AARHUS UNIVERSITY

Societal costs of shoulder disorders and quality of instruments to measure outcomes in shoulder patients

PhD dissertation

Lotte Sørensen

Health
Aarhus University
2022

Supervisors

Lisa Gregersen Oestergaard, associate professor, PhD (main supervisor)
Department of Public Health, Aarhus University, Denmark
DEFACTUM, Central Denmark Region, Denmark
Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Denmark

Maurits van Tulder, professor, PhD
Department of Human Movement Sciences, Faculty Behavioural & Movement Sciences, Amsterdam Movement Sciences research institute, Vrije Universiteit Amsterdam, the Netherlands
Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Denmark

Hans Viggo Johannsen, MD
Department of Orthopaedic Surgery, Aarhus University Hospital, Denmark

Assessment committee

David Høyrup Christiansen, associate professor, PhD (chairman of the committee)
Department of Clinical Medicine, Aarhus University, Denmark
Regional Hospital Central Jutland, Denmark

Margreth Grotle, professor, PhD
Department of Physiotherapy, Faculty of Health Science, Oslo, Norway

Jonas Bloch Thorlund, professor, PhD
Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Denmark
Research Unit for General Practice, Department of Public Health, University of Southern Denmark, Denmark

Financial support

The Health Research Foundation of Central Denmark Region
Aarhus University
The Health Foundation (Helsefonden)
Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital
The Danish Rheumatism Association

Acknowledgements

The work presented in this thesis was carried out between November 2018 and January 2022 during my employment at the Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital and in collaboration with the Department of Clinical Medicine, Aarhus University. I am grateful for the advice, help and support from numerous people which has made the work possible.

First, I wish to thank my supervisors Lisa Gregersen Østergaard, Maurits van Tulder and Hans Viggo Johannsen for guiding me on my academic journey and keeping me on track. I have learned a lot form our discussions and your constructive feedback. It has been a privilege to work with you.

I owe a grateful thanks to all the patients who participated in the study for spending time and effort assessing their health status.

A special thank goes to co-authors; Janne Ovesen for sharing knowledge from an orthopaedic surgeon's perspective, Ulrika Enemark for being helpful with the health economic theme, Emely Ek Blæhr for helping me out with the register-based data analyses and Annemette Krintel Petersen for her work with the two systematic reviews.

I would like to express a special thanks to my colleagues at DEFACTUM, MarselisborgCentret who welcomed me in their inspiring research environment. I have enjoyed your company and our academic discussions.

Also warm thanks to my colleagues at the Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital for good company, support and many laughs.

I would like to express my thanks to my employers at the Department of Physiotherapy and Occupational Therapy, Helle Kruse-Andersen and Morten Albæk Skrydstrup for supporting and believing in me.

Last but not least, I which to thank my family and friends for their interest and support. Especially, to Steffen for your humour and for making sure that I did not take it all too seriously, and to Jeppe and Ida for just being in my life. You are more important than anything.

This project was made possible by financial support from The Health Research Foundation of Central Denmark Region, Aarhus University, The Health Foundation (Helsefonden), Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, and The Danish Rheumatism Association.

April 2022
Lotte Sørensen

List of papers

Paper I **Cost of shoulder disorders in Denmark; a nationwide cost-of-illness study investigating 617,334 patients and matched controls.**
Sørensen L, van Tulder M, Johannsen HV, Ovesen J, Enemark U, Blæhr EE, Oestergaard LG.
Published ahead of print: Pain. 2022 Feb 7.
doi: 10.1097/j.pain.0000000000002610. Epub ahead of print. PMID: 35135991.

Paper II **Responsiveness and minimal important change of the Oxford Shoulder Score, EQ-5D, and the Fear-Avoidance Belief Questionnaire Physical Activity subscale in patients undergoing arthroscopic subacromial decompression.**
Sørensen L, van Tulder M, Johannsen HV, Ovesen J, Oestergaard LG.
Published: JSES Int. 2021 Jul 9;5(5):869-874.
doi: 10.1016/j.jseint.2021.05.008. PMID: 34505098; PMCID: PMC8411058.

Paper III **Measurement Properties of Isokinetic Dynamometry for Assessment of Shoulder Muscle Strength: A Systematic Review.**
Sørensen L, Oestergaard LG, van Tulder M, Petersen AK.
Published: Arch Phys Med Rehabil. 2021 Mar;102(3):510-520.
doi: 10.1016/j.apmr.2020.06.005. Epub 2020 Jun 30. PMID: 32619417.

Paper IV **Measurement properties of handheld dynamometry for assessment of shoulder muscle strength: A systematic review.**
Sørensen L, Oestergaard LG, van Tulder M, Petersen AK.
Publiched: Scand J Med Sci Sports. 2020 Dec;30(12):2305-2328.
doi: 10.1111/sms.13805. Epub 2020 Sep 17. PMID: 33463791.

List of abbreviations

AUC	Area Under Curve
CENTRAL	Cochrane Central Register of Controlled Trials
CI	Confidence interval
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
DAGS	Danish Outpatient Grouping System
DHSR	The Danish National Health Service Registry
DNPR	Danish National Patient Registry
DREAM	Danish Register for Evaluation of Marginalization
DRG	Diagnosis-Related-Grouping
EQ-5D	EuroQol – 5 Dimensions
ER	External rotation
FABQ	Fear-Avoidance Belief Questionnaire
FABQ-PA	Fear-Avoidance Belief Questionnaire for physical activity
FABQ-W	Fear-Avoidance Belief Questionnaire for work
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HHD	Handheld dynamometers
HR-QoL	Health-related quality of life
ICC	Intraclass correlation coefficient
ICD-10	The International Classification of Diseases, 10th edition
ID	Isokinetic dynamometers
IR	Internal rotation
MDC	Minimal Detectable Change
MIC	Minimal Important Change
NHSR	National Health Service Registry
OSS	Oxford Shoulder Score
OMERACT	Outcome Measures in Rheumatology
Pain VAS	Pain measured with visual analogue scale
PEDro	Physiotherapy Evidence Database
PRISMA	Preferred Reporting Items for Systematic Reviews and meta-Analysis statement
PROM	Patient Reported Outcome Measure
PROSPERO	International Prospective Register of Systematic Reviews
ROC	Receiver Operating Characteristics
SD	Standard deviation
SSV	Subjective Shoulder Value
VAS	Visual analogue scale

List of figures and tables

Figures

Tables

Table of contents

English summary

Introduction

Shoulder disorders are the third most common musculoskeletal condition. Besides having a major impact on individuals' daily life, shoulder disorders are often associated with considerable consumption of resources. Because of a high incidence and prevalence, the economic consequences of shoulder disorders are considerable, but little is known about the economic burden on society.
When conducting research and monitoring patients in clinical practice, assessing outcomes is essential to evaluate disease status, disease development and treatment efficacy. For this purpose, measurement instruments with satisfactory measurement properties are required. Thus, the aims of this thesis were:

1) to estimate the direct and indirect costs associated with shoulder disorders in patients compared with matched controls from the general population, and to investigate if costs were higher for specific subgroups of patients (Study I)

2) to evaluate the responsiveness and minimal important change (MIC) of patient-reported outcome measures (PROMs) commonly used in shoulder disorders; Oxford Shoulder Score (OSS), EuroQol – 5 Dimensions (EQ-5D) and Fear-Avoidance Belief Questionnaire for physical activity (FABQ-PA) in patients with subacromial impingement syndrome undergoing arthroscopic subacromial decompression (Study II)

3) to summarize evidence of the measurement properties of isokinetic dynamometers (ID) and handheld dynamometers (HHD) for assessment of shoulder muscle strength (Study III and IV)

Methods

Study I was a register-based cost-of-illness study including patients diagnosed with shoulder disorders between 2005 and 2017 and matched controls without shoulder disorders. Health care costs from the primary and secondary healthcare sector and costs of sick leave were included for a 6-year period, i.e. from one year before to up to five years after the time of diagnosis. Study II was a cohort study with 6-month follow-up including patients with subacromial impingement syndrome treated with decompression surgery. The responsiveness of the OSS, EQ-5D (EQ-$5D_{index}$ and EQ-$5D_{vas}$) and FABQ-PA was evaluated using the Global Rating of Change Scale as primary anchor and change in shoulder pain and the subjective shoulder value as secondary anchors. Study III and IV were systematic reviews of the literature including studies evaluating the reliability, measurement error or construct validity of ID and HHD. The reviews were performed in accordance with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology.

Results

In Study I, 617,334 unique individuals diagnosed with shoulder disorders and an equal number of matched controls were identified. The incidence rate increased by 26% from 966 per 100,000 person years in 2005 to 1,215 per 100,000 person years in 2017. Individuals aged ≥65 years had health care costs 83% higher than individuals <65 years. Costs of sick leave accounted for about 70% of the total costs for individuals <65 years. The mean additional total costs (health care costs and costs of sick leave) of patients with shoulder disorders compared with matched controls for the 6-year period were €25,771 (95% CI 25,531; 26,012) for individuals aged <65 years and €11,334 (95% CI 11,014; 11,654) for individuals aged ≥65 years. Furthermore, 20% of the cases accounted for 66% of the total costs. Estimated annual additional costs of patients with shoulder disorders compared with individuals without shoulder disorders were €1.21 billion.

In Study II, the area under the receiver-operating characteristic curve (ROC AUC) was 0.96 (95% CI 0.91;1.00) for OSS, 0.82 (95% CI 0.66;0.99) for EQ-$5D_{index}$, 0.73 (95% CI 0.58;0.87) for EQ-$5D_{vas}$ and 0.74 (95% CI 0.58;0.90) for FABQ-PA. The MIC results were 6.0 points for the OSS, 0.024 for EQ-$5D_{index}$, 10.0 for the EQ-$5D_{vas}$ and -5.0 for FABQ-PA.

In Study III and IV, the reliability of ID and HHD was overall sufficient with the majority of intraclass correlation coefficients ≥0.70. The quality of evidence was generally moderate or low for ID and high for HHD. The measurement error was rated not sufficient for either ID or HHD as none of the strata examined showed a sufficient proportion (≥75%) of minimal detectable change values ≤15%. The quality of evidence was generally moderate to very low for ID and high to moderate for HHD. The construct validity of HHD showed inconsistent results based on low quality of evidence.

Conclusion

Around 1.2% of the Danish population are seen in the secondary healthcare sector each year with a first-time diagnosis of shoulder disorder. The mean additional total costs for the 6-year period were €25,771 for individuals aged <65 years and €11,334 for individuals aged ≥65 years. Costs of sick leave accounted for 70% of the total costs for people in the working age, and a minor proportion of patients accounted for a substantial part of the total societal costs. In Denmark, expected annual costs associated with shoulder disorders were estimated at €1.21 billion. The OSS, EQ-5D and FABQ-PA revealed adequate responsiveness after arthroscopic decompression surgery and MIC estimates were established. The reliability of ID and HHD was sufficient for all positions and test modes. The measurement error was not sufficient and the ability of ID and HHD to measure changes less than 15% is questionable. In general, the quality of evidence was moderate to very low for ID and high to moderate for HHD.

Dansk resumé

Introduktion

Skulderlidelser er den tredjehyppigste muskuloskeletale lidelse. Udover at have stor betydning for patienternes dagligdag er skulderlidelser ofte forbundet med et stort ressourceforbrug, og de økonomiske konsekvenser er betydelige ikke mindst på grund af den høje incidens og prævalens. Den samfundsøkonomiske byrde ved skulderlidelser er imidlertid kun sparsomt belyst.

Det er essentielt at måle på udfald ved vurdering af sygdomsstatus, sygdomsudvikling samt effekt af behandling, både i forbindelse med forskning og monitorering af patienter i klinisk praksis. Til dette formål er det nødvendigt med måleinstrumenter med tilfredsstillende måleegenskaber. Formålet med denne afhandling var således:

1) at estimere de direkte og indirekte omkostninger forbundet med skulderlidelser hos patienter sammenlignet med matchede kontroller fra den generelle befolkning, samt at undersøge om omkostningerne var højere for specifikke patientgrupper (Studie I)

2) at evaluere følsomhed for ændring samt mindste relevante forskel (MIC) ved spørgeskemaerne Oxford Shoulder Score (OSS), EuroQol – 5 Dimensions (EQ-5D) og Fear-Avoidance Belief Questionnaire for fysisk aktivitet (FABQ-PA) hos patienter med afklemningssyndrom i skulderen, som gennemgår dekompressionskirurgi (Studie II)

3) at opsummere evidensen for måleegenskaber ved isokinetiske dynamometre (ID) og håndholdte dynamometre (HHD) til vurdering af skuldermuskelstyrke (Studie III og IV).

Metode

Studie I var et registerbaseret "cost-of-illness" studie, som inkluderede patienter diagnosticeret med skulderlidelser i perioden 2005 til 2017 samt køn- og aldersmatchede kontroller uden skulderlidelser. Sundhedsomkostninger fra den primære og sekundære sundhedssektor og omkostninger til sygefravær blev inkluderet for en 6-årig periode fra et år før til op til fem år efter diagnosetidspunktet. Studie II var et kohortestudie med 6-måneders follow-up, som inkluderede patienter med afklemningssyndrom i skulderen behandlet med dekompressionskirurgi. Følsomhed for ændring for OSS, EQ-5D ($EQ\text{-}5D_{index}$ og $EQ\text{-}5D_{vas}$) og FABQ-PA blev evalueret ved hjælp af Global Rating of Change Scale som primært anker og ændring i skuldersmerte og Subjective Shoulder Value som sekundære ankre. Studie III og IV var systematisk litteraturgennemgange, som inkluderede studier, der evaluerede reliabilitet, måleusikkerhed og konstruktionsvaliditet ved ID og HHD. Litteraturgennemgangene blev udført i overensstemmelse med metoden beskrevet af COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN).

Resultater

I studie I blev der identificeret 617,334 unikke individer diagnosticeret med skulderlidelser. Incidensen af skulderlidelser steg med 26% fra 966 per 100,000 person år i 2005 til 1.215 per 100.000 person år i 2017. For individer ≥65 år var sundhedsomkostninger 83% højere end for individer <65 år. Omkostningerne til sygefravær bidrog med omkring 70% af de totale omkostninger for individer <65 år. De gennemsnitlige totale meromkostninger (sundhedsomkostninger og omkostninger til sygefravær) for patienter med skulderlidelser sammenlignet med kontroller var for den 6-årige periode 25,771 € (95% CI 25,531; 26,012) for individer <65 år og 11,334 € (95% CI 11,014; 11,654) for individer ≥65 år. I alt tegnede 20% af patienterne med skulderlidelser sig for 66% af de totale omkostninger. De estimerede årlige meromkostninger for patienter med skulderlidelser sammenlignet med kontroller var 1.21 milliarder € inklusive sundhedsomkostninger og omkostninger til sygefravær.

I studie II var arealet under "receiver operating characteristic" kurven (ROC AUC) 0.96 (95% CI 0.91;1.00) for OSS, 0.82 (95% CI 0.66;0.99) for EQ-5D$_{index}$, 0.73 (95% CI 0.58;0.87) for EQ-5D$_{vas}$ og 0.74 (95% CI 0.58;0.90) for FABQ-PA. MIC-resultaterne var 6.0 points for OSS, 0.024 for EQ-5D$_{index}$, 10.0 for EQ-5D$_{vas}$ og -5.0 for FABQ-PA.

I studie III og IV var reliabiliteten af ID og HHD sufficient med størstedelen af "intraclass correlation coefficients" ≥0.70. Kvaliteten af evidensen var generelt moderat eller lav for ID og høj for HHD. Måleusikkerheden var insufficient eller ubestemmelig for både ID og HHD, idet ingen af de undersøgte strata havde en tilstrækkelig andel af resultater (≥75%), hvor den mindste målbare forskel var ≤15%. Kvaliteten af evidensen var generelt moderat til meget lav for ID og høj eller moderat for HHD. Konstruktionsvaliditeten af HHD viste inkonsistente resultater baseret på lav evidenskvalitet.

Konklusion

Omtrent 1.2% af den danske befolkning blev diagnosticeres med en skulderlidelse for første gang i den sekundære sundhedssektor. Omkostningerne til sygefravær udgjorde cirka 70% af de totale omkostninger for personer i den arbejdsdygtige alder, og en mindre andel af patienterne tegnede sig for en betydelig andel af de totale samfundsomkostninger. I Danmark blev de forventede årlige omkostninger forbundet med skulderlidelser anslået til 1.21 milliarder €. OSS, EQ-5D og FABQ-PA viste tilstrækkelig følsomhed for ændring efter artroskopisk dekompressionskirurgi og MIC-værdierne for de tre spørgeskemaer blev etableret. Reliabiliteten ved ID og HHD var tilstrækkelig. Målesikkerheden var ikke tilstrækkelig, og ID og HHDs evne til at måle ændringer på mindre end 15% er tvivlsom. Generelt var evidenskvaliteten moderat til meget lav for ID og høj eller moderat for HHD.

Introduction

The overall theme of this thesis is shoulder disorders. The included studies focus on different gaps in our knowledge about shoulder disorders. The first study adopts a societal perspective on shoulder disorders by mapping the resource utilization, health care costs and costs of sick leave associated with shoulder disorders. This is the first study worldwide estimating the costs of shoulder disorders based on nationwide databases. The other three studies evaluate the measurement properties of commonly used instruments for assessing patients with shoulder disorders in clinical practice and research. The second study focuses on a specific category of shoulder disorders, namely subacromial impingement syndrome, and the quality of some questionnaires, while the last two studies focus on the quality of instruments for measurement of shoulder muscle strength.

Musculoskeletal conditions and shoulder disorders

Musculoskeletal conditions are common among the adult population and up to one in seven of all consultations with the general practitioner is related to musculoskeletal complaints. (1, 2) In the UK, musculoskeletal conditions were the second-most common cause of sick leave in 2017, accounting for 18% of the total sickness absences. Hence, besides having consequences for the individual patient's health, musculoskeletal conditions often result in substantial utilisation of health care resources.

Shoulder complaints are the third most common musculoskeletal cause for seeking medical care (3, 4) with a prevalence between 7% and 26%. (4-6) Because the shoulder is involved in many upper-extremity activities, shoulder pain is associated with significant impairments in function and health-related quality of life (HR-QoL) and often limits the ability to work. (7, 8) Patients describe symptoms such as intense, unexpected pain and disturbed sleep leading to daytime tiredness. Furthermore, some patients suffer from restricted movement and reduced muscle strength leading to functional limitations. (3, 9)

In many western countries, the number of patients diagnosed with shoulder disorders and the number of shoulder surgeries have increased during the past decades. (8, 10-12) In Denmark, the incidence of rotator cuff-related lesions increased rapidly from 149 per 100,000 person years in 1996 to 715 per 100,000 person years in 2013. The increase was most prominent among people in the working age. (13) In the Netherlands, the incidence of upper extremity injuries increased by 13% from 970 to 1,098 per 100,000 persons in the period from 1986 to 2008. For shoulder injuries, the incidence differed depending on age and gender; in females, the highest incidence was seen from the age of 70 years onwards; in males, from age 10-35 years and from age 90 years onwards. Fractures were the most common injuries. (14)

Up to 70% of patients with shoulder pain are diagnosed as having shoulder impingement syndrome or rotator cuff-related lesions. (15, 16) First-line treatment of these disorders is conservative with exercise therapy, analgesia such as paracetamol, non-steroidal anti-inflammatory drugs or glucocorticoid injections. (17, 18) For patients with prolonged symptoms not responding to conservative treatment, surgical treatment with arthroscopic subacromial decompression may be an option. However, in recent years, evidence shows that subacromial decompression surgery is not more effective than conservative treatment including exercise therapy. (10, 15, 16, 18-21)

Economic burden

Management of shoulder disorders creates a large economic burden on society because of the high incidence and prevalence of these disorders. Few studies have investigated the use of health care services and productivity loss associated with shoulder disorders. An Australian cost-of-illness study focused on patients with shoulder pain on the waiting list at an orthopaedic department in a public hospital. The authors found that the mean annual per-patient costs of health care and domestic support were AU$7,563 (€4,826). When both absenteeism (absence from work) and presenteeism (loss of productivity while at work with a health condition) were included, the costs were AU$13,885 and AU$22,378 (€8,859 and €14,278, respectively). Of the included patients, 22% had surgical treatment and 51% of the hospital care costs were related to outpatient care. (22)

A Swedish cost-of-illness study investigating costs of primary health care in patients with shoulder pain revealed that the mean annual total costs were €4,139 per patient. Furthermore, the study showed that costs of sick leave contributed to more than 80% of the total costs, and a small proportion of the patients incurred very high costs. (23)

As with other musculoskeletal conditions, (24, 25) these two studies exposed that absence from work contributes to a substantial part to the costs associated with shoulder disorders. Work absence can be both temporary and permanent. A Danish register-based cohort study examined the risk of permanent work disability after surgery for rotator cuff-related disorders, frozen shoulder and osteoarthritis of the acromioclavicular joint. Results from this study showed that 10% became permanently work disabled within two years after surgery. (8)

Cost-of-illness studies

To assess the economic burden of health problems, e.g., shoulder disorders, cost-of-illness studies can be used as a descriptive analysis of the costs associated with the condition of interest. The traditional approach considers direct costs (health care costs) and indirect costs in terms of losses in productivity related to morbidity and mortality. All these costs are assessed in monetary values. (26) Other losses due to illness are related to the quality and length of life (intangible costs). However intangible costs are

rarely monetized, and these impacts may be best expressed in non-monetary values such as disability-adjusted life-years or quality-adjusted life-years. (26)

Cost-of-illness studies can aid our understanding of a health problem by assessing its impact on health care resource utilisation and labour market productivity. (26) However, cost-of-illness estimates quantify the economic burden associated with a disease and not how much of that burden could be saved if the disease was prevented. (26, 27)

The perspective of the study determines the type of costs included in the analysis. Perspectives range from the societal perspective, which is the broadest, to more narrow perspectives such as the hospital perspective. (26) Typically, cost-of-illness studies using a hospital perspective include as a minimum health care costs such as medical care expenditures for diagnosis, treatment and rehabilitation and sometimes also non-health care costs like transportation, household expenditures and informal care. Studies using a societal perspective usually include indirect costs from productivity losses in addition to health care costs. (28)

Furthermore, the approach used in a cost-of-illness study is relevant for the interpretation of the results. Studies using an incidence-based approach measure the costs of new (incident) cases and can show how costs vary with disease duration. Studies using a prevalence-based approach measure the costs of existing (prevalent) cases within a given time period. (26) For both approaches, the choice of time period in which the resource use is included needs to be considered, e.g., a long period is needed for long-lasting or chronic conditions.

Valuation of productivity loss is essential when calculating the costs associated with absence from work in a cost-of-illness study. Two of the most commonly used approaches are 1) the human capital approach which estimates the value of productivity loss as the value of an average individual's future earning, and 2) the friction cost approach which values the estimated actual production lost until the sick worker is replaced (friction period). (26, 27, 29)

When assessing the economic burden of shoulder disorders in a cost-of-illness study, it is important to take into account the most relevant approach, perspective, valuation of health care resources and productivity loss, and time frame for including costs for this specific disorder.

Outcome measures in patients with shoulder disorders

Measuring outcomes is a cornerstone in clinical practice and research to quantify the degree of impairment, guide treatment and evaluate treatment efficacy. In the literature, much variation in outcome measures is seen in studies examining benefits and harms of different interventions for shoulder disorders. The Outcome Measures in Rheumatology

(OMERACT) Shoulder Working Group was established in 2015. The group has developed and reached consensus on a core domain set to be included, as a minimum, in clinical trials of shoulder disorders in an attempt to harmonise outcome measures within this field of research. (30) Such harmonisation will enhance the ability to compare results across studies and pool data in meta-analyses.

The work of the OMERACT group was conducted in contained several steps. First, a systematic review exploring the outcome domains used in randomized trials of any interventions for shoulder disorders identified 409 studies, 32 outcome domains and 319 different measurement instruments. (31) Second, a two-round Delphi process including 91 (first round) and 96 (second round) clinicians, researchers and patients formed the basis for a preliminary core domain set. (32) Third, to determine if any potentially relevant outcomes were missing, a systematic review on qualitative studies investigating the patient's perspective of living with a shoulder disorder was performed. (33) Finally, at the OMERACT conference in 2018, the preliminary core domain set was presented, discussed and endorsed. (30) The final result from this process was a core domain set that includes four mandatory domains: pain, function, patient global shoulder, and adverse events including death; and four important but optional domains: participation (recreation/work), sleep, emotional well-being, and condition-specific domains such as pathophysiological manifestations (e.g., range of motion, muscle strength and radiographic outcome). (30)

These important outcomes can be measured using different instruments. The OMERACT Shoulder Working Group is currently working on defining a core outcome measurement set, which is a recommended collection of instruments to be used when measuring each of the core domains pointed out above. (30) Patient Reported Outcome Measures (PROMs) are simple and inexpensive methods to evaluate several of these domains. Furthermore, PROMs reflect the patient’s subjective assessment of the outcome of interest. The focus on the outcomes seen from a patient perspective when assessing the severity of diseases and limitations in daily life has increased in recent decades. (34) However, regardless of which instrument is used to measure these outcomes, it is crucial that the measurement properties of the instruments are adequate for the purpose for which they are used.

Measurement properties

When using outcome measurement instruments, both PROMs and performance-based ones, they must be valid, reliable and responsive.

The domain “validity” refers to “the degree to which an instrument truly measures the construct it purports to measure”. (35) This domain contains several different measurement properties reflecting different aspects of validity; content validity (including face validity), construct validity (including structural validity, hypotheses

testing and cross-cultural validity) and criterion validity. (35, 36) When assessing the construct validity of an instrument, the scores are compared with the scores of other similar instruments (convergent validity) or the scores are compared between subgroups of subjects (discriminative or known-groups validity). Criterion validity refers to a comparison with an instrument that is considered an accepted gold standard of the construct to be measured. (35, 36)

Scores of outcome measurement instruments are influenced by different sources of variation such as type of device, setting, instructions given to the participants and the tester's role. Standardisation of the procedure described in a test protocol can minimize the influence of these sources of variation. (37) However, variation may still exist. The domain "reliability" refers to the consistency of a measurement by considering "the degree to which the measurement is free from measurement error". (36) The overall reliability term includes the measurement properties; reliability, measurement error and internal consistency. (35, 36) Reliability is defined as "the proportion of the total variance in the measurements which is due to true differences between patients". (35) Reliability considers the instrument's ability to distinguish subjects from each other despite measurement error. (38-40) Measurement error is defined as "the systematic and random error of a patient's score that is not attributed to true change in the construct to be measured". (35) Measurement error assesses how close to each other the scores of repeated measurements are. (39-41) Internal consistency is defined as "the degree of the interrelatedness among items". (35) High-quality studies on reliability and measurement error are needed to understand the influence of different sources of variation on the measured scores. (37) Reliability and measurement error are examined in test-retest studies where no change is expected between the two test sessions. (36)

Assessing whether a disease status has changed over time is an important aspect of clinical practice and research. Therefore, measurement instruments must be responsive. (36) Responsiveness is defined as "the ability of an instrument to detect change over time in the construct to be measured". (35, 36) This means that if an instrument is responsive, the change measured by the instrument will reflect the change experienced by the patients in the construct of interest. (36) Responsiveness is an aspect of validity; the validity of a change score. It can be assessed using two different approaches: a criterion or construct approach. The criterion approach is appropriate when a gold standard exists for the construct to be measured; the construct approach, when no such gold standard exists. (36)

When evaluating interventions, it is relevant to know both whether results are statistically significant and clinically relevant. Therefore, another important aspect of change in a condition is the minimal important change (MIC). The MIC is defined as "the smallest change in score in the construct to be measured which patients perceive as important". (36, 42) Knowledge of the MIC is relevant when considering whether a

change is clinically relevant or not. For a given outcome measure, responsiveness and MIC may vary across patient groups and settings. Therefore, these measurement properties should be assessed in different populations and contexts. (42)

Measurement instruments used in patients with shoulder disorders

As described above, the OMERACT Shoulder Working Group recommends four mandatory and four important but optional domains to be included as a minimum in clinical trials of shoulder disorders. Besides these domains, studies focusing on other aspects related to shoulder disorders or comparing shoulder disorders with other conditions may choose to include domains not directly mentioned in the core domain set such as HR-QoL and fear-avoidance belief. These domains could either be considered domains that overlap with other domains, e.g., participation (recreation/work) or emotional well-being, or they could be considered independent domains. HR-QoL and fear-avoidance belief, along with pain, function and muscle strength, are often used in the literature reporting on shoulder disorders. Different instruments can be used to assess these outcome domains; pain, function, HR-QoL, fear-avoidance belief and muscle strength. Recommendations from the OMERACT Working Group regarding the choice of instruments for measuring each domain do not yet exist. (30) However, the following PROMs are frequently used in studies evaluating shoulder disorders: The Oxford Shoulder Score (OSS), EuroQol – 5 Dimensions (EQ-5D) and the Fear-Avoidance Belief Questionnaire (FABQ). Furthermore, dynamometers are commonly used as performance-based instruments to assess shoulder muscle strength.

The OSS is a commonly used PROM to assess pain and function in patients with shoulder disorders after both surgical and nonsurgical treatments. (43-48) The OSS has proven to be valid and reliable in patients with shoulder disorders. (43, 46) Additionally, the OSS has shown to be responsive in patients with rotator cuff disease receiving corticosteroid injection (49) and in patients with difficulty returning to usual activities after decompression surgery receiving occupational medical assistance or physical therapy. (45)

HR-QoL intends to capture an individual's perception of how an illness and its treatment affect the physical, mental and social aspect of his or her life. (36) Both disease-specific and generic instruments are commonly used to assess HR-QoL; however, the advantage of the generic instruments is that they allow comparisons across a variety of diseases. (50) Probably the most widely used generic instrument to assess HR-QoL is the EQ-5D, which is available in 170 languages. (51) Furthermore, the EQ-5D is frequently used to measure quality-adjusted life years in cost-effectiveness analyses. (51) Several studies examining the measurement properties of the EQ-5D in patients with upper-extremity orthopaedic disorders have found good validity and reliability and at least moderate responsiveness. (52-55)

In patients with musculoskeletal pain, fear-driven behaviours can potentially affect outcome negatively. (56, 57) The fear-avoidance model uses a biopsychosocial approach to provide an explanation of why acute pain can develop into chronic pain in a minority of patients. When acute pain is perceived as non-threatening, patients are likely to maintain engagement in daily activities, through which functional recovery is promoted. In contrast, if pain is misinterpreted in a catastrophizing manner, these catastrophizing thoughts can lead on to pain-related fear and associated safety-seeking behaviours such as avoidance. However, this behaviour could worsen the pain, have long-term consequences, such as disability and disuse, and may lower the threshold at which the person will experience pain. (57, 58) To assess this facet of behaviour, the FABQ was originally developed for patients with low-back pain. (59) Later, this questionnaire was adapted to patients with shoulder disorders. (56, 60-63) The FABQ consist of two subscales; physical activity (FABQ-PA) and work (FABQ-W). (59) The FABQ-PA has shown limited responsiveness in patients with subacromial impingement syndrome receiving physiotherapy treatment, but has not been evaluated in other shoulder patient groups. (63)

Management of shoulder pain and dysfunction often includes active exercise therapy with a focus on rotator cuff and scapular stabilisers. (10, 18, 64, 65) Muscle strength assessment can help quantify the degree of impairment and evaluate the effectiveness of a given treatment. Dynamometers are a useful objective method for clinicians to assess shoulder muscle strength. (66-68) Isokinetic dynamometers (ID) can measure muscle strength in different test modes; isometric, concentric or eccentric. (67) Furthermore, ID is capable of measuring muscle strength across a wide range of speeds with accommodating resistance at a constant angular velocity and assess the maximal torque production throughout a prescribed range of motion. (66, 67, 69) However, ID has the disadvantage of being expensive, time-consuming and stationary, and it occupies much space. In comparison, handheld dynamometers (HHD) are easier to use and therefore potentially more attractive for routine practice in a clinical setting and more attractive for research. (70, 71) Nevertheless, both types of dynamometers must present adequate reliability, measurement error and construct validity to be satisfactory for the intended use.

Rationale for the thesis

Shoulder disorders are common, are associated with pain and disability and health care utilisation, and often lead to sick leave from work. So, the consequences are considerable, but little is known about the economic burden of shoulder disorders. A nationwide register-based cost-of-illness study can provide evidence of the use of health care services, sick leave absence and accompanying societal costs associated with shoulder disorders. Results from such a study are valuable to decision makers in allocation of resources to different patient groups and can help identify areas for further research regarding preventive and health-promoting initiatives.

As stated in the Introduction, when conducting research and monitoring patients in clinical practice, evaluation of disease status, disease development and treatment efficacy by assessing outcomes is essential. For this purpose, measurement instruments with satisfactory measurement properties are required. The responsiveness and MIC of the OSS, EQ-5D and FABQ-PA are not well-established in patients with subacromial impingement syndrome undergoing arthroscopic subacromial decompression, even though adequate responsiveness and knowledge of the MIC is important for these PROMs to be valuable. Furthermore, shoulder muscle strength is commonly assessed with dynamometers, but systematic overviews of the reliability, measurement error and construct validity of ID and HHD have not been updated recently. Such an overview offers relevant evidence to clinicians who can use this information to plan research and routine monitoring, and interpret measured results.

Aims

The overall theme of the thesis was to fill two important gaps in the knowledge about patients with shoulder disorders. This will be done by estimating the economic burden associated with shoulder disorders and provide information on measurement properties of often used instruments to assess relevant outcomes in patients with shoulder disorders.

The thesis includes the following specific aims:

I. Estimate the direct and indirect costs associated with shoulder disorders in patients compared with matched controls from the general population, and to investigate if costs were higher for specific subgroups of patients **(Study I)**.

II. Evaluate the responsiveness and MIC of PROMs commonly used in shoulder disorders; OSS, EQ-5D and FABQ-PA in patients with subacromial impingement syndrome undergoing arthroscopic subacromial decompression **(Study II)**.

III. Summarize the evidence of the measurement properties of isokinetic and handheld dynamometry for assessment of shoulder muscle strength **(Study III and IV)**.

Methods

In this section, the methods of the of the studies included in the thesis will be described.

In Table 1, an overview of the study designs, populations and data analyses is presented. This is followed by a more detailed description of the studies; Study I and II are described separately and Study III and IV are described together as they are based on the same methods.

Table 1 Overview of study design, population, data source and analysis in the four studies in the thesis

		Study I	Study II	Study III and IV
Topic		Societal costs associated with shoulder disorders	The responsiveness and minimal important change of three questionnaires in patients treated with subacromial decompression	Measurement properties of handheld and isokinetic dynamometry
Design		Cost-of-illness study	Longitudinal cohort study with six month follow-up	Systematic review
Population		N = 617,334	N = 52	47 articles
Data source		Registers	Clinical data/PROMs	Electronic literature databases
Data analysis		Analysis of incidence, mean resource use and additional costs compared with matched controls	Receiver-operating curve statistics and correlation coefficients	Results from the included were summarised in an evidence synthesis

Study I

Design

The design was a retrospective register-based cost-off-illness study with a societal perspective on patients diagnosed with shoulder disorders.

Setting

For most patients with shoulder complaints, the general practitioner represents the first contact to the healthcare system, and acts as a gatekeeper for referral to further examination and treatment. If deemed necessary, the general practitioner can refer the patients to, e.g., physiotherapists, medical specialists and outpatient hospital care. (72) However, patients with traumatic shoulder injuries such as fractures or dislocation often have their first contact to the healthcare system at the emergency department. The

majority of the Danish health care expenditure is financed by the public healthcare system. Nevertheless, minor co-payments exist for some of the typical costs related to shoulder disorders, e.g., physiotherapy, chiropractor care and medication. Data on all contacts to both primary and secondary health care as well as weekly social security benefits are routinely collected and recorded in national registers. (72)

Study population

The International Classification of Diseases, 10th edition (ICD-10) diagnostic and treatment codes have been used since 1994. These codes are given by a medical doctor in the secondary healthcare sector. (73) Using data from the Danish National Patient Registry, we identified individuals aged 18 years or older and diagnosed with shoulder disorder in Denmark between 2005 and 2017.

The following primary diagnostic ICD-10 codes were included:

- Subacromial pain (DM751-9, DM709, DM791, DM795, DM799, DM255, DM629, DM796)
- Stiffness (DM190, DM191, DM198, DM750)
- Fracture (DS420, DS422)
- Dislocation (DS430, DS431, DS434, DS435, DS460)

The included diagnostic codes represent the majority of codes given to patients with shoulder disorders in clinical practice.

A comparison cohort of age- and gender-matched controls without shoulder disorders was randomly selected using a 1:1 ratio. Among the controls, no registration of shoulder disorders between 1996 and 2018 was allowed.

Both cases and controls were assigned a specific inclusion date in the study. For the cases, the inclusion date was the date when they received their first shoulder diagnosis within the inclusion period. For the controls, the inclusion date was the date equal to the inclusion date for the case to whom each control was matched.

The annual incidence rates per 100,000 person years was used to describe the number of new cases per year and to examine whether the incidence differed throughout the inclusion period. The incidence rates were reported for the total group of patients with shoulder disorders and for each of the four subcategories; subacromial pain, stiffness, fracture or dislocation.

All included cases and controls were alive and resident in Denmark on 31 December 2017.

Registers

The use of health care services and social security benefits was investigated using Danish longitudinal registers. Information from several national registers was linked at individual level with a unique personal identification number (CPR number). (72) Data were retrieved through Statistics Denmark.

To calculate the resource use, data from the following registers were included:

The Danish National Patient Registry (DNPR) began registrations in 1995. The register includes administrative information, primary and secondary diagnoses, treatment procedures and valuation of services from both public and private hospitals. The DNPR was used to extract information on primary diagnoses, visits to outpatient hospital care, hospital admissions, diagnostic tests and medical procedures. Valuation of these services was determined by the Diagnosis-Related-Grouping (DRG) reimbursement rate and the Danish Outpatient Grouping System (DAGS). (73) DNPR data were included from 2005 to 2018.

The Danish National Health Service Registry (NHSR) started in 1990. The register includes information of the activities of health professionals in the primary healthcare sector. In this study, data on visits to general practitioners, physiotherapists and chiropractors were extracted from the NHSR database, and valuation of these visits was calculated using the activity-based rates that are used for provider reimbursement. (74) NHSR data were included from 2005 to 2018.

The Danish Register for Evaluation of Marginalization (DREAM) database includes week-by-week registration on public transfer payments since 1991. The DREAM database was used to extract information on long-term sick leave. The number of sick leave days required to be defined as long-term sick leave has changed during the study period; >13 days up to 2007, >14 days in 2007-2008, >21 days in 2009-2011 and >30 days since 2012. (2, 75) The human capital approach was used to valuate costs of sick leave. (29) In this approach, the total number of weeks of sick leave were multiplied by age and gender-matched average gross salaries for each individual. Information on these average gross salaries was extracted from Statistics Denmark. Calculations were performed based on the assumption of full employment until the age of 65, which was the official retirement age in the period between 2005 and 2018. Therefore, only individuals under the age of 65 were included in the sick leave analyses. DREAM data were included from 2004 to November 2020.

At the time when data were retrieved from Statistics Denmark, 2005 was the first year when data from the NHSR were available, and 2018 was the last year when data from the DNPR were available. To have at least one year of follow-up, patients diagnosed between 2005 and 2017 were included.

Direct and indirect costs

Direct costs included costs of health care services from both the primary and secondary healthcare sector. Primary health care consists of visits to general practitioners, physiotherapists and chiropractors. Secondary health care consists of visits to medical specialist, outpatient hospital care, hospital admissions and diagnostic tests.

Indirect costs included in the present study were costs of sick leave.

All costs were calculated on a yearly basis and adjusted to 2020 prices using the general price index from Statistics Denmark. Costs were measured in Danish kroner and converted into euros at an exchange rate of 1€ = 7.44 Danish kroner (September 2020).

Follow-up period

The use of the registers provided a possibility to include data backwards and forwards in time from the date of the initial shoulder diagnosis. Symptoms of shoulder disorder can be both long-lasting and present some time before the first contact to the secondary healthcare sector where the diagnosis code is given. Therefore, we decided to include direct and indirect costs one year before and up to five years after the time of inclusion depending on data availability from each registry. This allowed for the possibility to follow the development in costs over time since diagnosis.

The decision to include data was based on availability of data from the registers. For individuals included in 2005, data on health care utilization the year before inclusion could not be added as data were not available for 2004. Data on sick leave were added for 2004. Data on the second to fifth year after inclusion are based on a declining number of individuals because individuals included in 2014 or later did not provide data for the full five-year follow-up period. Furthermore, the included number of individuals in the follow-up period was affected by censuring because of death.

Characteristics of study population

The characteristics of the study population were assessed at the inclusion date. Based on the first primary diagnosis, shoulder disorders were categorized into the four subcategories mentioned earlier (subacromial pain, stiffness, fracture and dislocation). The following characteristics are presented both for these subcategories and for the total group.

Table 2 Overview of outcomes, methods used for assessment and data availability

Outcome	Method	Registry
Health care use, primary healthcare sector	Valuation of visits using the activity-based rates used for provider reimbursement	The Danish National Health Service Registry (NHSR)
Health care use, secondary healthcare sector	Valuation of visits using the Diagnosis-Related-Grouping (DRG) reimbursement rates and the Danish Outpatient Grouping System (DAGS)	The Danish National Patient Register (DNPR)
Sick leave	Valuation by the number of weeks of sick leave multiplied by age- and gender-matched average gross salaries	The DREAM database, Statistics Denmark
Age and gender	All cases had a registration in the DNPR at the inclusion date. The majority of controls had a registration in the DREAM database (613,011 controls). A minority of controls (4,323, 0.7%) had no registration in the DREAM database, and data were imputed from the cases to whom they were matched	Cases: The Danish National Patient Register (DNPR) Controls: The DREAM database and imputation
Comorbidity	Measured using the Charlson Comorbidity Index based on primary ICD-10 codes from the 5-year period prior to inclusion. Categorized as 0, 1, 2-3, ≥4	The Danish National Patient Register (DNPR)
Marital status	Categorized as married, divorced, widowed, unmarried	The civil registry
Educational level	Highest level of education, categorized as low (primary, lower secondary, upper secondary), medium (vocational training, short-cycle higher education), high (bachelor, master, PhD) and not classified (no registration in the education register)	The education registry
Work status	Categorized as sick leave, subsidised employment, disability pension, employed, other (e.g., unemployed, education, maternity leave, retirement)	The DREAM database

Statistical analysis

Annual incidence rates of the Danish population are presented as the incidence rates per 100,000 person years for the four subgroups of patients with subacromial pain, stiffness, fracture or dislocation and for the total group. Statistics Denmark provided the

information on the size of the population aged ≥18 years for each year in the study period.

The mean (and 95% confidence interval (95% CI)) resource use and the related mean costs per person were estimated. These estimates were reported separately for general practice, physiotherapist, chiropractor, medical specialist care, hospital admission, outpatient visit and sick leave. The means are based on different numbers of individuals each year, as the availability of data for each individual covered by the registers differs through the follow-up period. The mean values account for these differences.

The different resource categories were summed into mean health care costs and reported separately for individuals aged <65 years and ≥65 years. Costs of sick leave and total costs (health care costs + costs of sick leave) were calculated for individuals aged <65 years. All costs were presented for each year (the year before and up to five years after inclusion) and for the total 6-year period.

Additional costs were estimated by subtracting the mean costs for controls from the mean costs for cases. Additional health care costs for individuals aged <65 years and ≥65 years, and additional costs of sick leave for individuals aged <65 years are illustrated separately for each of the four diagnosis categories. For patients diagnosed in 2017, the expected total costs were estimated by multiplying the mean additional costs by the number of patients in the age group <65 years and ≥65 years.

The impact of the change in the number of days required for sick leave to be defined as long-term sick leave and to be registered in the DREAM database was examined. This was done by calculating the proportion of individuals on sick leave, the mean number of weeks on sick leave and the costs of sick leave the first year after the diagnosis for each of the three subgroups; >13 or >14 days, >21 days, and >30 days.

The impact of comorbidity status on the total costs for the 6-year period was examined by calculating the additional costs of cases compared with controls for each Charlson Comorbidity Index score category; 0, 1, 2-3 and ≥4.

STATA 16.1 (StataCorp LLC, College Station, TX) was used for analyses.

Study II

Design

Study II was a longitudinal cohort study with a six-month follow-up.

Study population

Inclusion criteria were:

- patients diagnosed with subacromial impingement syndrome and treated with arthroscopic subacromial decompression
- age ≥18 years

Exclusion criteria were:

- frozen shoulder, full-thickness tear, osteoarthritis, trauma, cancer or neurological disorders
- previous surgical treatment in the affected shoulder
- inability to communicate in Danish.

The included patients were treated with decompression surgery between December 2018 and July 2020 at the Department of Orthopaedic Surgery, Aarhus University Hospital or Private Hospital Molholm, Vejle or Aarhus. Recruitment was first begun at Aarhus University Hospital; but due to organizational challenges, inclusion of patients went slower than planned and additional action was needed. Therefore, recruitment from Private Hospital Molholm commenced in September 2019 to increase the number of included patients.

We aimed at including a minimum of 50 subjects, as this number is recommended as being adequate for responsiveness studies. (76, 77)

Recruitment and data collection

Patients were recruited at the consultation with the orthopaedic surgeon when arthroscopic subacromial decompression was scheduled. Patients from Aarhus University Hospital completed the baseline assessment immediately after the consultation. The questionnaires were completed on a tablet with the possibility to use a paper form for those who had difficulties using the tablet. Patients recruited from Private Hospital Molholm received an email one or two days after the consultation with a link to the questionnaires. Two reminders were sent out to those who did not respond.

Six months after surgery, an email (and two reminders, if necessary) with a link to the follow-up questionnaires was sent out. Patients from Aarhus University Hospital were scheduled to an outpatient visit with a physical examination, and patients who did not complete the questionnaires in advance had the opportunity to complete them at the outpatient visit. The data were collected and managed using REDCap electronic data capture tool hosted at Aarhus University.

Measurement instruments

Baseline characteristics comprised age, sex, body mass index (BMI), comorbidity, duration of symptoms, working status, educational level and smoking status. This information was collected through electronic medical records and anamnesis.

Table 3 shows the measurement instruments used at baseline and the 6-month follow-up.

Table 3 Measurement instruments

Measurement instruments	Baseline	6-month follow-up
The Oxford Shoulder Score (OSS)	x	x
EuroQol – 5 Dimensions (EQ-5D)	x	x
Fear-Avoidance Belief Questionnaire Physical Activity subscale (FABQ-PA)	x	x
Pain at activity (pain visual analogue scale (VAS))	x	x
Subjective Shoulder Value (SSV)	x	x
Global Rating of Change Scale		x

Oxford Shoulder Score (OSS)

The OSS includes 12 items, four for pain and eight items for function. Items are scored from 0 (worst) to 4 (best) using a 5-point Likert scale. (44) A total score is calculated by summing up the scores of the 12 items, resulting in a range from 0 to 48. The total score is calculated if at least ten items have to be completed. (44)

EuroQoL – 5 Dimensions (EQ-5D)

In this study, the EQ-5D 5 level version was used. The EQ-5D questionnaire contains two parts, a descriptive part and a part assessing self-rated health. The descriptive part assesses five different aspects of health: mobility, self-care, usual activities, pain/discomfort, and depression/anxiety using a 5-point Likert scale. (53, 78) A key feature of the EQ-5D is the availability of "value sets" used to weigh the health states reported by patients into utility indexes ($EQ\text{-}5D_{index}$). In this study, we used the UK value set. These utility indexes range from -0.285 to 1.0. Full health corresponds to a value of 1.0, death corresponds to a value of 0, and negative values correspond to a health status considered to be worse than death. (79) The part assessing self-rated health used a visual analogue scale (VAS) (EQ_{vas}) ranging from 0 to 100. (78, 80)

Fear-Avoidance Belief Questionnaire Physical Activity subscale (FABQ-PA)

The FABQ-PA includes five items which are scored from 0 (strongly disagree) to 6 (completely agree) on a 7-point scale. A total score is calculated by summing up the scores of four of the five items, resulting in a range from 0 to 24. The higher the total score, the worse the outcome (more fear avoidance beliefs). (59)

Global Rating of Change Scale
The overall change of shoulder symptoms was rated using a 5-point Likert scale with the response options; much better, better, unchanged, worse or much worse. (81) Patients were asked "How would you describe your shoulder symptoms at present compared with before surgery?"

Pain measured with visual analogue scale (pain VAS)
Pain during activity in the affected shoulder was assessed using a VAS (pain VAS). The scale ranged from 0 to 100 with 0 corresponding to no pain and 100 corresponding to the worst possible pain. (82)

Subjective Shoulder Value (SSV)
The SSV rates the patients' subjective overall shoulder assessment expressed as a percentage of an entirely normal shoulder. (83) Patients were asked "What is the overall percent value of your shoulder if a completely normal shoulder represents 100%?"

To evaluate responsiveness, the Global Rating of Change Scale was treated as primary anchor and change in pain VAS and SSV as secondary anchors. The OSS measures shoulder-specific pain and function, which were considered to be constructs similar to the secondary anchors. The EQ-5D and FABQ-PA measure quality of life and fear avoidance behaviour, which were considered to be more complex constructs less similar to the secondary anchors.

Statistical analysis

The baseline demographic variables were presented as mean ± standard deviation (SD) or number and percentage. Change scores for all outcomes (OSS, EQ-$5D_{index}$, EQ_{vas}, FABQ-PA, pain VAS and SSV) were calculated by subtracting the baseline scores from the follow-up scores. These change scores are presented as mean with 95% CI. Patients were still included in the analyses if they had missing values in one outcome. Floor and ceiling effects (defined as more than 15% achieving the highest or lowest possible score) were assessed at baseline and follow-up. (36)

The responsiveness was evaluated using two approaches. Firstly, a receiver-operating characteristics (ROC) curve was made to assess the ability of the OSS, EQ-5D and FABQ-PA to correctly classify patients as improved (much better or better) or unimproved (unchanged, worse or much worse) according to the Global Rating of Change Scale. (76, 84) The ROC curve is a plot of the sensitivity against 1-specificity for different cut-off values. (85) An area under the curve (AUC) estimate of ≥0.70 was considered acceptable. (76, 77) Analysis of responsiveness of deterioration was not performed due to the low number of patients reporting worse or much worse outcomes.

Secondly, predefined hypotheses regarding expected correlations between the change scores of the OSS, EQ-$5D_{index}$, EQ_{vas} and FABQ-PA and the secondary anchors (change

scores of pain VAS and SSV) were tested. Because they were considered to measure similar constructs, higher correlations were expected between OSS vs. pain VAS and SSV. Lower correlations were expected between EQ-$5D_{index}$, EQ_{vas} and FABQ-PA vs. pain VAS and SSV since they were considered to measure overlapping but not similar constructs.

The following hypotheses were tested using Spearman rank correlation coefficients:

- the change score of the OSS has a positive correlation ≥0.5 when compared with SSV, and a negative correlation ≥0.5 when compared with pain VAS
- the change scores of the EQ-$5D_{index}$ and EQ_{vas} have a positive correlation ≥0.3 and <0.5 when compared with SSV, and a negative correlation ≥0.3 and <0.5 when compared with pain VAS
- the change score of the FABQ-PA has a negative correlation ≥0.3 and <0.5 when compared with SSV, and a positive correlation ≥0.3 and <0.5 when compared with pain VAS.

The MIC was defined as the change score in the OSS, EQ-$5D_{index}$, EQ_{vas} and FABQ-PA that best discriminated between the improved and not improved group of patients. The MIC was evaluated using the optimal cut-off point on the ROC curve, which is the value for which the sum of proportions of false positive and false negative classifications ((1-sensitivity) + (1-specificity)) is lowest. (84)

STATA 16.1 (StataCorp LLC, College Station, TX) was used for analyses. The statistical significance level was determined as $p<0.05$.

Study III and IV

Design

Study III and IV were both systematic reviews based on the same literature search. The following measurement properties were considered relevant for performance-based instruments like ID and HHD; reliability, measurement error and hypotheses testing for construct validity. We aimed to evaluate the measurement properties of both ID and HHD and report the results in a single review but separately for each type of dynamometer. A protocol was published describing this approach. (86) However, during the process, the data were considered too comprehensive for a single paper, and we decided to report the findings in two separate papers; one review focusing on ID and another on HHD.

The systematic reviews were performed in accordance with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology for systematic reviews of PROMs. (77) This methodology is based on existing guidelines

for reviews, such as the Preferred Reporting Items for Systematic Reviews and meta-Analysis statement (PRISMA), (87) the Cochrane Handbook for systematic reviews of interventions, (88) and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) principles. (89)

Search strategy and study selection

The electronic databases Cochrane Central Register of Controlled Trials (CENTRAL), PubMed/MEDLINE, EMBASE and Physiotherapy Evidence Database (PEDro) were search up to February 2020. A validated sensitive search filter to identify studies on measurement properties was used in combination with MeSH/Thesaurus and key words. No publication period or language restrictions were applied. In Paper IV, Appendix 1, (90) the search strategy for PubMed is presented.

Inclusion criteria were studies:

- evaluating ID or HHD used on the glenohumeral joint
- evaluating measurement properties
- including subjects aged ≥18 years with or without shoulder symptoms.

Exclusion criteria were studies:

- including patients with neurological, neuromuscular, systemic diseases or critical illness
- who did not separately report the results for each movement.

Data extraction

Characteristics of the instrument, study population (age, gender, healthy/symptomatic individuals), test procedure and results of the measurement properties were extracted from the included studies using a pre-tested form.

Quality assessment

First, the methodological quality of each study was assessed using the COSMIN Risk of Bias checklist. (77, 91) Second, the results of each study were rated as either sufficient (+), insufficient (-) or indeterminate (?) according to the described criteria for good measurement properties. (76, 77, 91)

Table 4 shows the described criteria and specifications for rating of the results in the two reviews for each assessed measurement properties.

Table 4 Described criteria for the included measurement properties

Measurement property	Described criteria	Specifications for rating of results
Reliability	ICC ≥0.70 (+) ICC not reported (?) ICC <0.70 (-)	As the described criteria.
Measurement error	MDC or LoA < MIC (+) MIC not defined (?) MDC or LoA > MIC (-)	No consensus exists on MIC in muscle strength testing. Based on available literature, we defined MIC as 15%. Sensitivity analyses were performed by setting the MIC to 10% and 20%, respectively.
Hypotheses testing for construct validity	The result in accordance with the hypothesis (+) No hypothesis defined (by the review team (?) The result is not in accordance with the hypothesis (-)	Hypothesis defined by the review team: Correlations between compared instruments (convergent validity) was ≥ 0.70. Comparison between groups expected to be different (discriminative validity) was significantly different.

ICC: intraclass correlation coefficient; MDC: minimal detectable change; LoA: limits of agreement; MIC: minimal important change

The study selection, data extraction, quality assessment and results rating were performed independently by two review authors. In case of disagreement that could not be resolved through discussion, a third review author was consulted.

Evidence synthesis

The results were summarized to determine the overall evidence of the measurement properties when test conditions were considered to be similar across studies. The results were summarized within the following strata; movement, (ID, HHD), test mode (ID), velocity (ID), position (ID) and intra- and inter-rater reliability (HHD).

To summarize, the reliability had to be measured using ICC; measurement error, using %MDC. If %MDC was not provided, calculation of %MDC was made if studies provided the required data.

For each stratum, the results were reported both as a range of minimum and maximum values and as the proportion of the results consistent with the described criteria. To determine the overall rating of results, the summarized results for each stratum were rated against the criteria for good measurement properties (Table 4). The rating was sufficient (+) if ≥75% of the results met the criteria, insufficient (-) if ≤25% met the

criteria, inconsistent (±) if 25-75% met the criteria and indeterminate (?) if the results of the individual studies were indeterminate. (92)

The GRADE approach, modified for reviews of measurement properties, was used to classify the quality of evidence as "high", "moderate", "low" or "very low". (77, 93) Risk of bias, inconsistency, imprecision and indirectness was used to downgrade the evidence. (77) Risk of bias refers to the methodological quality of the studies; inconsistency of the results of the studies within the pooled stratifications; imprecision to the total sample size of the included studies (downgraded with one level if the sample size was below 100, downgraded with two levels if the sample size was below 50); indirectness to the circumstance where included studies were partly performed in another population or context. (77)

Ethical approval and trial registration

The register-based study (Study I) was approved by the Danish Data Protection Agency (1-16-02-235-20). According to Danish legislation, register-based studies do not require individual consent or approval by Committees on Health Research Ethics.

The clinical cohort study (Study II) was approved by the Danish Data Protection Agency (1-16-02-534-18). Furthermore, the Regional Committee on Biomedical Research Ethics was invited to issue notification of the study, but further approval was not necessary (185/2018). Informed consent was given by all participants.

Before starting the reviews (Study III and IV), a protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (registration number CRD42017054027) and published. (86)

Results

In this section, the results of the individual studies are presented. Parts of Study III and IV are reported together.

Study I

Study I aimed to evaluate the costs associated with shoulder disorders. A total of 617,334 unique individuals with shoulder disorder (subacromial pain, stiffness, fracture or dislocation) were identified during the study period from 2005 to 2017. At the time of inclusion, the mean age was 50 years and 51% were females. Table 5 shows that patients with stiffness or fracture were older than patients with subacromial pain and dislocation and more patients with dislocation were males (70.5%). Compared with the controls, more cases had a low educational attainment (36.7% vs. 33.5%) and a Charlson Comorbidity Index score >0 (12.5% vs. 8.8%). Cases were more often on long-term sick leave (10.3% vs. 1.7%) and less often in jobs (45.7% vs. 55.1%) than the controls (Table 5).

Table 5 Characteristics of the study population at the time of inclusion (94)

	Controls	All cases	Cases Subacromial pain	Cases Stiffness	Cases Fracture	Cases Dislocation
	n=617,334	n=617,334	n=411,672	n=64,141	n=65,043	n=76,478
Age (years), mean (SD)	50.4 (16.4)	49.9 (16.4)	48.9 (15.6)	56.3 (13.4)	56.0 (18.1)	44.6 (17.9)
Gender, n (%)						
Male	301,376 (48.8)	301,376 (48.8)	189,675 (46.1)	28,874 (45.0)	29,681 (45.6)	53,146 (69.5)
Female	315,958 (51.2)	315,958 (51.2)	221,997 (53.9)	35,267 (55.0)	35,362 (54.4)	23,332 (30.5)
Charlson Comorbidity Index, n (%)						
0	562,736 (91.2)	540,296 (87.5)	361,201 (87.7)	54,443 (84.9)	54,670 (84.1)	69,982 (91.5)
1	32,325 (5.2)	47,261 (7.7)	31,134 (7.6)	5797 (9.0)	6133 (9.4)	4197 (5.5)
2-3	20,013 (3.2)	26,531 (4.3)	17,227 (4.2)	3511 (5.5)	3717 (5.7)	2076 (2.7)
≥4	2,260 (0.4)	3,246 (0.5)	2,110 (0.5)	390 (0.6)	523 (0.8)	223 (0.3)
Marital status, n (%)						
Married	370,440 (60.0)	368,579 (59.7)	252,612 (61.4)	41,999 (65.5)	32,569 (50.1)	41,399 (54.1)
Divorced	72,553 (11.8)	84,564 (13.7)	56,505 (13.7)	9462 (14.8)	9895 (15.2)	8702 (11.4)
Widowed	36,787 (6.0)	36,551 (5.9)	19,742 (4.8)	4783 (7.5)	8626 (13.3)	3400 (4.5)
Unmarried	137,554 (22.3)	127,640 (20.7)	82,813 (20.1)	7897 (12.3)	13,953 (21.5)	22,977 (30.0)
Educational level, n (%)						
Low	208,024 (33.5)	226,785 (36.7)	148,794 (36.1)	21,527 (33.6)	26,196 (40.3)	30,268 (39.6)
Medium	237,275 (38.4)	260,329 (42.2)	176,508 (42.9)	28,915 (45.1)	24,059 (37.0)	30,847 (40.3)
High	149,026 (24.1)	121,005 (19.6)	80,509 (19.6)	12,731 (19.9)	13,471 (20.7)	14,294 (18,7)
Not classified	24,009 (3.9)	9215 (1.5)	5861 (1.4)	968 (1.5)	1317 (2.0)	1069 (1.4)
Work status, n (%)						
Sick leave	10,563 (1.7)	63,665 (10.3)	41,448 (10.1)	6292 (9.8)	8048 (12.4)	7877 (10.3)
Subsidised employment	10,320 (1.7)	18,605 (3.0)	13,891 (3.4)	2196 (3.4)	1183 (1.8)	1335 (1.8)
Disability pension	30,062 (4.9)	39,129 (6.3)	26,234 (6.4)	4603 (7.2)	5338 (8.2)	2954 (3.9)
Employed	340,313 (55.1)	282,118 (45.7)	198,512 (48.2)	25,850 (40.3)	19,177 (29.5)	38,579 (50.4)
Other [a]	226,076 (36.6)	213,817 (34.6)	131,587 (32.0)	25,200 (39.3)	31,297 (48.1)	25,733 (33.7)

[a] Other include unemployed, education, maternity leave, early retirement, holiday

The number of new cases each year with a first-time primary diagnosis of shoulder disorder was 40,621 in 2005 rising to 55,669 in 2017. Thus, the annual incidence rate rose by 26% from 966 per 100,000 person years in 2005 to 1,215 per 100,000 person years in 2017. Subacromial pain was by far the most common condition, contributing with more than twice as high an incidence as the sum of stiffness, fracture and dislocation (Figure 1).

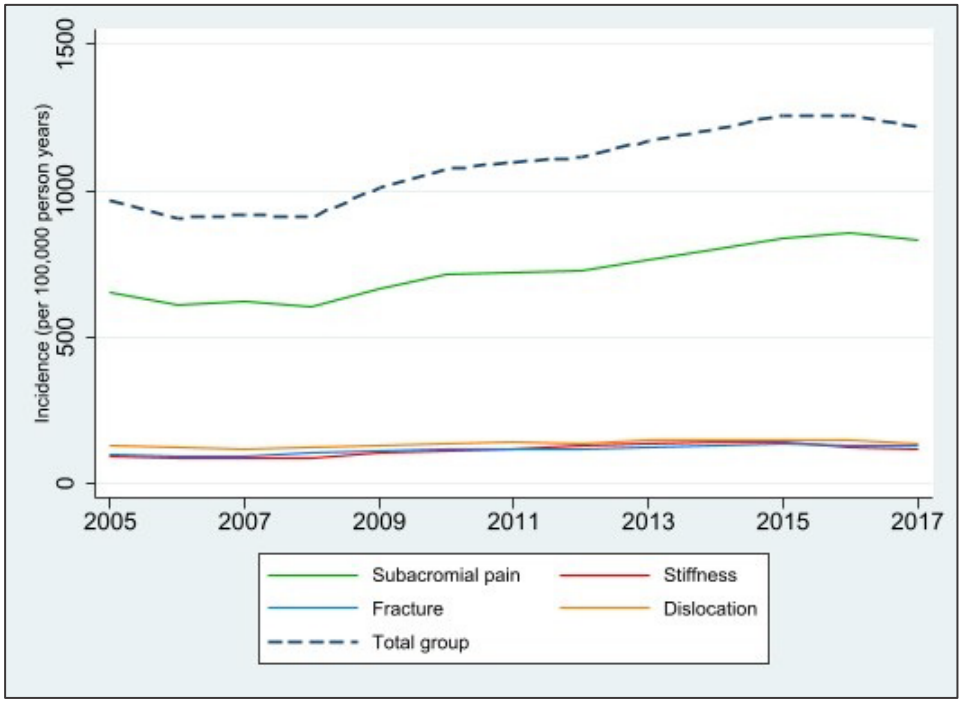

Figure 1 Annual incidence rates of shoulder disorders per 100,000 person years in individuals aged ≥18 years (94)

Table 6 shows the mean health care costs each year from the year before to five years after inclusion. The first year after the diagnosis, the health care costs increased by more than 50% compared with the year before. Secondary health care costs were much higher than primary health care costs with outpatient visits and hospital admission being the most expensive resource category. Compared with controls, cases continued having higher health care costs even five years after the diagnosis.

Table 7 details the mean costs per person for individuals aged <65 years and ≥65 years. Although individuals aged <65 years had much lower health care costs, the total costs for this age group were approximately 50% higher than for individuals aged ≥65 years because of the high costs of sick leave.

Table 6 Mean health care costs per person the year before up to five years after inclusion among patients diagnosed with shoulder disorders (cases) and age- and gender-matched controls. Costs are in euros in 2020 prices. (94)

Resource use category	Year before	1st year	2nd year	3rd year	4th year	5th year	In total (6-year period)
Primary health care, € (95% CI)	$n_{1,2}$=576,713	$n_{1,2}$=617,334	$n_{1,2}$=561,665	$n_{1,2}$=504,611	$n_{1,2}$=448,135	$n_{1,2}$=394,256	$n_{1,2}$=394,256
General practice							
Cases	195 (195;196)	189 (188;189)	179 (179;180)	179 (178;179)	180 (179;180)	181 (180;182)	1082 (1079;1085)
Controls	118 (118;119)	121 (120;121)	122 (122;123)	125 (124;125)	127 (126;127)	128 (128;129)	715 (713;718)
Physiotherapist							
Cases	60 (60;61)	68 (67;69)	60 (59;60)	60 (59;61)	62 (61;63)	63 (62;64)	354 (349;360)
Controls	28 (28;29)	30 (29;30)	31 (30;32)	32 (31;33)	33 (32;34)	34 (33;35)	172 (167;176)
Chiropractor							
Cases	6.3 (6.3;6.4)	5.4 (5.3;5.4)	5.2 (5.1;5.2)	5.2 (5.1;5.2)	5.1 (5.1;5.2)	5.1 (5.1;5.2)	32 (32;33)
Controls	3.2 (3.2;3.2)	3.2 (3.2;3.3)	3.2 (3.2;3.3)	3.3 (3.2;3.3)	3.3 (3.2;3.3)	3.3 (3.3;3.4)	20 (19;20)
Secondary health care, € (95% CI)							
Medical specialist							
Cases	126 (125;126)	113 (112;113)	116 (115;116)	117 (116;118)	119 (118;120)	120 (119;120)	696 (693;700)
Controls	78 (77;78)	79 (79;80)	81 (81;82)	83 (82;83)	84 (83;85)	85 (85;86)	473 (470;475)
Outpatient care							
Cases	1028 (1020;1036)	2010 (2001;2019)	1175 (1166;1183)	1138 (1129;1147)	1146 (1135;1157)	1165 (1154;1176)	7317 (7275;7360)
Controls	554 (548;560)	592 (586;598)	611 (604;618)	630 (623;637)	659 (650;667)	677 (668;686)	3414 (3383;3445)
Admission							
Cases	1264 (1247;1281)	1836 (1820;1852)	1268 (1253;1284)	1266 (1250;1283)	1295 (1277;1312)	1317 (1298;1336)	7756 (7696;7816)
Controls	673 (661;684)	703 (692;715)	726 (714;738)	754 (741;767)	785 (771;800)	808 (793;823)	4076 (4035;4117)
Total health care, € (95% CI)							
Cases	2680 (2660;2701)	4221 (4201;4241)	2803 (2783;2822)	2765 (2744;2786)	2806 (2784;2829)	2851 (2827;2876)	17,238 (17,152; 17,324)
Controls	1454 (1440;1469)	1528 (1514;1543)	1574 (1559;1590)	1626 (1610;1643)	1691 (1672;1709)	1736 (1716;1755)	8869 (8809;8930)

Abbreviations: $n_{1,2}$, number of cases and controls

Table 7 Mean costs per person the year before up to five years after inclusion for individuals aged ≥65 or <65 years for patients diagnosed with shoulder disorders (cases) and age- and gender-matched controls. Costs are in euros in 2020-prices. (94)

Resource use category	Year before	1st year	2nd year	3rd year	4th year	5th year	In total (6 year period)
Individuals aged ≥65 years	n_1=117,980 n_2=118,919	n_1=122,291 n_2=123,278	n_1=106,657 n_2=107,555	n_1=92,035 n_2=92,856	n_1=78,167 n_2=78,895	n_1=65,745 n_2=66,380	n_1=65,745 n_2=66,380
Total health care, € (95% CI)							
Cases	4446 (4389;4503)	6576 (6515;6636)	4628 (4570;4685)	4643 (4579;4707)	4727 (4656;4797)	4808 (4733;4884)	27,755 (27,499;28,010)
Controls	2660 (2617;2703)	2853 (2809;2898)	2981 (2931;3030)	3102 (3048;3156)	3177 (3117;3237)	3299 (3235;3362)	16,384 (16,194;16,575)
Individuals aged <65 years	n_1=458,733 n_2=457,794	n_1=495,043 n_2=494,056	n_1=455,007 n_2=454,110	n_1=412,575 n_2=411,755	n_1=369,966 n_2=369,239	n_1=328,509 n_2=327,875	n_1=328,509 n_2=327,875
Total health care, € (95% CI)							
Cases	2226 (2206;2247)	3639 (3619;3659)	2375 (2355;2395)	2346 (2325;2367)	2401 (2378;2424)	2460 (2435;2484)	15,133 (15,045;15,222)
Controls	1141 (1127;1155)	1197 (1183;1211)	1241 (1226;1257)	1294 (1278;1310)	1373 (1355;1391)	1419 (1399;1439)	7348 (7287;7409)
Sick leave, € (95% CI)							
Cases	4488 (4450;4525)	7988 (7937;8038)	4376 (4336;4417)	3401 (3365;3438)	2985 (2949;3021)	2713 (2677;2748)	27,542 (27,368;27,716)
Controls	1681(1658;1703)	1676 (1653;1699)	1628 (1605;1652)	1589 (1564;1613)	1539 (1513;1564)	1498 (1472;1524)	9621 (9523;9718)
Total costs, € (95% CI)							
Cases	6598 (6551;6645)	11,627 (11,569;11,684)	6751 (6704;6799)	5747 (5703;5792)	5385 (5340;5431)	5172 (5126;5219)	42,675 (42,468;42,882)
Controls	2838 (2808;2868)	2873 (2844;2902)	2870 (2839;2900)	2982 (2851;2914)	2911 (2878;2945)	2917 (2882;2953)	16,969 (16,842;17,095)

n_1, number of cases; n_2, number of controls

For both cases and controls, the majority of individuals had a low resource use, but few individuals had a very high resource use. As illustrated in Figure 2, the 20% of cases with the highest costs accounted for 66% of the total costs, and the 50% of cases with the highest costs accounted for 90% of the total costs. When examining sick leave, 52% of the cases and 73% of the controls had no registration of sick leave during the 6-year period. In total, 9.5% of the cases and 2.5% of the controls had costs of sick leave exceeding €100,000.

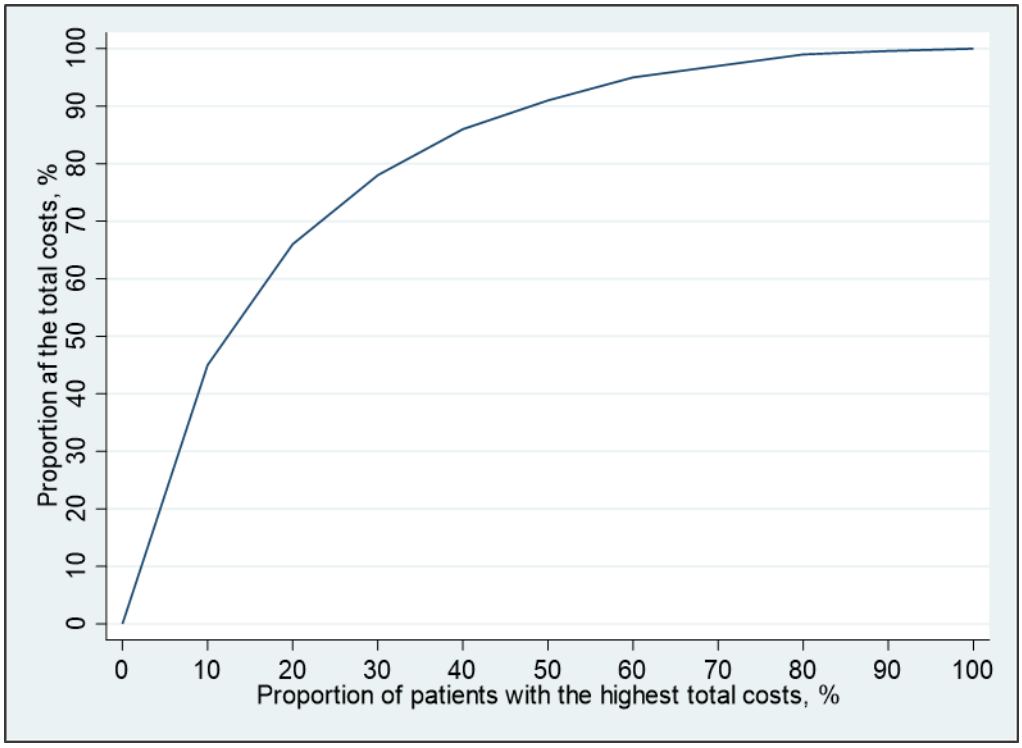

Figure 2 The added proportion of cases with the highest costs accounting for a given proportion of the total costs

When cases were compared with controls, the mean additional health care costs for the 6-year period were €7,760 (95% CI 7,654; 7,866) for individuals aged <65 years and €11,334 (95% CI 11,014; 11,654) for individuals aged ≥65 years (Figure 3). For individuals in the working age, the mean additional costs of sick leave were €18,011 (95% CI 17,813; 18,209) and the mean additional total costs were €25,771 (95% CI 25,531; 26,012). This means that for individuals aged <65 years, 70% of the additional total costs were costs of sick leave.

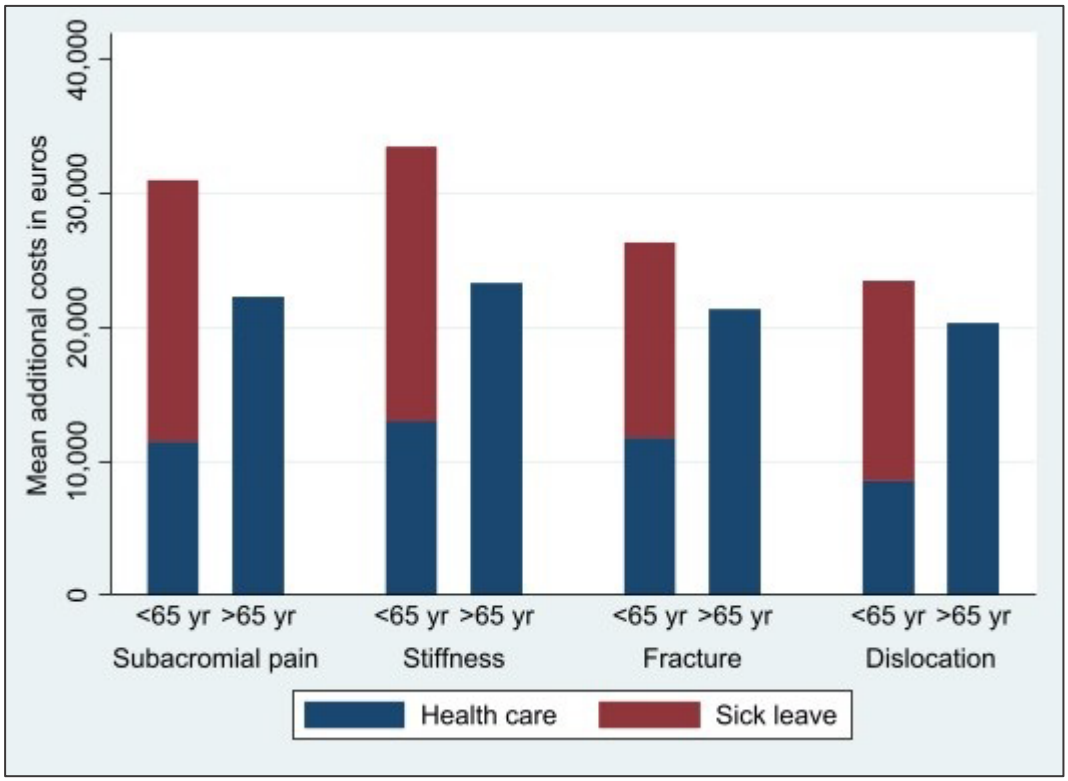

Figure 3 Mean additional costs per person for the 6-year period from the year before to five years after inclusion. Costs are in euros in 2020 prices. (94)

In 2017, which is the most recent year, 55,669 patients with shoulder disorders were identified. For these patients, the expected additional costs for the 6-year period were estimated to €1.21 billion; €0.49 billion to health care costs and €0.72 billion to costs of sick leave.

As shown in Table 8, the proportion of individuals on sick leave, the number of weeks on sick leave and the costs of sick leave decreased as the number of days required for long-term sick increased. About half of the individuals had been included since 2012 when >30 days were required to be defined as long-term sick leave.

Table 8 The proportion of individuals aged <65 years on sick leave, the mean number of weeks on sick leave and the mean costs of sick leave the first year after the diagnosis in groups depending on definition of long-term sick leave (>13 or 14 days, >21 days, >30 days)

Definition of long-term sick leave	Proportion of individuals on sick leave, %	Weeks on sick leave, weeks (95% CI)	Mean costs of sick leave, € (95% CI)
>13 or 14 days			
Cases, n=136,537	33.9	6.6 (6.5;6.7)	9508 (9406;9611)
Controls, n=136,312	10.5	1.14 (1.11;1.16)	1637 (1596;1679)
>21 days			
Cases, n=113,007	29.0	5.6 (5.5;5.7)	8635 (8523;8747)
Controls, n=112,742	9.4	1.14 (1.11;1.18)	1752 (1703;1802)
>30 days			
Cases, n=245,499	23.7	4.4 (4.3;4.4)	6843 (6776;6911)
Controls, 245,002	8.5	1.06 (1.04;1.08)	1662 (1630;1695)
In total			
Cases, n=495,043	27.7	5.3 (5.2;5.3)	7988 (7937;8038)
Controls, n=494,056	9.2	1.10 (1.09;1.12)	1676 (1653;1699)

Table 9 shows the mean additional costs for the 6-year period for each of the four comorbidity categories, Charlson Comorbidity Index score 0, 1, 2-3 and ≥4. The additional costs of cases compared with controls were almost similar up to index score 3 and somewhat higher for index score 4 or higher.

Table 9 The mean total costs and additional costs for the 6-year period from the year before up to five years after inclusion in groups depending on Charlson Comorbidity Index score (0, 1, 2-3, ≥4). Total costs include health care costs and costs of sick leave for individuals aged <65 years and health care costs for individuals aged ≥65 years

Charlson Comorbidity Index score	Mean total costs, € (95% CI)	Additional costs of cases compared with controls, € (95% CI)
0		
Cases, n=349,012	37,760 (37,580;37,942)	22,580 (22,556;22,604)
Controls, n=363,576	15,180 (15,074;15,286)	
1		
Cases, n=29,392	54,057 (53,331;54,781)	21,732 (21,680;21,784)
Controls, n=19,423	32,325 (31,648;33,001)	
2-3		
Cases, n=14,379	64,726 (63,424;66,028)	22,376 (22,307;22,445)
Controls, n=10,273	42,350 (41,207;43,493)	
≥4		
Cases, n=1,473	98,796 (93,006;104,587)	28,480 (28,330;28,630)
Controls, n=984	70,316 (64,595;76,038)	
In total		
Cases, n=394,256	40,187 (40,009;40,365)	23,317 (23,293;23,341)
Controls, n=394,256	16,870 (16,760;16,980)	

Study II

In Study II, the responsiveness and MIC of the OSS, EQ-5D and FABQ-PA were evaluated. A total of 58 patients were included in the study; 45 from Aarhus University Hospital and 13 from Private Hospital Molholm. Three patients withdrew from the study and three patients did not respond at follow-up, leaving 52 patients (90%) for analysis. The SSV and pain VAS were left unanswered for one patient, the OSS for two patients and the FABQ-PA for three patients. Table 10 shows the baseline characteristic of the study population.

Table 10 Baseline characteristic of the study population (n=52) (95)

Sex, n (%)	
Male	26 (50.0)
Age, y, mean (SD)	57.4 (10.1)
BMI, kg/m^2, mean (SD)	26.7 (4.3)
Number of comorbidities, n (%) [a]	
0	21 (40.4)
1	17 (32.7)
2	7 (13.5)
>2	6 (11.5)
Time with symptoms, n (%) [a]	
0-6 months	5 (9.6)
6-12 months	11 (21.2)
12-24 months	12 (23.1)
>24 months	24 (46.2)
Working status, n (%)	
No	13 (25.0)
Yes	39 (75.0)
Educational level, n (%) [a]	
Compulsory school	8 (15.7)
Skilled worker	23 (45.1)
Bachelor	14 (27.5)
Master's degree	6 (11.8)
Smoking status, n (%)	
No	38 (73.1)
Yes	14 (26.9)
OSS, mean (SD) [a]	29.4 (6.6)
EQ-5D$_{index}$, mean (SD)	0.71 (0.17)
EQ$_{VAS}$, mean (SD)	68.7 (19.5)
FABQ-PA, mean (SD) [a]	15.5 (5.9)
Pain VAS, mean (SD)	61.8 (23.3)
SSV, mean (SD) [a]	55.5 (18.4)

BMI, Body Mass Index; OSS, Oxford Shoulder Score; EQ-5D$_{index}$, EQ-5D utility index; EQ$_{vas}$, EQ visual analogue scale; FABQ-PA, Fear-Avoidance Belief Questionnaire for physical activity; VAS, visual analogue scale; SSV, Subjective Shoulder Value,[a] data missing for one patient

Answering the Global Rating of Change Scale, 25 patients reported much better, 15 patients better, 5 patients unchanged, 7 patients worse and no patients reported much worse. Altogether, 40 patients were categorized as improved (77%) and 12 patients (23%) as unimproved. Furthermore, no floor or ceiling effect was observed in the three questionnaires.

The mean change scores from baseline to 6-month follow-up of all outcomes are shown in Table 11. The differences between the improved and unimproved group of patients were statistically significant for all outcomes.

Table 11 Mean change scores of the Oxford Shoulder Score, EQ-5D utility index, EQ visual analogue scale, Fear-Avoidance Belief Questionnaire for physical activity, Subjective Shoulder Value and shoulder pain on VAS scale for the total group, the improved and unimproved group, and the difference between groups (95)

	Total group (n=52) mean (95% CI)	**Improved group (n=40) mean (95% CI)**	**Unimproved group (n=12) mean (95% CI)**	**Difference between groups mean (95% CI)**
OSS (0-48)	9.5 (6.9;12.1)[b]	12.7 (10.3;15.0)[a]	-1.7 (-4.8;1.4)[a]	14.4 (9.7;19.1)*
$EQ\text{-}5D_{index}$ (-0.205-1.0)	0.10 (0.06;0.14)	0.13 (0.10;0.17)	-0.03 (-0.13;0.07)	0.16 (0.08;0.25)*
$EQ\text{-}5D_{vas}$ (0-100)	1.3 (-3.4;5.9)	3.9 (-1.6;9.3)	-7.3 (-15.5;1.0)	11.1 (0.4;21.8)*
FABQ-PA (0-24)	-4.1 (-6.0;-2.2)[c]	-5.3 (-7.5;-3.1)[c]	-0.3 (-3.9;3.2)	-4.9 (-9.2;-0.7)*
SSV (0-100)	19.1 (12.6,25.7)[a]	25.5 (18.8;32.3)[a]	-1.7 (-13.1;9.8)	27.2 (13.7;40.7)*
Pain VAS (0-100)	-38.0 (-46.5;-29.5)[a]	-43.2 (-52.6;-33.7)[a]	-21.3 (-39.2;-3.3)	-21.9 (-41.0;-2.8)*

OSS, Oxford Shoulder Score; $EQ\text{-}5D_{index}$, EQ-5D utility index; EQ_{vas}, EQ visual analogue scale; FABQ-PA, Fear-Avoidance Belief Questionnaire for physical activity; SSV, Subjective Shoulder Value; VAS, visual analogue scale

[a] data missing for one patient, [b] data missing for two patients, [c] data missing for three patients

* significant at $P<0.05$

Table 12 presents the correlations between the change scores of the investigated PROMS and the secondary anchors that were used to test the predefined hypotheses. The results showed that:

- The change scores of the OSS revealed correlations as hypothesized when compared with the change scores of the SSV and pain VAS

- The change scores of the EQ-5D$_{index}$ revealed correlations as hypothesized when compared with the change scores of the SSV and pain VAS; the EQ-5D$_{vas}$ revealed correlations lower than expected
- The change scores of the FABQ-PA revealed correlations as hypothesized when compared with the change scores of pain VAS; when compared with the SSV, the correlation was lower than expected.

Table 12 Spearman correlation coefficients for the Oxford Shoulder Score, EQ-5D utility index, EQ visual analogue scale and Fear-Avoidance Belief Questionnaire for physical activity compared with the change scores of the Subjective Shoulder Value and shoulder pain on VAS scale

	OSS (0-48)[b]	EQ-5D$_{index}$ (-0.205-1.0)	EQ-5D$_{vas}$ (0-100)	FABQ-PA (0-24)[c]
SSV [a]	0.67	0.50	0.28	-0.28
Pain VAS [a]	-0.58	-0.47	-0.23	0.39

OSS, Oxford Shoulder Score; EQ-5D$_{index}$, EQ-5D utility index; EQ$_{vas}$, EQ visual analogue scale; FABQ-PA, Fear-Avoidance Belief Questionnaire for physical activity; SSV, Subjective Shoulder Value; VAS, visual analogue scale [a] data missing for one patient, [b] data missing for two patients, [c] data missing for three patients

Figure 4 illustrates the ROC curve analysis. The ROC AUC results were 0.96 (95% CI 0.91;1.00) for OSS, 0.82 (95% CI 0.66;0.99) for EQ-5D$_{index}$, 0.73 (95% CI 0.58;0.87) for EQ-5D$_{vas}$ and 0.74 (95% CI 0.58;0.90) for FABQ-PA.

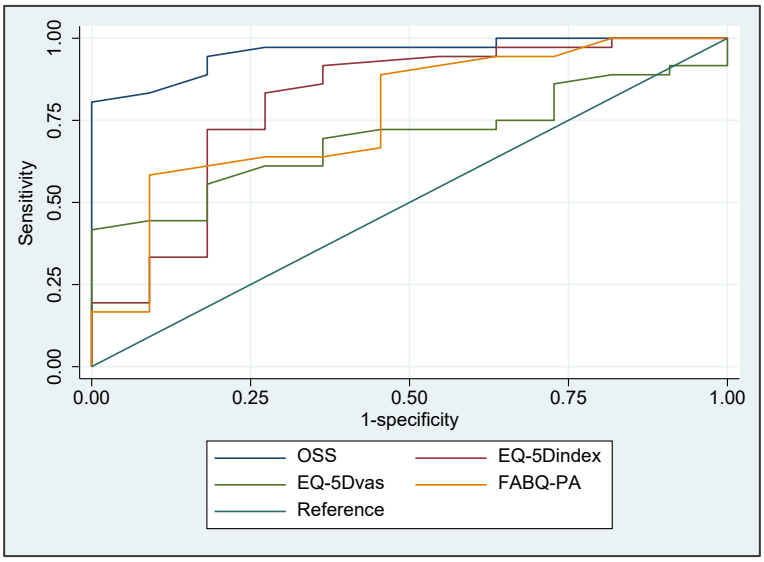

Figure 4 Receiver-operating characteristic (ROC) curve

The MIC ROC cut off points were 6.0 points for the OSS, 0.024 for the EQ-5D$_{index}$, 10.0 for the EQ-5D$_{vas}$ and -5.0 for the FABQ-PA.

Study III and IV

Study III and IV aimed to summarize the evidence of the measurement properties of ID and HHD for the assessment of shoulder muscle strength in two systematic reviews of the literature. As seen from Figure 5, the electronic search strategy identified 8,054 records. In total, 47 studies were included in the two reviews. Of these, 21 studies with 963 participants evaluated ID, 28 studies with of 597 participants evaluated HHD and two studies evaluated both types of dynamometers.

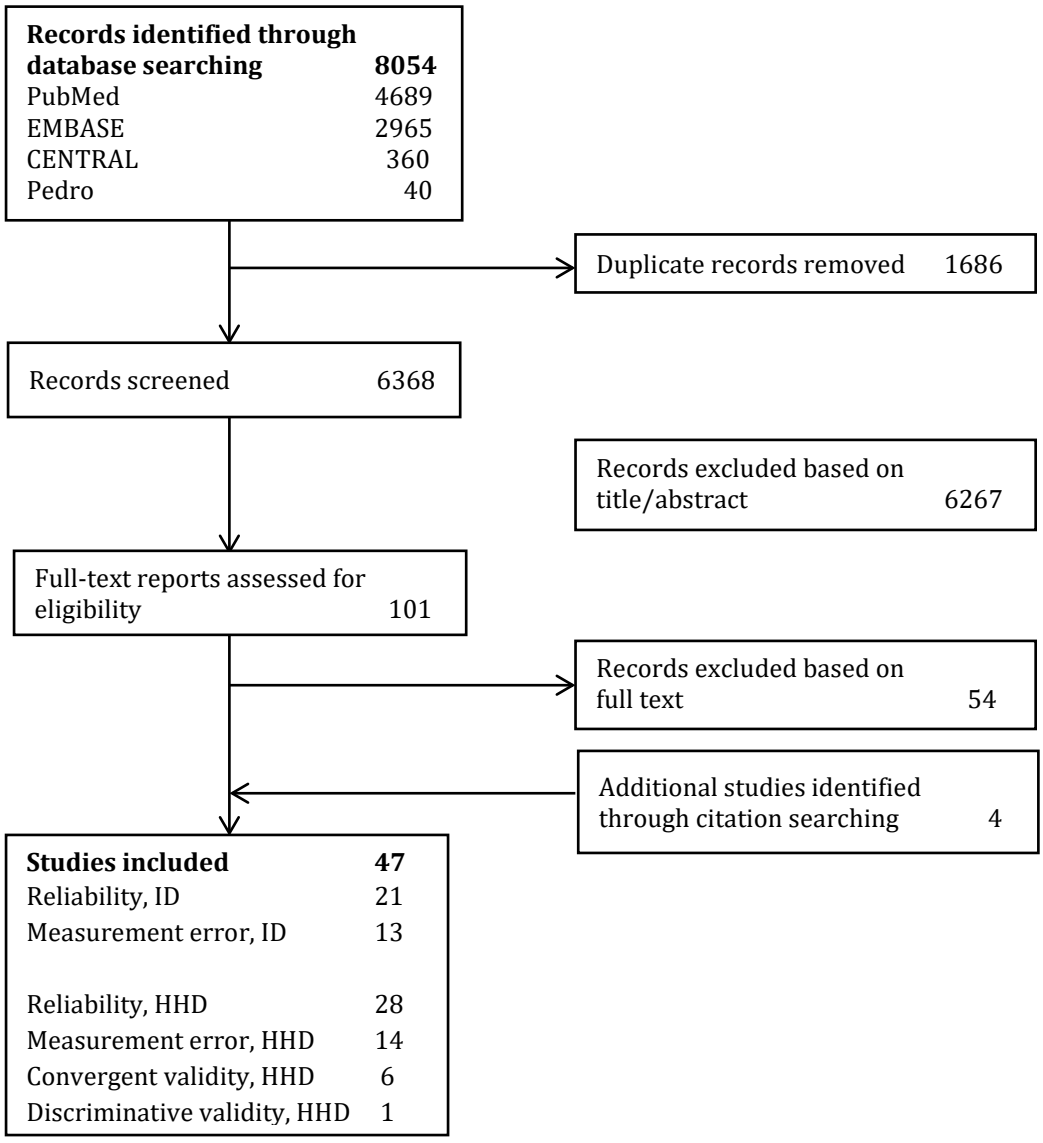

Figure 5 Flowchart of studies through the selection process. (90, 96)
CENTRAL: Cochrane Central Register of Controlled Trials; PEDro: Physiotherapy Evidence Database; HHD: handheld dynamometers; ID: isokinetic dynamometers

Characteristics of the studies included in the two papers are shown in Paper III, Table 1 (95) and Appendix B (supplementary material Paper IV). The results of the measurement properties extracted from each study are shown in Appendix A (supplementary material Paper III) and Paper IV, Appendix 2 and 3. (90)

Isokinetic dynamometry (Study III)

The population included 563 healthy subjects in 19 studies and 34 subjects with shoulder disorders in two studies. Results are summarized and presented separately for the isometric, concentric and eccentric test mode; for the velocities 30-60°/s, 90°/s, 120°/s, and 240°/s; for the seated, supine and standing position; and for the movements internal rotation (IR), external rotation (ER) and the ER/IR ratio (ER divided by IR). Other movements were evaluated, but in few studies, and these results are not summarized but presented in Appendix D.

Table 13 shows the rating of results and risk of bias assessment for the reliability and measurement error.

Table 13 Rating of results and risk of bias assessment for reliability and measurement error of isokinetic dynamometry (96)

Author/year	Reliability		Measurement error	
	Rating of results	Risk of bias assessment	Rating of results	Risk of bias assessment
Anderson, 2006 (97)	+	D	-	A
Dauty, 2003 (98)	+	A	?	A
Cavuoto, 2019 (99)	+	A	+	A
Edouard, 2013 (100)	+	A	-	A
Forthomme, 2011 (101)	?	I	-	A
Frisiello, 1994 (102)	+	D		
Grabowski, 2017 (103)	+	V	-	V
Habets, 2018 (69)	+	A	-	A
Hill, 2005 (104)	?	D		
Kramer, 1996 (105)	?	D	-	D
Leggin, 1996 (106)	+	D		
Lindström, 2003 (107)	?	D	-	V
Magnusson, 1990 (108)	+	D		
Malerba, 1993 (109)	?	D		
Mandalidis, 2001 (110)	+	D	-	A
Mayer, 1994 (111)	?	I		
Meeteren, 2002 (112)	+	I	-	I
Papotto, 2016 (113)	+	D	-	A
Plotnikoff, 2002 (114)	+	D	-	D
Smith, 2001 (115)	+	D		
Sullivan, 1988 (116)	?	D		

+, sufficient; -, insufficient; ?, indeterminate; V, very good; A, adequate; D, doubtful; I, inadequate

Reliability

Reliability was evaluated in all 21 studies. The summarized results, the rating of results and the quality of evidence are shown in the summary of findings table (Table 14). Of the 30 strata examined, 28 had their results rated as sufficient and two as indeterminate. The quality of evidence varied depending on the strata examined; "high" in 7% of strata, "moderate" in 33%, "low" in 33% and "very low" in 27%.

Table 14 Summary of findings for the reliability of isokinetic dynamometry (96)

	Number of studies, number of participants	**Summary of results ICC, range**	**Summary of results Proportion of ICC≥0.70**	**Overall rating of results**	**Quality of evidence**
IR isometric					
Seated	2 studies, n=27	0.82-0.99	100%	sufficient	low
Supine	1 study, n=17	0.87-0.89	100%	sufficient	very low
Standing	1 study, n=17	0.96-0.97	100%	sufficient	very low
IR concentric					
Seated, 30-60°/s	8 studies, n=176	0.39-0.97	94%	sufficient	moderate
Seated, 90°/s	2 studies, n=27	0.32-0.98	66%	indeterminate	low
Seated, 120°/s	5 studies, n=132	0.11-0.98	90%	sufficient	moderate
Supine, 60°/s	2 studies, n=66	0.86-0.94	100%	sufficient	moderate
Supine, 90°/s	1 study, n=17	0.87-0.93	100%	sufficient	very low
Supine, 120°/s	2 studies, n=66	0.88-0.94	100%	sufficient	moderate
IR eccentric					
Seated, 30-60°/s	4 studies, n=115	0.70-0.96	100%	sufficient	high
Seated, 120°/s	1 study, n=31	0.77-0.86	100%	sufficient	very low
Standing, 90°/s	1 study, n=18	0.75-0.78	100%	sufficient	low
Standing, 120°/s	1 study, n=18	0.83	100%	sufficient	low
ER isometric					
Seated	3 studies, n=37	0.85-0.99	100%	sufficient	low
Supine	1 study, n=17	0.73-0.88	100%	sufficient	very low
Standing	2 studies, n=61	0.96-0.97	100%	sufficient	moderate
ER concentric					
Seated, 30-60°/s	8 studies, n=176	0.70-0.95	100%	sufficient	high
Seated, 90°/s	2 studies, n=27	0.74-0.99	100%	sufficient	low
Seated, 120°/s	5 studies, n=132	0.62-0.92	80%	sufficient	moderate
Supine, 60°/s	2 studies, n=66	0.84-0.93	100%	sufficient	moderate
Supine, 90°/s	1 study, n=17	0.75-0.82	100%	sufficient	very low
Supine, 120°/s	2 studies, n=66	0.74-0.94	100%	sufficient	low
ER eccentric					
Seated, 30-60°/s	6 studies, n=139	0.44-0.98	83%	sufficient	moderate
Seated, 120°/s	2 studies, n=45	0.86-0.96	100%	sufficient	low
Supine, 60°/s	1 study, n=49	0.78-0.91	100%	sufficient	very low
Supine, 120°/s	1 study, n=49	0.72-0.80	100%	sufficient	very low
Standing, 90°/s	1 study, n=18	0.78-0.86	100%	sufficient	low
Standing, 120°/s	1 study, n=18	0.83	100%	sufficient	low
ER/IR concentric					
Seated, 60°/s	2 studies, n=60	0.50-0.79	75%	sufficient	moderate
Seated, 120°/s	2 studies, n=60	0.53-0.81	50%	indeterminate	moderate

n, number of participants

Measurement error

Measurement error was evaluated in 13 studies. The summarized results, the rating of results and the quality of evidence are shown in the summary of findings table (Table 15). All strata were rated as insufficient and the %MDC values ranged from 12% to 77%. Of the 22 strata examined, the quality of evidence was "high" in 9% of strata, "moderate" in 27%, "low" in 23% and "very low" in 41%.

Table 15 Summary of findings for measurement error of isokinetic dynamometry (96)

	Number of studies, number of participants	**Summary of results %MDC, range**	**Summary of results Proportion of %MDC≤15%**	**Overall rating of results**	**Quality of evidence**
IR isometric					
Seated	1 study, n=40	38.8	0%	insufficient	very low
IR concentric					
Seated, 30-60°/s	7 studies, n=173	11.0-77.1	0%	insufficient	high
Seated, 120°/s	3 studies, n=117	23.3-73.7	0%	insufficient	moderate
Seated, 240°/s	1 study, n=12	26.1	0%	insufficient	very low
Supine, 60°/s	2 studies, n=61	19.7-29.7	0%	insufficient	moderate
Supine, 120°/s	1 study, n=49	27.2-40.2	0%	insufficient	very low
Supine, 240°/s	1 study, n=12	22.2-34.1	0%	insufficient	very low
IR eccentric					
Seated, 30-60°/s	4 studies, n=131	25.5-69.0	0%	insufficient	moderate
Seated, 120°/s	2 studies, n=71	46.3-69.0	0%	insufficient	low
ER isometric					
Seated	2 studies, n=50	12.2-40.2	13%	insufficient	low
ER concentric					
Seated, 30-60°/s	7 studies, n=173	27.2-72.1	0%	insufficient	high
Seated, 120°/s	3 studies, n=117	23.0-50.2	0%	insufficient	moderate
Seated, 240°/s	1 study, n=12	49.3	0%	insufficient	very low
Supine, 60°/s	2 studies, n=61	19.1-25.5	0%	insufficient	moderate
Supine, 120°/s	1 study, n=49	20.2-33,5	0%	insufficient	very low
Supine, 240°/s	1 study, n=12	21.3-22.5	0%	insufficient	very low
ER eccentric					
Seated, 30-60°/s	5 studies, n=141	14.4-58.2	11%	insufficient	moderate
Seated, 120°/s	2 studies, n=71	34.6-49.6	0%	insufficient	low
Supine, 60°/s	1 study, n=49	23.8-32.7	0%	insufficient	very low
Supine, 120°/s	1 study, n=49	33.8-36.0	0%	insufficient	very low
ER/IR concentric					
Seated, 60°/s	2 studies, n=86	33.5-69.3	0%	insufficient	low
Seated, 120°/s	2 studies, n=86	33.8-74.8	0%	insufficient	low

n, number of participants

When MIC was set to 15%, all results were rated as "insufficient". Therefore, the sensitivity analysis was made only by increasing MIC to 20% and not by decreasing it to 10%. When increasing MIC, all results were still rated as "insufficient".

The quality of evidence was highest for the seated position, the velocities 30°/s-60°/s or 120°/s and the concentric test mode because these conditions were the most frequently evaluated conditions in the included studies.

Handheld dynamometry (Study IV)

The population consisted of 806 healthy subjects in 25 studies; 145 subjects with shoulder symptoms in five studies and 12 subjects with chronic obstructive pulmonary disease in one study. (117) Two studies included both healthy subjects and subjects with shoulder symptoms. (118, 119) Results were summarized and presented separately for intra- and inter-rater assessment and for each movement of the shoulder joint.

Table 16 shows the rating of results and risk of bias assessment for reliability, measurement error and construct validity.

Table 16 Rating of results and risk of bias assessment for reliability, measurement error and construct validity of handheld dynamometry (90)

Author/year	Reliability		Measurement error		Construct validity	
	Rating of results	Risk of bias assessment	Rating of results	Risk of bias assessment	Rating of results	Risk of bias assessment
Awatani[a], 2016 (120)	+	A	-	A		
Awatani[b], 2016 (121)	+	A	?	A		
Balogun, 1998 (122)	+	D	-	D		
Beshay, 2011 (118)	+	D				
Bohannon, 1997 (123)	+	V	?	I		
Burnham, 1995 (124)	?	D			-	A
Byl, 1988 (125)	?	I				
Cadogan, 2010 (126)	+	A	?	A		
Cools, 2014 (127)	+	A	?	A		
Dollings, 2012 (128)	+	A	-	A		
Donatelli, 2000 (129)	+	D				
Fieseler, 2015 (130)	+	A	+	A		
Fieseler, 2017 (119)	+	D	-	D		
Hayes, 2002 (131)	+	D			+	D
Holt, 2016 (132)	+	V	-	V	?	I
Johansson, 2005 (133)	+	A			+	I
Johansson, 2015 (134)	+	I	-	I		
Leggin, 1996 (106)	+	D				
Magnusson, 1990 (108)	+	D				
McLaine, 2016 (135)	+	A	?	A		
McMahon, 1992 (136)	+	A				
O´Shea, 2007 (117)	+	A				
Ottenbacher, 2002 (137)	+	D				
Phillips, 2000 (138)	+	A				
Riemann, 2010 (139)	?	A				
Sciascia, 2015 (140)	+	D	-	D	-	D
Sullivan, 1988 (116)	?	A			?	A
Vermeulen, 2005 (141)	+	A	?	A	?	D

+, sufficient; -, insufficient; ?, indeterminate; V, very good; A, adequate; D, doubtful; I, inadequate

Reliability

Reliability was evaluated in all 28 studies. The summarized results, the rating of results and the quality of evidence are shown in the summary of findings table (Table 17). Of all reported ICC values, 98% were ≥0.70. Consequently, for all movements and types of reliability examined, the overall rating of the results was sufficient. The quality of evidence varied depending on the movement examined; "high" for IR, ER, abduction and flexion; and "high" to "very low" for adduction and extension.

Table 17 Summary of findings for reliability of handheld dynamometry (90)

Movement and type of reliability	Number of studies, number of participants	Summary of results: ICC, range	Summary of results: Proportion of ICC ≥0.70, %	Overall rating of results	Quality of evidence
IR					
Intra-rater	12 studies, n=237	0.57-1.00	97	sufficient	high
Inter-rater	9 studies, n=243	0.72-0.99	100	sufficient	high
ER					
Intra-rater	16 studies, n=561	0.70-1.00	100	sufficient	high
Inter-rater	11 studies, n=316	0.64-0.98	91	sufficient	high
Abd					
Intra-rater	12 studies n=472	0.55-0.98	95	sufficient	high
Inter-rater	10 studies n=340	0.77-0.98	100	sufficient	high
Add					
Intra-rater	2 study n=41	0.87-0.98	100	sufficient	very low
Inter-rater	2 study n=117	0.86-0.98	100	sufficient	moderate
Flex					
Intra-rater	7 studies n=124	0.86-0.97	100	sufficient	high
Inter-rater	6 studies n=208	0.86-0.97	100	sufficient	high
Ex					
Intra-rater	5 studies, n=296	0.77-0.99	100	sufficient	high
Inter-rater	2 studies, n=113	0.87-0.97	100	sufficient	moderate

ER, external rotation; IR, internal rotation; Flex, flexion; Ex, extension; Abd, abduction; Add, adduction; ICC, intraclass correlation coefficient

Measurement error

Measurement error was evaluated in 14 studies. The summarized results, the rating of results and the quality of evidence are shown in the summary of findings table (Table 18). The %MDC values ranged from 0% to 51%, and all movements and types of reliability examined were rated as either "insufficient" (7 strata) or "indeterminate" (5 strata). Except for adduction, the quality of evidence was graded as "high" and "moderate".

Table 18 Summary of findings for measurement error of handheld dynamometry (90)

Movement and type of reliability	Number of studies, number of participants	Summary of results: %MDC, range	Summary of results: Proportion of %MDC≤15%	Overall rating of results	Quality of evidence
IR					
Intra-rater	7 studies, n=172	0-29.0	40	indeterminate	high
Inter-rater	5 studies, n=186	12.0-42.7	19	insufficient	high
ER					
Intra-rater	7 studies, n=185	0-48.5	40	indeterminate	high
Inter-rater	5 studies, n=199	15.0-33.3	6	insufficient	high
Abd					
Intra-rater	3 studies, n=68	15.0-35.2	11	insufficient	moderate
Inter-rater	3 studies, n=144	16.1-32.4	0	insufficient	high
Abd					
Intra-rater	1 study, n=25	51.0	0	insufficient	very low
Inter-rater	1 study, n=101	19.1-24.7	0	insufficient	low
Flex					
Intra-rater	4 studies, n=83	11.6-35.8	27	indeterminate	moderate
Inter-rater	4 studies, n=180	16.1-39.9	0	insufficient	high
Ex					
Intra-rater	4 studies, n=65	2.8-30.5	39	indeterminate	moderate
Inter-rater	2 studies, n=113	8.6-25.2	50	indeterminate	moderate

ER, external rotation; IR, internal rotation; Flex, flexion; Ex, extension; Abd, abduction; Add, adduction

The sensitivity analysis revealed that if MIC was set to 10%, the results of all strata were rated as "insufficient". If MIC was set to 20%, two strata were rated as "sufficient" and the rest as "insufficient" or "indeterminate".

Construct validity

The quality of evidence for the comparisons of HHD with other types of dynamometers were all low or very low. The results were rated as both "sufficient", "insufficient" and "indeterminate" depending on comparison examined. In one study examining discriminative validity, no difference between individuals with and without shoulder symptoms was found (Table 19).

Table 19 Summary of findings for construct validity of handheld dynamometry (90)

	Number of studies, number of participants	Summary of results	Overall rating of results	Quality of evidence
HHD/ID	3 studies, n=54	*r*, range 0.28-0.85	insufficient	very low
HHD/EFD	1 study, n=20	Mean diff., range (N) -6.5 to 29.9	indeterminate	low
HHD/Spring scale	2 studies, n=18	*r*, range 0.77-0.99	sufficient	very low
With/Without symptoms	1 studies, n=36	P-value, range 0.89-0.99	insufficient	very low

HHD, handheld dynamometer; ID, isokinetic dynamometer; EFD, externally fixed dynamometer; r, Pearson correlation coefficient; diff., difference; N, Newton

Discussion

This chapter will discuss the main findings of the four individual studies; study I and II separately and study III and IV together. First, the main findings will be presented and compared with findings from the literature. Second, some methodological considerations will be addressed, and the strengths and limitations will be discussed. Finally, the external validity of the findings will be discussed.

Summary of the main findings and comparison with the literature

In the register-based cost-of-illness study (**Study I**), the total costs associated with shoulder disorders was evaluated. This study showed that shoulder disorders are not only a burden to patients but also a large economic burden to society. During the 13-year inclusion period, 617,334 unique individuals with a primary diagnosis of shoulder disorders were identified. We found that the incidence of shoulder disorders rose from 966 per 100,000 person years in 2005 to 1,215 per 100,000 person years in 2017, an increase of 26%. The mean total costs for the 6-year period were €25,771 higher for individuals aged <65 years and €11,334 higher for individuals aged ≥65 years with shoulder disorders than for matched controls without shoulder disorders. The health care costs were roughly twice as high for individuals aged ≥65 years as for individuals aged <65 years. Both health care costs and costs of sick leave were highest the first year after the diagnosis. Costs of sick leave accounted for about 70% of the total costs for individuals <65 years. Furthermore, 20% of the cases accounted for 66% of the total costs. Expected annual costs associated with shoulder disorders were estimated at €1.21 billion.

Another Danish register-based study, including 244,519 patients with rotator cuff-related lesions, found that the incidence increased by more than 400% from 1996 to 2013. (13) This is much higher than the 26% increase reported in our study. Furthermore, in our study it appears that the incidence rate has stabilised since 2014. Several possible explanations may account for the increase in the incidence of shoulder disorders; earlier underreporting of diagnoses, a growing pain awareness in society, and easier accessibility to magnetic resonance imaging and ultrasound may have led to an increasing number of patients receiving a diagnosis of shoulder disorders. (13)

With regard to estimating the costs associated with shoulder disorders, a Swedish cost-of-illness study based on 204 patients with shoulder pain found mean primary health care costs of €326 for the first six months after the diagnosis. (23) This was higher than the €262 found in our study covering the first 12 months after the diagnosis. However, the study population in the Swedish study was much smaller and consisted of patients with shoulder pain seen by either a general practitioner or physiotherapist in two municipalities, which could explain the differences. Costs of sick leave were also included in the Swedish study, but comparing these costs with our costs can be

misleading as different registration methods were used. (23) Several studies on both shoulder pain and back pain support our findings, viz. that costs of sick leave were markedly higher than health care costs and that a small proportion of patients accounted for the majority of the total costs. (23, 142, 143) This indicates that interventions focusing on return to work could be key elements in minimizing societal costs of shoulder disorders.

In Studies II, III and IV, we evaluated measurement properties of instruments that are commonly used in the evaluation of shoulder disorders; questionnaires in Study II and dynamometers in Study III and IV. This is important, because satisfactory measurement properties are essential for instruments that are being used in clinical practice and research.

The responsiveness and the MIC of the OSS, EQ-5D$_{index}$, EQ-5D$_{vas}$ and FABQ-PA were evaluated in 52 patients undergoing arthroscopic subacromial decompression in the clinical cohort study (**Study II**). Patients categorised as improved according to the Global Rating of Change Scale had a statistically significantly better change score in all outcomes than patients categorised as unimproved. The ROC AUC values ranged from 0.73 to 0.96, highest for the OSS. The change scores of the OSS, EQ-5D$_{index}$ and partly the FABQ-PA showed correlations as hypothesized when compared with the change scores of the SSV and pain VAS. However, the change scores of EQ-5D$_{vas}$ and partly the FABQ-PA showed correlations slightly lower than hypothesized. The primary anchor was considered most important, and the secondary anchors were used to support expected associations with other instruments measuring overlapping but not similar constructs. Based on all ROC AUC values being ≥0.70, the responsiveness was concluded to be adequate for all three PROMs. The determined MIC values were 6.0 points for the OSS, 0.024 for the ED-5D$_{index}$, 10.0 for the EQ-5D$_{vas}$ and -5.0 for the FABQ-PA.

In the present study, the Global Rating of Change Scale showed that 77% of the included patients had improved six months after decompression surgery. This result is in line with results reported in other studies. Ketola et al. found that 65% of surgery-treated patients were pain free after 24 months (144); and Davis et al. performed a systematic review and found that ≥68% of the patients were satisfied after open or arthroscopic decompression. (145) The adequate responsiveness of the OSS, EQ-5D and FABQ-PA found in our study is largely supported by others. Similar ROC AUC estimates for the OSS were seen in patients with rotator cuff disease treated with glucocorticoid injections and patients having difficulty returning to normal activity levels after decompression surgery who received occupational medical assistance or physiotherapy. (45, 49) In addition, ROC AUC estimates for the EQ-5D were similar to ours in patients with shoulder disorders treated with surgery or physiotherapy. (52, 55) Opposite our findings, the responsiveness of the FABQ-PA was concluded to be inadequate in physiotherapy-treated patients with subacromial impingement syndrome. However, a

different methodological approach used in this study could explain the different result. (63) Our finding for the MIC ROC cut-off point for the OSS was identical to what was found in a previous study assessing the MIC in a population of patients receiving occupational medical assistance or physical therapy due to difficulty returning to usual activities after decompression surgery. (45) No studies were found that estimated the MIC of the FABQ-PA or the EQ-5D 5-level version in patients with shoulder disorders. The validity and reliability of the OSS, EQ-5D and FABQ-PA were investigated and found adequate by others. (43, 46, 52, 53, 56, 60-63) Now the responsiveness and MIC have been established and the PROMs are now considered useful for measuring changes in the outcomes; pain and function, HR-QoL and fear-avoidance belief in patients with subacromial impingement syndrome.

The measurement properties of ID and HHD were investigated based on systematic reviews of the literature (**Study III and IV**). Overall, the reliability of ID and HHD was sufficient and these instruments are useful to distinguish between individuals at group level. (36) The quality of evidence was generally moderate or low for ID and high for HHD, especially for internal and external rotation. The measurement error was rated as insufficient or indeterminate for both ID and HHD, and the quality of evidence was generally moderate to very low for ID and high or moderate for HHD. Less than 75% of the included studies found MDC values ≤15%, which was the threshold to be rated as sufficient. The sensitivity analysis revealed that even MDC values ≤20% were sparse. Therefore, to be sure that change in muscle strength exceeds the measurement error of the dynamometer, it must be larger than 15% and probably even above 20%. Whether a change of this size is clinically meaningful depends of the expected magnitude of change in the context, intervention and population under study, e.g., individuals with severe limitations may expect large improvements after an effective intervention. The construct validity of HHD showed inconsistent results based on low quality of evidence when compared with other dynamometer types or when different populations were compared.

Although exercise therapy is widely used as treatment strategy in patients with shoulder disorders, few systematic reviews have focused on the measurement properties of dynamometers for assessment of shoulder muscle strength. To evaluate the reliability of ID, a review of patients with post-stroke hemiparesis found ICC values ranging from 0.87 to 0.92 and standard error of measurement from 15% to 24% (equal to %MDC from 46% to 67%), which was comparable to our findings. (66) Another review focused on the effect of position on the reliability of HHD and found the seated position to be more reliable than the standing and supine position. (67) For HHD evaluations, one review evaluating the reliability found ICC values similar to those found in our study. (70)

Methodological considerations

Study I

Perspective

We used a societal perspective when estimating the costs of shoulder disorders including health care costs from the primary and secondary sector and the costs of sick leave. Which costs to include when using a societal perspective depends on the disorder or disease investigated and the availability of data. For some more life-threatening diseases, for example cancer, neurological diseases or heart failure, costs of medication, costs of home care, out-of-pocket costs and costs of the time spent on treatment for both the patient and relatives may be substantial and therefore relevant to include in a cost-of-illness study. (26, 28)

In the present study, we included health care costs and costs of sick leave. Costs of subsidised employment and disability pension could have been included as data on these benefits are available in the DREAM register. However, these benefits are granted only if all treatment and rehabilitation efforts have been exhausted and there is a permanent loss of function. This process can take up to five years. (146) Because the follow-up period in the present study was up to five years after the diagnosis, we believed that relatively few patients would be granted subsidised employment or disability pension because of their shoulder disorder within this timeframe. Therefore, we decided to focus only on costs of sick leave. Furthermore, costs of lost life years and home care were not included as shoulder disorders were considered to be a disease with low mortality and low risk of need for home care. Data on prescription medication are available from a separate medication register. We chose not to apply for data from this register because we expected the costs of medication for shoulder disorders to be very low. A cost-of-illness study on back pain found that costs of prescription medication accounted for only 0.6% of the total costs (143), and we expected the medication costs for shoulder disorders to be at a similarly low level. Furthermore, over-the-counter drugs are commonly used for pain management in patients with shoulder disorders, and data on these costs are not available in any register. In addition, data on costs of time spent on treatment, for patient and relatives, and out-of-pocket costs were not available in the registers and could not be included. Overall, some costs were not included as they were considered not to be essential for patients with shoulder disorders. Other costs, some of which were relevant (e.g., over-the-counter drugs, time spent on treatment and out-of-pocket costs) could not be included because data were not available, which may have led to an underestimation of the total costs associated with shoulder disorders. However, we consider this underestimation to be of minor magnitude for shoulder disorders. A major advantage of using register-based data is the high number of included individuals and the high data quality and coverage. A disadvantage is the lack of information on a more detailed personal level, like time and out-of-pocket costs.

Matching

The controls were matched on age and gender. Other matching variables could potentially also have been relevant. However, given the high number of included cases and the relatively small Danish population (approximately 4,300,000 persons ≥18 years on average during the study period), it was challenging to match on more than two variables, and age and gender were considered the most important ones. Nevertheless, there were minor differences between cases and controls regarding comorbidity status, educational level and work status at the time of inclusion.

Compared with the controls, more cases had a Charlson Comorbidity Index score above zero; 12.5% of the cases compared with 8.8% of the controls. The Charlson Comorbidity Index score was calculated based on the primary diagnosis codes from the DNPR after receiving the data from Statistic Denmark, (147); therefore, matching on comorbidity was not possible. Some of the differences in both health care costs and costs of sick leave between cases and controls may be explained by the minor difference in comorbidity. Therefore, a subgroup analysis was made to compare the costs of cases and controls in each Charlson Comorbidity Index score category: 0, 1, 2-3 or ≥4. This subgroup analysis provides an estimate of the costs associated with shoulder disorders when taking the costs of comorbidity into account (Table 9). Findings from this subgroup analysis showed that the mean additional costs of patients with shoulder disorders compared with controls were similar within each comorbidity category except for the group with an index score of 4 or higher which had increased mean additional costs. However, the Charlson Comorbidity Index includes chronic, potentially life-threatening comorbidities such as cancer, AIDS, cardiac diseases, diabetes, liver and kidney diseases and cerebral diseases. (147) The index does not include other musculoskeletal disorders (e.g., low back and neck pain), which may be more common among patients with shoulder disorder than among the controls, and may contribute to the additional costs.

Compared with the controls, more cases had a low or medium educational level. Some of the jobs associated with these educational levels could be more physically demanding, and work-related factors could expose to shoulder complaints. (75) Besides, for patients with shoulder complaints, it can be more difficult to return to a physically demanding job than to an office job. Likewise, the risk of postoperative permanent work disability has been found to be associated with low educational level. (8) This may have led to higher costs in the cases. However, if controls were matched on educational level, they would have been more similar to the cases but less similar to the general population without shoulder disorders. Hence, some of the costs associated with factors that characterises patients with shoulder disorders could have been eliminated, leading to underestimating of the costs associated with shoulder disorders.

Cases were more often on sick leave and less often in jobs at the time of inclusion. Cases may have had symptoms for some time before they are referred to a physician in the

secondary healthcare sector where the diagnosis code was given. These symptoms could explain the greater proportion of cases on sick leave at baseline. Like with educational attainment, having a control group matching the cases too well on factors associated with shoulder disorders may lead to an underestimation of the additional costs.

Study II

The responsiveness of the OSS, EQ-5D and FABQ-PA was assessed using anchor-based methods in accordance with recommendations. (36) The use of a Global Rating of Change Scale as an external anchor has been questioned as doubt has been expressed about the reliability and validity of such retrospective measures of change. (81) However, Global Rating of Change Scales are commonly used in the assessment of responsiveness and MIC, and no better alternative seems to exist that reliably measures change in the patients' health condition. (36, 42) Furthermore, as recommended, several anchors were used to cover different aspects of the constructs of interest. (42) The MIC was assessed using a method integrating both anchor- and distribution-based approaches, which is in accordance with recommendations. (84) We expected higher correlations between the OSS compared with SSV and pain VAS than between the EQ-5D and FABQ-PA compared with SSV and pain VAS because they were considered to be more overlapping constructs. Our findings supported these expectations.

Study III and IV

Based on the GRADE approach, the quality of evidence was downgraded if necessary. (77, 93) The most common causes for downgrading the quality of evidence were risk of bias (few studies or the methodological quality doubtful or inadequate) and imprecision (one level if sample size <100 and two levels if sample size <50). We rarely downgraded for inconsistency of the results and not at all for indirectness (studies performed in another population or context). The grading "very low" quality of evidence only occurred if the stratum examined was evaluated in few studies of doubtful quality with a total sample size <50.

In 2021, the COSMIN initiative published a new risk of bias tool to assess the quality of studies on reliability and measurement error of all types of outcome measurement instruments. This extended version was developed for clinician-reported outcome measures, performance-based outcome measurement instruments and laboratory values. (37) In the new version, two questions were added to the risk of bias tool for both reliability and measurement error: 1) Did the professional(s) administer the measurement without knowledge of scores or values of other repeated measurement(s) in the same patients? and 2) Did the professional(s) assign scores or determine values without knowledge of the scores or values of other repeated measurement(s) in the same patients? This new risk of bias checklist was published after we had performed the quality assessment in the present studies, and assessment of these two questions was not included. However, using the new version would have had no impact on the results

because other strength measurements than dynamometers were not performed in any of the included studies.

For assessment of clinically relevant changes with a given instrument, the measurement error needs to be smaller than the MIC. (36, 77) However, no widely accepted definition of the MIC for muscle strength was identified. Some studies indicated a change in muscle strength of 10% to 15% as being clinically relevant (68, 132), and other studies found changes ranging from 7% to 23% in patients with shoulder disorder after a strength exercise intervention. (21, 148) Based on these findings, a criterion with a MIC of 15% was deemed adequate criterion against which to rate the measurement error results. To examine the robustness of the results, a sensitivity analysis was made by setting the MIC to 10% and 20%, respectively; and rate the results against these criteria. However, the sensitivity analysis did not change the results substantially.

Strengths

Study I

One important strength of this register-based study is that the incidence and costs associated with shoulder disorders are based on the entire Danish population and not just a subsample that needs to be extrapolated. In addition, Denmark is considered an optimal country for making cost-of-illness studies because the healthcare system is tax-funded, because government-maintained nationwide registers cover routinely collected administrative and health care data and because a unique personal identification number permits linkages between registers on an individual level. (72) The coverage of data registration in the registers is assumed to be high since registrations are linked to reimbursements. Additionally, the quality of data, e.g., the ICD-10 codes, is generally very high. (72, 149, 150)

Another strength is that a comparison cohort of age-and gender-matched controls was included. This allowed for calculation of the additional costs of patients with shoulder disorders compared with individuals without shoulder disorders. If controls were not included, the total costs would not necessarily reflect only the costs of shoulder disorders but also costs of other health problems, because all health care and sick leave utilisations are registered. By comparing cases to matched controls, we could identify costs attributed to shoulder disorders.

Furthermore, the included individuals were followed for a period of up to five years after inclusion (first day of receiving a diagnosis in the secondary healthcare system), allowing for investigating the development in costs over time. Because of this relatively long follow-up period, we were able to show that cases continued to have higher costs than controls for both health care (59%) and sick leave (81%). The included patients represent the whole continuum of patients ranging from those with a short duration and

only one single contact to the healthcare system (low-cost patients) to those with a long duration and many visits to different providers and a long period of sick leave (high-cost patients). The study period with costs included up to five years after the first diagnosis allowed us to estimate costs related to both short-term symptoms, long-term symptoms, and repeated shoulder conditions.

Study II

A strength of this clinical cohort study is that data were collected prospectively. Baseline data were collected prior to the surgical treatment without knowledge of the outcome. Furthermore, during the study period, few patients dropped out or were lost to follow-up and few items were left unanswered, leading to low risk of bias because of missing data. Additionally, no floor or ceiling effects were seen at either baseline or follow-up in any of the outcome measures. If floor or ceiling effects were present at baseline, patients would be limited in assessing change at follow-up as they could not use the full scale. If floor or ceiling effects were present at follow-up, this could have indicated limited ability to assess change correctly as the limit of the scale was already reached. Both scenarios could have affected the evaluation of responsiveness and MIC.

Study III and IV

The two reviews were based on a protocol registered in PROSPERO and published prior to the study start. (86) The benefits of registering a protocol are transparency of literature search, data collection and evidence synthesis. Furthermore, the studies were performed according to COSMIN methodology. This guideline ensured a systematic and transparent approach throughout the different steps of the review process of assessing the measurement properties of ID and HHD. Two review authors performed the study selection, data extraction and quality assessment to minimize the risk of missing relevant information. Furthermore, we used a validated and sensitive published search filter developed by Terwee et al. (151) to identify studies on measurement properties to ensure that relevant studies were included. Thus, the methodological quality of the two reviews is considered to be high. Furthermore, the high number of included studies provided valid estimates and generally high or moderate quality of evidence, especially for some strata. The high quality of the evidence presented suggests that we are confident in the estimates and conclusions; and no further research is recommended.

Limitations

Before drawing conclusions of this thesis, some important limitations must be kept in mind as they may distort the results.

Study I

The classification of cases and controls is based on the ICD-10 codes given by physicians at the hospital. Individuals with shoulder disorders seen only by a general practitioner

are not included as cases, and the estimates presented in this study do not reflect all individuals suffering from shoulder pain. In contrast, individuals visiting a general practitioner with shoulder complaints could potentially be included as controls. However, we expect these biases to be limited as patients with long-lasting symptoms and needing several visits are typically referred to a physician in the secondary healthcare sector. Besides, even though the quality of data is considered to be high, some incorrect coding may have occurred. However, we assume that the risk of misclassification between cases and controls and between diagnosis categories is low and that information bias is of minor concern.

In Denmark, for individuals who are working, the first period of sick leave is paid by the employer. When this period exceeds a given number of days, the employer can apply for reimbursement in the form of sickness benefits, in which case the sick leave period is registered in the DREAM database. (2) Short-term sick leave is not a part of this registration and health conditions typically resulting in short-term sick leave will therefore be underestimated. During the 13-year inclusion period, the number of days required for the employer to apply for sickness benefits was extended several times; it started with >13 days up to 2007, but increased from >14 days in 2007-2008, >21 days in 2009-2011 to >30 days since 2012. (2, 75) Consequently, a decrease over time was seen in the proportion of individuals on sick leave, the mean number of weeks on sick leave and the costs of sick leave in the first year after inclusion. This decrease was seen both among cases and controls, but the largest decrease was seen among cases. Productivity loss due to sick leave is likely to be underestimated, and this underestimation became more pronounced in more recent years.

A comparison cohort of controls was included to make an estimate of the additional costs of patients with shoulder disorders compared with similar individuals without a history of shoulder disorders. Preferably, the controls should be free from the included diagnosis codes. However, the data extraction and matching process was performed by Statistics Denmark, a third party not directly involved in the study. They checked the controls for the relevant shoulder diagnosis codes above age 18, just like the inclusion criteria for the cases. During the analysis, we found that some controls (2,428 persons equal to 0.4%) had a shoulder diagnosis in their childhood, but on average 9 years before they were included as controls. It was not possible to replace these controls. We considered the period since the diagnosis to be long enough to assume that the former childhood shoulder diagnosis did not incur costs anymore.

Age 65 was the official retirement age in Denmark throughout the whole study period. From this age, people receive public pension payment from the government. Retirement before age 65 requires self-financing; in Denmark, only a small minority can afford this. However, some individuals in both groups will properly choose to retire earlier or later than at the age of 65. One scenario is that sick leave is underestimated as some patients

could choose to retire earlier than planned because of shoulder problems. Another scenario is that people close to the retirement age reduced their working hours. In this case, using full employment until age 65 will overestimate the costs of sick leave. The proportion of Danes who continue to work beyond age 65 is approximately 10%, but about half of those work less than 20 hours per week. (152) However, using sick leave after age 65 when people can retire was considered challenging as several factors can affect people's choice for leaving the labour market when their working ability is limited. In our data, only 0.02% of cases and 0.01% of controls aged ≥65 had a sick leave registration the year after inclusion. This supports that sick leave in older workers is a minor factor.

Study II

The clinical cohort study has three main limitations. About 25% of the included patients were on retirement or out of job, and completing the FABQ work subscale was not relevant for these patients. The remaining sample size (39 patients) was considered too small to assess the responsiveness and MIC of the FABQ work subscale with adequate quality as a sample of at least 50 individuals is recommended. Therefore, we were only able to assess these measurement properties of the FABQ physical activity subscale.

Furthermore, the plan was to assess the MIC using both the ROC cut-off point and 95% limit cut-off point, but we were able to assess only the first. The 95% limit cut-off point uses the distribution of unchanged patients according to the anchor. (84) Only five patients responded unchanged, which was considered too few to assess the MIC using this method.

Finally, we had to change recruitment strategy during the study period and included patients from both private and public hospitals. However, we believe this strategy has had only minor impact on the results.

Study III and IV

Although we have used an exhaustive literature search and the study selection was performed by two authors independently, some relevant studies may have been missed, causing potential selection bias. Especially if studies of high quality have been missed, this could have affected the evidence of the measurement properties provided in these two studies. Other limitations are that some of the strata were examined only in very few studies and some of the included studies had high risk of bias. These limitations reduce the quality of evidence but do not affect the estimates. Only few studies of low risk of bias examined the isometric and eccentric test mode, the supine and standing position and the velocity 90°/s for ID, and the movements abduction and adduction for HHD, leading to lower the quality of evidence for these strata.

External validity

External validity refers to the extent to which the results of the studies included in the present thesis are applicable to other settings and populations. The cost-of-illness study (**Study I**) includes the majority of shoulder diagnoses seen in clinical practice, and the costs are presented for both the total group of all shoulder disorders and for four specific diagnosis categories; i.e. subacromial pain, stiffness, fracture and dislocation. The register-based approach is independent of active participation; all patients visiting the secondary healthcare sector for shoulder disorders are included in the registries. Patients cannot refuse being included in the Danish national databases. Consequently, the results are representative for patients with shoulder disorders using services offered by the Danish healthcare system. Whether the costs are similar in other countries will depend on differences in healthcare and social security systems.

The clinical cohort study (**Study II**) specifically included patients with subacromial impingement syndrome undergoing decompressing surgery. Whether the responsiveness and MIC results are similar in other patient groups is uncertain. However, to some extent the results found in this thesis may be transferable to similar shoulder populations receiving treatments with similar effects.

The two systematic reviews (**Study III and IV**) assessing the measurement properties of ID and HHD are mainly based on studies including healthy subject. The few studies including patients with shoulder disorders did not indicate that the reliability or measurement error should be different in patients with shoulder disorders, but this is based on very limited evidence.

Conclusion

The overall findings of this thesis have filled in some knowledge gaps by mapping the resource usage, health care costs and costs of sick leave associated with shoulder disorders and by providing evidence of the measurement properties of commonly used shoulder outcome measurement instruments.

Using national registers, we found that the mean additional total costs (health care costs and costs of sick leave) of patients with shoulder disorders compared with matched controls for the 6-year period were €25,771 for individuals aged <65 years and €11,334 for individuals aged ≥65 years. Furthermore, 1.2% of the Danish population are seen in the secondary health care sector each year with a first-time diagnosis of shoulder disorder. Individuals aged 65 years or older had health care costs 83% above those of individuals younger than 65 years. Costs of sick leave accounted for 70% of the total costs for people in the working age. In addition, the 20% of patients with shoulder disorders with the highest costs accounted for 66% of total societal costs. Estimated additional annual costs of patients with shoulder disorders compared with individuals without shoulder disorders were €1.21 billion.

In patients with subacromial impingement syndrome, the responsiveness of the OSS, EQ-5D and FABQ-PA was adequate, and these PROMs are considered suitable for assessing changes over time after arthroscopic decompression surgery. The MIC ROC cut-off points of the three PROMs were established, which may assist clinicians and researchers when interpreting the results of these outcome measures.

The reliability of ID and HHD was sufficient according on systematic reviews. The measurement error was not sufficient, and the ability of ID and HHD to measure changes below 15% and even 20% is questionable. It depends on the context, intervention and population whether a change of this size is expected and the instruments thereby suitable. In general, the quality of evidence was moderate to very low for ID and high to moderate for HHD.

Perspectives and future research

National population-based registers, like the Danish health registers and the DREAM database, provide an exclusive data source with which to map the economic burden associated with shoulder disorders. (72) Findings from this thesis may be used to demonstrate to the political level of the Danish healthcare and social security system the high number of individuals affected by shoulder disorders each year and the large economic burden in terms of costs of health care and sick leave accompanying these disorders. Furthermore, the findings provide a detailed overview of the distribution of costs in different resource categories. Such findings are important to decisions makers deciding on the allocation of resources. Furthermore, this mapping of resource usage may be used to identify specific target areas for future research. In particular, we are planning to conduct a research project in 2022 aiming to investigate the subgroup of patients with the highest costs. A detailed overview of the use of health care services and social security benefits among this subgroup has never been established. Knowledge of the characteristics of these patients, their pathway through the healthcare system and their use of social security benefits could serve as a tool to aid early identification. Furthermore, development of tailored interventions with a focus on return to work, e.g., earlier and stronger collaboration between the patient, the employer, the healthcare system and the occupational medical assistant, may be a focus for future research in order to improve treatment and minimize costs.

The PROMs evaluated in this thesis (the OSS, EQ-5D and FABQ-PA) demonstrated adequate responsiveness in patients with subacromial impingement syndrome after decompression surgery. We also established the MIC values in the same population. When measuring changes in clinical practice and research, it is essential to consider whether the measured changes are important to the patients whether they are statistically significant or not. Therefore, the adequate responsiveness and the MIC values are valuable for interpretation of individual change after surgery, and it is recommended to use the OSS, EQ-5D and FABQ-PA. Which outcomes to measure depends on the context. In routine clinical practice, it may be relevant to measure pain and function with the OSS as these outcomes are often essential for patients with subacromial impingement syndrome, whereas EQ-5D and FABQ-PA may be relevant to use when there is a particular interest in HR-QoL and fear-avoidance belief. However, further research evaluating the responsiveness and MIC of the OSS, EQ-5D and FABQ-PA are recommended if the population or treatment is considered to have a considerably different change response in the constructs assessed with these PROMs. (36, 91)

The findings of the measurement properties of ID and HHD revealed sufficient reliability and insufficient or indeterminate measurement error of the ID and HHD. The sufficient reliability of the ID and HHD indicates that the dynamometers can be used to distinguish between individuals at group level for comparing the muscle strength in two or more

groups of subjects. (36) However, the measurement error was not sufficient when evaluated against a described criterion for MIC of either 15% or 20%. Therefore, evaluation at the individual level should be interpreted with caution. (36) Dynamometers may be useful in clinical practice only if the change in muscle strength is expected to exceed the measurement error of the instruments. Whether a 20% change is realistic depends of the clinical context, e.g. type of shoulder disorder and intervention provided. The results from the two systematic reviews are based mainly on healthy subjects as only few studies evaluated the reliability and measurement error in subjects with shoulder disorders. As we have discussed above, we have limited indications of whether the findings are representative of other populations. To confirm the findings in patients with shoulder disorders, high-quality research is needed. Furthermore, currently no consensus exists on the MIC of muscle strength. Studies assessing which change in muscle strength is perceived as important from a patient perspective could add valuable information to this field, e.g., studies using a Global Rating of Change Scale together with muscle strength assessment before and after an intervention that results in a change in muscle strength.

References

1. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord. 2010;11:144.
2. Flachs EM EL, Koch MB, Ryd JT, Dibba E, Skov-Ettrup L, Juel K. Sygdomsbyrden i Danmark - sygdomme. Kbh: Statens Institut for Folkesundhed, Syddansk Universitet; 2015. Report No.: 9788771046045.
3. Coghlan JA, Buchbinder R, Green S, Johnston RV, Bell SN. Surgery for rotator cuff disease. Cochrane Database Syst Rev. 2008;2008(1):Cd005619.
4. Picavet HSJ, Schouten JSAG. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC3-study. Pain. 2003;102(1):167-78.
5. Mäkelä M, Heliövaara M, Sainio P, Knekt P, Impivaara O, Aromaa A. Shoulder joint impairment among Finns aged 30 years or over: prevalence, risk factors and co-morbidity. Rheumatology (Oxford, England). 1999;38(7):656-62.
6. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. Annals of the Rheumatic Diseases. 1998;57(11):649-55.
7. Roquelaure Y, Ha C, Leclerc A, Touranchet A, Sauteron M, Melchior M, et al. Epidemiologic surveillance of upper-extremity musculoskeletal disorders in the working population. Arthritis and Rheumatism. 2006;55(5):765-78.
8. Svendsen SW, Frost P, Jensen LD. Time trends in surgery for non-traumatic shoulder disorders and postoperative risk of permanent work disability: A nationwide cohort study. Scandinavian Journal of Rheumatology. 2012;41(1):59-65.
9. Minns Lowe CJ, Moser J, Barker K. Living with a symptomatic rotator cuff tear 'bad days, bad nights': a qualitative study. BMC Musculoskelet Disord. 2014;15:228.
10. Holmgren T, Hallgren HB, Öberg B, Adolfsson L, Johansson K. Effect of specific exercise strategy on need for surgery in patients with subacromial impingement syndrome: randomised controlled study. BMJ: British Medical Journal. 2012;344(7846):15-.
11. Judge A, Murphy RJ, Maxwell R, Arden NK, Carr AJ. Temporal trends and geographical variation in the use of subacromial decompression and rotator cuff repair of the shoulder in England. The bone & joint journal. 2014;96-B(1):70-4.
12. Vitale MA, Arons RR, Hurwitz S, Ahmad CS, Levine WN. The Rising Incidence of Acromioplasty. The Journal of Bone & Joint Surgery. 2010;92(9):1842-50.
13. Monrad N, Ganestam A, Kallemose T, Barfod KW. Alarming increase in the registration of degenerative rotator cuff-related lesions a nationwide epidemiological study investigating 244,519 patients. Knee Surgery, Sports Traumatology, Arthroscopy. 2018;26(1):188-94.
14. Polinder S, Iordens GI, Panneman MJ, Eygendaal D, Patka P, Den Hartog D, et al. Trends in incidence and costs of injuries to the shoulder, arm and wrist in The Netherlands between 1986 and 2008. BMC Public Health. 2013;13:531.
15. Vandvik PO, Lähdeoja T, Ardern C, Buchbinder R, Moro J, Brox JI, et al. Subacromial decompression surgery for adults with shoulder pain: a clinical practice guideline. Bmj. 2019;364:l294.
16. Paavola M, Malmivaara A, Taimela S, Kanto K, Inkinen J, Kalske J, et al. Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial. BMJ (Clinical research ed). 2018;362:k2860.
17. Svendsen SW. National klinisk retningslinje for diagnostik og behandling af patienter med udvalgte skulderlidelser. Sundhedsstyrelsen; 2013. Contract No.: Report.
18. Diercks R, Bron C, Dorrestijn O, Meskers C, Naber R, de Ruiter T, et al. Guideline for diagnosis and treatment of subacromial pain syndrome: a multidisciplinary review by the Dutch Orthopaedic Association. Acta Orthop. 2014;85(3):314-22.

19. Beard DJ, Rees JL, Cook JA, Rombach I, Cooper C, Merritt N, et al. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. Lancet. 2018;391(10118):329-38.
20. Karjalainen TV, Jain NB, Page CM, Lähdeoja TA, Johnston RV, Salamh P, et al. Subacromial decompression surgery for rotator cuff disease. Cochrane Database Syst Rev. 2019;1(1):Cd005619.
21. Lombardi JI, Magri AG, Fleury AM, Da Silva AC, Natour J. Progressive resistance training in patients with shoulder impingement syndrome: a randomized controlled trial. Arthritis and Rheumatism. 2008;59(5):615-22.
22. Marks D, Comans T, Bisset L, Thomas M, Scuffham PA. Shoulder pain cost-of-illness in patients referred for public orthopaedic care in Australia. Australian health review : a publication of the Australian Hospital Association. 2019;43(5):540-8.
23. Virta L, Joranger P, Brox JI, Eriksson R. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. BMC musculoskeletal disorders. 2012;13(1):17-.
24. Health matters: health and work GOV.UK: Public Health England; 2019 [Guidance]. Available from: https://www.gov.uk/government/publications/health-matters-health-and-work/health-matters-health-and-work#resources.
25. Davis K, Dunning K, Jewell G, Lockey J. Cost and disability trends of work-related musculoskeletal disorders in Ohio. Occup Med (Lond). 2014;64(8):608-15.
26. Larg A, Moss JR. Cost-of-illness studies: a guide to critical evaluation. Pharmacoeconomics. 2011;29(8):653-71.
27. Finkelstein E, Corso P. Cost-of-illness analyses for policy making: a cautionary tale of use and misuse. Expert Rev Pharmacoecon Outcomes Res. 2003;3(4):367-9.
28. Jo C. Cost-of-illness studies: concepts, scopes, and methods. Clin Mol Hepatol. 2014;20(4):327-37.
29. Pike J, Grosse SD. Friction Cost Estimates of Productivity Costs in Cost-of-Illness Studies in Comparison with Human Capital Estimates: A Review. Applied health economics and health policy. 2018;16(6):765-78.
30. Ramiro S, Page MJ, Whittle SL, Huang H, Verhagen AP, Beaton DE, et al. The OMERACT Core Domain Set for Clinical Trials of Shoulder Disorders. The Journal of rheumatology. 2019:jrheum.181070.
31. Page MJ, Huang H, Verhagen AP, Gagnier JJ, Buchbinder R. Outcome Reporting in Randomized Trials for Shoulder Disorders: Literature Review to Inform the Development of a Core Outcome Set. Arthritis Care Res (Hoboken). 2018;70(2):252-9.
32. Page MJ, Huang H, Verhagen AP, Buchbinder R, Gagnier JJ. Identifying a core set of outcome domains to measure in clinical trials for shoulder disorders: a modified Delphi study. RMD Open. 2016;2(2):e000380.
33. Page MJ, O'Connor DA, Malek M, Haas R, Beaton D, Huang H, et al. Patients' experience of shoulder disorders: a systematic review of qualitative studies for the OMERACT Shoulder Core Domain Set. Rheumatology (Oxford, England). 2019.
34. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of Patient-Reported Outcomes in Routine Medical Care. Am Soc Clin Oncol Educ Book. 2018;38:122-34.
35. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther. 2016;20(2):105-13.
36. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide: Cambridge University Press; 2011.
37. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or

measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol. 2020;20(1):293.
38. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1-15.
39. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. Journal of clinical epidemiology. 2006;59(10):1033-9.
40. Weir JP. QUANTIFYING TEST-RETEST RELIABILITY USING THE INTRACLASS CORRELATION COEFFICIENT AND THE SEM. Journal of Strength and Conditioning Research. 2005;19(1):231-40.
41. Kottner J, Gajewski BJ, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). International journal of nursing studies. 2011;48(6):659-60.
42. 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;61(2):102-9.
43. Angst F, Schwyzer HK, 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 assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). Arthritis Care Res (Hoboken). 2011;63 Suppl 11:S174-88.
44. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. Arch Orthop Trauma Surg. 2009;129(1):119-23.
45. Christiansen DH, Frost P, Falla D, Haahr JP, Frich LH, Svendsen SW. Responsiveness and Minimal Clinically Important Change: A Comparison Between 2 Shoulder Outcome Measures. J Orthop Sports Phys Ther. 2015;45(8):620-5.
46. St-Pierre C, Desmeules F, Dionne CE, Frémont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: A systematic review. Disability and rehabilitation. 2016;38(2):103-22.
47. Varghese M, Lamb J, Rambani R, Venkateswaran B. The use of shoulder scoring systems and outcome measures in the UK. Annals of the Royal College of Surgeons of England. 2014;96(8):590-2.
48. van Kampen DA, Willems WJ, van Beers LWAH, Castelein RM, Scholtes VAB, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). Journal of orthopaedic surgery and research. 2013;8:40.
49. Ekeberg OM, Bautz-Holter E, Keller A, Tveitå EK, Juel NG, Brox JI. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. J Clin Epidemiol. 2010;63(5):575-84.
50. Jain SS, DeFroda SF, Paxton ES, Green A. Patient-Reported Outcome Measures and Health-Related Quality-of-Life Scores of Patients Undergoing Anatomic Total Shoulder Arthroplasty. Journal of bone and joint surgery American volume. 2020;102(13):1186-93.
51. Räsänen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynänen O-P, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. International journal of technology assessment in health care. 2006;22(2):235-41.
52. Evans J, Dattani R, Ramasamy V, Patel V. Responsiveness of the EQ-5D-3L in elective shoulder surgery: Does it adequately represent patient experience? Journal of orthopaedic surgery (Hong Kong). 2018;26(2):2309499018774922-.
53. Grobet C, Marks M, Tecklenburg L, Audigé L. Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity

orthopaedic disorders: a systematic literature review. Arch Orthop Trauma Surg. 2018;138(7):953-61.
54. Olerud P, Tidermark J, Ponzer S, Ahrengart L, Bergström G. Responsiveness of the EQ-5D in patients with proximal humeral fractures. J Shoulder Elbow Surg. 2011;20(8):1200-6.
55. Skare Ø, Mowinckel P, Schrøder CP, Liavaag S, Reikerås O, Brox JI. Responsiveness of outcome measures in patients with superior labral anterior and posterior lesions. Shoulder & elbow. 2014;6(4):262-72.
56. George SZ, Stryker SE. Fear-Avoidance Beliefs and Clinical Outcomes for Patients Seeking Outpatient Physical Therapy for Musculoskeletal Pain Conditions. Journal of Orthopaedic & Sports Physical Therapy. 2011;41(4):249-59.
57. Leeuw M, Goossens ME, Linton SJ, Crombez G, Boersma K, Vlaeyen JW. The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. Journal of behavioral medicine. 2007;30(1):77-94.
58. Vlaeyen JWS, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain. 2000;85(3):317-32.
59. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain (Amsterdam). 1993;52(2):157-68.
60. Inrig T, Amey B, Borthwick C, Beaton D. Validity and Reliability of the Fear-Avoidance Beliefs Questionnaire (FABQ) in Workers with Upper Extremity Injuries. Journal of Occupational Rehabilitation. 2012;22(1):59-70.
61. Mintken PE, Cleland JA, Whitman JM, George SZ. Psychometric properties of the Fear-Avoidance Beliefs Questionnaire and Tampa Scale of Kinesiophobia in patients with shoulder pain. Arch Phys Med Rehabil. 2010;91(7):1128-36.
62. Riley SP, Tafuto V, Cote M, Brismée JM, Wright A, Cook C. Reliability and relationship of the fear-avoidance beliefs questionnaire with the shoulder pain and disability index and numeric pain rating scale in patients with shoulder pain. Physiother Theory Pract. 2019;35(5):464-70.
63. Trolle N, Christiansen DH. Measurement properties of the Fear-Avoidance Belief Questionnaire for physical activity in patients with shoulder impingement syndrome. Patient Relat Outcome Meas. 2019;10:83-7.
64. Klintberg IH, Cools AM, Holmgren TM, Holzhausen AC, Johansson K, Maenhout AG, et al. Consensus for physiotherapy for shoulder pain. Int Orthop. 2015;39(4):715-20.
65. Behandling af patienter med subakromielt smertesyndrom i skulderen (Impingement syndrom/rotator-cuff syndrom), National klinisk retningslinje. Sundhedsstyrelsen; 2021.
66. Rabelo M, Nunes GS, da Costa Amante NM, de Noronha M, Fachin-Martins E. Reliability of muscle strength assessment in chronic post-stroke hemiparesis: a systematic review and meta-analysis. Topics in stroke rehabilitation. 2016;23(1):26-35.
67. Edouard P, Samozino P, Julia M, Gleizes Cervera S, Vanbiervliet W, Calmels P, et al. Reliability of isokinetic assessment of shoulder-rotator strength: a systematic review of the effect of position. Journal of Sport Rehabilitation. 2011;20(3):367-83.
68. Chamorro C, Armijo-Olivo S, Fuente CDl, Fuentes J, Chirosa LJ. Absolute reliability and concurrent validity of hand held dynamometry and isokinetic dynamometry in the hip, knee and ankle joint: systematic review and meta-analysis. Open Medicine. 2017;12(1):359-75.
69. Habets B, Staal JB, Tijssen M, van Cingel R. Intrarater reliability of the Humac NORM isokinetic dynamometer for strength measurements of the knee and shoulder muscles. BMC research notes. 2018;11(1):15.
70. Schrama PPM, Stenneberg MS, Lucas C, Van Trijffel E. Intraexaminer reliability of hand-held dynamometry in the upper extremity: A systematic review. Archives of Physical Medicine and Rehabilitation. 2014;95(12):2444-69.

71. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held Dynamometry Correlation With the Gold Standard Isokinetic Dynamometry: A Systematic Review. PM&R. 2011;3(5):472-9.
72. Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, et al. The Danish health care system and epidemiological research: from health care contacts to database records. Clin Epidemiol. 2019;11:563-91.
73. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol. 2015;7:449-90.
74. Andersen JS, Olivarius Nde F, Krasnik A. The Danish National Health Service Register. Scand J Public Health. 2011;39(7 Suppl):34-7.
75. Burr H, Pedersen J, Hansen JV. Work environment as predictor of long-term sickness absence: linkage of self-reported DWECS data with the DREAM register. Scand J Public Health. 2011;39(7 Suppl):147-52.
76. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology. 2007;60(1):34-42.
77. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation - Official Journal of the International Society of Quality of Life Research. 2018;27(5):1147-57.
78. Devlin NJ, Krabbe PFM. The development of new research methods for the valuation of EQ-5D-5L. The European journal of health economics. 2013;14(S1):1-3.
79. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ. 2018;27(1):7-22.
80. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Applied health economics and health policy. 2017;15(2):127-37.
81. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-70.
82. Karcioglu O, Topacoglu H, Dikme O, Dikme O. A systematic review of the pain scales in adults: Which to use? Am J Emerg Med. 2018;36(4):707-14.
83. Gilbart MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. J Shoulder Elbow Surg. 2007;16(6):717-21.
84. de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007;16(1):131-42.
85. Kirkwood BR, Sterne JAC. Essential medical statistics. 2. ed. ed. Malden, Mass.: Blackwell Science; 2007. 501 p.
86. Sørensen L, Oestergaard LG, van Tulder M, Petersen AK. Measurement properties of objective methods to assess shoulder muscle strength (a systematic review protocol). Physical Therapy Reviews. 2017;22(5-6):238-42.
87. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews. 2015;4(1):1-.
88. Higgins JPT GSe. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. ; [Available from: Available from www.handbook.cochrane.org.
89. Schünemann H. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. 2013 [updated October 2013. Available from: https://gdt.gradepro.org/app/handbook/handbook.html.

90. Sørensen L, Oestergaard LG, van Tulder M, Petersen AK. Measurement properties of handheld dynamometry for assessment of shoulder muscle strength: A systematic review. Scand J Med Sci Sports. 2020;30(12):2305-28.
91. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation - Official Journal of the International Society of Quality of Life Research. 2018;27(5):1171-9.
92. Mokkink LB, Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., Terwee, C. B. . COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manuel 2018 [Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf.
93. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.
94. Sørensen L, van Tulder M, Johannsen HV, Ovesen J, Enemark U, Oestergaard LG, et al. Costs of shoulder disorders in Denmark; a nationwide cost-off-illness study investigating 617,334 patients and matched controls. Submitted to PAIN-S-21-01260, October 2021 2021.
95. Sørensen L, van Tulder M, Johannsen HV, Ovesen J, Oestergaard LG. Responsiveness and minimal important change of the Oxford Shoulder Score, EQ-5D, and the Fear-Avoidance Belief Questionnaire Physical Activity subscale in patients undergoing arthroscopic subacromial decompression. JSES Int. 2021;5(5):869-74.
96. Sørensen L, Oestergaard LG, van Tulder M, Petersen AK. Measurement Properties of Isokinetic Dynamometry for Assessment of Shoulder Muscle Strength: A Systematic Review. Arch Phys Med Rehabil. 2021;102(3):510-20.
97. Anderson VB, Bialocerkowski AE, Bennell KL. Test-retest reliability of glenohumeral internal and external rotation strength in chronic rotator cuff pathology. Physical Therapy in Sport. 2006;7(3):115-21.
98. Dauty M, Delbrouck C, Huguet D, Rousseau B, Potiron-Josse M, Dubois C. Reproducibility of concentric and eccentric isokinetic strength of the shoulder rotators in normal subjects 40 to 55 years old. Isokinetics and Exercise Science. 2003;11(2):95-100.
99. Cavuoto LA, Pajoutan M, Mehta RK. Reliability analyses and values of isometric shoulder flexion and trunk extension strengths stratified by body mass index. PLoS One. 2019;14(7):e0219090.
100. Edouard P, Codine P, Samozino P, Bernard PL, Herisson C, Gremeaux V. Reliability of shoulder rotators isokinetic strength imbalance measured using the Biodex dynamometer. Journal of science and medicine in sport. 2013;16(2):162-5.
101. Forthomme B, Dvir Z, Crielaard JM, Croisier JL. Isokinetic assessment of the shoulder rotators: a study of optimal test position. Clinical physiology and functional imaging. 2011;31(3):227-32.
102. Frisiello S, Gazaille A, O'Halloran J, Palmer ML, Waugh D. Test-retest reliability of eccentric peak torque values for shoulder medial and lateral rotation using the Biodex isokinetic dynamometer. The Journal of orthopaedic and sports physical therapy. 1994;19(6):341-4.
103. Grabowski P, Narveson M, Siegle S. Reliability, Responsiveness, and Criterion Validity of the Kiio Sensor, a New Tool for Assessment of Muscle Function. Military medicine. 2017;182(S1):26-31.
104. Hill AM, Pramanik S, McGregor AH. Isokinetic dynamometry in assessment of external and internal axial rotation strength of the shoulder: Comparison of two positions. Isokinetics and Exercise Science. 2005;13(3):187-95.

105. Kramer JF, Ng LR. Static and dynamic strength of the shoulder rotators in healthy, 45- to 75-year-old men and women. The Journal of orthopaedic and sports physical therapy. 1996;24(1):11-8.
106. Leggin BG, Neuman RM, Iannotti JP, Williams GR, Thompson EC. Intrarater and interrater reliability of three isometric dynamometers in assessing shoulder strength. Journal of shoulder and elbow surgery. 1996;5(1):18-24.
107. Lindström B, Waling K, Sundelin G, Ahlgren C. Test-retest Reliability of Biomechanical Output and Subjective Ratings of Exertion in Isometric and Isokinetic Shoulder Forward Flexion in Healthy Subjects. Advances in Physiotherapy. 2003;5(4):169-78.
108. Magnusson SP, Gleim GW, Nicholas JA. Subject variability of shoulder abduction strength testing. Am J Sports Med. 1990;18(4):349-53.
109. Malerba JL, Adam ML, Harris BA, Krebs DE. Reliability of dynamic and isometric testing of shoulder external and internal rotators. The Journal of orthopaedic and sports physical therapy. 1993;18(4):543-52.
110. Mandalidis DG, Donne B, O'Regan M, O'Brien M. Reliability of isokinetic internal and external rotation of the shoulder in the scapular plane. Isokinetics and Exercise Science. 2001;9(1):65-72.
111. Mayer F, Horstmann T, Kranenberg U, Rocker K, Dickhuth HH. Reproducibility of isokinetic peak torque and angle at peak torque in the shoulder joint. International Journal of Sports Medicine. 1994;15 Suppl 1:S26-31.
112. Meeteren J, Roebroeck ME, Stam HJ. Test-retest reliability in isokinetic muscle strength measurements of the shoulder. Journal of Rehabilitation Medicine. 2002;34(2):91-5.
113. Papotto BM, Rice T, Malone T, Butterfield T, Uhl TL. Reliability of Isometric and Eccentric Isokinetic Shoulder External Rotation. Journal of Sport Rehabilitation. 2016;Technical Report:10-1123/jsr.2015-0046.
114. Plotnikoff NA, MacIntyre DL. Test-retest reliability of glenohumeral internal and external rotator strength. Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine. 2002;12(6):367-72.
115. Smith J, Padgett DJ, Kotajarvi BR, Eischen JJ. Isokinetic and isometric shoulder rotation strength in the protracted position: A reliability study. Isokinetics and Exercise Science. 2001;9(2-3):119-27.
116. Sullivan SJ, Chesley A, Hebert G, McFaull S, Scullion D. The validity and reliability of hand-held dynamometry in assessing isometric external rotator performance. The Journal of orthopaedic and sports physical therapy. 1988;10(6):213-7.
117. O'Shea SD, Taylor NF, Paratz JD. Measuring muscle strength for people with chronic obstructive pulmonary disease: retest reliability of hand-held dynamometry. Archives of Physical Medicine and Rehabilitation. 2007;88(1):32-6.
118. Beshay N, Lam PH, Murrell GAC. Assessing the Reliability of Shoulder Strength Measurement: Hand-Held versus Fixed Dynamometry. Shoulder & Elbow. 2011;3(4):244-51.
119. Fieseler G, Laudner KG, Irlenbusch L, Meyer H, Schulze S, Delank KS, et al. Inter- and intrarater reliability of goniometry and hand held dynamometry for patients with subacromial impingement syndrome. Journal of exercise rehabilitation. 2017;13(6):704-10.
120. Awatani T, Mori S, Shinohara J, Koshiba H, Nariai M, Tatsumi Y, et al. Same-session and between-day intra-rater reliability of hand-held dynamometer measurements of isometric shoulder extensor strength. Journal of Physical Therapy Science. 2016;28(3):936-9.
121. Awatani T, Morikita I, Shinohara J, Mori S, Nariai M, Tatsumi Y, et al. Intra- and inter-rater reliability of isometric shoulder extensor and internal rotator strength measurements performed using a hand-held dynamometer. Journal of Physical Therapy Science. 2016;28(11):3054-9.

122. Balogun JA, Powell R, Trullender B, Olson S, Balogun AO. Intra- and inter-tester reliability of the Nicholas hand-held dynamometer during evaluation of upper extremity isometric muscle strength. European Journal of Physical Medicine and Rehabilitation. 1998;8(2):48-53.
123. Bohannon RW. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil. 1997;78(1):26-32.
124. Burnham SR, Bell CG, Olenik CL, Reid CD. Shoulder Abduction Strength Measurement in Football Players: Reliability and Validity of Two Field Tests. Clinical Journal of Sport Medicine. 1995;5(2):90-4.
125. Byl NN, Richards S, Asturias J. Intrarater and interrater reliability of strength measurements of the biceps and deltoid using a hand held dynamometer. The Journal of orthopaedic and sports physical therapy. 1988;9(12):395-8.
126. Cadogan A, Laslett M, Hing W, McNair P, Williams M. Reliability of a new hand-held dynamometer in measuring shoulder range of motion and strength. Manual therapy. 2011;16(1):97-101.
127. Cools AM, De Wilde L, Van Tongel A, Ceyssens C, Ryckewaert R, Cambier DC. Measuring shoulder external and internal rotation strength and range of motion: comprehensive intra-rater and inter-rater reliability study of several testing protocols. Journal of shoulder and elbow surgery. 2014;23(10):1454-61.
128. Dollings H, Sandford F, O'Conaire E, Lewis JS. Shoulder strength testing: the intra- and inter-tester reliability of routine clinical tests, using the PowerTrack™ II Commander. Shoulder & Elbow. 2012;4(2):131-40.
129. Donatelli R, Ellenbecker TS, Ekedahl SR, Wilkes JS, Kocher K, Adam J. Assessment of shoulder strength in professional baseball pitchers. The Journal of orthopaedic and sports physical therapy. 2000;30(9):544-51.
130. Fieseler G, Molitor T, Irlenbusch L, Delank KS, Laudner KG, Hermassi S, et al. Intrarater reliability of goniometry and hand-held dynamometry for shoulder and elbow examinations in female team handball athletes and asymptomatic volunteers. Archives of orthopaedic and trauma surgery. 2015;135(12):1719-26.
131. Hayes K, Walton JR, Szomor ZL, Murrell GAC. Reliability of 3 methods for assessing shoulder strength. Journal of Shoulder and Elbow Surgery. 2002;11(1):33-9.
132. Holt KL, Raper DP, Boettcher CE, Waddington GS, Drew MK. Hand-held dynamometry strength measures for internal and external rotation demonstrate superior reliability, lower minimal detectable change and higher correlation to isokinetic dynamometry than externally-fixed dynamometry of the shoulder. Physical therapy in sport : official journal of the Association of Chartered Physiotherapists in Sports Medicine. 2016;21:75-81.
133. Johansson KM, Adolfsson LE. Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment. Journal of Shoulder and Elbow Surgery. 2005;14(3):273-8.
134. Johansson FR, Skillgate E, Lapauw ML, Clijmans D, Deneulin VP, Palmans T, et al. Measuring Eccentric Strength of the Shoulder External Rotators Using a Handheld Dynamometer: Reliability and Validity. Journal of athletic training. 2015;50(7):719-25.
135. McLaine SJ, Ginn KA, Kitic CM, Fell JW, Bird ML. The Reliability of Strength Tests Performed In Elevated Shoulder Positions Using a Handheld Dynamometer. Journal of Sport Rehabilitation. 2016;Technical Report:10-1123/jsr.2015-0034.
136. McMahon LM, Burdett R, Whitney SL. Effects of muscle group and placement site on reliability of hand-held dynamometry strength measurements. The Journal of orthopaedic and sports physical therapy. 1992;15(5):236-42.
137. Ottenbacher KJ, Branch LG, Ray L, Gonzales VA, Peek MK, Hinman MR. The reliability of upper- and lower-extremity strength testing in a community survey of older adults. Archives of Physical Medicine and Rehabilitation. 2002;83(10):1423-7.

138. Phillips BA, Lo SK, Mastaglia FL. Muscle force measured using "break" testing with a hand-held myometer in normal subjects aged 20 to 69 years. Arch Phys Med Rehabil. 2000;81(5):653-61.
139. Riemann BL, Davies GJ, Ludwig L, Gardenhour H. Hand-held dynamometer testing of the internal and external rotator musculature based on selected positions to establish normative data and unilateral ratios. Journal of Shoulder and Elbow Surgery. 2010;19(8):1175-83.
140. Sciascia A, Uhl T. Reliability of Strength and Performance Testing Measures and their Ability to Differentiate Persons with and without Shoulder Symptoms. International journal of sports physical therapy. 2015;10(5):655-66.
141. Vermeulen HM, de Bock GH, van Houwelingen HC, van der Meer RL, Mol MC, Plus BT, et al. A comparison of two portable dynamometers in the assessment of shoulder and elbow strength. Physiotherapy. 2005;91(2):101-12.
142. Kuijpers T, van Tulder MW, van Der Heijden GJMG, Bouter LM, van Der Windt DAWM. Costs of shoulder pain in primary care consulters: a prospective cohort study in The Netherlands. BMC musculoskeletal disorders. 2006;7:83-.
143. Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. Spine (Phila Pa 1976). 2011;36(13):1050-8.
144. Ketola S, Lehtinen J, Arnala I, Nissinen M, Westenius H, Sintonen H, et al. Does arthroscopic acromioplasty provide any additional value in the treatment of shoulder impingement syndrome?: a two-year randomised controlled trial. The Journal of bone and joint surgeryBritish volume. 2009;91-B(10):1326-34.
145. Davis AD, Kakar S, Moros C, Kaye EK, Schepsis AA, Voloshin I. Arthroscopic Versus Open Acromioplasty. The American Journal of Sports Medicine. 2010;38(3):613.
146. Indsatser ved sygdom, nedslidning mv.: Styrelse for arbejdsmarked of rekruttering; 2021 [updated 22-04-2020. Available from: https://star.dk/da/indsatser-og-ordninger/indsatser-ved-sygdom-nedslidning-mv/.
147. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sørensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol. 2011;11:83.
148. Bae YH, Lee GC, Shin WS, Kim TH, Lee SM. Effect of Motor Control and Strengthening Exercises on Pain, Function, Strength and the Range of Motion of Patients with Shoulder Impingement Syndrome. Journal of Physical Therapy Science. 2011;23(4):687-92.
149. Hjollund NH, Larsen FB, Andersen JH. Register-based follow-up of social benefits and other transfer payments: accuracy and degree of completeness in a Danish interdepartmental administrative database compared with a population-based survey. Scand J Public Health. 2007;35(5):497-502.
150. Stapelfeldt CM, Jensen C, Andersen NT, Fleten N, Nielsen CV. Validation of sick leave measures: self-reported sick leave and sickness benefit data from a Danish national register compared to multiple workplace-registered sick leave spells in a Danish municipality. BMC Public Health. 2012;12:661.
151. Terwee CB, Jansma EP, Riphagen II, De Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Quality of Life Research. 2009;18(8):1115-23.
152. Næsten halvdelen af de beskæftigede over pensionsalderen arbejder under 20 timer om ugen: Danmarks Statistik; 2019 [updated 03.05.2019. Available from: https://www.dst.dk/da/Statistik/nyheder-analyser-publ/bagtal/2019/2019-02-12-naesten-halvdelen-af-beskaeftigede-over-pensionsalderen-arbejder-under-20-timer-om-ugen.

Appendices

Paper I Cost of shoulder disorders in Denmark; a nationwide cost-of-illness study investigating 617,334 patients and matched controls.

Paper II Responsiveness and minimal important change of the Oxford Shoulder Score, EQ-5D, and the Fear-Avoidance Belief Questionnaire Physical Activity subscale in patients undergoing arthroscopic subacromial decompression.

Paper III Measurement Properties of Isokinetic Dynamometry for Assessment of Shoulder Muscle Strength: A Systematic Review.

Paper IV Measurement properties of handheld dynamometry for assessment of shoulder muscle strength: A systematic review.

Appendix A Supplementary material Paper III: Reliability and measurement error results from the included studies

Appendix B Supplementary material Paper IV: Characteristics of the included studies using handheld dynamometry