Voor de DASH en de *Quick*DASH (range 0–100), was dit 16 punten. Voor de OSS (range 12–60), was dit 6 punten.

DEEL II: DE DIAGNOSTISCHE WAARDE VAN DE COMBINATIE VAN PATIËNT INFORMATIE EN LICHAMELIJK ONDERZOEK TESTEN

Vraag 3:Kan de diagnostische waarde voor het detecteren van een rotator cuff scheur worden verbeterd door het combineren van patiënt karakteristieken, anamnese en meerdere lichamelijk onderzoek testen?

Rotator cuff scheuren zijn moeilijk klinisch te diagnosticeren. Het doel van het onderzoek in **Hoofdstuk 4** was om de diagnostische waarde van de individuele lichamelijk onderzoek testen te bepalen en om daarnaast een predictie model te ontwikkelen welke zowel patiënt karakteristieken, anamnese en lichamelijk onderzoek testen combineert om zo beter tot de diagnose rotator cuff scheur te komen. Hierbij diende de magnetische resonantie met intra articulair contrast (MRA) als goudenstandaard. In totaal zijn 169 patiënten met schouder klachten geïncludeerd voor deze studie. Eén ervaren orthopedisch chirurg voerde in het totaal 25 lichamelijk onderzoek testen uit per patient, waarvan er 9 specifiek voor rotator cuff scheuren zijn. Vervolgens kregen alle patiënten een MRA om de definitieve diagnose vast te stellen. In dit cohort, werden 43 patiënten (26%) gediagnostiseerd met een rotator cuff scheur op basis van de MRA. We hebben de diagnostische waarde van de verschillende lichamelijk onderzoek testen bepaald op basis van sensitiviteit, specificiteit en accuratesse.

Onze studie toont aan dat individuele lichamelijk onderzoek testen voor rotator cuff scheuren een beperkt diagnostische waarde hebben: variërend van 68-82% in accuraatheid. Door het combineren van patiënt informatie en klinische testen is het mogelijk om de diagnostische waarde voor rotator cuff scheuren te verhogen. Dit leidt tot een Area Under the Curve (AUC) van 0.81. In onze populatie was het hebben van een rotator cuff scheur geassocieerd met een hogere leeftijd, geen eerdere schouder luxatie en een positieve Neer test.

Vraag 4: Kan de diagnostische waarde voor het detecteren van traumatische anterieure schouder instabiliteit worden verbeterd door het combineren van patiënt karakteristieken, anamnese en meerdere lichamelijk onderzoek testen?

Er is weinig bekend over de diagnostische waarde van de combinatie van patiënt karakteristieken, anamnese en lichamelijk onderzoek testen voor het stellen van de diagnose traumatische anterieure schouder instabiliteit. Het doel van de studie in **Hoofdstuk 5** was om de diagnostische waarde van de individuele lichamelijk onderzoek testen te bepalen en om daarnaast een predictie model te ontwikkelen die zowel patiënt karakteristieken, anamnese en lichamelijk onderzoek testen combineert om zo beter tot de diagnose traumatische anterieure schouder instabiliteit te komen. Hierbij diende MRA als goudenstandaard.

We hebben de zelfde patiënt populatie gebruikte als in **Hoofdstuk 4**. Eén ervaren orthopedisch chirurg voerde in het totaal 25 lichamelijk onderzoek testen uit, waarvan er 6 specifiek voor traumatische anterieure schouder instabiliteit zijn. Vervolgens kregen alle patiënten een MRA om de definitieve diagnose vast te stellen. In dit cohort, werden 60 patiënten (36%) gediagnostiseerd met traumatische anterieure schouder instabiliteit op basis van de MRA. We hebben de diagnostische waarde van de verschillende lichamelijk onderzoek testen bepaald op basis van sensitiviteit, specificiteit en de accuratesse.

Onze studie toont aan de individuele lichamelijk onderzoek testen voor traumatische anterieure schouder instabiliteit een goede diagnostische waarde hebben, variërend van 81-86% in accuraatheid. Door het combineren van patiënt informatie en klinische testen is het mogelijk om de diagnostische waarde voor traumatische anterieure schouder instabiliteit te verhogen met een AUC van 0.95. In onze populatie was het hebben traumatische anterieure schouder instabiliteit geassocieerd met een lage leeftijd, een voorgeschiedenis van schouder luxaties, plotseling ontstaan van de klachten en een positieve release test. Dit predictie model kan potentieel de noodzaak voor het maken een MRA voor het stellen van de diagnose traumatische schouder instabiliteit overbodig maken. Deze studie onderstreept de grote waarde van een anamnese en een lichamelijk onderzoek voor het stellen van de diagnose traumatische anterieure schouder instabiliteit.

DEEL III: DE TOEPASSING VAN MEETINSTRUMENTEN IN DE KLINISCHE PRAKTIJK

Vraag 5: Wat zijn de langer termijn resultaten van een arthroscopische schouder stabilisatie?

Traumatische anterieure schouder instabiliteit komt veel voor, en veel patiënten worden uiteindelijk behandeld door middel van een operatie. Meestal betreft dit een arthroscopische schouder stabilisatie. Deze behandeling lijkt op het eerste gezicht succesvol, echter er zijn weinig langetermijnresultaten in de literatuur. In **Hoofdstuk 6** hebben we de resultaten van een case serie van 67 patiënten (70 schouders) geëvalueerd. Deze patiënten hadden een gemiddelde follow-up duur van 8 tot 10 jaar na hun arthroscopische schouder stabilisatie.

Onze studie resultaten tonen een hoog percentage falen (gedefinieerd als opnieuw een schouder luxatie) van 35% na 8-10 jaar. Dit is een hoger percentage dan wij verwacht hadden op basis van de huidige literatuur en dit benadrukt het belang van lange termijn evaluatie van de behandelingsresultaten. Het bleek daarnaast dat patiënten ook nog na meer dan 2 jaar na de chirurgische behandeling nog een eerste recidief schouder luxatie opliepen. De meeste gepubliceerde resultaten over schouder stabilisatie hebben maar een follow-up van maximaal 2 jaar, dit lijkt dan ook een te korte follow-up. Over het algemeen hadden alle patiënten een

goede schouder functie op basis van hun score op de PROMs. Alleen de PROMs welke specifiek zijn ontwikkeld voor schouder instabiliteit toonde een duidelijk verschil in punten aan tussen de succesvol behandelde patiënten groep en de gefaalde groep.

Vraag 6: Zijn neurocognitieve testen resultaten een waardevolle aanvulling naast de zelf-gerapporteerde klachten van patiënten voor het detecteren van post-hersenschudding afwijkingen na een sport gerelateerde hersenschudding?

Op het gebied van de sportgeneeskunde is er behoefte aan een meer gestandaardiseerde manier voor het diagnosticeren van sport gerelateerde hersenschuddingen bij sporters. Historisch gezien lag de nadruk sterk op het zelf rapporteren van klachten. Echter, er is rede tot twijfel of sporters wel accuraat hun klachten rapporteren aan hun arts. Vandaar dat er behoefte is aan een meer objectieve informatie om zo beter adviezen te kunnen geven ten aanzien van belasting en het weer gaan sporten. Neurocognitieve testen lijken potentieel een waardevol aanvullend instrument voor het diagnosticeren van hersenschudding. In **Hoofdstuk 7** hebben we de diagnostische waarde van de combinatie van de zelf-gerapporteerde symptomen en de neurocognitieve test resultaten gebruikt voor het detecteren van hersenschuddingen. Dit hebben we geëvalueerd in een groep van 122 jongeren gediagnosticeerd met een sport gerelateerde hersenschudding.

Onze studie toont aan dat op basis van een significante verandering in zelf-gerapporteerde klachten, 36% van de patiënten met een hersenschudding worden gemist (niet gediagnostiseerd). Als men een combinatie van een significante verandering in klachten dan wel een significante verslechtering in neurocognitieve test resultaten gebruikt voor het diagnosticeren van een hersenschudding dan leidt dit tot een stijging van de sensitiviteit tot 93%. Wij concluderen dan ook dat het gebruik van neurocognitieve testen een meerwaarde heeft voor het diagnosticeren van post-hersenschudding afwijkingen. Deze studie onderstreept het belang van het gebruikt van zowel patiënt- en klinische informatie.

CONCLUSIE

In dit proefschrift hebben we gekeken naar de waarde van zowel klinische informatie en PROMs voor het evalueren van patiënten in de orthopedische chirurgie en de sportgeneeskunde. We hebben laten zien dat beide belangrijke instrumenten zijn voor het diagnosticeren van de oorzaak van de schouder klachten en voor het evalueren van de uitkomst van de patiënt. PROMs zijn een waardevolle instrument om een beter beeld te krijgen vanuit het perspectief van de patiënt.

Dankwoord

DANKWOORD

Terugkijkend is het ongelooflijk hoeveel mensen er betrokken zijn geweest om dit proefschrift tot een goed einde te brengen. Ik ben al die mensen dan ook zeer dankbaar.

Allereerst alle patiënten die in Nederland en Amerika alle vragenlijsten, vaak zelfs meerdere malen hebben ingevuld. Zonder hun informatie was er nooit een proefschrift geweest, veel dank daarvoor.

Prof. Dr. R.M. Castelein, beste René, ik voel mij zeer vereerd om bij jou te mogen promoveren. Je hebt erg goed sturing gegeven aan mijn promotieplan. Je kritisch oordeel op de verschillende artikelen hebben ervoor gezorgd dat de teksten leesbaar waren voor de algemene orthopeed. Zeer veel dank voor alle tijd die je erin hebt gestoken ondanks je zeer volle agenda.

Dr. J.W. Willems, beste Jaap, in 2008 kwam ik voor het eerst bij je met het plan om een studie op te zetten in het OLVG. Je was direct enthousiast en bent dat ook altijd gebleven. Heel veel dank voor het includeren van al jouw patiënten in onze studies. Je bent een van de allerbelangrijkste schakels in mijn hele onderzoeksperiode geweest. Heel veel dank daarvoor.

Dr. C.B. Terwee, beste Caroline, het feit dat jij betrokken bent geraakt bij dit onderzoek heeft er uiteindelijk voor gezorgd dat het onderzoek ook echt een promotietraject kon worden. Jouw kennis en kunde, maar vooral je enthousiasme om mij als 'simpele dokter' alles uit te leggen en te begeleiden, is fantastisch. Ik ben je zeer dankbaar voor alles wat je voor mij gedaan hebt.

Geachte leden van de Promotiecommissie, prof. dr. F.J.G. Backx, Prof. dr. R.L. Diercks, Prof. dr. F.P.J.G. Lafeber en Prof. dr. K.C.B. Roes, veel dank voor het kritisch lezen en positief beoordelen van het manuscript.

Beste Mary Boekhorst, ook jij hebt ongelooflijk veel tijd in dit project gestoken. Ik ben je zeer dankbaar voor het invoeren van de patiënten gegevens, e-mailen van alle patiënten en de controle of alle vragenlijsten wel netjes waren ingevuld. Veel dank, en geniet van je welverdiende pensioen.

Dr. H.J. van der Woude, beste Henk-Jan veel dank voor het beoordelen voor de grote hoeveelheid MR arthro's en je hulp bij het schrijven van de verschillende artikelen.

Beste Tobias van der Berg, ik heb erg prettig met je samengewerkt in het VUMC. Heel erg veel dank voor jouw hulp bij het ontwikkelen van de predictie modellen.

Beste Just van der Linde, ik heb met zeer veel plezier met jou samen aan onze eerste studie gewerkt, met een mooi resultaat. Ik ben dan ook zeer vereerd dat je mij hebt benaderd om mijn onderzoek voort te zetten voor patiënten met anterieure schouder instabiliteit klachten. Het gaat zeker lukken om over een paar jaar jouw proefschrift af te ronden, veel succes!

Prof. Dr. D.B.F. Saris, beste Daan, ik heb in mijn anderhalf jaar orthopedie opleiding zeer veel van je geleerd. Veel dank voor je zeer betrokken begeleiding bij mij als AIOS en als onderzoeker.

Dr. R.W. Poolman, beste Rudolf, je manier van werken en je wetenschappelijke ambities hebben de afgelopen jaren als een voorbeeld en bron van inspiratie voor mij gediend. Dank voor een erg goede tijd in het OLVG.

Prof. dr. F.H. Fu, dear Freddy, thank you for the great time in Pittsburgh and your help with my ACL - and concussion research. You are a great mentor.

Dr. M.R. Lovell, dear Mark, thank you for the fantastic guidance's during my time in Pittsburgh. Your concussion program has been very inspiring to me. You have been one of the first sparks in my interest in clinical epidemiology.

Loes van Beers en Thomas Nijman, het duo van de Joint Research. Erg veel dank voor jullie hulp bij mijn onderzoek en vooral de gezelligheid op N2.

Dr. V.A.B. Scholtes, beste Vanessa jouw rol binnen de Joint Research van de orthopedie OLVG, is een erg belangrijke. Jij hebt mij enorm geholpen met het structureren van mijn onderzoek en het tekstueel verbeteren van meerdere van mijn artikelen. Veel dank daarvoor.

Beste Simone Sienema, dank voor je ondersteuning bij het regelen van afspraken in de altijd overvolle agenda van Prof. Castelein en in het wegwijs maken in de formulieren en procedures die met een promotie gepaard gaan.

Dr. V.P.N. van der Hulst, beste Victor, dank voor jouw rol als derde partij bij de beoordeling van de MR arthro.

De maatschap Orthopedie van het Onze Lieve Vrouwe Gasthuis (OLVG). Al vele jaren werk ik met ongelooflijk veel plezier in het OLVG. Het is een bijzonder inspirerende en open werkomgeving waar ik veel heb kunnen leren over de algemene orthopedie. Erg veel dank voor een zeer goede tijd.

De medewerkers van de medische bibliotheek OLVG, jullie hebben zeer veel artikelen voor mij opgevraagd de afgelopen jaren. Erg veel dank daarvoor.

Joep Maeijer en Rik Nienhuis. Dank voor het opnemen, model staan en bewerken van alle lichamelijk onderzoek filmpjes voor in dit proefschrift.

Zonder de steun van Stichting Marti Keuning Eckhardt, het Anna Fonds, the research grant van the European Society for Surgery of the Shoulder and the Elbow en de andere sponsoren was dit proefschrift niet tot stand gekomen. Nogmaals veel dank voor financiële de steun.

Dr. E.J. Derksen en Dr. B.J. Dwars, best Eric en Boudewijn, mijn opleiders duo uit mijn vooropleiding. Ik heb echt een fantastische vooropleiding gehad in het Slotervaart. Dank voor de zeer prettige open sfeer waarin ik mij heb kunnen ontwikkelen. Daarnaast hebben jullie mij de ruimte gegeven om al tijdens mijn vooropleiding de verschillende lijnen uit te zetten voor wat uiteindelijk mijn proefschrift is geworden.

De Carpool: Remco, Sybrand, Blom, Marwaan, Bastiaan Heere. Stipt om 7:04 een date op een zanderige carpoolstrook bij Abcoude aan de A2. Om vervolgens met 3-4 man opgepropt in een oude 206 of Honda Civic (met kapotte voorruit) richting het UMC te rijden. Het had een dubieus Pools klusbedrijfje kunnen zijn, maar het was een team van aanstormende medische specialisten. En het klinkt misschien onaantrekkelijk, maar integendeel: terug in de file stoom afblazen na een weekje zaal in het UMCU is onbetaalbaar. Dank voor jullie morele steun, ik heb er van genoten!

Mijn dank gaat uit naar al mijn collega's uit vooropleiding en orthopedieopleiding. Ik heb een fantastische periode achter de rug. Ik heb er erg veel goede vrienden aan overgehouden.

Mijn paranimfen: Marijn en Rogier. Mijn buddy's uit Groningen! Bij jullie promotie trajecten kon ik natuurlijk niet achterblijven, dus ik ben ook maar aan de slag gegaan. Top dat jullie me willen bijstaan op 6 juni.

Mijn ouders, al 32 jaar besteden jullie ongelooflijk veel tijd en liefde aan mij. Altijd maar weer die dyslectische teksten controleren en meedenken over de toekomst. Heel veel dank voor jullie onvoorwaardelijke liefde.

Mijn schoonfamilie, mijn broer Pim en zijn vriendin Renske en al mijn andere vrienden: veel dank voor jullie steun en interesse, het vele oppassen op Famke en alle gezellig momenten.

Lieve Elske, uiteindelijk sta jij aan de basis van al dit onderzoek. Een epidemioloog thuis op de bank is voor een simpele orthopeed een wetenschappelijke droomsituatie. Je hebt me fantastische geholpen in alle fases van het proefschrift. Maar het meest gelukkig ben ik met ons heerlijke leven samen en onze prachtige dochter Famke.

About the author

ABOUT THE AUTHOR

Derk van Kampen was born on March 15th in Enschede, the Netherlands. In 1999 he graduated from high school (NSG Groenewoud, Nijmegen) and started to study medicine the same year at the University of Groningen. In 2004 he went to Pittsburgh for a sport medicine research fellowship of 6 month with Prof. Dr. FH Fu. Afterwards he did his clinical rotations in the Deventer Hospital. In 2006 he did his final internship at the Department of Orthopaedic Surgery at the University Medical Center Utrecht. After finishing medical school, he moved to Amsterdam, to start his residency for general surgery at the Slotervaart hospital, Amsterdam (head: Dr. BJ Dwars). During this time he started his research project together with Dr. WJ Willems and Dr. CB Terwee, EMGO-VUMC. In 2009 he started his training for orthopaedic surgery at the Onze Lieve Vrouwe Gasthuis, Amsterdam (head: Dr. WJ Willems). He was awarded with fellowships grants from Stichting Marti Keuning Eckhardt, Anna Fonds and the European Society for Surgery of the Shoulder and the Elbow. In July 2010 he continued his training at the University Medical Center Utrecht (head: Prof. Dr. DBF Saris). In January 2012 he returned to the Onze Lieve Vrouwe Gasthuis, Amsterdam (head: Dr. RW Poolman) to finish his training. During this last period he was able to spend 6 months full time on his research project to finish this thesis. The author is expected to finish his orthopaedic training in July 2013. During his orthopaedic surgery residency he participated in the training for clinical epidemiology. After completing his PhD he will be registered as clinical epidemiologist. Derk lives in Amsterdam together with his wife Elske Sieswerda and their daughter Famke.

Abbreviation list

ABBREVIATION LIST

AUC	Area under the ROC curve
CM	Constant-Murley shoulder assessment
DASH	Disability of the Arm, Shoulder and Hand
ImPACT	Immediate postconcussion assessment and cognitive testing
LOC	Loss of conscious
MIC	Minimal important change
OSIS	Oxford Shoulder Instability Score
OSS	Oxford Shoulder Score
PCS	Postconcussion symptoms
*Quick*DASH	Shortened Disabilities of the Arm, Shoulder, and Hand Questionnaire
RCI	Reliable change index (this is the same as SDC)
ROC	Receiver operating characteristic
SDC	Smallest detectable change
SF-36	Short Form (36) Health Survey
SST	Simple Shoulder Test
WOSI	Western Ontario Shoulder Instability Index

List of the 26 clinical shoulder tests used in this thesis

LIST OF THE 26 CLINICAL SHOULDER TESTS USED IN THIS THESIS

Anterior shoulder instability

1. Load and shift test

The load and shift test, also known as the push-pull test, was carried out with the patient in a sitting position.[17] The examiner grasped the patient's elbow with the corresponding hand. The examiner's other hand grasped the patient's upper arm. The examiner then positioned the patient's arm in 90° abduction in the scapular plane, in neutral rotation, and centred the patient's humeral head on the glenoid by applying a load along the axis of the humerus with the hand that was grasping the patient's elbow. The examiner then attempted to shift the patient's humeral head off the glenoid in the anterior direction. The test was considered positive when the humeral head could be shifted anteriorly off the glenoid (grade II and III) or when the patients displayed apprehension.[17]

2. Apprehension test

The apprehension test was carried out with the patient supine with the arm in 90° abduction, the elbow in 90° flexion, and maximum external rotation.[22] The examiner applied an anterior, external, rotatory force. The test was considered positive when the patient demonstrated an apprehensive feeling; when only pain was experienced, the result was not considered positive.

3. Relocation test

The relocation test was carried out immediately after the apprehension test.[9] The patient remained in the position that evoked symptoms, and the humeral head was depressed in the posterior direction with direct force to the humeral head. The test was considered positive when the applied posterior-directed force provided relief of the apprehensive feeling, and the patient was able to tolerate maximal external rotation.

4. Anterior release test

The anterior release test, also known as the surprise test, was carried out immediately after the relocation test.[5,23] The examiner suddenly released the pressure on the humeral head (while maintaining the patient's arm in the position of apprehension). The test was considered positive when the patient experienced a sudden apprehensive feeling

5. Anterior drawer test

The anterior drawer test was carried out with the patient in the supine position.[3] The patient's hand was positioned on the examiner's axilla. The patient's arm was abducted to 80° to 120°, 0° to 20° forward flexion, and 0° to 30° external rotation. With one hand, the examiner stabilised the scapula by applying pressure on the coracoid process. With the other hand, the examiner grasped the humerus and drew it out anteriorly. The test was considered positive when there was an increased translation of the humeral head compared to the other shoulder, or when the patient became apprehensive.

Posterior shoulder instability

6. Jerk test

The Jerk test was carried out with the patient in standing or sitting position with the patients arm abducted to 90° and internally rotated.[11] The examiner axially loads the humerus while the arm was moved horizontally across the body. The other hand stabilizes the scapula; A sharp pain with or without posterior clunk or click suggested a positive test result.

Shoulder laxity

7. Anterior drawer test

The anterior drawer test was carried out with the patient in the supine position.[3] The patient's hand was positioned on the examiner's axilla. The patient's arm was abducted to 80° to 120°, 0° to 20° forward flexion, and 0° to 30° external rotation. With one hand, the examiner stabilised the scapula by applying pressure on the coracoid process. With the other hand, the examiner grasped the humerus and drew it out anteriorly. The test was considered positive when there was an increased translation of the humeral head compared to the other shoulder, or when the patient became apprehensive.

8. Sulcus sign

The sulcus sign was carried out with the patient sitting with his or her arms hanging relaxed by the side. The examiner grasped the patient's elbows and pulled down on them. The sulcus, if any, that appeared in the subacromial area of the patient's shoulder was measured in centimeters. This can be performed in neutral and 90* of external rotation.

9. Hyperabductie test

The hyperabduction test was carried out with the patient standing, and the examiner standing behind the patient.[2] With the examiner's forearm, the shoulder girdle was pushed down firmly, while the examiner's other hand lifted the patient's upper limb, which was relaxed in abduction. During the test, the elbow was flexed at 90°, and the forearm was horizontal. The test was considered positive when the arm could be hyperabducted above 105°, or when the patient displayed apprehension.

SLAP lesion and biceps pathology

10. Clunk test

The Clunk test was performed with the patient supine, with the arm in full abduction.[15] The arm was turned into full external rotation while pressure was applied in anterior direction to the humeral head. The test was considered positive if the examiner heard a clunck or grinding.

11. Crank test

The crank test was performed with the patient supine or sitting with the arm elevated to 160 degrees in the scapular plane of the body, loaded axially along the humerus, and with maximal internal and external rotation.[14,24] The test was considered positive if the external rotation was painful or if a click was felt that reproduced the patient's symptoms of pain or catching.

12. New pain provocation test

The new pain provocation test was performed with the patient in the sitting position.[19] The patients arm was brought to 90° of abduction and the elbow flexed in 90°. Shoulder was rotated externally (comparable with the apprehension test). This was done in maximum supination- and pronation of the forearm. The test was considered positive when pain was provoked only when the forearm was in the pronated position or when pain was more severe in this position than with the forearm supinated.

13. O'Brien test

The O'Brien test was performed with the patient standing.[21] The patient's arm was in 90° of anteflexion, 10° adduction, and maximum internal rotation (thumb pointing downward), the elbow was extended. The examiner attempted to press the arm downward against the patient's resistance. The test was considered positive if the test provoked pain in side the shoulder.

14. Biceps load test

The biceps load test was performed with the patient in the supine position.[12] The examiner sits adjacent to the patient on the same side as the affected shoulder and gently grasps the patient's wrist and elbow. The arm was abducted to 90°, with the forearm in the supinated position. The patient was allowed to relax, and an anterior apprehension test was performed. When the patient becomes apprehensive during the external rotation of the shoulder, external rotation was stopped. The patient was then asked to flex the elbow while the examiner resists the flexion with one hand and asks how the apprehension has changed, if at all. If the apprehension was lessened, or if the patient feels more comfortable than before the test, the test was negative for a SLAP lesion. If the apprehension has not changed, or if the shoulder becomes more painful, the test was positive.

15. Speed Test

The Speed test was performed with the patient sitting or standing position with the elbow extended and the forearm supinated.[16] The patient was asked to elevate his/her arm forward to approximately 60 degrees while the examiner was resisted this motion. The test was considered positive if this elicits pain was localized around the bicipital groove area.

16. Yergason's Test

The Yergason's test was performed with the patient in sitting or standing position.[25] The elbow was flexed to 90° and the forearm was pronated. With the examiner holding the patient's wrist, the patient was asked to actively supinate against resistance. The test was considered positive if pain was localized around the bicipital groove area. This suggests disorder in the long head of the biceps tendon in its sheath.

Shoulder impingement and rotator cuff tear

17. Neer test

The Neer test was performed with the patient sitting or standing.[20] The ipsilateral scapula was stabilised with the examiner's hand, and the patient's arm was passively elevated forward. The test was considered positive when the patient experienced pain. In the original description, Neer advised giving an injection of lidocaine in the subacromial space to relieve pain. Due to time limitations in the orthopaedic outpatient setting, we decided not to give patients a lidocaine injection.

18. Empty can

The empty can test, also known as the Jobe test, was performed with the patient standing, the shoulder in 90° abduction in the scapular plane, and with full internal rotation.[8] The thumbs were pointing toward the floor. The patient maintained this position against downward resistance applied by the examiner. The test was considered positive when the patient demonstrated weakness or pain during the applied resistance.

19. Hawkins-Kennedy test

The Hawkins Kennedy test was performed with the examiner facing the seated or standing patient.[6] The patient's arm was elevated forward at 90°, and the elbow was flexed at 90°. The test was considered positive when pain occurred with passive internal rotation.

20. Painful arc

The painful arc test was performed with the patient standing.[10] The patient was asked to elevate the arm actively in the scapular plane, until the arm was fully elevated, and then, to let the arm down in the same arc. The test was considered positive when the patient demonstrated pain, or reported a painful catching between 60° and 120° elevation.

21. Drop-arm test

The drop arm test, also known as Codman's sign, was performed with the patient standing.[1] The patient was asked to abduct the arm fully, and then, to reverse the motion slowly, in the same arc. When the arm dropped suddenly, the test was considered positive.

22. Infraspinatus muscle strength

The infraspinatus muscle strength test was performed in sitting or standing position with the elbow flexed to 90° and the arm was adducted to the trunk in neutral rotation.[13] The examiner then applied an internal rotation force to the arm while the patient resisted. The test was considered positive if the patient gave way because of weakness or pain.

23. Exorotation lag sign

The external rotation lag sign was performed with the patient seated.[7] The elbow was passively flexed to 90°, and the examiner held the shoulder at 20° elevation (in the scapular plane), near maximal external rotation (i.e., maximum external rotation minus 5, to avoid elastic recoil in the shoulder). The patient was then asked to maintain the external rotation in elevation as the examiner released the wrist, but maintained support of the limb at the elbow. The sign was considered positive when a lag, or angular drop occurred.

24. Drop sign

The dropsign, also known as the infraspinatus drop sign, was similar to the ERLS, but the arm was held at 90° elevation (in the scapular plane) by the examiner, instead of the 20° elevation.[7]

25. Lift of test

The lift off test, also known as the Gerber test, was performed with the patient standing.[4] The patient was asked to place their hand on their back for maximum internal rotation, and then, to lift their hand off their back. The test was considered positive when the patient was not able perform this.

Acromio clavicular joint

26. cross-body adduction

The cross-body adduction test was performed with the patient in sitting or standing position with the arm in 90° of forward flexion.[18] The examiner then adducted the arm across the body of the patient in the horizontal plane. The test was considered to be positive if it caused pain in the shoulder.

REFERENCES:

1. Codman EA. The shoulder. Rupture of the supraspinatus tendon and other lesions in or about the subacromial bursa. Brooklyn, New York, USA: G Miller & Co Medical Publishers; 1934.
2. Gagey OJ, Gagey N. The hyperabduction test. J Bone Joint Surg Br. 2001 Jan;83(1):69–74.
3. Gerber C, Ganz R. Clinical assessment of instability of the shoulder. With special reference to anterior and posterior drawer tests. J Bone Joint Surg Br. 1984 Aug;66(4):551–556.
4. Gerber C, Krushell RJ. Isolated rupture of the tendon of the subscapularis muscle. Clinical features in 16 cases. J Bone Joint Surg Br. 1991 May;73(3):389–394.
5. Gross ML, Distefano MC. Anterior release test. A new test for occult shoulder instability. Clin. Orthop. Relat. Res. 1997 Jun;(339):105–108.
6. Hawkins RJ, Kennedy JC. Impingement syndrome in athletes. Am J Sports Med. 1980 May;8(3):151–158.
7. Hertel R, Ballmer FT, Lombert SM, Gerber C. Lag signs in the diagnosis of rotator cuff rupture. J Shoulder Elbow Surg. 1996 Jul;5(4):307–313.
8. Jobe FW, Jobe CM. Painful athletic injuries of the shoulder. Clin. Orthop. Relat. Res. 1983 Mar;(173):117–124.
9. Jobe FW, Kvitne RS, Giangarra CE. Shoulder pain in the overhand or throwing athlete. The relationship of anterior instability and rotator cuff impingement. Orthop Rev. 1989 Sep;18(9):963–975.
10. Kessel L, Watson M. The painful arc syndrome. Clinical classification as a guide to management. J Bone Joint Surg Br. 1977 May;59(2):166–172.
11. Kim S-H, Park J-C, Park J-S, Oh I. Painful jerk test: a predictor of success in nonoperative treatment of posteroinferior instability of the shoulder. Am J Sports Med. 2004 Dec;32(8):1849–1855.
12. Kim SH, Ha KI, Han KY. Biceps load test: a clinical test for superior labrum anterior and posterior lesions in shoulders with recurrent anterior dislocations. Am J Sports Med. 1999 May;27(3):300–303.
13. Leroux JL, Thomas E, Bonnel F, Blotman F. Diagnostic value of clinical tests for shoulder impingement syndrome. Rev Rhum Engl Ed. 1995 Jun;62(6):423–428.
14. Liu SH, Henry MH, Nuccion SL. A prospective evaluation of a new physical examination in predicting glenoid labral tears. Am J Sports Med. 1996 Nov;24(6):721–725.
15. Luime JJ, Verhagen AP, Miedema HS, Kuiper JI, Burdorf A, Verhaar JAN, et al. Does this patient have an instability of the shoulder or a labrum lesion? JAMA. 2004 Oct 27;292(16):1989–1999. doi:10.1001/jama.292.16.1989
16. Magee DJ. Orthopedic Physical Assessment. Elsevier Health Sciences; 1992.
17. McFarland EG, Torpey BM, Curl LA. Evaluation of shoulder laxity. Sports Med. 1996;22:264–272.
18. McLaughlin HL. On the frozen shoulder. Bulletin of the Hospital for Joint Diseases. 1951;
19. Mimori K, Muneta T, Nakagawa T, Shinomiya K. A new pain provocation test for superior labral tears of the shoulder. Am J Sports Med. 1999 Mar;27(2):137–142.
20. Neer CS. Impingement lesions. Clin. Orthop. Relat. Res. 1983 Mar;(173):70–77.
21. O'Brien SJ, Pagnani MJ, Fealy S, McGlynn SR, Wilson JB. The active compression test: a new and effective test for diagnosing labral tears and acromioclavicular joint abnormality. Am J Sports Med. 1998 Sep; 26(5):610–613.
22. Rowe CR, Zarins B. Recurrent transient subluxation of the shoulder. J Bone Joint Surg Am. 1981 Jul; 63(6):863–872.
23. Silliman JF, Hawkins RJ. Classification and physical diagnosis of instability of the shoulder. Clin. Orthop. Relat. Res. 1993 Jun;(291):7–19.
24. Stetson WB, Templin K. The crank test, the O'Brien test, and routine magnetic resonance imaging scans in the diagnosis of labral tears. Am J Sports Med. 2002 Nov;30(6):806–809.
25. Yergason RM. Supination sign. The Journal of Bone and Joint Surgery. 1931 Nov 20;(131):160.

The Patient Reported Outcome Measures used in this thesis

Constant score ***Functie bereik: (max 50 punten)***

Pijn: ***(max. 15 punten)***

geen	= 15
mild, geen effect op normaal functioneren	= 10
aanzienlijke beperking	= 5
ernstige beperking	= 0

Score pijn **Li..... Re:…… .**

Abductie (dorsum hand naar boven)

		Li	Re
0-30	= 0		
31-60	= 2		
61-90	= 4		
91-120	= 6		
121-150	= 8		
151-180	=10		
Totaal		.....	.....

Anteversie:

		Li	Re
0-30	= 0		
31-60	= 2		
61-90	= 4		
91-120	= 6		
121-150	= 8		
151-180	=10		
		.....	.....

Interne rotatie	**Li**	**Re**
Het dorsum van de hand tussen schouderbladen		10 punten
Het dorsum van de hand op de 12e thoracale wervel		8 punten
Het dorsum van de hand tot het middel (3e lumbale wervel)		6 punten
Het dorsum van de hand ter hoogte SI gewricht		2 punten
Het dorsum van de hand op het laterale bovenbeen		0 punten
Externe rotatie		
Hand achter hoofd met de elleboog naar voren		2 punten
Hand achter hoofd met de elleboog naar achteren		2 punten
Hand op het hoofd met de elleboog naar voren		2 punten
Hand op het hoofd met de elleboog naar achteren		2 punten
Volledige elevatie hand vanaf hoofd		2 punten
Totaal…….	**…… ……**	

Score functie bereik: **Li …… Re: …… .**

Kracht: (max 25 pounds)

Met een veer met de arm in 90º abductie of bij minder indien dit niet gehaald wordt. Maximale score is 25 pounds !!!

	Li	Re	
	.....	.....	(Kg)
x 2.2046	.....	.....	(pounds)

Activiteiten in het dagelijkse leven:

Activiteiten niveau: (max 10)

Volledig werkzaam	= 4
Volledig sport/ recreatie	= 4
Ongestoorde slaap	= 2
Totaal	.....

Arm heffen bij activiteiten in dagelijks leven waarbij patiënt GEEN hinder ervaart (max 30)

Boven het hoofd	= 10
Tot hoofd	= 8
Tot nek	= 6
Tot Xiphoid	= 4
Tot middel	= 2
Totaal	**Li:..... Re:……**

Totale score **Links: ..... Rechts.....**

[1] Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop 1987; 214: 160-64.

Simple Shoulder Test (SST)

1. Wanneer uw schouder in rust langs het lichaam hangt bent u dan vrij van klachten?

$_1$☐**ja** $_0$☐ **nee**

2. Kunt u ondanks de schouderklachten ongestoord slapen?

$_1$☐**ja** $_0$☐ **nee**

3. Kunt u met de aangedane zijde uw hemd achter in uw broek of rok stoppen?

$_1$☐**ja** $_0$☐ **nee**

4. Kunt u de hand achter het hoofd brengen met de elleboog opzij gehouden?

$_1$☐**ja** $_0$☐ **nee**

5. Kunt u een muntje op een plank op schouderhoogte leggen zonder de elleboog te buigen?

$_1$☐**ja** $_0$☐ **nee**

6. Kunt u een halve kilo optillen tot schouderhoogte met gestrekte arm?

$_1$☐**ja** $_0$☐ **nee**

7. Kunt u 4 kilo optillen (4 pakken suiker) tot schouderhoogte met gestrekte arm?

$_1$☐**ja** $_0$☐ **nee**

8. Kunt u 10 kilo dragen (zware boodschappentas) aan de aangedane zijde?

$_1$☐**ja** $_0$☐ **nee**

9. Denkt u dat u met de aangedane arm een tennisbal onderhands 10 meter ver kunt gooien?

$_1$☐**ja** $_0$☐ **nee**

10. Denkt u dat u met de aangedane arm een bal bovenhands 20 meter ver kunt gooien?

$_1$☐**ja** $_0$☐ **nee**

11. Kunt u met de aangedane arm de achterkant van uw andere schouder wassen?

$_1$☐**ja** $_0$☐ **nee**

12. Stelt uw schouder u in staat om volledig uw dagelijkse werk te verrichten?

$_1$☐**ja** $_0$☐ **nee**

Oxford Shoulder Score (OSS)

1. Hoe zou u de ergste pijn , die u in uw schouder heeft gehad, willen beschrijven?

- ❑ Ondraaglijk
- ❑ Erg pijnlijk
- ❑ Nogal pijnlijk
- ❑ Beetje pijnlijk
- ❑ Helemaal niet pijnlijk

2. Hoe zou U de pijn, die U meestal in uw schouder heeft, willen beschrijven ?

- ❑ Ondraaglijk
- ❑ Erg pijnlijk
- ❑ Nogal pijnlijk
- ❑ Beetje pijnlijk
- ❑ Helemaal niet pijnlijk

3. Hoeveel beinvloedt de pijn aan de schouder uw dagelijkse werkzaamheden? (ook het dagelijkse huiswerk).

- ❑ Totaal
- ❑ Grotendeels
- ❑ Matig
- ❑ Klein beetje
- ❑ Geheel niet

4. Heeft u 's nachts als U in bed ligt pijn in de schouder ?

- ❑ Elke nacht
- ❑ De meeste nachten
- ❑ Sommige nachten
- ❑ 1 of 2 nachten
- ❑ Nooit

5. Bent u in staat u aan- en uit te kleden met uw aangedane arm?

- ❑ Nee, onmogelijk
- ❑ Heel erg beperkt
- ❑ Matig beperkt
- ❑ Nagenoeg niet beperkt
- ❑ Geen beperking

6. Bent u in staat in - en uit een auto te stappen, of gebruik te maken van het openbaar vervoer met uw aangedane arm?

- ❑ Nee, onmogelijk
- ❑ Heel erg beperkt
- ❑ Matig beperkt
- ❑ Nagenoeg niet beperkt
- ❑ Geen beperking

7. Kunt U op hetzelfde moment mes en vork gebruiken?

- ❑ Nee, onmogelijk
- ❑ Met veel moeite
- ❑ Enigszins moeilijk
- ❑ Zonder veel moeite
- ❑ Ja, eenvoudig

8. Kunt u de boodschappen voor het huishouden zelfstandig doen?

- ❑ Nee, onmogelijk
- ❑ Met veel moeite
- ❑ Enigszins moeilijk
- ❑ Zonder veel moeite
- ❑ Ja, eenvoudig

9. Kon (Kunt U) u een dienblad met daarop een bord eten door de kamer dragen?

- ❑ Nee, onmogelijk
- ❑ Met extreem veel problemen
- ❑ Met nogal wat problemen
- ❑ Met lichte problemen
- ❑ Ja, eenvoudig

10. Kunt U met de aangedane arm uw haar borstelen of kammen ?

- ❑ Nee, onmogelijk
- ❑ Met veel moeite
- ❑ Enigszins moeilijk
- ❑ Zonder veel moeite
- ❑ Ja, eenvoudig

11. Kunt u uw kleding in de kledingkast hangen met de aangedane arm?

- ❑ Nee, onmogelijk
- ❑ Met veel moeite
- ❑ Enigszins moeilijk
- ❑ Zonder veel moeite
- ❑ Ja, eenvoudig

12. Kunt U zichzelf onder beide oksels wassen en drogen ?

- ❑ Nee, onmogelijk
- ❑ Met veel moeite
- ❑ Enigszins moeilijk
- ❑ Zonder veel moeite
- ❑ Ja, eenvoudig

Oxford Shoulder Instability Score (OSIS)

1. Hoe vaak is de schouder gedurende **de laatste 6 maanden** uit de kom of bijna uit de kom geschoten ?

    - ❑ Niets
    - ❑ 1 tot 2 keer in de laatste 6 maanden
    - ❑ 1 tot 2 keer per maand
    - ❑ 1 tot 2 keer per week
    - ❑ meer dan 2 keer per week

2. Heeft u **de afgelopen drie maanden** bij het aankleden last gehad (of u zorgen gemaakt) vanwege uw schouder?

    - ❑ Helemaal geen last
    - ❑ Weinig last
    - ❑ Matig lastig
    - ❑ Veel last
    - ❑ Onmogelijk te doen

3. Als u gedurende de **laatste 3 maanden** pijn had, hoe zou U de pijn beschrijven?

    - ❑ Geen pijn
    - ❑ Lichte pijn
    - ❑ Matige pijn
    - ❑ Ernstige pijn
    - ❑ Ondraaglijke pijn

4. In hoeverre heeft het probleem met uw schouder u **de afgelopen drie maanden** belemmerd in uw gewone werkzaamheden?

    - ❑ Helemaal niet
    - ❑ Een klein beetje
    - ❑ Matig
    - ❑ In grote mate
    - ❑ Totaal

5. Heeft u gedurende de **laatste 3 maanden** aktiviteiten vermeden omdat U bang was, dat dan de schouder uit de kom zou schieten?

    - ❑ Niets vermeden
    - ❑ af en toe
    - ❑ soms
    - ❑ meestal
    - ❑ bijna alle aktiviteiten vermeden

6. Heeft het probleem met uw schouder u **de afgelopen drie maanden** belet dingen te doen die belangrijk voor u zijn?

   - ❑ Helemaal niet
   - ❑ Heel soms
   - ❑ Sommige dagen
   - ❑ Meeste dagen of meer dan één activiteit
   - ❑ Elke dag of veel activiteiten

7. In hoeverre heeft het probleem met uw schouder u **de afgelopen drie maanden** belemmerd in uw sociale leven?
   - ❑ Helemaal niet
   - ❑ Heel soms
   - ❑ Sommige dagen
   - ❑ Meeste dagen of meer dan één activiteit
   - ❑ Iedere dag

7. In hoeverre heeft het probleem met uw schouder u **de afgelopen vier weken** belemmerd in sport of hobbies? (inclusief sexuele activiteiten – indien van toepassing)

   - ❑ Helemaal niet
   - ❑ Een beetje/ wel eens
   - ❑ Af en toe
   - ❑ Meestal
   - ❑ Altijd

9. Hoe vaak is uw schouder in beeld geweest of hoe vaak heeft u aan de schouder gedacht **gedurende de laatste 4 weken**?

   - ❑ Nooit, alleen als er iemand naar vraagt
   - ❑ Wel eens
   - ❑ Sommige dagen
   - ❑ De meeste dagen
   - ❑ Iedere dag

10. In hoeverre heeft het probleem met uw schouder u **de afgelopen vier weken** belemmerd in het optillen van zware voorwerpen?

    - ❑ Nooit, alleen als er iemand naar vraagt
    - ❑ Wel eens
    - ❑ Sommige dagen
    - ❑ De meeste dagen
    - ❑ Iedere dag

11. Als u kijkt naar **de afgelopen vier weken**, hoe zou u dan de pijn beschrijven die u doorgaans in uw schouder had?

- ❑ Geen
- ❑ Erg licht
- ❑ Licht
- ❑ Gemiddeld
- ❑ Heftig

12. Heeft u de afgelopen vier weken 's nachts in bed bepaalde slaaphoudingen vermeden vanwege uw schouder?

- ❑ Nee
- ❑ maar 1 of 2 nachten
- ❑ Sommige nachten
- ❑ De meeste nachten
- ❑ Iedere nacht

AANWIJZINGEN VOOR DE PATIËNT

In de onderdelen A, B, C en D wordt u verzocht vragen te beantwoorden in het onderstaande vragenformulier. Beantwoord iedere vraag door op de horizontale lijn een schuine streep (“/”) te plaatsen.

LET OP: Voorbeeld

1. Als u aan het linkeruiteinde van de lijn een schuine streep (“/”) plaatst, geeft u aan dat u
   geen pijn heeft.

geen pijn /_____/extreme pijn

2. Als u aan het rechteruiteinde van de lijn een schuine streep (“/”) plaatst, geeft u aan dat u
   extreme pijn heeft.

geen pijn /_____/ extreme pijn

3. Houd ook rekening met het volgende:

a) hoe verder u de schuine streep naar rechts plaatst (“/”), hoe sterker u het betreffende symptoom ervaart.

b) hoe verder u de schuine streep naar links plaatst (“/”), hoe minder u het betreffende symptoom ervaart.

c) plaats de schuine streep (“/”) s.v.p. niet buiten de streepjes die de uiteinden van de lijn aangeven.

In deze vragenlijst wordt u verzocht aan te geven in hoeverre u een bepaald symptoom de afgelopen week hebt ervaren in verband met uw problematische schouder. Als u niet weet om welke schouder het gaat of als niet alles duidelijk is, wordt u verzocht navraag te doen voordat u de vragenlijst invult.

Als u om welke reden dan ook een vraag niet goed begrijpt, lees dan de toelichting aan het eind van de vragenlijst. Plaats vervolgens de schuine streep (“/”) op de juiste plek op de horizontale lijn.
Als een bepaalde vraag niet op u van toepassing is of als het betreffende symptoom zich de afgelopen week niet heeft voorgedaan, probeert u de vraag dan te beantwoorden op basis van een zo goed mogelijke schatting. U mag de vraag niet overslaan.

ONDERDEEL A
Fysieke symptomen

AANWIJZINGEN VOOR PATIËNTEN

De onderstaande vragen hebben betrekking op de fysieke symptomen die u ervaart als gevolg van uw schouderprobleem. In alle gevallen verzoeken wij u uw antwoord te baseren op de intensiteit van de symptomen van de afgelopen week (door op de juiste plek op de horizontale lijn een schuine streep "/" te plaatsen).

1. Hoeveel pijn heeft u aan uw schouder als u activiteiten verricht waarbij u uw arm boven uw hoofd moet verheffen?

geen pijn /_____/ extreme pijn

2. Hoeveel last hebt u van een pijnlijke of kloppende schouder?

geen pijn/kloppen/_____/ extreme pijn/ kloppen

3. Hoeveel last hebt u van een zwakke of verzwakte schouder?

geen zwakte /_____/ extreme zwakte

4. Hoeveel last hebt u van een vermoeide of krachteloze schouder?

geen vermoeidheid /_____/extreme vermoeidheid

5. Hoeveel last hebt u van klikken, kraken of knakken in uw schouder?

geen klikken /_____/extreem klikken

6. Hoeveel last hebt u van stijfheid in uw schouder?

geen stijfheid /_____/extreme stijfheid

7. Hoeveel last hebt u van pijn in de nekspieren in verband met uw schouder?

geen last /_____/extreme last

8. Hoeveel last hebt u van een instabiele of los zittende schouder?

geen instabiliteit /_____/extreme instabiliteit

9. In hoeverre compenseert u uw schouderprobleem door andere spieren te gebruiken?

in het geheel niet/_____/ in extreme mate

10. In hoeverre is de beweeglijkheid van uw schouder beperkt?

niet beperkt /_____/extreem beperkt

ONDERDEEL B:
Sport/Recreatie/Werk

AANWIJZINGEN VOOR PATIËNTEN

Het onderstaande gedeelte gaat over de mate waarin uw schouderprobleem de afgelopen week een rol heeft gespeeld bij het sporten, in uw vrijetijdsbesteding of op het werk. Beantwoord iedere vraag door op de juiste plek op de horizontale lijn een schuine streep (“/”) te plaatsen.

11. In hoeverre werd u door uw schouderprobleem beperkt bij het verrichten van sportieve of recreatieve activiteiten?

niet beperkt /_____/extreem . beperkt

12. In hoeverre werd u door uw schouder gehinderd bij het verrichten van de specifieke activiteiten die voor uw werk of sport nodig zijn? (Als uw schouderprobleem zowel uw werk als sportieve activiteiten beïnvloedt, beantwoord de vraag dan voor de activiteit waarbij die invloed het grootst is.)

niet gehinderd /_____/extreem . gehinderd

13. In hoeverre bent u geneigd uw arm tijdens activiteiten te beschermen?

in het geheel niet /_____/ in . extreme mate

14. Hoeveel moeite hebt u met het optillen van zware voorwerpen onder uw schouderhoogte?

geen moeite /_____/extreme . moeite

AANWIJZINGEN VOOR PATIËNTEN

Het volgende gedeelte gaat over de mate waarin uw schouderprobleem uw levensstijl heeft beïnvloed of veranderd. Ook hier wordt u verzocht iedere vraag voor de afgelopen week te beantwoorden, door op de juiste plek op de horizontale lijn een schuine streep (“/”) te plaatsen.

15. In hoeverre bent u bang om op uw schouder te vallen?

geen angst /_____/ extreme . angst

16. In hoeverre hebt u moeite om in conditie te blijven?

geen moeite /_____/ extreme . moeite

17. In hoeverre kost het u moeite te stoeien e.d. met familie of vrienden?

geen moeite /_____/ extreme . moeite

18. In hoeverre leidt uw schouderprobleem ook tot een slaapprobleem?

geen slaapprobleem /_____/ extreem slaapprobleem

ONDERDEEL D
Emoties

AANWIJZINGEN VOOR PATIËNTEN

In de onderstaande vragen wordt u verzocht aan te geven hoe u zich in verband met uw schouderprobleem de afgelopen week hebt gevoeld. Beantwoord iedere vraag door op de juiste plek op de horizontale lijn een schuine streep (“/”) te plaatsen.

19. In hoeverre bent u zich bewust van uw schouder?

niet bewust /_____/ extreem . bewust

20. In hoeverre maakt u zich zorgen dat uw schouderprobleem verergert?

niet bezorgd /_____/extreem . bezorgd

21. In hoeverre raakt u gefrustreerd door uw schouderprobleem?

niet gefrustreerd/_____/extreem . gefrustreerd

Onderdeel A: Fysieke symptomen

Vraag 1.
Heeft betrekking op iedere activiteit waarbij u uw arm boven schouderniveau moet verheffen, zoals borden in een keukenkastje zetten, uw haar op orde brengen, borstcrawl in het zwembad, het plafond schilderen, een bal gooien in een bovenhandse worp enz.

Vraag 2.
Heeft betrekking op een doffe pijn op de achtergrond, in tegenstelling tot een plotselinge stekende pijn.

Vraag 3.
Heeft betrekking op het gebrek aan kracht als u uw arm gebruikt om een handeling te verrichten.

Vraag 4.
Heeft betrekking op de mate waarin uw schouder bij inspanning vermoeid raakt.

Vraag 5.
Heeft betrekking op de geluiden die zich in uw schouder voordoen als u hem gebruikt.

Vraag 6.
Heeft betrekking op het gevoel van stroefheid in het schoudergewricht; een veel voorkomend verschijnsel bij het opstaan, na inspanning of juist na een periode van rust. Deze vraag gaat niet over beperkte beweeglijkheid.

Vraag 7.
Heeft betrekking op spanning, pijn of kramp in uw nekspieren die het gevolg lijkt te zijn van uw schouderprobleem.

Vraag 8.
Heeft betrekking op het gevoel dat uw schouder half los zit, volledig uit de kom is geraakt, naar onderen glijdt of in allerlei richtingen verschuift.

Vraag 9.
Heeft betrekking op het gebruik van arm- of nekspieren bij bewegingen of handelingen, ter compensatie van uw schouderprobleem.

Vraag 10.
Heeft betrekking op het gebrek aan beweeglijkheid van uw schouder in verschillende richtingen.

Onderdeel B: Sport/Recreatie/Werk

Vraag 11.
Heeft betrekking op de mate waarin uw schouder het sporten of recreëren belemmert of geheel onmogelijk maakt.

Vraag 12.
Heeft betrekking op de moeite die u hebt bij het verrichten van handelingen die voor werk, sport of recreatie noodzakelijk zijn.

Toelichting op de vragen (vervolg)

Vraag 13.
Heeft betrekking op het bewust of onbewust beschermen van uw arm door hem dicht tegen uw lichaam aan te houden, af te schermen of in een beugel te dragen.

Vraag 14.
Deze vraag gaat niet over het boven uw hoofd tillen van voorwerpen, maar over het optillen van zware objecten *onder* schouderniveau, zoals een tas met boodschappen, apparatuur op het werk, boeken of een bowlingbal.

Onderdeel C: Levensstijl

Vraag 15.
Heeft betrekking op de angst die u hebt om op uw schouder te vallen of, aan die zijde, op uw uitgestrekte hand terecht te komen.

Vraag 16.
Heeft betrekking op uw conditie voordat u een schouderprobleem kreeg. Houd hierbij rekening met uw cardiovasculaire conditie en met de kracht en spanning in uw spieren.

Vraag 17.
Heeft betrekking op ruwe of speelse activiteiten die u normaal gesproken onderneemt met familie of vrienden.

Vraag 18.
Heeft betrekking op de mate waarin u, als gevolg van uw schouderprobleem, uw slaaphouding hebt moeten aanpassen, 's nachts wakker wordt, moeite hebt bij het inslapen of vermoeid bent bij het opstaan.

Onderdeel D: Emoties

Vraag 19.
Heeft betrekking op de mate waarin u zich bewust bent van uw schouder of bij iedere activiteit eerst aan uw schouder denkt.

Vraag 20.
Heeft betrekking op de mate waarin u zich zorgen maakt dat uw schouderprobleem niet afneemt, maar stabiel blijft of zelfs erger wordt.

Vraag 21.
Heeft betrekking op de frustratie die u voelt omdat u dingen die u eerst wel kon doen, of die u wilt doen, nu vanwege uw schouder niet meer kunt doen.

Disability of the Arm, Shoulder and Hand (DASH)

Beperkingen van arm, schouder en hand.

Instructies:

Deze vragenlijst heeft betrekking op zowel uw symptomen als uw mogelijkheid om bepaalde handelingen te verrichten.

Beantwoord alle vragen door het juiste vakje aan te kruisen, gebaseerd op uw conditie van de afgelopen week.

Als u de afgelopen week geen activiteiten heeft uitgevoerd, schat dan het meest nauwkeurige antwoord.

Het maakt niet uit welke hand, of arm u gebruikt om de handeling te verrichten: baseer uw antwoord alstublieft op de mogelijkheid een opdracht uit te voeren ongeacht de manier waarop.

	Geen moeite 1	Geringe moeite 2	Meer moeite 3	Zeer veel moeite 4	Niet instaat 5
1. Een dichte of nieuwe pot openen	☐	☐	☐	☐	☐
2. Schrijven	☐	☐	☐	☐	☐
3. Een sleutel omdraaien	☐	☐	☐	☐	☐
4. Koken	☐	☐	☐	☐	☐
5. Een zware deur openduwen	☐	☐	☐	☐	☐
6. Een voorwerp op een plank boven uw hoofdplaatsen	☐	☐	☐	☐	☐
7. Zwaar huishoudelijk werk doen. (bv. tegels afwassen, vloeren schrobben)	☐	☐	☐	☐	☐
8. Tuinieren	☐	☐	☐	☐	☐
9. Bed opmaken	☐	☐	☐	☐	☐
10. Boodschappentas of aktetas dragen	☐	☐	☐	☐	☐
11. Een zwaar voorwerp dragen (>5kg)	☐	☐	☐	☐	☐
12. Een lamp boven uw hoofd verwisselen	☐	☐	☐	☐	☐
13. Haren wassen of Föhnen	☐	☐	☐	☐	☐
14. Uw rug wassen	☐	☐	☐	☐	☐
15. Een trui aantrekken	☐	☐	☐	☐	☐
16. Met een mes eten snijden	☐	☐	☐	☐	☐
17. Recreactieve activiteiten die weinig moeite kosten (bv. kaarten, breien etc.)	☐	☐	☐	☐	☐
18. Recreatieve activiteiten die kracht of druk uitoefenen op arm, schouder of hand (golfen, timmeren, tennissen etc.).	☐	☐	☐	☐	☐
19. Recreatieve activiteiten waarbij de arm vrij beweegt (bv. frisbee, badminton, etc).	☐	☐	☐	☐	☐
20. Van de ene naar de andere plaats gaan	☐	☐	☐	☐	☐
21. Seksuele activiteiten	☐	☐	☐	☐	☐

	1 **Helemaal niet**	2 **In geringe mate**	3 **Matig**	4 **Aardig wat**	5 **Zeer veel**
22. Heeft uw probleem aan uw arm, hand of schouder u de afgelopen week belemmerd in uw normale sociale activiteiten met familie, vrienden, buren of groepen?	☐	☐	☐	☐	☐

	Helemaal niet beperkt	**In geringe mate beperkt**	**Matig beperkt**	**Aardig wat beperkt**	**Zeer veel beperkt**
23. Was u de afgelopen week beperkt in uw werk of andere dagelijkse activiteiten als gevolg van uw probleem aan uw arm, hand of schouder?	☐	☐	☐	☐	☐

	1 **Geen**	2 **Licht**	3 **Matig**	4 **Ernstig**	5 **Extreem**
24. Pijn aan arm, schouder of hand	☐	☐	☐	☐	☐
25. Pijn aan arm, schouder o hand bij welke activiteit dan ook	☐	☐	☐	☐	☐
26. Tintelingen (slapend gevoel) in de arm, schouder of hand	☐	☐	☐	☐	☐
27. Zwakheid in uw arm, schouder of hand	☐	☐	☐	☐	☐
28. Stijfheid in uw arm, schouder of hand	☐	☐	☐	☐	☐

	Geen moeite	**Geringe moeite**	**Meer moeite**	**Zeer veel moeite**	**Niet instaat**
29. Hoeveel moeite heeft u de afgelopen week gehad met slapen vanwege de pijn in uw arm, schouder of hand?	☐	☐	☐	☐	☐

	Sterk mee oneens	**oneens**	**Niet eens, niet oneens**	**Mee eens**	**Sterk mee eens**
30. Ik voel me minder bekwaam, minder zeker of minder nuttig door de problemen aan mijn arm, schouder of hand.	☐	☐	☐	☐	☐

Sport/podiumkunsten module (naar keuze)

De volgende vragen hebben betrekking op de impact die het probleem aan uw arm, schouder of hand heeft op het bespelen van een muziekinstrument of het beoefenen van een sport, of beide.
Als u meer dan één sport beoefent of instrument bespeelt (of beide), antwoordt dan uitgaande van de activiteit die het belangrijkste voor u is.

Geeft u alstublieft aan welke sport of welk muziekinstrument het belangrijkste voor u is:

……..

- ❑ Ik beoefen geen sport, of ik bespeel geen instrument. (U mag dit gedeelte overslaan).

Omcirkel het getal dat uw fysieke vermogen van de afgelopen week het beste beschrijft.
Had u moeite met:

Sport/podiumkunsten module	**Geen moeite** 1	**Geringe moeite** 2	**Meer moeite** 3	**Zeer veel moeite**	**Niet instaat** 5
1. Het toepassen van uw gebruikelijke techniek om uw Instrument te bespelen of uw sport te beoefenen?	☐	☐	☐	☐	☐
2. Het bespelen van uw instrument of beoefenen van uw sport vanwege pijn aan arm, schouder of hand?	☐	☐	☐	☐	☐
3. Het bespelen van uw instrument of het beoefenen van uw sport zo goed als u zou willen?	☐	☐	☐	☐	☐
4. Het besteden van uw gebruikelijke hoeveelheid tijd aan het bespelen van uw instrument of beoefenen van uw sport?	☐	☐	☐	☐	☐

Werkmodule (naar keuze)

De volgende vragen gaan over de invloed van uw probleem aan arm, schouder of hand op uw mogelijkheid om te werken (inclusief huishouden als dat uw hoofdtaak is).

Geeft u alstublieft aan wat uw beroep/werk is:

……

- ❑ Ik werk niet. (U mag dit gedeelte overslaan).

Omcirkel het getal dat uw fysieke vermogen van de afgelopen week het beste beschrijft.
Had u moeite met:

Werk module	**Geen moeite** 1	**Geringe moeite** 2	**Meer moeite** 3	**Zeer veel moeite**	**Niet instaat** 5
1. Het toepassen van uw gebruikelijke techniek om uw werk?	☐	☐	☐	☐	☐
2. Het doen van uw normale werk door de pijn aan arm, schouder of hand?	☐	☐	☐	☐	☐
3. Het doen van uw werk zo goed als u dat zou willen?	☐	☐	☐	☐	☐
4. Het doen van uw werk binnen normale tijd?	☐	☐	☐	☐	☐

Short Form 36 health survey (SF-36)

1) Hoe zou u over het algemeen uw gezondheid noemen?

☐ 1 Uitstekend ☐ 2 Zeer goed ☐ 3 Goed ☐ 4 Matig ☐ 5 Slecht

2) Hoe beoordeelt u nu uw gezondheid over het algemeen, vergeleken met een jaar geleden?

☐ 1 Veel beter nu dan een jaar geleden
☐ 2 Wat beter nu dan een jaar geleden
☐ 3 Ongeveer hetzelfde nu als een jaar geleden
☐ 4 Wat slechter nu dan een jaar geleden
☐ 5 Veel slechter nu dan een jaar geleden

3) De volgende vragen gaan over de bezigheden die u misschien doet op een doorsnee dag. Wordt u door uw gezondheid op dit moment beperkt bij deze bezigheden? Zo ja in welke mate

BEZIGHEDEN ***Kruis ÉÉN hokje per vraag aan***	1 **Ja, ernstig beperkt**	2 **Ja, een beetje beperkt**	3 **Nee, helemaal niet beperkt**
a) Forse inspanning, zoals hardlopen, tillen van zware voorwerpen, een veeleisende sport beoefenen	☐	☐	☐
b) Matige inspanning, zoals een tafel verplaatsen, stofzuigen, zwemmen of fietsen	☐	☐	☐
c) Boodschappen tillen of dragen	☐	☐	☐
d) Een paar trappen oplopen	☐	☐	☐
e) Bukken knielen of hurken	☐	☐	☐
f) Meer dan een kilometer lopen	☐	☐	☐
g) Een paar honderd meter lopen	☐	☐	☐
h) Ongeveer honderd meter lopen	☐	☐	☐
i) Uzelf wassen of aankleden	☐	☐	☐

4) Heeft u de afgelopen 4 weken, een van de volgende problemen bij uw werk of andere dagelijkse bezigheden gehad, ten gevolge van uw lichamelijke gezondheid?

Kruis ÉÉN hokje per vraag aan	**Ja** 1	**Nee** 2
a) U besteedde minder tijd aan werk of andere bezigheden	☐	☐
b) U heeft minder bereikt dan u zou willen	☐	☐
c) U was beperkt in het soort werk of andere bezigheden	☐	☐
d) U had moeite om uw werk of andere bezigheden uit te voeren (het kostte u bijv. extra inspanning)	☐	☐

5) Heeft u de afgelopen 4 weken, een van de volgende problemen bij uw werk of andere dagelijkse bezigheden gehad, ten gevolge van emotionele problemen (bv door dat u zich depressief of angstig voelde)?

Kruis ÉÉN hokje per vraag aan	**Ja** 1	**Nee** 2
a) U heeft minder tijd kunnen besteden aan werk of andere bezigheden	☐	☐
b) U heeft minder bereikt dan u zou willen	☐	☐
c) U heeft uw werk of andere bezigheden niet zo zorgvuldig gedaan als u normaal gewend bent te doen	☐	☐

6) In hoeverre hebben uw lichamelijke gezondheid of emotionele problemen u gedurende de afgelopen 4 weken gehinderd in uw normale omgang met de familie, vrienden of buren, of bij activiteiten in groepsverband?

☐1 Helemaal niet ☐2Enigszins ☐3Nogal ☐4Veel ☐5Heel erg veel

7) Hoeveel lichamelijke pijn heeft u de afgelopen 4 weken gehad?

☐1Geen ☐2Heel licht ☐3Licht ☐4Nogal ☐5Ernstig ☐6Heel ernstig

8) In welke mate bent u de afgelopen 4 weken door pijn gehinderd in uw normale werk? Zowel werk buitenshuis als huishoudelijk werk.

☐1 Helemaal niet ☐2Enigszins ☐3Nogal ☐4Veel ☐5Heel erg veel

9) Deze vragen gaan over hoe u zich voelt en hoe het met u ging in de afgelopen 4 weken. Wilt u a.u.b. bij elke vraag het antwoord geven dat het best benadert hoe u zich voelde. Hoe vaak gedurende de afgelopen 4 weken.

Kruis ÉÉN hokje per vraag aan	1 **Altijd**	2 **Meestal**	3 **Vaak**	4 **Soms**	5 **Zelden**	6 **Nooit**
a) voelde u zich levenslustig?	☐	☐	☐	☐	☐	☐
b) was u erg zenuwachtig?	☐	☐	☐	☐	☐	☐
c) Zat u zo in de put dat niets u kon opvrolijken?	☐	☐	☐	☐	☐	☐
d) Voelde u zich rustig en tevreden?	☐	☐	☐	☐	☐	☐
e) Had u veel energie?	☐	☐	☐	☐	☐	☐
f) Voelde u zich somber?	☐	☐	☐	☐	☐	☐
g) Voelde u zich uitgeput?	☐	☐	☐	☐	☐	☐
h) Was u een gelukkig mens?	☐	☐	☐	☐	☐	☐
i) Voelde u zich moe	☐	☐	☐	☐	☐	☐

10) Hoe vaak hebben uw lichamelijke gezondheid of emotionele problemen u gedurende de afgelopen 4 weken gehinderd bij uw activiteiten (zoals vrienden of familie bezoeken etc.)

☐ 1 Altijd ☐ 2 Meestal ☐ 3 Soms ☐ 4 Zelden ☐ 5 Nooit

11)***Kruis ÉÉN hokje per vraag aan***	1 **Volkomen juist**	2 **Groten deels juist**	3 **Weet ik niet**	4 **Groten deels onjuist**	5 **Volkomen onjuist**
a) Ik lijk wat gemakkelijker ziek te worden dan andere mensen	☐	☐	☐	☐	☐
b) Ik ben even gezond als andere mensen die ik ken	☐	☐	☐	☐	☐
c) ik verwacht dat mijn gezondheid achteruit zal gaan	☐	☐	☐	☐	☐
d) Mijn gezondheid is uitstekend	☐	☐	☐	☐	☐