BMJ Open

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT): A protocol for a randomized trial comparing arthroscopic subacromial decompression and diagnostic arthroscopy (placebo control), with an exercise therapy control, in the treatment of shoulder impingement syndrome

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014087
Article Type:	Protocol
Date Submitted by the Author:	30-Aug-2016
Complete List of Authors:	Paavola, Mika; Helsinki University Central Hospital, Orthopaedics and traumatology Malmivaara, Antti; National Institute for Health and Welfare, Centre for Health and Social Economics Taimela, Simo; University of Helsinki, Department of Orthopedics and Traumatology Kanto, Kari; Hatanpaan kantasairaala Jarvinen, Teppo; University of Helsinki, Department of Orthopaedics and Traumatology
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Evidence based practice, Rehabilitation medicine, Sports and exercise medicine, General practice / Family practice
Keywords:	Orthopaedic & trauma surgery < SURGERY, Clinical trials < THERAPEUTICS, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, REHABILITATION MEDICINE, SPORTS MEDICINE

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT): A protocol for a randomized trial comparing arthroscopic subacromial decompression and diagnostic arthroscopy (placebo control), with an exercise therapy control, in the treatment of shoulder impingement syndrome

Mika Paavola[1], Antti Malmivaara[2], Simo Taimela[1], Kari Kanto[3], and Teppo L N Järvinen[1], for the FIMPACT investigators*

[1] Department of Orthopedics and Traumatology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

[2] National Institute for Health and Welfare, Center for Health and Social Economics, Helsinki, Finland

[3] Department of Orthopedics and Traumatology, Hatanpää City Hospital, Tampere, Finland

Address of corresponding author:
Teppo Järvinen, MD, PhD
Department of Orthopedics and Traumatology
Helsinki University Central Hospital / Töölö Hospital
Topeliuksenkatu 5
P.O. Box 266, 00029 HUS,
Helsinki, Finland
Phone: +358-50-4272481
Email: teppo.jarvinen@helsinki.fi

* FIMPACT investigators, see appendix 1.

ABSTRACT

Introduction: Arthroscopic subacromial decompression (ASD) is the most commonly performed surgical intervention for shoulder pain, yet evidence on its efficacy is limited. The rationale of the surgery rests on the tenet that symptom relief is achieved through removal of a bony acromial spur and the resulting decompression of the tendon passage. Acknowledging the potential placebo effect of surgery, the primary objective of this superiority trial is to compare the efficacy of ASD versus diagnostic arthroscopy (DA) in patients with SIS, the latter procedure differing from the former by only lacking subacromial decompression. As a non-surgical treatment option, a third group of supervised progressive exercise therapy (ET) is also included for allowing pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies.

Methods and Analysis: FIMPACT trial is an ongoing multicentre, three-group randomised controlled study to assess the efficacy of the ASD vs. DA. We performed two-fold concealed allocation, first by randomizing patients to surgical (ASD or DA) or conservative (ET) treatment in 2:1 ratio and then those allocated to surgery further to ASD or DA in 1:1 ratio. Our two primary outcomes are pain at rest and arm activity assessed with 100 mm visual analog scales (VASs), while the secondary outcomes are functional assessment (Constant score and Simple shoulder test), global assessment of change, proportion of recovered patients, quality of life (15D), reoperations/treatment conversions, adverse effects and complications, all at 2 years post randomization. We recruited a total of 210 patients from 3 tertiary referral centres. We will conduct the primary analysis on the intention-to-treat basis.

Ethics and Dissemination: The study was approved by the institutional review board of the Pirkanmaa Hospital District and duly registered at ClinicalTrials.gov. The findings of this study will be disseminated widely through peer-reviewed publications and conference presentations.

Trial registration: ClinicalTrials.gov NCT00428870 (first registered January 29, 2007).

Keywords: Acromion; Acromioplasty; Arthroscopy; Impingement; Physiotherapy; Placebo; Sham; Shoulder; Syndrome; Randomised; Trial

Strengths of this study

- Efficacy design: Strict eligibility criteria
- Placebo-surgery controlled trial: Blinding of both the participants and the outcome assessors in the comparison between index surgery and control (placebo surgery)
- Inclusion of a non-surgical treatment option to allow a pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies

Limitations of this study

- Potential confounding due to participants' knowledge of the treatment delivered (in comparing surgical vs. non-operative treatment strategies)

INTRODUCTION

Subacromial decompression is one of the most frequently performed procedures in orthopaedics[1 2]. It is carried out to treat patients with shoulder pain attributed to "subacromial impingement syndrome" (SIS). Conventional wisdom dictates that SIS is caused by 'impingement' of the rotator cuff (RC) between the humeral head and the overlying acromion while lifting the arm. The appropriateness of this mechanistic explanation has been challenged lately and accordingly, a more generic label of "subacromial pain syndrome" (SAPS) is currently advocated[3]. The aim of subacromial decompression procedure, typically carried out arthroscopically, is to decompress the RC tendon passage through the subacromial space through resection and smoothening of the hypertrophied or prominent anterolateral undersurface of the acromion. Management of shoulder pain has been estimated to account for 4.5 million visits annually to physicians in the USA alone[4], accounting for US$3 billion financial burden each year[5]. Since 44-65% of all shoulder complains are related to SIS - the rest to other shoulder pathologies, particularly to repair the RC tendons - it can be estimated that annual direct medical costs of SIS are over $1 billion in the USA[6 7].

Since the introduction of subacromial decompression surgery in the early 1970s[8], the incidence (volume) of this procedure has shown a steady increase across the entire western world. Recent statistics show that with the advent of arthroscopy, the number of these surgeries has increased dramatically -- 5-fold from 1980s to 2005 in the US[9] and 700% from 2000 to 2010 in the UK[10]. Remarkably, there is dire absence of evidence from high-quality controlled trials to support the existing practice of performing subacromial decompression for patients with SIS. In fact, two recent systematic reviews/meta-analyses concluded that subacromial decompression provides no superior benefits in terms of pain relief, function, or quality of life to conservative treatment[11 12]. However, the proponents of the procedure have argued that the evidence is skewed with respect to the therapeutic potential of surgery due to a significant cross-over (5-15%) from conservative treatment to surgery[13-15]. Although such concern is obviously warranted, it should also be recalled that surgeons' own perceptions on the success of any surgery might similarly be biased due to a considerable surgical placebo effect.

The outcome of any medical (surgical) intervention – particularly when treating primarily subjective symptoms – is a cumulative effect of three main elements: placebo effects, critical therapeutic (surgical) element, and non-specific effects, most importantly, the normal variation in the course of the disease and the regression-to-the-mean phenomenon[16 17]. Conceding that the act of surgery *per se* produces a profound placebo response, a 'true' treatment effect is impossible to disentangle from the nonspecific (placebo) effects – such as the patients' or researchers' expectations of benefit – without a placebo

comparison group[18]. The critical therapeutic (surgical) element is the component of the surgical procedure that is believed to provide the therapeutic effect (here, subacromial decompression), being distinct from aspects of the procedures that are diagnostic or required to access the disease being treated (here, shoulder arthroscopy).

To the best of our knowledge, there is only one other ongoing study aiming to assess the true efficacy of subacromial decompression surgery in patients with SIS using a placebo controlled study design. According to the published protocol of this CSAW trial[19], the investigators have chosen a highly similar approach to that of our FIMPACT trial. In brief, the CSAW trial is a three-group pragmatic RCT comparing arthroscopic acromioplasty, active monitoring with specialist reassessment, and investigational shoulder arthroscopy only. CSAW aims for recruitment of 300 patients with SIS to assess the efficacy of the surgery against no surgery, the need for a specific component of the surgery (acromioplasty), and the quantification of the possible placebo effect.

The primary hypothesis of our FIMPACT trial is that ASD is superior to DA in patients with SIS. In addition, we will perform a pragmatic comparison of surgical and non-surgical treatment options (ASD vs. ET). The relative benefits of ASD and ET will be assessed without a priori hypothesis on the superiority of one or the other.

MATERIALS AND METHODS

Overview of study design

FIMPACT trial is an ongoing multicentre, three group randomised controlled superiority study with a primary objective to assess the efficacy of the ASD vs. DA in patients diagnosed with SIS. The primary objective of the trial is to assess the efficacy of the ASD vs. DA. Our design also enables the pragmatic comparison of surgical and non-surgical treatment strategies (ASD vs. ET) (Figure 1). We performed a two-fold concealed allocation, first by randomizing patients to surgical or conservative treatment in 2:1 ratio. We then randomized those allocated to surgery to ASD or DA in 1:1 ratio. The initial patient screening for the trial began at one site (Tampere) in February 1, 2005 and was then expanded to two additional tertiary referral centres in March 2006 and December 2006 to improve recruitment and to ensure multicentre design with its obvious benefits to the generalisability of the results. The recruitment was completed (all 210 required patients enrolled) in August 2013.

Ethical approval

Ethical approval was obtained on December 28, 2004 from the institutional review board (IRB) of the Pirkanmaa Hospital District (R04200). Local research and development approvals were gained for each recruiting centre.

Patient/Participant selection

We assessed patients referred to any of the participating clinics and complaining of subacromial shoulder pain for eligibility. All potential participants were screened to determine eligibility according to the inclusion and exclusion criteria. A consultant surgeon confirmed the clinical diagnosis of SIS. To qualify as a recruiting surgeon, all trial surgeons had to have experience of more than 500 shoulder arthroscopies before the start of the trial. Detailed clinical examination of the shoulder was performed to rule out possible instability, clinical signs of rotator cuff rupture, frozen shoulder or other causes of symptoms. Standard x-rays and MRI were obtained from all potential participants and assessed by both a musculoskeletal radiologist and an orthopaedic surgeon. If patient was found eligible for this study (fulfilling indications for ASD) and a written informed consent was obtained, participants were randomised into non-operative or operative group (1:2) immediately after the baseline appointment.

Eligibility criteria

We used specific eligibility criteria to ascertain that the participants recruited represented only those with SIS. Accordingly, a standardized clinical examination was first performed, followed by a subacromial injection test. To exclude patients with concomitant pathology, particularly rotator cuff rupture, standard x-rays and magnetic resonance imaging with intra-articular contrast injection (MRA) were carried out on all potential participants.

Inclusion criteria

1) Adult men or women ages 35 to 65 years

2) Subacromial pain for greater than 3 months with no relief from non-operative means (physiotherapy, non-steroidal anti-inflammatory medication, corticosteroid injections, and rest)

3) Pain provoked by abduction and positive painful arc -sign

4) Positive impingement test (temporary relief of pain by subacromial injection of lidocaine)

5) Pain in at least 2 out of 3 of isometric tests (abduction 0° and 30° or external rotation)

6) Provision of informed consent from the participant

7) Ability to speak, understand and read in the language of the clinical site.

Exclusion criteria

1. Full thickness tear of the rotator cuff tendons diagnosed on clinical examination or magnetic resonance imaging with intra-articular contrast (MRA)

2. Osteoarthritis of the glenohumeral and/or acromioclavicular joint diagnosed on clinical examination or on x-rays

3. Previous surgical procedure on the affected shoulder

4. Evidence of shoulder instability (positive apprehension/positive sulcus sign)

5. Symptomatic cervical spine pathology

6. History of alcoholism, drug abuse, psychological or other emotional problems that are likely to invalidate informed consent

Recruitment process

Consultant orthopaedic surgeons carried out eligibility screening among patients referred to the study centres through standard clinical practice for shoulder pain. Patients meeting the eligibility criteria were introduced to the study. If patients expressed interest in participating, written information about the study was provided and they were asked to opt in. If the interest continued, arrangements were made for obtaining required imaging (x-rays and MRA) and for a separate baseline appointment.

Informed consent

At the first appointment, all participants were introduced to the detailed written information about the study and asked to sign a written informed consent form. At the baseline appointment (arranged within 45 days of initial contact), baseline data was completed and participant's willingness to participate in the study was confirmed. This procedure ensured that all potential participants had a reflection period for consent of at least 48 hours before giving their final consent to participate. Particular attention was paid to ensure that the participants realized that on entering the study they may receive only diagnostic arthroscopy, in which case the subacromial decompression would not be performed. They were also informed that participation in the study is entirely voluntary and the decision they make would not affect their possible future care in case of refusal. In addition, every participant was informed of their right to withdraw from the trial whenever they desire without giving the researchers any reason for such decision.

Baseline assessment

Baseline assessment included documentation of sex, birth date, education, employment, hand dominance, time from the onset of symptoms, recreational habits, and employment status. We asked participants to assess their general heath and usage of pain medication. Modalities of the prior conservative treatment were also recorded (Table 1).

Table 1: Baseline characteristics

	ASD	DA	ET
Age (years), mean (SD)			
Gender (female/male), n (%)			
Dominant hand affected, n (%)			
Social economic status/ work load			
Heavy manual labor (construction work etc.), n (%)			
Heavy manual labor (variable workload), n (%)			
Mostly manual labor including daily office work, n (%)			

Mostly office work with occasional manual assignments, n (%)	
Full-time office work, n (%)	
Unemployed, n (%)	
Pensioner/disability pensioner, n (%)	
Student, n (%)	
Homemaker/housewife/other, n (%)	
Subjective health	
Duration of symptoms (Months), mean (SD)	
Ability to work normally regardless of the shoulder symptoms ? (yes/no), n (%)	
Recreational ability regardless of the shoulder symptoms ? (yes/no), n (%)	
Prior treatments	
Rest, n (%)	
Pain medication, n (%)	
Topical pain medication, n (%)	
Corticosteroid injection, n (%)	
Ultrasound, laser or any other similar therapies, n (%)	
Physiotherapy including exercise therapy, n (%)	
Other, n (%)	
Generic health states	
15D	
SF-36	
Pain measurements/Shoulder scores	
Pain at rest (100mm VAS scale), mean (SD)	
Pain during activity (100mm VAS scale), mean (SD)	
Constant- Murley score (CM), mean (SD)	
The simple shoulder test (SST), mean (SD)	

Baseline clinical symptoms

The recruiting surgeon carried out a clinical history and a clinical examination related to shoulder pain. Other possible shoulder complaints than SIS, such as full-thickness rotator cuff tears, frozen shoulder, osteoarthritis of the acromioclavicular joint and instability, were ruled out as much as clinical diagnosis allows.

Baseline imaging

Standard x-rays of the shoulder were obtained to assess possible glenohumeral or acromioclavicular osteoarthritis. A magnetic resonance image with intra-articular contrast medium (MRA) was also obtained to rule out any other intra- or extra-articular pathologies. A musculoskeletal radiologist and an orthopaedic surgeon assessed all the images.

Randomisation and concealment

We used a two-phase sequential randomization. In the Phase I, the participants were randomized into non-surgical or surgical treatment with allocation ratio 1:2. In the Phase II, those allocated to surgical treatment in the Phase I were further randomized to ASD or DA with 1:1 ratio (Figure 1).

An independent statistician with no involvement in the execution of the trial prepared separate randomization lists for each study centre using a computer-generated schedule. Randomization was carried out using sequentially numbered sealed opaque envelopes. The envelopes were kept in a secure, agreed location at each centre. To ensure concealment, block randomization was applied using blocks varying in size randomly, the block size known only by the statistician.

To initially enter a participant into the study (Phase I), an envelope containing the treatment assignment (non-surgical (ET) or surgery (ASD or DA), ratio 1:2) was opened during the baseline appointment. Participants randomized to ET started standardized physiotherapy within 2 weeks of the baseline appointment. Participants allocated to surgical treatment were scheduled for surgery with the aim to carry out the procedure within 12 weeks of randomization.

At the day of surgery, a diagnostic arthroscopy was first carried out to confirm the eligibility of the participant (to rule out full-thickness RC tear and other obvious intra-articular pathology). Research/staff nurse then completed the randomization procedure (Phase II) by opening an envelope containing the surgical treatment allocation (ASD or DA, ratio 1:1). The allocation was revealed to the surgeon by showing the paper, but not expressed verbally.

Interventions

Diagnostic arthroscopy (DA)

All participants in the two operative groups first underwent arthroscopic examination of the shoulder with the use of standard posterior and lateral portals and a 4-mm arthroscope. To maintain concealment, the surgery was carried out under general anesthesia. The orthopaedic surgeon evaluated and graded possible intra-articular pathologic changes. The rotator cuff integrity was evaluated also from the subacromial space without performing routine bursectomy. If the integrity of the rotator cuff could not be assessed, bursal tissue was bluntly stretched with troachar or resected on the tendon side to allow visualisation. If arthroscopic examination revealed any unexpected pathology (such as capsular pathology, full-thickness rotator cuff tear, or osteoarthritis), the patient was treated according current clinical prac-

tice guidelines for the given pathology while under the same anesthesia. In such a case, the participant was excluded from the trial.

After the arthroscopic examination of the glenohumeral joint and subacromial space, confirming the eligibility of the participant, the participants were randomly assigned to receive either ASD or DA only. If the patient was allocated to the DA group, the operation was terminated. To ensure concealment of the participants and the staff other than those in the operating theatre, the participants were kept in the operating theatre for the required time to perform the subacromial decompression. DA group had all the same essential operative components and risks of ASD, but it did not involve any surgical procedure on the bony acromion.

Arthroscopic subacromial decompression (ASD)

Debridement of the subacromial bursa was performed with a shaver and/or electrocoagulation, followed by the resection of the bony spurs and projecting anterolateral undersurface of the acromion by a shaver as described by Ellman[20].

Postoperative care

In both ASD and DA groups, the postoperative rehabilitation was identical, and carried out according to the standardized rehabilitation protocols of the participant centres. Since the initial rehabilitation after a surgery needs to be "tempered" due to surgical trauma/tissue/joint irritation, the rehabilitation protocol of the operatively treated groups (ASD and DA) was not identical to the ET group.

Exercise therapy (ET)

In the exercise therapy (ET) group, supervised progressive physiotherapy was started within 2 weeks of randomization using a standardized protocol. The protocol was based on the same principles as the regimen shown effective for the treatment of SIS earlier[15], but was updated – with the help of the principal investigator of the original study[15] – to conform with the state-of-the-art exercise therapy for SIS. The regimen was based on daily home exercises, but also included 15 visits to an independent physiotherapist for guidance and monitoring of the progress. The aim of the supervised exercise treatment was to restore painless, normal mobility of the shoulder girdle, eliminate any capsular tightness and to increase the dynamic stability of the glenohumeral joint and the scapula.

Compliance to treatment allocation and possible crossover

Participants allocated to ET group were told at the time of giving consent that they would be allowed to consider crossing over to the ASD group if adequate relief of symptoms was not achieved by conservative means (preferably no sooner than 6 months post randomization). Similarly, in the two surgical treatment groups, the participants were informed of the possibility of unblinding if deliberating symptoms persisted 6 months or more after operation. If the participant was allocated to DA group, ASD was then offered.

Outcome measures

The outcomes used in this study and the timetable for follow-up assessments are summarised in Table 2.

Table 2: Outcomes and follow-up time points

Assesment	Screening	Enrollment (Baseline)	Surgery	3 Months	6 Months	12 Months	24 Months	5 years	10 years
Screening form	X								
Informed consent		X							
Baseline characteristics form		X							
X-ray and MRI	X								X
Randomisation		X (1st)	X (2nd)						
Arthroscopic findings form			X						
Follow-up form*				X		X			
Clinical examination					X		X	X	X
Complications form**			(X)	(X)	(X)	(X)	(X)	(X)	(X)
VAS, at rest		X		X	X	X	X	X	X
VAS, at arm activity		X		X	X	X	X	X	X
Constant- Murley Score		X			X		X	X	X
Simple Shoulder Test (SST)		X			X		X	X	X
SF-36		X			X		X	X	X
15D		X		X	X	X	X	X	X
Patients satisfaction to the treatment				X	X	X	X	X	X
Patients assessment of the treatment allocation				X					
Health resource utilization				X	X	X	X	X	X

* Letter/telephone interview

** If required

Primary outcome measure

VAS

As the primary outcome measure, we used a visual analogue scale (0-100 mm) to measure the patient's perceived pain intensity at rest and at arm activity during the 24 hours preceding the assessment. We considered the minimal clinically important difference (MCID) for VAS 15 mm on a 100 mm VAS scale and the patient acceptable symptom state (PASS), the score below which patients consider themselves well, was considered 30 mm.[21]

Secondary outcome measures

Constant-Murley score

Constant-Murley score (CS) is the most commonly used scoring system for evaluation of various disorders of the shoulder[22]. It consists of both objective (range of motion and strength) and subjective measurements (pain assessment, work load, and leisure time activities), which are summarized in a score between 0 and 100. A higher score indicates better shoulder function. The minimal detectable change (MDC) of the Constant score is 17 for patients with SIS[23].

SST

The simple shoulder test (SST) was developed to assess the functional limitations of the patient's activities of daily living [24]. The SST consists of 12 questions with yes (1) or no (0) response options. The maximum SST score is 12 indicating normal shoulder function, minimum score of 0 points refers severely diminished shoulder function. The SST has good reliability and responsiveness in patients with rotator cuff symptoms[25]. The MCID for the SST in rotator cuff disease is 2 points[26].

15D

The 15D instrument is a generic health-related quality of life (HRQoL) instrument comprising 15 dimensions[27]. For each dimension, the respondent must choose one of the five levels that best describes his/her state of heal that the moment (the best level being 1 and the worst level being 5). A set of utility or preference weights is used in an addition aggregate formula to generate a single index number, the utility or 15D score. The maximum 15D score is 1 (no problems on any dimension) and the minimum

score is 0 (being dead). The responsiveness, reliability and validity of 15D have been thoroughly established, and this instrument has been used extensively in clinical and healthcare research[28 29].

SF-36

Short form or SF-36 is a generic HRQoL instrument to quantify the physical, functional, and psychological aspects of health related quality of life. It consists of 36 questions in eight subscales that assess physical, functional, social, and psychological well-being[30]. Score ranges from 0 to 100, a higher score is associated with better health. The physical and mental component summary scales (PCS and MCS, respectively) are then calculated as composites of the related subscales. SF-36 is one of most widely used measure of health-related quality of life[31].

Patient satisfaction

Patients' global assessment of satisfaction was elicited using the following question: "How satisfied are you with the treatment given?" on a 5-item scale at each follow-up timepoint (Table 2). As before[32], the responses "Very satisfied", "Satisfied" and "Somewhat satisfied" were categorized as satisfied, while responses, "Dissatisfied" and "Very dissatisfied" were categorized as dissatisfied.

Return to previous leisure activities

Similarly, at each follow-up (Table 2), participants responded to the following question: "Have you been able to return to their previous leisure activities?" ("yes" or "no").

Patients' perception of operative treatment-group assignment

At the 3-month follow-up point, the patients in the two operative groups were asked to guess whether they had undergone ASD or DA.

Health resource utilization and costs

For the cost-effectiveness analysis, at each follow-up visit the participants were asked to fill in a questionnaire inquiring the use of healthcare resources. The questionnaire contains a list of items of

healthcare resources available and the participants were asked to fill in the number of visits per item during the recall period of each follow-up time point. The resource use will be calculated based on the number of visits times unit cost per item and expressed as mean costs by items of resource use, and the mean direct total health care resource costs. All costs will be discounted to the 2016 price level.

Time to return to work

Information about return to work was recorded at each follow-up time point (Table 2).

Complications and adverse events

The participants were encouraged to contact the participating hospitals if any adverse events occurred and contacts to the health care system were monitored at every follow-up visit. Potential adverse events (AE) were categorized to serious adverse events (SAE) and minor adverse events (MAE). Death, cardiovascular or gastrointestinal events, deep venous thrombosis, pulmonary embolism, systemic or local infection were categorised as SAEs. Shoulder symptoms like pain, swelling and decreased range of motion were categorised as MAEs if the participants sought treatment. Data on complications and adverse events were recorded and their severity and frequency will be assessed.

Follow-up

The full follow-up process is shown in figure 1. In brief, the participants filled in the above noted (mailed) outcome questionnaires at 3, 6, 12 and 24 months post randomization, in addition to which they were also assessed clinically at 6 and 24 months (and 5 and 10 years) post randomisation by a study physiotherapist unaware of treatment allocation.

Adherence and loss to follow-up

Several procedures were implemented to limit loss to follow-up, including exclusion of individuals likely to pose suboptimal adherence to follow-up from the study, obtaining of a verified contact information from each consented participant, and having local research nurse remind participants of upcoming follow-up/clinic visits. All attempts were made to also make the follow-up as convenient for the patients as possible. Participants were required to visit the outpatient clinic only at 6 months and 24 months (and 5

and 10 years) post randomisation, while the 3- and 12-month follow-ups were carried out using mailed questionnaires to minimize inconvenience to the participants. The follow-up visits involved no discomfort for the participant than the routine clinical shoulder examinations. The follow up schedule did not provide extra costs to the participants. Follow-up rate is monitored throughout the trial. Patients who do not return follow-up questionnaires will receive/have received reminder telephone calls. Using strategies highly similar to these in our previous placebo-surgery controlled trial[33], a 99% follow-up rate was achieved.

The number and proportion of individuals eligible for and compliant with each follow-up was documented. Individuals who died during the study (from causes unrelated to the study or procedure) will be tabulated. An analysis of the demographic and prognostic characteristics will be carried out between the individuals who withdrew and those who remained in the study. For continuous variables, parametric or non-parametric analysis of variance will be used. For categorical variables, $\chi 2$ or Fisher's exact test will be applied.

Missing items

We will use multiple imputation to handle missing data for those statistical analyses that cannot handle occasional missing values. All variables which to be included in the final analyses will be included in the chained equations imputation model. The imputation algorithm, fully conditional specification (FCS), uses a specific univariate model for each variable and, for each specific imputed dataset, iteratively imputes each variable with missing values and uses the imputed values in the imputation of other variables.

Sample size

The sample size calculation was based on the two primary outcome measures, VAS at rest and at arm activity, at 24 months post randomization. FIMPACT trial was powered to detect a minimal clinically important improvement (MCII) in a VAS pain score (improvement of at least 15mm; assumed standard deviation 25 mm) between ASD and DA (or ET). To achieve a somewhat unconventional (stringent) 90% study power and using a two-sided Type I error rate (5%), our trial requires 68 patients per study group to show clinically meaningful advantage of ASD over DA (or ET). Acknowledging the stringent power threshold, we reserved only 3% surplus for potential loss to follow up/crossovers (3%), and accordingly, we set the recruitment target at 70 patients per treatment group.

Recruitment rate

A total of 210 patients were recruited between February 1, 2005 and August 6, 2013 from three tertiary referral centres. The recruitment rate was similar to our previous placebo-surgery controlled trial with similar, highly specific eligibility criteria (efficacy trial)[33].

Safety analysis

There are no anticipated safety issues with the FIMPACT Study. Identically to our previous placebo-surgery controlled trial[33], an interim analysis, as requested by the ethics board, was carried out after the enrolment of 45 participants by an independent data and safety monitoring board (the National Institute for Health and Welfare) to ensure that the rates of complications or reoperations were within acceptable limits (within the normal rate of complications and/or reoperations related to shoulder arthroscopy). As no marked discrepancy was found in the crude assessment of the incidence of complications/reoperations, no unsealing of group assignments (unblinding) was carried out. No other interim analysis was carried out.

Data management

Questionnaire forms on paper were the primary data collection tools for the study. Upon receipt of the questionnaire forms, a study nurse made a visual check of the responses and queried missing data when possible. Research assistants, blinded to the group allocation, stored the forms into an electronic database by double data entry to minimize typing errors. The researchers, blinded to the group allocation, are currently (July 2016) making a visual check of the data in the electronic database and will then query all missing, implausible, and inconsistent data. Patient records in the participating hospitals are also used when collecting missing data or interpreting inconsistent or implausible data. The final analysis will be performed on data transferred to the file "FIMPACT-full data_final", having been documented as meeting the cleaning and approval requirements of our independent statistician and after the finalisation and approval of the accompanying statistical analysis plan (SAP) document. Participant files will be maintained in storage (both in electronic and paper format) at the coordinating centre for a period of 10 years after completion of the study (10 year follow-up visits).

STATISTICAL METHODS

Statistical Analysis plan (SAP)

A statistical analysis plan (SAP) is published along this protocol. An independent statistician who is unaware of the group assignments will perform all the analyses.

We will summarise the baseline characteristics of the participants by group, reported as a mean (standard deviation) or median (first quartile, third quartile) for continuous variables and count (percent) for categorical variables.

We will analyse the data in a blinded manner. All p-values will be reported to 3 decimal places with those less than 0.001 reported as $p < 0.001$. The criterion for statistical significance will be set at alpha = 0.05.

Primary analysis

We will carry out the primary analysis according to the intention-to-treat (ITT) principle: participants are retained in the groups to which they were initially randomized.

The primary comparison will be on the efficacy of ASD (ASD vs. DA). We will perform the primary comparison on the efficacy of ASD (ASD vs. DA) as a between-group comparison using a repeated measures mixed-effects model (RMMM). Study group and time of assessment (baseline, 3, 6, 12 and 24 months) will be included as fixed factors and patient as a random factor. The model will include interactions between study group and time of assessment. The baseline value will be included as a covariate. The RMMM model will be used to quantify the treatment effect as the difference between the groups in pain scores (VAS) with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization. To safeguard against potential multiplicity bias[34], we will require a statistically significant treatment effect on both of our primary outcome variables, i.e., pain at rest and pain at activity.

The same statistical model will also apply to the pragmatic comparison of the relative benefits of surgical vs. non-operative treatment strategies on SIS (ASD vs. ET).

Secondary analyses

We will also use the RMMM model to analyse secondary outcomes where applicable. The results will be reported as the differences between the groups with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization.

We will also carry out a responder analysis, in which the proportions of patients reaching the patient-acceptable symptom state (PASS) and those ending up with a patient-disappointing symptom state (PDSS) will be determined. According to Tashjian et al.[35], a VAS score < 30 mm represent an appropriate cut-off for determining PASS in patients treated for rotator cuff disease. Accordingly, this threshold will be used for "responders" (VAS $\leq$ 30 mm). As regards a disappointing response to treatment, there exist – to our best knowledge - no criteria for PDSS in the context of subacromial pain syndrome. Therefore we will explore patient satisfaction with treatment, arm pain at rest and at activity, and night pain as the criteria for determining the PDSS, without *a priori* set cut-offs. Categorical variables, re-operations or treatment conversions, and complications as well as adverse effects will be analysed using logistic regression analysis or Poisson regression dependent on whether subjects with complications or (multiple) complications (per subject) are analysed.

These secondary analyses will be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not a problem[2]

Sensitivity analyses

We will carry out the following sensitivity analyses: 1) per-protocol analyses, in which the above noted primary and secondary analyses will be carried out again with patients who received the interventions as allocated; 2) and potential effects due to the treatment providing centres.

Subgroup analyses and Hypothesized Effects

We have identified three important subgroups. We will perform these three subgroup analyses with the primary endpoint as the outcome and the direction of hypothesized effect described[36]:

1) Duration of symptoms – Neer originally suggested that ASD should be considered for patients with persistent symptoms despite over one year of conservative treatment[37]. Recent RCTs failing to find efficacy on ASD (vs. conservative treatment) have prompted arguments that ASD should be reserved to situations when long-term conservative treatment has failed[38]. Although a recent study specifically addressed this question and failed to support this hypothesis[39], we still intend to com-

pare the treatment effects of participants stratified based on the duration of symptoms. Accordingly, we will compare those with symptoms less than 12 months to those with symptoms longer than 12 months. We hypothesize that subacromial decompression will work better in patients with duration of symptoms > 12 months than for patients with symptoms < 12 months.

2) Severity of symptoms - A subgroup analysis will also be conducted comparing the treatment effects in patients with severe (VAS 70 or more), moderate (VAS 55 to 69), and mild (VAS less than 55) symptoms at baseline. We hypothesize that subacromial decompression will work better in patients with more severe (VAS 70 or more) than moderate (VAS 55 to 69) or mild (VAS less than 55) symptoms at baseline.
3) Acromial anatomy - A hook-type acromion has been suggested as an independent risk factor for subacromial impingement[40]. To assess the validity of this suggestion, a subgroup analysis will be conducted comparing the treatment effects in patients with flat (type I), curved (type II), or hooked (type III) acromion according to classification by Bigliani et al.[41] We hypothesize that subacromial decompression will work better in patients with hooked (type III) than curved (type II) or flat (type I) acromion at baseline.

Effect modifying and mediating factors

Multiple regression models will be used to assess the potential effect modifying factors (e.g., age, gender, psychological well-being, mental health, occupational shoulder load, education level, and hand dominance) and effect mediating factors (e.g., absence of complications and adherence to rehabilitation) on pain, functional disability and quality of life. These analyses will be supportive, explanatory and/or hypothesis generating.

Blinded data interpretation

To safeguard against potential risk of bias during interpretation, we will use our recently introduced method of "blinded data interpretation"[42]. So far, this method has been successfully used at least on three previous occasions[33 43 44]. In brief, an independent statistician will provide the Writing committee of the FIMPACT trial (authors of this protocol) with blinded results from the analyses with study groups labelled as group A, group B, and group C. The Writing Committee will then contemplate on the interpretation of the results until a consensus is reached and agree in writing on all alternative interpretations of the findings. Once reaching a consensus, we will record the minutes of this meeting as a

statement of interpretation document signed by all members of the Writing Committee. Only after this common agreement will the data manager and independent statistician break the randomization code.

DISCUSSION

In this protocol paper, we describe the execution of a randomised, placebo-surgery controlled trial for the assessment of the efficacy of arthroscopic subacromial decompression (ASD) in patients with subacromial impingement syndrome (SIS). Acknowledging the potential of surgery to produce powerful placebo effects[45], a control group of diagnostic arthroscopy, differing from the ASD only by lacking the critical therapeutic element of the ASD (subacromial decompression), is used as the primary comparator. We will also conduct the pragmatic comparison of surgical and non-surgical treatment options of SIS by including a third group of progressive exercise therapy (ET) (Figure 1, ASD vs. ET).

Interpretations and generalizability

Our interpretation scheme rests on the primary tenet that the minimum requirement for the clinical viability of ASD is that it needs to show superiority to DA (a therapeutically inert and thus a clinically non-viable option). To test this, we have chosen a classic *efficacy* or *"can it work" (proof-of-concept)* design[46-48]: The recruited participants are those who - according to current evidence - should have an "optimal response" to ASD and the participants and outcome assessors are blinded to the interventions given. This design should thus yield findings that are widely applicable to patients with characteristic clinical signs and symptoms of SIS. We will also compare ASD with non-operative treatment option for SIS, the progressive ET, in a more pragmatic comparison, which is confounded by the lack of blinding of the participants. (Figure 2)

The generalizability of our primary (efficacy) comparison may be questioned as the patients are carefully selected (strict eligibility criteria) and treated by experienced shoulder surgeons. Nevertheless, the eligibility criteria are in agreement with the existing treatment guidelines on SIS [3]. The results should thus be applicable to the populations currently receiving treatment for their SIS. As for the skill-level of the surgeons, the index surgical procedure (ASD) is a relatively simple procedure and thus likely not very sensitive to surgeons' experience. For example, the amount of bone removed from the undersurface of the acromion seem to have at best a marginal effect on the outcome. Even bursectomy alone has been shown to produce the same therapeutic effect as standard acromioplasty[49].

Rationale for outcome assessment and statistical analysis

Traditionally, the assessment of the treatment effects of two or more interventions has relied primarily on the statistical significance of the mean differences of the intervention groups. However, as attentively described in a recent paper[50], to truly assess the clinical relevance of a treatment, one also needs information about the distribution of individual responses. In essence, one needs to look at how many people on treatment and on comparator group(s) had a response at least as great as the minimum (clinically) important difference (MCID). Such individuals have been described as "responders," and this approach of comparing treatment groups as a "responder analysis"[51 52]. The authors[50] suggested that "*Clinical trials should specify in their protocol that they will report the distribution of results in individual participants as well as the mean difference. Researchers should publish plots of individual results and responder analyses in clinical trial reports.*" The FIMPACT trial adheres to this suggested action. Accordingly, we will elaborate several relevant and often interrelated issues, such as the study power, the primary outcomes and their interpretation, the minimal clinically important difference (MCII), the patient-acceptable symptom state (PASS), and patient-disappointing symptoms state (PDSS).

Study power

Traditionally the sample size is calculated based on the minimal clinically important difference or change (MCID or MCII), i.e., the smallest change in measurement that signifies an important/detectable improvement in a patient's symptom(s). MCII/-D is not a static value even for one outcome instrument, but rather can have different values when assessed with different methods or *in different patient populations*. We chose VAS at rest and during arm activity as our primary outcomes, because shoulder pain is the primary complaint of patients with SIS. The FIMPACT trial was powered to detect an improvement of at least 15mm (on a 100 mm VAS scale)[35] between ASD and ET. This yielded a sample size estimate of 70 participants per group. To safeguard against lack of study power, we chose a statistical threshold of 90% instead of the more conventional 80%. In this context, Norman et al.[53] recently introduced a thought-provoking proposal by arguing that a standard ('off-the-peg') sample size of 64 per group would be just as valid an estimate as one obtains by more traditional ('made-to-measure') sample size calculations[53]. Finally, although the statistical power is a vital step in the *planning phase* of any clinical trial, the actual quality of evidence (certainty in the obtained estimates) can only be appropriately assessed from the confidence intervals (CI) of the data obtained[54].

Responder analysis

As noted above, instead of focusing only on the statistical significance of the mean differences between treatment groups in the VAS (i.e., the mean improvement from baseline to 24 months), we will also carry out "a responder analysis". In principle, this analysis allows physicians to inform a patient of his or her chance of experiencing a clinically meaningful improvement from the treatment, both in absolute terms and in comparison to a control group. The difference between responders and non-responders can be considered the net-benefit of the treatment. Responder analysis requires the assessment of the proportion of patients reaching the patient-acceptable symptom state (PASS) and the patient-disappointing symptoms state (PDSS). Tashjian et al. have recently proposed that a VAS score < 30 mm represent an appropriate cut-off for determining PASS in patients treated for rotator cuff disease and as VAS is also our primary outcome, we chose to use this threshold for "responders" (VAS ≤ 30 mm). Regarding the opposite, a disappointing response to treatment, we are not aware of any study defining the PDSS in the context of subacromial pain syndrome. Therefore we will plan to determine the PDSS by exploring patient satisfaction with treatment, arm pain at rest and at activity, and night pain.

Ethics of placebo surgery

Recent systematic review of the use of surgical placebo shows that in more than half of these studies the treatment group that included critical surgical/therapeutic element had no greater effect than a placebo group [17]. The review also showed that risks of adverse effects were small and the placebo group was safer than surgery under investigation. These findings make a compelling case for the use of surgical placebo controls when a placebo effect may be present. Regarding the ethics of surgical placebo controls, the authors of the review state "*Placebo controlled surgical trials raise important ethical concerns but are justified when there is a genuine equipoise; that is, a disagreement in the medical community about whether one treatment is superior to another, because standard treatment does not exist or its efficacy is questioned.*" They continue by concluding: "*Placebo controlled trials in surgery are as important as they are in medicine, and they are justified in the same way. They are powerful, feasible way of showing the efficacy of surgical procedures. They are necessary to protect the welfare of present and future patients as well as to conduct proper cost effectiveness analyses. Only then may publicly funded surgical interventions be distributed fairly and justly. Without such studies ineffective treatment may continue unchallenged.*" Our views regarding the ethics of using a surgical placebo group are perfectly aligned with these notions.

Limitations of the study

One possible confounder in our trial is that subacromial pain is also the hallmark symptom of a rotator cuff tear, although the latter patients usually also represent with muscle weakness. To exclude patients with a (clinically-relevant) rotator cuff tear, our eligibility screening included two preoperative assessments: (a) clinical exams targeted at finding obvious weakness of the rotator cuff muscles and (b) MRA, an imaging modality with a shown 92 specificity and 94 sensitivity for "full-thickness" RC tears[55]. In addition to these, we also carried out (c) a diagnostic arthroscopy in the ASD and DA groups prior to randomisation. Despite the thorough *preoperative* screening, 10% (14/136) of the participants allocated to the two surgical groups had to be excluded because of AC-arthrosis (n=1) or intra-articular pathology found at diagnostic arthroscopy (n=13). Although this does not have any effect on our primary comparison (ASD vs. DA), one could argue that the ET and operatively treated groups (ASD and DA) are not fully comparable. However, one should also recall that the clinical relevance of small RC tears or SLAP lesions, those not resulting in obvious muscle weakness and/or not apparent in MRA, is unknown. In the end, if this bias proves clinically relevant in our analysis, it will skew our results by favouring the ASD group in the pragmatic comparison (ASD vs. ET). Another concern related to the pragmatic comparison (ASD vs. ET) is that the progressive exercise therapy regimen carried out in the ET group is different from the postoperative rehabilitation carried out by patients in the ET group, for obvious reasons; surgically treated patients need time to recover from the initial surgical trauma. Furthermore, ASD patients are also subject to some degree of postoperative immobilization, extended sick leave, and modifications in pain medication and activities, all of which potentially have an effect on the outcome of treatment.

Another obvious concern related to our study design is the discrepant timing of the start of the actual treatment between the ET and the two surgical groups due to the time required to arrange the surgery. Acknowledging this, the two-year follow-up was chosen as our primary time point for assessing the benefits of treatment, as we assume that by this time the potential confounding effect of slightly different follow-up times should be diluted to a minimum. This is also the reason why we use the shorter-term follow-up visits data (follow-up visits performed at 3, 6 and 12 months after randomization) primarily to illustrate the trajectory of the treatment response in the three groups only. The same concern of varying time span from the randomization of the patients to the trial to the actual induction of treatment (due to delay in surgery) also applies to the CSAW trial[19]. To compensate for the waiting list effects, the CSAW investigators have chosen a slightly different strategy: Although the primary outcome assessment is performed at 6 months after randomization in CSAW trial, they have introduced addition-

al follow-up assessments, referenced from surgery, for patients waiting for longer than 4 months for their surgery after randomization. They have also set a secondary outcome measurement point at 1-year post randomization.

Author affiliations

Mika Paavola
Helsinki University Central Hospital,
Department of Orthopaedics and Traumatology,
Töölö hospital,
Topeliuksenkatu 5
P.O. Box 266, 00029 HUS,
Helsinki, Finland
Email: mika.paavola@hus.fi

Antti Malmivaara
National Institute for Health and Welfare
Center for Health and Social Economics
Mannerheimintie 166
00270 Helsinki
Finland
Email:antti.malmivaara@thl.fi

Simo Taimela
University of Helsinki,
Department of Orthopaedics and Traumatology,
Töölö hospital, Topeliuksenkatu 5,
00014 University of Helsinki, Helsinki, Finland
Email: simo.taimela@evalua.fi

Kari Kanto
Hatanpään sairaala, Tampere
Hatanpäänkatu 24, 33900 Tampere, Finland
Email: kari.kanto@fimnet.fi

Teppo L N Järvinen
University of Helsinki,
Department of Orthopaedics and Traumatology,
Töölö hospital, Topeliuksenkatu 5,
00014 University of Helsinki, Helsinki, Finland
Email: teppo.jarvinen@helsinki.fi

Acknowledgements

The FIMPACT investigators would like to thank Professor Jonas Ranstam (Lund University) for the statistical analysis plan and acknowledge Dr. Pekka Paavolainen and Professor Markku Järvinen for their critical contributions and support on the early phases of the planning and execution of the trial. Virginia Mattila is acknowledged for her linguistic expertise and language revisions.

Contributions

Concept and design: Mika Paavola (MP), Antti Malmivaara (AM), Simo Taimela (ST) and Teppo Järvinen (TJ).

Drafting and critical revision of the article for important intellectual content: MP, AM, ST, TJ and Kari Kanto (KK).

Final approval of the article: MP, AM, ST, TJ, KK.

Ensuring the accuracy of the work: MP, AM, ST, TJ, KK.

Statistical expertise: Jonas Ranstam (JR).

Obtaining of funding: TJ and Markku Järvinen (MJ).

Funding

The FIMPACT trial was supported by the Sigrid Juselius Foundation, the Competitive Research Fund of Pirkanmaa and Helsinki University Central Hospital Districts, the Academy of Finland, and the Jane and Aatos Erkko Foundation. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of this protocol.

Figure 1: Flowchart of the trial: enrolment, assigned intervention and follow-up scheme.

Figure 2: Study design and interpretation of results.

Table 1: Baseline characteristics.

Table 2: A diagram outlining the follow-up scheme used.

Appendix 1: FIMPACT investigators.

Appendix 2: Statistical analysis plan (SAP).

Appendix 3: Blinded data interpretation plan.

REFERENCES

1. Paloneva J, Lepola V, Karppinen J, et al. Declining incidence of acromioplasty in Finland. Acta Orthop 2015;**86**(2):220-4.
2. Vitale MA, Arons RR, Hurwitz S, et al. The rising incidence of acromioplasty. J Bone Joint Surg Am 2010;**92**(9):1842-50.
3. Diercks R, Bron C, Dorrestijn O, 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.
4. Oh LS, Wolf BR, Hall MP, et al. Indications for rotator cuff repair: a systematic review. Clinical orthopaedics and related research 2007;**455**:52-63.
5. Aurora A, McCarron J, Iannotti JP, et al. Commercially available extracellular matrix materials for rotator cuff repairs: state of the art and future trends. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2007;**16**(5 Suppl):S171-8.
6. Michener LA, McClure PW, Karduna AR. Anatomical and biomechanical mechanisms of subacromial impingement syndrome. Clinical biomechanics 2003;**18**(5):369-79.
7. Koester MC, George MS, Kuhn JE. Shoulder impingement syndrome. The American journal of medicine 2005;**118**(5):452-5.
8. Neer CS, 2nd. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. The Journal of bone and joint surgery American volume 1972;**54**(1):41-50.
9. Yu E, Cil A, Harmsen WS, et al. Arthroscopy and the dramatic increase in frequency of anterior acromioplasty from 1980 to 2005: an epidemiologic study. Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association 2010;**26**(9 Suppl):S142-7.
10. Judge A, Murphy RJ, Maxwell R, et al. 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.
11. Dong W, Goost H, Lin XB, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. Medicine 2015;**94**(10):e510.
12. Shi LL, Edwards TB. The role of acromioplasty for management of rotator cuff problems: where is the evidence? Advances in orthopedics 2012;**2012**:467571.
13. Haahr JP, Ostergaard S, Dalsgaard J, et al. Exercises versus arthroscopic decompression in patients with subacromial impingement: a randomised, controlled study in 90 cases with a one year follow up. Ann Rheum Dis 2005;**64**(5):760-4.
14. Ketola S, Lehtinen J, Rousi T, et al. No evidence of long-term benefits of arthroscopicacromioplasty in the treatment of shoulder impingement syndrome: Five-year results of a randomised controlled trial. Bone Joint Res 2013;**2**(7):132-9.
15. Brox JI, Staff PH, Ljunggren AE, et al. Arthroscopic surgery compared with supervised exercises in patients with rotator cuff disease (stage II impingement syndrome). Bmj 1993;**307**(6909):899-903.
16. Ernst E, Resch KL. Concept of true and perceived placebo effects. Bmj 1995;**311**(7004):551-3.
17. Wartolowska K, Judge A, Hopewell S, et al. Use of placebo controls in the evaluation of surgery: systematic review. Bmj 2014;**348**:g3253.
18. Buchbinder R. Meniscectomy in patients with knee osteoarthritis and a meniscal tear? N Engl J Med 2013;**368**(18):1740-1.
19. Beard D, Rees J, Rombach I, et al. The CSAW Study (Can Shoulder Arthroscopy Work?) - a placebo-controlled surgical intervention trial assessing the clinical and cost effectiveness of arthroscopic subacromial decompression for shoulder pain: study protocol for a randomised controlled trial. Trials 2015;**16**:210.
20. Ellman H. Arthroscopic subacromial decompression: analysis of one- to three-year results. Arthroscopy 1987;**3**(3):173-81.
21. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in

patients treated for rotator cuff disease. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2009;**18**(6):927-32.
22. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clinical orthopaedics and related research 1987(214):160-4.
23. Henseler JF, Kolk A, van der Zwaal P, et al. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2015;**24**(3):376-81.
24. Matsen FA, 3rd, Ziegler DW, DeBartolo SE. Patient self-assessment of health status and function in glenohumeral degenerative joint disease. J Shoulder Elbow Surg 1995;**4**(5):345-51.
25. Godfrey J, Hamman R, Lowenstein S, et al. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. J Shoulder Elbow Surg 2007;**16**(3):260-7.
26. Tashjian RZ, Deloach J, Green A, et al. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am 2010;**92**(2):296-303.
27. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001;**33**(5):328-36.
28. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res 2008;**17**(3):485-95.
29. Bowling A. *Measuring health. A review of quality of life measurement scales.* 3rd ed: Berkshire: Open University Press, 2004.
30. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;**30**(6):473-83.
31. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. Bmj 2002;**324**(7351):1417.
32. Hamilton DF, Lane JV, Gaston P, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. BMJ Open 2013;**3**(4).
33. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. N Engl J Med 2013;**369**(26):2515-24.
34. Ranstam J. Multiple P-values and Bonferroni correction. Osteoarthritis Cartilage 2016;**24**(5):763-4.
35. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. J Shoulder Elbow Surg 2009;**18**(6):927-32.
36. Sun X, Ioannidis JP, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. Jama 2014;**311**(4):405-11.
37. Neer CS, 2nd. Impingement lesions. Clin Orthop Relat Res 1983(173):70-7.
38. Dong W, Goost H, Lin XB, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. Medicine (Baltimore) 2015;**94**(10):e510.
39. Ketola S, Lehtinen J, Rousi T, et al. Which patients do not recover from shoulder impingement syndrome, either with operative treatment or with nonoperative treatment? Acta Orthop 2015;**86**(6):641-6.
40. Tangtrakulwanich B, Kapkird A. Analyses of possible risk factors for subacromial impingement syndrome. World J Orthop 2012;**3**(1):5-9.
41. Bigliani LU MD, April EW The morphology of the acromion and its relationship to rotator cuff tears. Orthop Trans 1986;**10**:228.
42. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol 2014;**67**(7):769-72.
43. Skou ST, Roos EM, Laursen MB, et al. A Randomized, Controlled Trial of Total Knee Replacement. N Engl J Med 2015;**373**(17):1597-606.
44. Investigators F, Bhandari M, Jeray KJ, et al. A Trial of Wound Irrigation in the Initial Management of Open Fracture Wounds. N Engl J Med 2015;**373**(27):2629-41.

45. Jonas WB, Crawford C, Colloca L, et al. To what extent are surgery and invasive procedures effective beyond a placebo response? A systematic review with meta-analysis of randomised, sham controlled trials. BMJ Open 2015;**5**(12):e009655.
46. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving. Bmj 1999;**319**(7211):652-3.
47. Djulbegovic B, Paul A. From efficacy to effectiveness in the face of uncertainty: indication creep and prevention creep. Jama 2011;**305**(19):2005-6.
48. Jarvinen TL, Sievanen H, Kannus P, et al. The true cost of pharmacological disease prevention. Bmj 2011;**342**:d2175.
49. Donigan JA, Wolf BR. Arthroscopic subacromial decompression: acromioplasty versus bursectomy alone--does it really matter? A systematic review. Iowa Orthop J 2011;**31**:121-6.
50. Cates C, Karner C. Clinical importance cannot be ruled out using mean difference alone. Bmj 2015;**351**:h5496.
51. Guyatt GH, Juniper EF, Walter SD, et al. Interpreting treatment effects in randomised trials. Bmj 1998;**316**(7132):690-3.
52. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. Trials 2007;**8**:31.
53. Norman G, Monteiro S, Salama S. Sample size calculations: should the emperor's clothes be off the peg or made to measure? BMJ 2012;**345**:e5278.
54. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med 1994;**121**(3):200-6.
55. Lenza M, Buchbinder R, Takwoingi Y, et al. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. Cochrane Database Syst Rev 2013;**9**:CD009020.

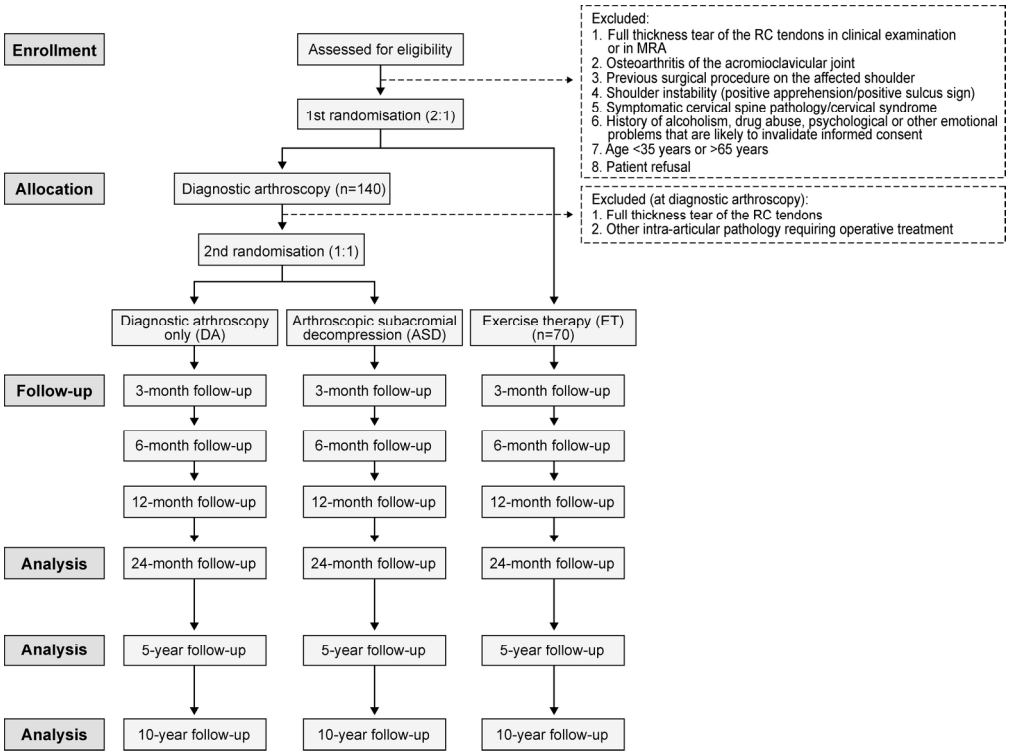


Figure 1.

179x133mm (300 x 300 DPI)

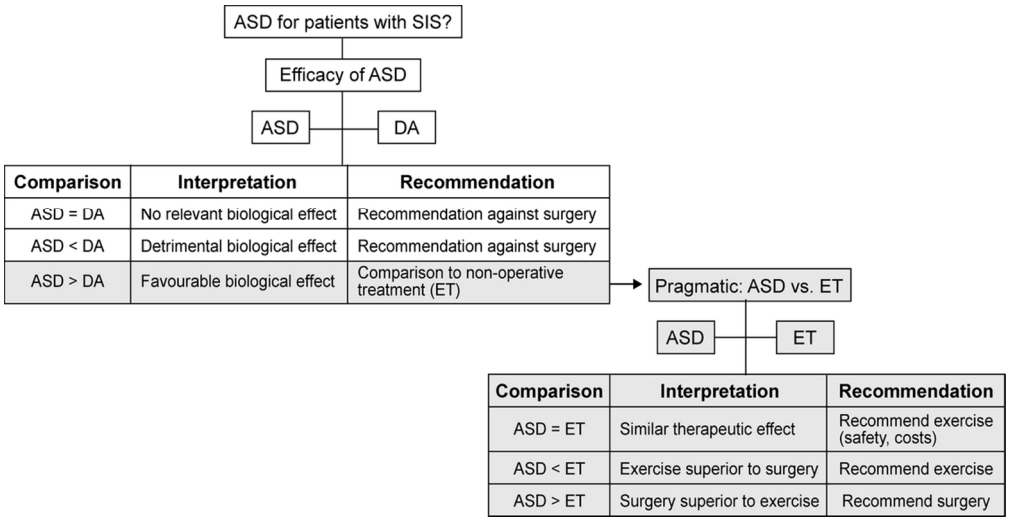


Figure 2.

107x54mm (300 x 300 DPI)

FIMPACT INVESTIGATORS

Steering Committee: Mika Paavola (Chair, Helsinki University Central Hospital), Teppo Järvinen (Co-Chair, University of Helsinki and Helsinki University Central Hospital), Antti Malmivaara (National Institute for Health and Welfare), Simo Taimela (University of Helsinki)

Writing Committee: Teppo Järvinen (Co-Chair, University of Helsinki and Helsinki University Central Hospital), Simo Taimela (Co-Chair, University of Helsinki), Antti Malmivaara (National Institute for Health and Welfare), Mika Paavola (Helsinki University Central Hospital), Kari Kanto (Hatanpää City Hospital, Tampere).

Participating Clinical Sites: *Helsinki University Central Hospital/Jorvi Hospital*: Kalevi Hietaniemi, Juha Kalske, Vesa Lepola, Jyrki Salmenkivi, Sikri Tukiainen. – *Helsinki University Central Hospital/Herttoniemi Hospital*: Jarkko Pajarinen, Mikko Salmela, Vesa Savolainen, Ilkka Sinisaari. – *Hatanpää City Hospital, Tampere*: Timo Järvelä, Kari Kanto, Janne Lehtinen, Mikael Salmela.

FIMPACT Methods Center: Leena Caravitis, Sari Karesvuori, Pirjo Toivonen (Project management), Mathias Bäck (Data management), Ville Haapamäki (Imaging) (Helsinki University Central Hospital and University of Helsinki); Jari Inkinen (Physiotherapy) (Kunnon Klinikka, Tampere), Esa Läärä (Randomization) (University of Oulu), Harri Sintonen (QoL outcomes) (National Institute for Health and Welfare).

Statistical Analysis Plan (SAP) for:

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT), 2-year follow-up

Simo Taimela[1] and Teppo L N Järvinen[1]

Statistical advisor:

Jonas Ranstam

[1] Department of Orthopedics and Traumatology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

Table of Contents

STUDY SYNOPSIS

Introduction: Arthroscopic subacromial decompression (ASD) is the most commonly performed surgical intervention for shoulder pain. Conventional wisdom dictates that subacromial pain syndrome, and particularly its' sub-type 'shoulder impingement syndrome' (SIS), is due to 'impingement' of the rotator cuff tendons on the overlying acromion while passing through the subacromial space. The rationale of the ASD procedure rests on the tenet that symptom relief can be achieved through removal of a bony acromial spur and the resulting decompression of the tendon passage. However, evidence on the efficacy of this procedure is limited. Acknowledging the potential placebo effect of surgery, the primary objective of this superiority trial is to compare the efficacy of ASD versus diagnostic arthroscopy (DA) in patients with SIS, the latter procedure differing from the former by only lacking subacromial decompression. As a non-surgical treatment option, a third group of supervised progressive exercise therapy (ET) is also included for allowing pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies.

Methods/Design: FIMPACT trial is an ongoing multicentre, three-group randomised controlled study with a primary objective to assess the efficacy of the ASD vs. DA. Our design also enables the pragmatic comparison of surgical and non-surgical treatment options (ASD vs. ET). Two-fold concealed allocation was performed, first by randomizing patients to surgical (ASD or DA) or conservative (ET) treatment groups in 2:1 ratio. Those allocated to surgery were then further randomized to ASD or DA groups in 1:1 ratio/fashion. The two primary outcome measures are pain at rest and activity assessed with 100 mm visual analog scales (VASs) at 2 years post randomization. Secondary outcome measures are functional assessment (Constant score and Simple shoulder test), global assessment of change, proportion of recovered patients, quality of life (15D), reoperations/treatment conversions, complications and adverse effects, all at 2 years. A total of 210 patients were recruited from 3 tertiary referral centres between February 1, 2005 and August 6, 2013. The study was powered to detect a difference of 15mm on the VAS scale (standard deviation 25mm, $\beta = 0.1$ and $\alpha = 0.05$) with 90% power and to allow for 3% loss to follow-up. The primary analysis will be conducted on the intention-to-treat analysis.

TRIAL REGISTRATION

ClinicalTrials.gov NCT00428870 (first registered January 29, 2007).

STUDY OBJECTIVES AND OUTCOMES

This statistical analysis plan (SAP) is accompanying the actual study protocol of the FIMPACT trial, a document that elaborates the methods used in detail. All outcomes were inquired from participants at baseline and follow-ups (6 and 24 months) and selected additional measures at 3 and 12 months (for details, see Table 1). The last patient reached the primary endpoint, the 24-month follow-up, in September 2015.

Table 1: Outcomes and follow-up time points

Assesment	Screening	Enrollment (Baseline)	Surgery	3 Months	6 Months	12 Months	24 Months	5 years	10 years
Screening form	X								
Informed consent		X							
Baseline characteristics form		X							
X-ray and MRI	X								X
Randomisation		X (1st)	X (2nd)						
Arthroscopic findings form			X						
Follow-up form*				X		X			
Clinical examination					X		X	X	X
Complications form**			(X)	(X)	(X)	(X)	(X)	(X)	(X)
VAS, at rest		X		X	X	X	X	X	X
VAS, at arm activity		X		X	X	X	X	X	X
Constant- Murley Score		X			X		X	X	X
Simple Shoulder Test (SST)		X			X		X	X	X
SF-36		X			X		X	X	X
15D		X		X	X	X	X	X	X
Patients satisfaction to the treatment				X	X	X	X	X	X
Patients assessment of the treatment allocation				X					
Health resource utilization				X	X	X	X	X	X

* Letter/telephone interview

** If required

DESCRIPTIVE OUTCOMES

At screening, the participants filled out a questionnaire to record gender, age, hand dominance, weight, height, level of education (socioeconomic status), workload (type of work), physical activity level, sports discipline, subjective health, symptoms (onset, frequency, and severity), use of pain medications, prior treatments, expectations to treatment, generic health state, and disease-specific scores. To exclude patients with concomitant shoulder pathology (particularly rotator cuff rupture), magnetic resonance imaging with contrast (MRA) was acquired for each participant.

OBJECTIVES AND PRIMARY OUTCOME

The primary objective of this trial is to compare the efficacy of arthroscopic subacromial decompression (ASD) versus diagnostic arthroscopy (DA) in patients with SIS. The trial is designed as a superiority trial, i.e. we expected in the power calculation that the ASD will result in greater pain relief at 24-month follow-up than DA (or ET). The 24-month follow-up was chosen as the primary endpoint, since this time point is a commonly held "minimal requirement" for any procedure in the field (orthopaedics) and most commonly used in the trials assessing the treatment of SIS.

The primary hypothesis: The primary hypothesis of our FIMPACT trial is that ASD is superior to DA in patients with SIS.

To enable pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies on SIS, a non-surgical (third) treatment option of supervised progressive exercise therapy (ET) is also included (ASD vs. ET).

Additional hypothesis: The relative benefits of ASD and ET will be assessed without a priori hypothesis on the superiority of one or the other.

As the primary outcome measure, a visual analogue scale (0-100 mm) was used to measure the patient's perceived pain intensity at rest and at arm activity during the 24 hours preceding the assessment. The minimal clinically important difference (MCID) for VAS was considered 15 mm on a 100 mm VAS scale and the patient acceptable symptom state (PASS), the score below which patients consider themselves well, was considered 30 mm.[1]

SECONDARY OUTCOMES

Our secondary outcome measures are listed below. These outcomes will only be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not considered to be a problem[2].

Constant-Murley score

Constant-Murley score (CS) is the most commonly used scoring system for evaluation of various disorders of the shoulder [3]. It consists of both objective (range of motion and strength) and subjective measurements (pain assessment, work load, and leisure time activities), which are summarized in a score between 0 and 100. A higher score indicates better shoulder function. The minimal detectable change (MDC) of the Constant score is 17 for patients with SIS[4]

SST

The simple shoulder test (SST) was developed to assess the functional limitations of the patient's activities of daily living[5]. The SST consists of 12 questions with yes (1) or no (0) response options. The maximum SST score is 12 indicating normal shoulder function, minimum score of 0 points refers severely diminished shoulder function. The SST has good reliability and responsiveness in patients with rotator cuff symptoms[6]. The MCID for the SST in rotator cuff disease is 2 points[7].

15D

The 15D instrument is a generic health-related quality of life (HRQoL) instrument comprising 15 dimensions [8]. For each dimension, the respondent must choose one of the five levels that best describes his/her state of heal that the moment (the best level being 1 and the worst level being 5). A set of utility or preference weights is used in an addition aggregate formula to generate a single index number, the utility or 15D score. The maximum 15D score is 1 (no problems on any dimension) and the minimum score is 0 (being dead). The responsiveness, reliability and validity of 15D have been thoroughly established, and this instrument has been used extensively in clinical and healthcare research[9 10].

SF-36

Short form or SF-36 is a generic HRQoL instrument to quantify the physical, functional, and psychological aspects of health related quality of life. It consists of 36 questions in eight subscales that assess physical, functional, social, and psychological well-being [11]. Score ranges from 0 to 100, a higher score is associated with better health. The physical and mental component summary scales (PCS and MCS, respectively) are then calculated as composites of the related subscales. SF-36 is one of most widely used measure of health-related quality of life[12].

Patient satisfaction

Patients' global assessment of satisfaction was elicited using the following question: "How satisfied are you with the treatment given?" on a 5-item scale at each follow-up timepoint (Table 1). As before[13], the responses "Very satisfied", "Satisfied" and "Somewhat satisfied" were categorized as satisfied, while responses, "Dissatisfied" and "Very dissatisfied" were categorized as dissatisfied.

Return to previous leisure activities

Similarly, at each follow-up (Table 1), participants responded to the following question: "Have you been

able to return to their previous leisure activities?" ("yes" or "no").

Patients' perception of operative treatment-group assignment

At the 3-month follow-up point, the patients in the two operative groups were asked to guess whether they had undergone ASD or DA.

Health resource utilization and costs

For the cost-effectiveness analysis, at each follow-up visit the participants were asked to fill in a questionnaire inquiring the use of healthcare resources. The questionnaire contains a list of items of healthcare resources available and the participants were asked to fill in the number of visits per item during the recall period of each follow-up time point. The resource use will be calculated based on the number of visits times unit cost per item and expressed as mean costs by items of resource use, and the mean direct total health care resource costs. All costs will be discounted to the 2016 price level.

Time to return to work

Information about return to work was recorded at each follow-up time point (Table 1).

Complications and adverse events

The participants were encouraged to contact the participating hospitals if any adverse events occurred and contacts to the health care system were monitored at every follow-up visit. Potential adverse events (AE) were categorized to serious adverse events (SAE) and minor adverse events (MAE). Death, cardiovascular or gastrointestinal events, deep venous thrombosis, pulmonary embolism, systemic or local infection were categorised as SAEs. Shoulder symptoms like pain, swelling and decreased range of motion were categorised as MAEs if the participants sought treatment. Data on complications and adverse events were recorded and their severity and frequency will be assessed.

EXPLORATORY OUTCOMES

We have identified three potentially important effect modifying factors. We will perform subgroup analyses with the primary endpoint as the outcome and the direction of hypothesized effect described as below[14].

Duration of symptoms

We will compare the treatment effects stratified based on the duration of symptoms (those with < 6/12 months vs. those > 6/12 months). We hypothesize that subacromial decompression will work better in patients with duration of symptoms > 6 months than for patients with symptoms < 6 months.

Severity of symptoms

We will compare the treatment effects in patients with severe (VAS 70 or more), moderate (VAS 55 to 69), and mild (VAS less than 55) symptoms at baseline. We hypothesize that subacromial decompression will work better in patients with more severe (VAS 70 or more) than moderate (VAS 55 to 69) or mild (VAS less than 55) symptoms at baseline.

Acromial anatomy

We will compare the treatment effects in patients with flat (type I), curved (type II), or hooked (type III) acromion according to classification by Bigliani et al.[15] We hypothesize that subacromial decompression will work better in patients with hooked (type III) than curved (type II) or flat (type I) acromion at baseline.

STUDY DESIGN

Sample size

The sample size calculation was based on the two primary outcome measures, VAS at rest and at arm activity, at 24 months post randomization. FIMPACT trial was powered to detect a minimal clinically important improvement (MCII) in a VAS pain score (improvement of at least 15mm; assumed standard deviation 25 mm) between ASD and DA (or ET). To achieve a somewhat unconventional (stringent) 90% study power and using a two-sided Type I error rate (5%), our trial requires 68 patients per study group to show clinically meaningful advantage of ASD over DA (or ET). Acknowledging the stringent power threshold, only 3% surplus was reserved for potential loss to follow up/crossovers (3%), and accordingly, the recruitment target was set at 70 patients per treatment group.

Randomization and blinding

A two-phase sequential randomization was used. In the Phase I, the participants were randomized into non-surgical or surgical treatment with allocation ratio 1:2. In the Phase II, those allocated to surgical treatment were further randomized to ASD or DA with allocation ratio 1:1. An independent statistician with no clinical involvement in the execution of the trial prepared separate randomization lists for each study centre using a computer-generated schedule. Randomization was carried out using sequentially numbered sealed opaque envelopes. The envelopes were kept in a secure, agreed location at each centre. To ensure concealment, block randomization was applied using blocks varying in size randomly (block size known only by the independent statistician).

To initially enter a participant into the study (Phase I), an envelope containing the treatment assignment (non-surgical (ET) or surgery (ASD or DA), ratio 1:2) was opened during the baseline appointment. Participants randomized for ET started standardized physiotherapy within 2 weeks of the baseline appointment. Participants allocated for surgical treatment were scheduled for surgery with the aim to carry out the procedure within 12 weeks of randomization.

At the day of surgery, an arthroscopic examination was first carried out to confirm the eligibility of the participant (to rule out full-thickness RC tear and other obvious intra-articular pathology). Research/staff nurse then completed the randomization procedure (Phase II) by opening an envelope containing the surgical treatment allocation (ASD or DA, ratio 1:1). The allocation was revealed to the surgeon by showing the paper, but not expressed verbally.

The full follow-up process is shown in Table 1. In brief, the participants filled in the above noted (mailed) outcome questionnaires at 3, 6, 12 and 24 months post randomization, in addition to which they were also assessed clinically at 6 and 24 months post randomisation (and will be assessed at 5 and 10 years) by a study physiotherapist unaware of treatment allocation.

Data analysis will be done in a blinded manner by the study statistician (JR) not directly involved in the study.

STUDY POPULATION

Subject disposition

Study procedures, including recruitment strategies and inclusion and exclusion criteria, are presented in detail in the accompanying actual study protocol.

STATISTICAL ANALYSIS

Data will be analysed in a blinded manner. All p-values will be reported to 3 decimal places with those less than 0.001 reported as $p < 0.001$. The criterion for statistical significance will be set at alpha = 0.05.

Primary analysis

The primary analysis will be carried out according to the intention-to-treat (ITT) principle: participants are retained in the groups to which they were initially randomized. The primary comparison on the efficacy of ASD (ASD vs. DA) will be performed as a between-group comparison using a repeated measures mixed-effects model (RMMM). Study group and time of assessment (baseline, 3, 6, 12 and 24 months) will be included as fixed factors and patient as a random factor. The model will include interactions between study group and time of assessment. The baseline value will be included as a covariate. An unstructured covariance structure will be assumed. If the model cannot be fitted, compound symmetry will be assumed instead. The number of degrees of freedom will be assessed using Satterthwaite's method. The RMMM model will be used to quantify the treatment effect as the difference between the groups in pain scores (VAS) with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization. To safeguard against potential multiplicity bias[2], we will require a statistically significant treatment effect on both of our primary outcome variables, i.e., pain at rest and pain at activity (Table 2).

Table 2: Primary comparison ASD vs ET: Outcomes of the trial at 24 months follow-up.

	ASD ET	**Improvement from baseline**	**Between-Group Difference in Improvement from Baseline**
Primary outcomes		**ASD ET**	
VAS (rest)			
VAS (at arm activity)			
Secondary outcomes			
Constant-Murley Score			
SST			

SF-36			
15D			
Patients' satisfaction to the treatment			
Return to previous leisure activities			
Health resource utilization and costs			
Patients' assessment of the treatment allocation			
Time to return to work			

Abbreviations: VAS, visual analogue scale (0-100 mm); SST, Simple Shoulder Test; SF-36, Short form- 36

The same statistical model will also apply to the pragmatic comparison of the relative benefits of surgical vs. non-operative treatment strategies on SIS (ASD vs. ET) (Table 3).

Table 3. Secondary comparison ASD vs ET: Outcomes of the trial at 24 months follow-up.

	ASD	**ET**	**Improvement from baseline**		**Between-Group Difference in Improvement from Baseline**
Primary outcomes			**ASD**	**ET**	
VAS (rest)					
VAS (at arm activity)					
Secondary outcomes					
Constant-Murley Score					
SST					
SF-36					
15D					
Patients' satisfaction to the treatment					
Return to previous leisure activities					
Health resource utilization and costs					

Time to return to work			

Abbreviations: VAS, visual analogue scale (0-100 mm); SST, Simple Shoulder Test; SF-36, Short form- 36

Secondary analyses

The RMMM model will also be used to analyse secondary outcomes (Table 2 and 3) where applicable. The results will be reported as the differences between the groups with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization.

We will also carry out a responder analysis, in which the proportion of patients reaching the patient-acceptable symptom state (PASS) will be determined. According to Tashjian et al.1, a VAS score < 30 mm represent an appropriate cut-off for determining PASS in patients treated for rotator cuff disease. Accordingly, this threshold will be used for "responders" (VAS < 30 mm). As regards a disappointing response to treatment, there exist – to our best knowledge - no criteria for PDSS in the context of subacromial pain syndrome. Therefore we will explore patient satisfaction with treatment, arm pain at rest and at activity, and night pain as the criteria for determining the PDSS, without *a priori* set cut-offs.

Categorical variables, the rates of unblinding, reoperation, treatment conversion, complications and adverse effects will be analysed using logistic regression analysis or Poisson regression dependent on whether subjects with complications or (multiple) complications (per subject) are analysed.

These secondary analyses will be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not a problem[2].

Sensitivity analyses

The following two sensitivity analyses will be carried out: 1) per-protocol analysis, in which the above noted primary analyses will be carried out again with patients who received the interventions as allocated will be redone; 2) and potential effects due to the treatment providing centres.

As all the participants in the ASD group have received the critical therapeutic element (subacromial decompression), no treatment group conversion is possible in this group.

In the per-protocol comparison of the efficacy of ASD (ASD vs. DA), we define the DA per-protocol population as those participants who have not received ASD during the 24-month follow-up (who have not crossed over to ASD).

In the per-protocol comparison of the effectiveness of ASD (ASD vs. ET), we define the ET per-protocol population as those participants who have not received ASD during the 24-month follow-up

(who have not crossed over to ASD).

INTERPRETATION OF RESULTS

To safeguard against potential risk of bias during interpretation, a method of "blinded data interpretation" will be used[17]. In brief, an independent statistician will provide the Steering/Writing committee of the FIMPACT trial with blinded results from the analyses with study groups labelled as group A, group B, and group C. This data will be presented to the Steering/Writing Committee, who will then contemplate on the interpretation of the results until a consensus is reached and agree in writing on all alternative interpretations of the findings. Once reaching a consensus, the minutes of this meeting are recorded as a statement of interpretation document signed by all members of the Steering/Writing Committee. Only after this common agreement will the data manager (independent statistician) break the randomization code.

There was also variation in the actual execution of the follow-up assessments, particularly in the earlier time-points (3- and 6-month follow-up visits).

IMPLEMENTATION OF ANALYSIS PLAN

This SAP will be used as a work description for the statistician performing the analyses. All analyses will be performed by the same statistician and none of the investigators involved in this trial will perform any of the statistical analyses.

The implementation of the SAP will be as follows:

1. A 'data collection form' will be outlined in a collaboration between the database manager (Leena Caravitis), statistician and principal investigators (Mika Paavola and Teppo Järvinen).
2. The database manager will code each treatment arm into 'treatment A', 'treatment B' and 'treatment C', thus leaving all others blinded to group assignment during the analyses.
3. Blinded data will be delivered to the statistician according to the 'data collection form'.
4. Primary, secondary and exploratory endpoint analyses will be made blinded to group assignment.
5. Results will be presented to the trial Writing and Steering committee, any uncertainties will be clarified and blinded interpretations of the primary endpoint results will be conducted prior to unblinding of data.

REFERENCES

1. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2009;**18**(6):927-32.
2. Ranstam J. Multiple P-values and Bonferroni correction. Osteoarthritis Cartilage 2016;**24**(5):763-4.
3. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clinical orthopaedics and related research 1987(214):160-4.
4. Henseler JF, Kolk A, van der Zwaal P, et al. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2015;**24**(3):376-81.
5. Matsen FA, 3rd, Ziegler DW, DeBartolo SE. Patient self-assessment of health status and function in glenohumeral degenerative joint disease. J Shoulder Elbow Surg 1995;**4**(5):345-51.
6. Godfrey J, Hamman R, Lowenstein S, et al. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. J Shoulder Elbow Surg 2007;**16**(3):260-7.
7. Tashjian RZ, Deloach J, Green A, et al. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am 2010;**92**(2):296-303.
8. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001;**33**(5):328-36.
9. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res 2008;**17**(3):485-95.
10. Bowling A. *Measuring health. A review of quality of life measurement scales.* 3rd ed: Berkshire: Open University Press, 2004.
11. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;**30**(6):473-83.
12. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. Bmj 2002;**324**(7351):1417.
13. Hamilton DF, Lane JV, Gaston P, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. BMJ Open 2013;**3**(4).
14. Sun X, Ioannidis JP, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. Jama 2014;**311**(4):405-11.
15. Bigliani LU MD, April EW The morphology of the acromion and its relationship to rotator cuff tears. Orthop Trans 1986;**10**:228.
16. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc., 1987.
17. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol 2014;**67**(7):769-72.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Interpretation of Blinded Data, Statement of Interpretation

Background assumptions regarding our primary comparison (ASD vs. DA)

1) This superiority RCT is designed to address the true *efficacy* of arthroscopic subacromial decompression (ASD), i.e., can ASD theoretically work? Accordingly, we have chosen patients that – based on the existing literature – represent optimal responders to this index surgical procedure.
2) Conceding that the act of surgery *per se* produces a profound placebo response, a 'true' treatment effect is impossible to disentangle from the nonspecific (placebo or meaning) effects – such as the patients' or researchers' expectations of benefit – without a placebo comparison group.
3) The only difference between ASD and DA treatment groups is that the subacromial decompression, the critical therapeutic (surgical) element, has been carried out for patients in the ASD group.
    a. The critical therapeutic (surgical) element is the component of the surgical procedure that is believed to provide the therapeutic effect (here, subacromial decompression), being distinct from aspects of the procedures that are diagnostic or required to access the disease being treated (here, shoulder arthroscopy).
    b. Apart from the critical therapeutic element, the treatment of the ASD and DA groups is identical, i.e., all "placebo or meaning effect" related to the entire treatment and care is identical.
4) To be deemed effective, ASD should provide a statistically significant benefit over DA in both of the two primary outcomes, pain at rest and activity assessed with a 100 mm visual analog scale (VAS), as determined by the mean VAS difference between the groups. This is to safeguard against potential multiplicity bias[2].

If ASD is found effective (see above), it should also provide a clinically relevant benefit over DA according to following rationale:

1) There is a proven benefit as follows: Mean VAS-difference between ASD and DA shall exceed the threshold for the minimal clinically important difference (MCID) in VAS. We will consider 15 mm on a 100 mm VAS scale as the threshold for the minimal clinically important difference (MCID).

AND,

2) There is NO proven harm. If there is a proven benefit of ASD but significantly higher proportion of patients show adverse effects, the amount of benefits will be discussed in relation to the frequency and seriousness of the adverse effects.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Statistical commitments:

a) I-T-T is the primary data analysis, but per-protocol analysis will also be carried out.
b) The pre-specified time point of primary interest is 24 months after randomisation.
c) In addition to the two primary outcome parameters, we will also take into account the number of treatment conversions and re-operations, the incidence and seriousness of adverse effects between the ASD and DA groups, and the responder analysis.

Based on these theoretical commitments, our interpretation of the findings will be as follows:

a) If ASD is found superior to DA, the critical therapeutic element of the ASD procedure (subacromial decompression) has a clinically relevant effect on patients with symptoms consistent with SIS.
b) If ASD is not found superior to DA, the critical therapeutic element of the ASD procedure (subacromial decompression) does not have a clinically relevant effect on patients with symptoms consistent with SIS. Considering our efficacy design (study participants are 'optimal responders to ASD' and the surgeons are highly experienced), such finding would imply that ASD does not work at all.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Background assumptions regarding our secondary (independent) comparison (ASD vs. ET)

1) This pragmatic comparison is designed to address whether arthroscopic subacromial decompression (ASD) followed by postoperative rehabilitation is superior to supervised progressive exercise therapy (ET). We recognize that in this pragmatic comparison (ASD vs. ET) the supervised progressive exercise therapy regimen carried out in the ET group is different from the postoperative rehabilitation carried out by patients in the ASD group. In addition, the timing of the start of the actual treatment between the ET and ASD groups was somewhat discrepant due to the time required to arrange the surgery. The ASD patients are also subject to some degree of postoperative immobilization, sick-leave, and modification of pain medication and activities, unlike the patients in the ET group, all of which may also have an effect on the treatment outcome. However, these concord with the current best practice recommendations and the two-year follow-up chosen as our primary time point should dilute the effects of somewhat discrepant timing of the interventions.
2) To be deemed effective, either ASD or ET should provide a statistically significant benefit over ET or ASD, respectively, in both of our two primary outcomes, pain at rest and activity assessed with a 100 mm visual analog scale (VAS), as determined by the mean VAS difference between the two treatment groups. This is to safeguard against potential multiplicity bias[2].
3) The following concern (apparent confounding) needs to be taken into account in the interpretation. Despite the thorough *preoperative* screening, 10% (14/136) allocated to the two surgical groups had to be excluded because of pathology found after the 1st random allocation. Although this does not have any effect on our primary comparison (ASD vs. DA), the ET and ASD groups are not fully comparable. This discrepancy will possibly skew our results by favouring the ASD group.

Acknowledging all this, if ASD (or ET) is found effective (statistically significant difference in both primary outcomes), it should also provide a clinically relevant benefit over ET (or ASD) according to following rationale:

1) There is a proven benefit as follows: Mean VAS-difference between ASD and ET shall exceed the threshold for the minimal clinically important difference (MCID) in VAS. We will consider 15 mm on a 100 mm VAS scale as the threshold for the minimal clinically important difference (MCID).

AND,

2) There is NO proven harm. If there is a proven benefit of ASD (or ET) but significantly higher proportion of patients show adverse effects, the amount of benefits will be discussed in relation to the frequency and seriousness of the adverse effects.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Statistical commitments:

a) I-T-T is the primary data analysis, but per-protocol analysis will also be carried out.
b) The pre-specified time point of primary interest is 24 months after randomisation.
c) In addition to the two primary outcome parameters, we will also take into account the number of treatment conversions and re-operations, the incidence and seriousness of adverse effects between the ASD and ET groups, and the responder analysis.
d) Given the Background assumption 3.a) (above, the discrepant number of patients with a shoulder pathology other than SIS), we will carry out a worst-case analysis by creating a subgroup of the ET group by removing seven (an equal number of patients excluded from both surgical treatment arms due to pathology found after 1st randomization) worst-cases/highest VAS-pain scores at the primary analysis time-point (24 months). The number of removed cases is based on the assumption that the prevalence of shoulder pathology is identical in the randomized population, while the decision to remove the individual with the highest VAS-pain scores at the end of the study basis on the assumption that shoulder pathology is an effect-modifying factor, predicting poor outcome.

Based on these theoretical commitments, our interpretation of the findings will be as follows:

a) If ASD is found superior to ET in both the complete case and the sensitivity (subgroup) analyses, ASD is a more effective treatment option than ET for patients with SIS.
b) If ET is found superior to ASD in both the complete case and the sensitivity (subgroup) analyses, our results suggests that ET is a more effective treatment option than ASD for patients with subacromial pain syndrome.
c) If there are no statistically significant differences between ASD and ET, ASD and ET are equally effective.

BMJ Open

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT): A protocol for a randomized trial comparing arthroscopic subacromial decompression and diagnostic arthroscopy (placebo control), with an exercise therapy control, in the treatment of shoulder impingement syndrome

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014087.R1
Article Type:	Protocol
Date Submitted by the Author:	05-Dec-2016
Complete List of Authors:	Paavola, Mika; Helsinki University Central Hospital, Orthopaedics and traumatology Malmivaara, Antti; National Institute for Health and Welfare, Centre for Health and Social Economics Taimela, Simo; University of Helsinki, Department of Orthopedics and Traumatology Kanto, Kari; Hatanpaan kantasairaala Jarvinen, Teppo; University of Helsinki, Department of Orthopaedics and Traumatology
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Evidence based practice, Rehabilitation medicine, Sports and exercise medicine, General practice / Family practice
Keywords:	Orthopaedic & trauma surgery < SURGERY, Clinical trials < THERAPEUTICS, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, REHABILITATION MEDICINE, SPORTS MEDICINE

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT): A protocol for a randomized trial comparing arthroscopic subacromial decompression and diagnostic arthroscopy (placebo control), with an exercise therapy control, in the treatment of shoulder impingement syndrome

Mika Paavola[1], Antti Malmivaara[2], Simo Taimela[1], Kari Kanto[3], and Teppo L N Järvinen[1], for the FIMPACT investigators*

[1] Department of Orthopedics and Traumatology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

[2] National Institute for Health and Welfare, Center for Health and Social Economics, Helsinki, Finland

[3] Department of Orthopedics and Traumatology, Hatanpää City Hospital, Tampere, Finland

Address of corresponding author:
Mika Paavola, MD, PhD
Department of Orthopedics and Traumatology
Helsinki University Hospital / Töölö Hospital
Topeliuksenkatu 5
P.O. Box 266, 00029 HUS,
Helsinki, Finland
Phone: +358-50-4272481
Email: mika.paavola@hus.fi

* FIMPACT investigators, see appendix 1.

ABSTRACT

Introduction: Arthroscopic subacromial decompression (ASD) is the most commonly performed surgical intervention for shoulder pain, yet evidence on its efficacy is limited. The rationale for the surgery rests on the tenet that symptom relief is achieved through removal of a bony acromial spur and the resulting decompression of the tendon passage. Acknowledging the potential placebo effect of surgery, the primary objective of this superiority trial is to compare the efficacy of ASD versus diagnostic arthroscopy (DA) in patients with shoulder impingement syndrome (SIS), where DA differs only by the lack of subacromial decompression. As a non-surgical treatment option, a third group of supervised progressive exercise therapy (ET) will allow for pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies.

Methods and Analysis: FIMPACT trial is an ongoing multicentre, three-group randomised controlled study with a primary objective of assessing the efficacy of the ASD vs. DA. We performed two-fold concealed allocation, first by randomizing patients to surgical (ASD or DA) or conservative (ET) treatment in 2:1 ratio and then those allocated to surgery further to ASD or DA in 1:1 ratio. Our two primary outcomes are pain at rest and arm activity assessed with visual analog scale (VAS), while the secondary outcomes are functional assessment (Constant score and Simple shoulder test), quality of life (15D and SF-36), patient satisfaction, proportions of responders and non-responders, reoperations/treatment conversions, all at 2 years post-randomization, as well as adverse effects and complications. We recruited a total of 210 patients from 3 tertiary referral centres. We will conduct the primary analysis on the intention-to-treat basis.

Ethics and Dissemination: The study was approved by the institutional review board of the Pirkanmaa Hospital District and duly registered at ClinicalTrials.gov. The findings of this study will be disseminated widely through peer-reviewed publications and conference presentations.

Trial registration: ClinicalTrials.gov NCT00428870 (first registered January 29, 2007).

Keywords: Acromion; Acromioplasty; Arthroscopy; Impingement; Physiotherapy; Placebo; Sham; Shoulder; Syndrome; Randomised; Trial

Strengths of this study

- Efficacy design: Strict eligibility criteria
- Placebo-surgery controlled trial: Blinding of both the participants and the outcome assessors in the comparison between index surgery and control (placebo surgery)
- Inclusion of a non-surgical treatment option to allow a pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies

Limitations of this study

- Potential confounding due to participants' knowledge of the treatment delivered in our secondary comparison between surgical and non-operative treatments

INTRODUCTION

Subacromial decompression is one of the most frequently performed procedures in orthopaedics[1 2]. It is carried out to treat patients with shoulder pain attributed to "subacromial impingement syndrome" (SIS). Conventional wisdom dictates that SIS is caused by 'impingement' of the rotator cuff (RC) between the humeral head and the overlying acromion while lifting the arm. The appropriateness of this mechanistic explanation has been challenged lately where the generic label of "subacromial pain syndrome" (SAPS) is currently advocated[3 4]. The aim of the subacromial decompression procedure, typically carried out arthroscopically, is to decompress the RC tendon passage through the subacromial space through resection and smoothening of the hypertrophied or prominent anterolateral undersurface of the acromion. Management of shoulder pain has been estimated to account for 4.5 million visits annually to physicians in the USA alone[5], accounting for US$3 billion in costs each year[6]. Since 44-65% of all shoulder complains are related to SIS, it is estimated that annual direct medical costs of SIS are over $1 billion in the USA[7 8].

Since the introduction of subacromial decompression surgery in the early 1970s[9], the number of procedures has steadily increased across the entire western world. With the advent of arthroscopy, the number of these surgeries has increased dramatically -- 5-fold from the 1980s to 2005 in the US[10] and 700% between 2000 and 2010 in the UK[11]. Remarkably, there is a stark absence of evidence from high-quality controlled trials to support the existing practice of performing subacromial decompression for patients with SIS. Two recent systematic reviews concluded that subacromial decompression provides no superior benefits in terms of pain relief, function, or quality of life compared to non-surgical treatment[12 13]. There is even a placebo controlled trial to show the beneficial effect of exercise therapy over placebo physiotherapy[14]. However, the proponents of the procedure have argued that the evidence is skewed in favour of the therapeutic potential of surgery due to a significant cross-over (5-15%) from conservative treatment to surgery[14-16]. Although such concern is obviously warranted, it should also be recalled that surgeons' own perceptions on the success of any surgery might similarly be biased due to a considerable surgical placebo effect.

The outcome of any medical (surgical) intervention – particularly when treating primarily subjective symptoms – is a cumulative effect of three main elements: placebo effects, critical therapeutic (surgical) element, and non-specific effects, most importantly, the normal variation in the course of the disease and the regression-to-the-mean phenomenon[17 18]. Conceding that the act of surgery *per se* produces a profound placebo response, a 'true' treatment effect is impossible to disentangle from the nonspecific (placebo) effects – such as the patients' or researchers' expectations of benefit – without a placebo

comparison group[19]. The critical therapeutic element is the component of the surgical procedure that is believed to provide the therapeutic effect (here, subacromial decompression), which are distinct from aspects of the procedures that are diagnostic or required to access the disease being treated (here, shoulder arthroscopy).

To the best of our knowledge, there is only one other ongoing study aiming to assess the true efficacy of subacromial decompression surgery in patients with SIS using a placebo controlled study design. According to the published protocol of this CSAW trial[20], the investigators have chosen a very similar approach to that of our FIMPACT trial. In brief, the CSAW trial is a three-group pragmatic RCT comparing arthroscopic acromioplasty, active monitoring with specialist reassessment, and investigational shoulder arthroscopy only. CSAW aims for recruitment of 300 patients with SIS to assess the efficacy of the surgery against no surgery, the need for a specific component of the surgery (acromioplasty), and the quantification of the possible placebo effect. As readily apparent, the two trials (FIMPACT vs. CSAW) are very similar in design with the only notable differences being the primary outcome measure (Pain at rest and after activity vs. Oxford Shoulder Score, a score that assesses both pain and ADL impairment), the primary outcome assessment point (24 months vs. 6 months), and the intervention delivered for the third group (exercise therapy vs. active monitoring with specialist reassessment), respectively.

The primary hypothesis of our FIMPACT trial is that ASD is superior to DA in patients with SIS. In addition, we will perform a pragmatic comparison of surgical and non-surgical treatment options (ASD vs. ET). The relative benefits of ASD and ET will be assessed without a priori hypothesis of the superiority of one or the other.

MATERIALS AND METHODS

Overview of study design

FIMPACT trial is an ongoing multicentre, three group randomised controlled superiority study with a primary objective to assess the efficacy of the ASD vs. DA in patients diagnosed with SIS. Our design also enables the pragmatic comparison of surgical and non-surgical treatment strategies (ASD vs. ET) (Figure 1). To obtain three balanced study groups (of similar group size), we performed a two-fold, sequential randomization as follows: First, we randomized patients to surgical or conservative treatment in 2:1 ratio and then randomized those allocated to surgery to ASD or DA in 1:1 ratio. The initial patient screening for the trial began at one site (Tampere) in February 1, 2005 and was then expanded to two additional tertiary referral centres in March 2006 and December 2006 to improve recruitment and overall generalisability of the results. The recruitment was completed (all 210 required patients enrolled) in August 2013.

Ethical approval

Ethical approval was obtained on December 28, 2004 from the institutional review board (IRB) of the Pirkanmaa Hospital District (R04200). Local research and development approvals were gained for each recruiting centre.

Participant selection

We assessed for eligibility all patients complaining of subacromial shoulder pain to any of the participating clinics. These participants were screened according to the inclusion and exclusion criteria and a recruitment surgeon confirmed the clinical diagnosis of SIS. To qualify as a recruitment surgeon, all trial surgeons had to have experience of more than 500 shoulder arthroscopies before the start of the trial. Detailed clinical examination of the shoulder was performed on all referred patients to rule out possible instability, clinical signs of rotator cuff rupture, frozen shoulder or other causes of symptoms. Standard x-rays and MRI were obtained from all potential participants and assessed by both a musculoskeletal radiologist and an orthopaedic surgeon. For patients found eligible for this study (fulfilling indications for ASD), we obtained written informed consent and randomised them into non-operative or operative group (1:2) immediately after the baseline appointment. If patient had bilateral symptoms, only one shoulder was included in the study.

Eligibility criteria

We used specific eligibility criteria to ensure that recruited participants were only those with SIS. Accordingly, a standardized clinical examination was first performed, followed by a subacromial injection test. To exclude patients with concomitant pathology, particularly rotator cuff rupture, standard x-rays and magnetic resonance imaging with intra-articular contrast injection (MRA) were carried out on all potential participants.

Inclusion criteria

1) Adult men or women ages 35 to 65 years
2) Subacromial pain for greater than 3 months with no relief from non-operative means (physiotherapy, non-steroidal anti-inflammatory medication, corticosteroid injections, and rest)
3) Pain provoked by abduction and positive painful arc -sign
4) Positive impingement test (temporary relief of pain by subacromial injection of lidocaine)
5) Pain in at least 2 out of 3 of isometric tests (abduction 0° and 30° or external rotation)
6) Provision of informed consent from the participant
7) Ability to speak, understand and read in the language of the clinical site

Exclusion criteria

1) Full thickness tear of the rotator cuff tendons diagnosed on clinical examination (marked weakness in any of the examined muscles) or magnetic resonance imaging with intra-articular contrast (MRA)
2) Osteoarthritis of the glenohumeral and/or acromioclavicular joint diagnosed on clinical examination or on x-rays
3) Substantial calcific deposits in the rotator cuff tendons found in the preoperative imaging
4) Previous surgical procedure on the affected shoulder
5) Evidence of shoulder instability (positive apprehension/positive sulcus sign)
6) Symptomatic cervical spine pathology
7) History of alcoholism, drug abuse, psychological or psychiatric problems that are likely to invalidate informed consent
8) Patient declined to participate

Recruitment process

Consultant orthopaedic surgeons carried out eligibility screening among patients referred to the study centres through standard clinical practice for shoulder pain. Patients meeting the eligibility criteria were introduced to the study. If patients expressed interest in participating, written information about the study was provided and they were asked to opt in. If the interest continued, arrangements were made for obtaining required imaging (x-rays and MRA) and for a separate baseline appointment.

Informed consent

At the first appointment, all participants were introduced to the detailed written information about the study and asked to sign a written informed consent form. At the baseline appointment (arranged within 45 days of initial contact), baseline data was completed and participant's willingness to participate in the study was confirmed. This procedure ensured that all potential participants had a reflection period for consent of at least 48 hours before giving their final consent to participate. Particular attention was paid to ensure that the participants realized that on entering the study they may receive only diagnostic arthroscopy, in which case the subacromial decompression would not be performed. They were also informed that participation in the study is entirely voluntary and any decision they make would not affect their possible future care. In addition, every participant was informed of their right to withdraw from the trial whenever they desire without the need to supply any reason for such decision.

Baseline assessment

Baseline assessment included documentation of gender, birth date, education, employment, hand dominance, time from the onset of symptoms, recreational habits, and employment status. We asked participants to assess their general heath and usage of pain medication. Modalities of any prior conservative treatment were also recorded (Table 1).

Table 1: Baseline characteristics

	ASD	DA	ET
Age (years), mean (SD)			
Gender (female/male), n (%)			
Dominant hand affected, n (%)			
Social economic status/ work load			
Heavy manual labor (construction work etc.), n (%)			
Heavy manual labor (variable workload), n (%)			
Mostly manual labor including daily office work, n (%)			
Mostly office work with occasional manual assignments, n (%)			
Full-time office work, n (%)			

Unemployed, n (%)	
Pensioner/disability pensioner, n (%)	
Student, n (%)	
Homemaker/housewife/other, n (%)	
Subjective health	
Duration of symptoms (Months), mean (SD)	
Ability to work normally regardless of the shoulder symptoms? (yes/no), n (%)	
Recreational ability regardless of the shoulder symptoms? (yes/no), n (%)	
Prior treatments	
Rest, n (%)	
Pain medication, n (%)	
Topical pain medication, n (%)	
Corticosteroid injection, n (%)	
Ultrasound, laser or any other similar therapies, n (%)	
Physiotherapy including exercise therapy, n (%)	
Other, n (%)	
Generic health states	
15D	
SF-36	
Pain measurements/Shoulder scores	
Pain at rest (0-100 VAS scale), mean (SD)	
Pain during activity (0-100 VAS scale), mean (SD)	
Constant- Murley score (CM), mean (SD)	
The simple shoulder test (SST), mean (SD)	

Baseline clinical symptoms

The recruiting surgeon carried out a clinical history and a clinical examination related to shoulder pain. Shoulder complaints other than SIS, such as full-thickness rotator cuff tears, frozen shoulder, osteoarthritis of the acromioclavicular joint and instability were ruled out as much as clinical diagnosis allows.

Baseline imaging

Standard x-rays of the shoulder were obtained to assess possible glenohumeral or acromioclavicular osteoarthritis. A magnetic resonance image with intra-articular contrast medium (MRA) was also obtained to rule out any other intra- or extra-articular pathologies. A musculoskeletal radiologist and an orthopaedic surgeon assessed all the images.

Randomisation and concealment

We used a two-phase sequential randomization. In Phase I, the participants were randomized into non-surgical or surgical treatment with allocation ratio 1:2. In the Phase II, those allocated to surgical treatment were further randomized to ASD or DA with 1:1 ratio (Figure 1).

An independent statistician with no involvement in the execution of the trial prepared separate randomization lists for each study centre using a computer-generated algorithm. Randomization was carried out using sequentially numbered sealed opaque envelopes. The envelopes were kept in a secure, agreed location at each centre. To ensure concealment, block randomization was applied using blocks varying in size randomly, the block size known only by the statistician.

To initially enter a participant into the study (Phase I), an envelope containing the treatment assignment [non-surgical (ET) or surgery (ASD or DA), ratio 1:2] was opened during the baseline appointment. Participants randomized to ET started standardized physiotherapy within 2 weeks of the baseline appointment. Participants allocated to surgical treatment were scheduled for surgery aimed to be completed within 12 weeks of randomization.

At the day of surgery, a diagnostic arthroscopy was first carried out to confirm the eligibility of the participant (to rule out full-thickness RC tear and other obvious intra-articular pathology). Research/staff nurse then completed the randomization procedure (Phase II) by opening an envelope containing the surgical treatment allocation (ASD or DA, ratio 1:1). The allocation was revealed to the surgeon by showing the paper, but not expressed verbally.

Interventions

Diagnostic arthroscopy (DA)

All participants in the two operative groups first underwent arthroscopic examination of the shoulder with the use of standard posterior and lateral portals and a 4-mm arthroscope. To maintain concealment, the surgery was carried out under general anesthesia. The orthopaedic surgeon evaluated and graded possible intra-articular pathologic changes. The rotator cuff integrity was evaluated also from the subacromial space without performing routine bursectomy. If the integrity of the rotator cuff could not be assessed, bursal tissue was bluntly stretched with troachar or resected on the tendon side to allow visualisation. If arthroscopic examination revealed any unexpected pathology (such as capsular pathology, full-thickness rotator cuff tear, or osteoarthritis), the patient was treated according current clinical practice guidelines for the given pathology while under the same anesthesia. In such a case, the participant

was excluded from the trial. Patients with partial tears were included in the study, while patients with a full-thickness tear were excluded and rotator cuff repair was carried out.

After the arthroscopic examination of the glenohumeral joint and subacromial space, confirming the eligibility of the participant, the participants were randomly assigned to receive either ASD or DA only. If the patient was allocated to the DA group, the operation was terminated. To ensure concealment of the participants and the staff other than those in the operating theatre, the participants were kept in the operating theatre for the required time to perform subacromial decompression.

Arthroscopic subacromial decompression (ASD)

Debridement of the subacromial bursa was performed with a shaver and/or electrocoagulation, followed by the resection of the bony spurs and projecting anterolateral undersurface of the acromion by a shaver as described by Ellman[21].

Postoperative care

In both the ASD and the DA group, the postoperative rehabilitation was identical. All surgically treated participants received one visit to an independent physiotherapist for guidance and instructions for home exercises. Subsequent rehabilitation was carried out according to the standardized rehabilitation protocols of the participant centres. Since the initial rehabilitation after a surgery needs to be "tempered" due to joint irritation, the rehabilitation protocol of the operatively treated groups (ASD and DA) was not identical to the ET group.

Exercise therapy (ET)

In the exercise therapy (ET) group, supervised progressive physiotherapy was started within 2 weeks of randomization using a standardized protocol. The protocol was based on the same principles as the regimen shown effective for the treatment of SIS earlier[14], but was updated – with the help of the principal investigator of the original study[14] – to conform with the state-of-the-art exercise therapy for SIS. The regimen was based on daily home exercises, but also included 15 visits to an independent physiotherapist for guidance and monitoring of the progress, carried out approximately once a week. The aim of the supervised exercise treatment was to restore painless, normal mobility of the shoulder girdle, eliminate any capsular tightness and to increase the dynamic stability of the glenohumeral joint and the scapula.

Compliance to treatment allocation and possible crossover

Participants allocated to ET group were told at the time of giving consent that they would be allowed to consider crossing over to the ASD group if they didn't get adequate relief of symptoms (preferably no sooner than 6 months post randomization). Similarly, in the two surgical treatment groups, the participants were informed of the possibility of unblinding if debilitating symptoms persisted 6 months or more after operation. If the participant was allocated to DA group, ASD was then offered. If the participant had undergone ASD, he/she was offered extended physiotherapy. No pre-specified criteria were used for determining "inadequate relief of symptoms/debilitating symptoms", rather it was left to the participants and the study physicians to make the clinical judgment together.

Outcome measures

The outcomes used in this study and the timetable for follow-up assessments are summarised in Table 2.

Table 2: Outcomes and follow-up time points

Assessment	Screening	Enrolment (Baseline)	Surgery	3 Months	6 Months	12 Months	24 Months	5 years	10 years
Screening form	X								
Informed consent		X							
Baseline characteristics form		X							
X-ray and MRI	X								X
Randomisation		X (1st)	X (2nd)						
Arthroscopic findings form			X						
Follow-up form*				X		X			
Clinical examination		X			X		X	X	X
Complications/adverse effects form**			(X)	(X)	(X)	(X)	(X)	(X)	(X)
VAS, at rest		X		X	X	X	X	X	X
VAS, at arm activity		X		X	X	X	X	X	X
Constant- Murley Score		X			X		X	X	X
Simple Shoulder Test (SST)		X			X		X	X	X
SF-36		X		X	X	X	X	X	X
15D		X		X	X	X	X	X	X
Return to work				X	X	X	X	X	X
Return to previous leisure activities				X	X	X	X	X	X
Responder analysis				X	X	X	X	X	X
Patients satisfaction to the treatment				X	X	X	X	X	X
Patients assessment of the treatment allocation				X					
Health resource utilization				X	X	X	X	X	X

* Letter/telephone interview

** If required

Primary outcome measure

VAS

As the primary outcome measure, we used a visual analogue scale (VAS) to measure the patient's perceived pain intensity at rest and at arm activity during the 24 hours preceding the assessment. Shoulder pain was assessed on a 100 mm scale ranging from 0 (no pain) to 100 (extreme pain). We considered 15 as the minimal clinically important difference (MCID) for VAS[22].

Secondary outcome measures

Constant-Murley score

Constant-Murley score (CS) is the most commonly used scoring system for evaluation of various disorders of the shoulder[23]. It consists of both objective (range of motion and strength) and subjective measurements (pain assessment, work load, and leisure time activities), which are summarized in a score between 0 and 100. A higher score indicates better shoulder function. The minimal detectable change (MDC) of the Constant score is 17 for patients with SIS[24].

In addition, as night pain is considered one of the hallmark symptoms in patients with SIS and our two primary outcome measures (patient's perceived pain intensity at rest and at arm activity in the last 24 hours) do not specifically address this issue, a specific question from the Constant-Murley score (unaffected sleep: "Yes" or "No") will be analysed separately.

SST

The simple shoulder test (SST) was developed to assess any impairment of the patient's activities of daily living [25]. The SST consists of 12 questions with yes (1) or no (0) response options. The maximum SST score is 12 indicating normal shoulder function, minimum score of 0 points refers severely diminished shoulder function. The SST has good reliability and responsiveness in patients with rotator cuff symptoms[26]. The MCID for the SST in rotator cuff disease is 2 points[27].

15D

The 15D instrument is a generic health-related quality of life (HRQoL) instrument comprising 15 dimensions[28]. For each dimension, the respondent must choose one of the five levels that best describes his/her state of health at that moment (the best level being 1 and the worst level being 5). A set of utility or preference weights is used in an addition aggregate formula to generate a single index number, the utility or 15D score. The maximum 15D score is 1 (no problems on any dimension) and the minimum score is 0 (being dead). The responsiveness, reliability and validity of 15D have been thoroughly established, and this instrument has been used extensively in clinical and healthcare research[29 30].

SF-36

Short form or SF-36 is a generic HRQoL instrument to quantify the physical, functional, and psychological aspects of health-related quality of life. It consists of 36 questions in eight subscales that assess physical, functional, social, and psychological well-being[31]. Score ranges from 0 to 100, where a higher score is associated with better health. The physical and mental component summary scales (PCS and MCS, respectively) are then calculated as composites of the related subscales. SF-36 is one of most widely used measure of health-related quality of life[32].

Patient satisfaction and Responder analysis

We elicited patients' global assessment of satisfaction to the treatment with this question: "Are you satisfied with the treatment you have received?" We used a VAS scale ranging from 0 (completely disappointed) to 100 (completely satisfied).

Additionally, we elicited patient satisfaction to the treatment outcome with the following question at each follow-up time point (Table 2): "How satisfied are you with the outcome of your treatment?" on a 5-item scale. Participants who reported very satisfied or satisfied will be categorized as "Responders" and patients who responded very dissatisfied or dissatisfied as "Non-responders".

Return to previous leisure activities

Similarly, at each follow-up (Table 2), participants were asked to respond to the following question: "Have you been able to return to your previous leisure activities?" ("yes" or "no").

Patients' perception of operative treatment-group assignment

At the 3-month follow-up point, the patients in the two operative groups were asked to guess whether they had undergone ASD or DA.

Health resource utilization and costs

For the cost-effectiveness analysis, at each follow-up visit the participants were asked to fill in a questionnaire inquiring about the use of healthcare resources. The questionnaire contains a list of items of healthcare resources available and the participants were asked to fill in the number of visits per item during the recall period of each follow-up time point. The resource use will be calculated based on the number of visits times unit cost per item and expressed as mean costs by items of resource use, and the mean direct total health care resource costs. All costs will be discounted to the 2016 price level.

Time to return to work

Information about return to work was recorded at each follow-up time point (Table 2).

Complications and adverse effects

Complications directly related to the interventions were registered. The participants were also encouraged to contact the participating hospitals if any adverse effects occurred and contacts to the health care system were monitored at every follow-up visit. Potential adverse effects (AE) were categorized to serious adverse effects (SAE) and minor adverse effects (MAE) if the participants sought treatment. Death, cardiovascular or gastrointestinal effects, deep venous thrombosis, pulmonary embolism, systemic or local infection were categorised as SAEs and shoulder symptoms like pain, swelling and decreased range of motion were categorised as MAEs. The number and severity of complications and adverse effects will be assessed.

Follow-up

The full follow-up process is shown in figure 1. In brief, the participants filled in the above noted (mailed) outcome questionnaires at 3, 6, 12 and 24 months post randomization, in addition to which they were also assessed clinically at 6 and 24 months (and 5 and 10 years) post randomisation by a

study physiotherapist unaware of treatment allocation, treatment given or possible unblinding. Outcome assessors were instructed not to inquire anything about prior treatment. Further, participants wore a t-shirt on all follow-up examinations.

Adherence and loss to follow-up

Several procedures were implemented to limit loss to follow-up, including excluding individuals likely to pose suboptimal adherence to study follow-up, obtaining verified contact information from each consented participant, and having a local research nurse remind participants of upcoming follow-up/clinic visits. All attempts were made to make follow-up as convenient for the patients as possible. Participants were required to visit the outpatient clinic only at 6 months and 24 months (and 5 and 10 years) post randomisation, while the 3- and 12-month follow-ups were carried out using mailed questionnaires to minimize inconvenience to the participants. The follow-up visits had no more discomfort for the participant than the routine clinical shoulder examinations. The follow up schedule did not involve extra costs to the participants. Follow-up rate was monitored throughout the trial and patients who did not return follow-up questionnaires would receive reminder telephone calls. Using strategies highly similar to these in our previous placebo-surgery controlled trial[33], a 99% follow-up rate was achieved.

The number and proportion of individuals eligible for and compliant with each follow-up was documented. Individuals who died during the study (from causes unrelated to the study or procedure) will be tabulated. An analysis of the demographic and prognostic characteristics will be carried out between the individuals who withdrew and those who remained in the study. For continuous variables, parametric or non-parametric analysis of variance will be used. For categorical variables, $\chi 2$ or Fisher's exact test will be applied.

Missing items

We will use multiple imputation to handle missing data for those statistical analyses that cannot handle occasional missing values. All variables to be included in the final analyses will be included in the chained equations imputation model. The imputation algorithm, fully conditional specification (FCS), uses a specific univariate model for each variable and, for each specific imputed dataset, iteratively imputes each variable with missing values and uses the imputed values in the imputation of other variables.

Sample size

The sample size calculation was based on the two primary outcome measures, VAS at rest and at arm activity, at 24 months post randomization. FIMPACT trial was powered to detect a minimal clinically important improvement (MCII) in a VAS pain score (improvement of at least 15; assumed standard deviation 25) between ASD and DA (or ET). To achieve a somewhat unconventional (stringent) 90% study power and using a two-sided Type I error rate (5%), our trial requires 68 patients per study group to show clinically meaningful advantage of ASD over DA (or ET). Acknowledging the stringent power threshold, we reserved only 3% surplus for potential loss to follow up/crossovers (3%), and accordingly, we set the recruitment target at 70 patients per treatment group.

Safety analysis

There are no anticipated safety issues with the FIMPACT Study. Identically to our previous placebo-surgery controlled trial[33], an interim analysis, as requested by the ethics board, was carried out after the enrolment of 45 participants by an independent data and safety monitoring board (the National Institute for Health and Welfare) to ensure that the rates of complications or reoperations were within acceptable limits (within the normal rate of complications and/or reoperations related to shoulder arthroscopy). Since we found no marked discrepancy in our crude assessment of the incidence of complications/reoperations, no unsealing of group assignments (unblinding) was carried out. No other interim analysis was carried out.

Data management

Questionnaire forms on paper were the primary data collection tools for the study. Upon receipt of the questionnaire forms, a study nurse made a visual check of the responses and queried missing data when possible. Research assistants, blinded to the group allocation, stored the forms into an electronic database by double data entry to minimize typing errors. The researchers, blinded to the group allocation, perform a visual check of the data in the electronic database and then queried all missing, implausible, and inconsistent data. Patient records in the participating hospitals were used when collecting missing data or interpreting inconsistent or implausible data. The final analysis was performed on data transferred to the file "FIMPACT-full data_final", having been documented as meeting the cleaning and approval requirements of our independent statistician and after the finalisation and approval of the accompanying statistical analysis plan (SAP) document. Participant files will be maintained in storage (both in

electronic and paper format) at the coordinating centre for a period of 10 years after completion of the study (10 year follow-up visits).

STATISTICAL METHODS

Statistical Analysis plan (SAP)

A statistical analysis plan (SAP) is published along this protocol. An independent statistician who is unaware of the group assignments will perform all the analyses.

We will summarise the baseline characteristics of the participants by group, reported as a mean (standard deviation) or median (first quartile, third quartile) for continuous variables and count (percent) for categorical variables.

We will analyse the data in a blinded manner. All p-values will be reported to 3 decimal places with those less than 0.001 reported as $p < 0.001$. The criterion for statistical significance will be set at alpha = 0.05.

Primary analysis

We will carry out the primary analysis according to the intention-to-treat (ITT) principle: participants are retained in the groups to which they were initially randomized.

The primary comparison will be on the efficacy of ASD (ASD vs. DA). We will perform the primary comparison on the efficacy of ASD (ASD vs. DA) as a between-group comparison using a repeated measures mixed-effects model (RMMM). Study group and time of assessment (baseline, 3, 6, 12 and 24 months) will be included as fixed factors and patient as a random factor. The model will include interactions between study group and time of assessment. The baseline value will be included as a covariate. The RMMM model will be used to quantify the treatment effect as the difference between the groups in pain scores (VAS) with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization. To safeguard against potential multiplicity bias[34], we will require a statistically significant treatment effect on both of our primary outcome variables, i.e., pain at rest and pain at activity.

The same statistical model will also apply to the pragmatic comparison of the relative benefits of surgical vs. non-operative treatment strategies on SIS (ASD vs. ET).

Secondary analyses

We will also use the RMMM model to analyse secondary outcomes where applicable. The results will be reported as the differences between the groups with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization.

Categorical variables, reoperations or treatment conversions, and complications as well as adverse effects will be analysed using logistic regression analysis or Poisson regression dependent on whether subjects with complications or (multiple) complications (per subject) are analysed.

These secondary analyses will be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not a problem[2]

Sensitivity analyses

We will carry out the following sensitivity analyses: 1) per-protocol analyses, in which the above noted primary and secondary analyses will be carried out again with patients who received the interventions as allocated; 2) and potential effects due to the treatment providing centres.

Subgroup analyses and Hypothesized Effects

We have identified three important subgroups. We will perform these three subgroup analyses with the primary endpoint as the outcome and the direction of hypothesized effect described[35]:

1) Duration of symptoms – Neer originally suggested that ASD should be considered for patients with persistent symptoms despite over one year of conservative treatment[36]. Recent RCTs failing to find efficacy on ASD (vs. conservative treatment) have prompted arguments that ASD should be reserved to situations when long-term conservative treatment has failed[37]. Although a recent study specifically addressed this question and failed to support this hypothesis[38], we still intend to compare the treatment effects of participants stratified based on the duration of symptoms. Accordingly, we will compare those with symptoms less than 12 months to those with symptoms longer than 12 months. We hypothesize that subacromial decompression will work better in patients with duration of symptoms > 12 months than for patients with symptoms < 12 months.
2) Severity of symptoms - A subgroup analysis will also be conducted comparing the treatment effects in patients with severe (VAS 70 or more), moderate (VAS 55 to 69), and mild (VAS less than 55) symptoms at baseline. We hypothesize that subacromial decompression will work better in patients

with more severe (VAS 70 or more) than moderate (VAS 55 to 69) or mild (VAS less than 55) symptoms at baseline.

3) Acromial anatomy - A hook-type acromion has been suggested as an independent risk factor for subacromial impingement[39]. To assess the validity of this suggestion, a subgroup analysis will be conducted comparing the treatment effects in patients with flat (type I), curved (type II), or hooked (type III) acromion according to classification by Bigliani et al.[40] We hypothesize that subacromial decompression will work better in patients with hooked (type III) than curved (type II) or flat (type I) acromion at baseline.

Effect modifying and mediating factors

Multiple regression models were used to assess the potential effect modifying factors (e.g., age, gender, psychological well-being, mental health, occupational shoulder load, education level, and hand dominance) and effect mediating factors (e.g., absence of complications and adherence to rehabilitation) on pain, functional disability and quality of life. These analyses are supportive, explanatory and/or hypothesis generating.

Blinded data interpretation

To safeguard against potential risk of bias during interpretation, we will use our recently introduced method of "blinded data interpretation"[41]. So far, this method has been successfully applied to three previous trials[33 42 43]. In brief, an independent statistician will provide the Writing committee of the FIMPACT trial (authors of this protocol) with blinded results from the analyses with study groups labelled as group A, group B, and group C. The Writing Committee will then contemplate on the interpretation of the results until a consensus is reached and agree in writing on all alternative interpretations of the findings. Once reaching a consensus, we will record the minutes of this meeting as a statement of interpretation document signed by all members of the Writing Committee. Only after reaching this common agreement will the data manager and independent statistician break the randomization code.

DISCUSSION

In this protocol paper, we describe the execution of a randomised, placebo-surgery controlled trial for the assessment of the efficacy of arthroscopic subacromial decompression (ASD) in patients with sub-acromial impingement syndrome (SIS). Acknowledging the potential of surgery to produce powerful placebo effects[44], our primary comparator is diagnostic arthroscopy, differing from the ASD only by lacking the critical therapeutic element of the ASD (subacromial decompression). We will also conduct the pragmatic comparison of surgical and non-surgical treatment options of SIS by including a third group of progressive exercise therapy (ET) (Figure 1, ASD vs. ET).

Interpretations and generalizability

Our interpretation scheme primarily rests on the tenet that the minimum requirement for the clinical viability of ASD is that it needs to show superiority to DA - a therapeutically inert and thus a clinically non-viable option. To test this, we have chosen a classic *efficacy* or *"can it work" (proof-of-concept)* design[45-47]: The recruited participants are those who - according to current evidence - should have an "optimal response" to ASD and the participants and outcome assessors are blinded to the interventions given. This design should thus yield findings that are widely applicable to patients with characteristic clinical signs and symptoms of SIS. We will also compare ASD with non-operative treatment option for SIS, the progressive ET, in a more pragmatic comparison, which is confounded by the lack of blinding of the participants. (Figure 2)

The generalizability of our primary (efficacy) comparison may be questioned as the patients are carefully selected (strict eligibility criteria) and treated by experienced shoulder surgeons. Nevertheless, the eligibility criteria are in agreement with the existing treatment guidelines on SIS [4]. The results should thus be applicable to the specific populations currently receiving treatment for their SIS. As for the skill-level of the surgeons, the index surgical procedure (ASD) is a relatively simple procedure and thus likely not very sensitive to individual surgeons' experience. For example, the amount of bone removed from the undersurface of the acromion seems to have at best a marginal effect on the outcome. Even bursectomy alone has been shown to produce the same therapeutic effect as standard acromioplasty[48].

Rationale for outcome assessment and statistical analysis

Traditionally, the assessment of the treatment effects of two or more interventions has relied primarily on the statistical significance of the mean differences of the intervention groups. However, as described in a recent paper[49], to truly assess the clinical relevance of a treatment, one also needs information about the distribution of individual responses. In essence, one needs to look at how many people on treatment and on comparator group(s) had a response at least as great as the minimum (clinically) important difference (MCID). Such individuals have been described as "responders," and this approach of comparing treatment groups as a "responder analysis"[50 51]. The authors[49] suggested that "*Clinical trials should specify in their protocol that they will report the distribution of results in individual participants as well as the mean difference. Researchers should publish plots of individual results and responder analyses in clinical trial reports.*" The FIMPACT trial adheres to this suggested approach. Accordingly, we will elaborate several relevant and often interrelated issues, such as the study power, the primary outcomes and their interpretation, the minimal clinically important difference (MCII), as well as the approach we have chosen for carrying out a responder analysis.

Study power

Traditionally the sample size is calculated based on the minimal clinically important difference or change (MCID or MCII), i.e., the smallest change in measurement that signifies an important/detectable improvement in a patient's symptom(s). MCII/-D is not a static value even for one outcome instrument, but rather can have different values when assessed with different methods or *in different patient populations*. We chose VAS at rest and during arm activity as our primary outcomes, because shoulder pain is the primary complaint of patients with SIS. The FIMPACT trial was powered to detect an improvement of at least 15 on a 0-100 VAS scale[52] between ASD and ET. This yielded a sample size estimate of 70 participants per group. To safeguard against lack of study power, we chose a statistical threshold of 90% over the more conventional 80%. In this context, Norman et al.[53] recently introduced a thought-provoking proposal arguing that a standard ('off-the-peg') sample size of 64 per group would be just as valid an estimate as one obtains by more traditional ('made-to-measure') sample size calculations[53]. Finally, although the statistical power is a vital step in the *planning phase* of any clinical trial, the actual quality of evidence (certainty in the obtained estimates) can only be appropriately assessed from the confidence intervals (CI) of the data obtained[54].

Responder analysis

As noted above, instead of focusing only on the statistical significance of the mean differences between treatment groups in the VAS (i.e., the mean improvement from baseline to 24 months), we will also carry out "a responder analysis". In principle, this analysis allows physicians to inform a patient of his or her chance of experiencing a clinically meaningful improvement from the treatment, both in absolute terms and in comparison, to a control group. The difference between responders and non-responders can be considered the net-benefit of the treatment. One proposed means to carry out a responder analysis relies on the assessment of the proportion of patients reaching the patient-acceptable symptom state (PASS) and the patient-disappointing symptoms state (PDSS). As no universal consensus exists on either the PASS or the PDSS in the context of SIS, we chose to anchor our responder analysis to the patient's assessment of satisfaction with the shoulder treatment outcome: Patients reporting very satisfied or satisfied will be categorized as "Responders" and those reporting very dissatisfied or dissatisfied as "Non-responders". Given the obvious coarseness of this approach, we plan to evaluate the appropriate criteria for PASS and PDSS in more detail in the future, exploring the potential contribution of, e.g., arm pain at rest and at activity, shoulder function, and night pain.

Ethics of placebo surgery

A recent systematic review of the use of surgical placebo shows that in more than half of these studies the treatment group that included critical surgical/therapeutic element had no greater effect than a placebo group [18]. The review also showed that risks of adverse effects were small and the placebo group was safer than the surgery under investigation. These findings make a compelling case for the use of surgical placebo controls when a placebo effect may be present. Regarding the ethics of surgical placebo controls, the authors of the review state "*Placebo controlled surgical trials raise important ethical concerns but are justified when there is a genuine equipoise; that is, a disagreement in the medical community about whether one treatment is superior to another, because standard treatment does not exist or its efficacy is questioned.*" They continue by concluding: "*Placebo controlled trials in surgery are as important as they are in medicine, and they are justified in the same way. They are powerful, feasible way of showing the efficacy of surgical procedures. They are necessary to protect the welfare of present and future patients as well as to conduct proper cost effectiveness analyses. Only then may publicly funded surgical interventions be distributed fairly and justly. Without such studies ineffective treatment may continue unchallenged.*" Our views regarding the ethics of using a surgical placebo group are perfectly aligned with these notions.

Limitations of the study

One possible confounder in our trial is that subacromial pain is also the hallmark symptom of a rotator cuff tear, although the latter patients usually also represent with muscle weakness. To exclude patients with a (clinically-relevant) rotator cuff tear, our eligibility screening included two preoperative assessments: (a) clinical exams targeted at finding obvious weakness of the rotator cuff muscles and (b) MRA, an imaging modality with a shown 92 specificity and 94 sensitivity for "full-thickness" RC tears[55]. In addition to these, we also carried out (c) a diagnostic arthroscopy in the ASD and DA groups prior to randomisation. Despite the thorough *preoperative* screening, 10% (14/136) of the participants allocated to the two surgical groups had to be excluded because of AC-arthrosis (n=1) or intra-articular pathology found at diagnostic arthroscopy (n=13). Although this does not have any effect on our primary comparison (ASD vs. DA), one could argue that the ET and operatively treated groups (ASD and DA) are not fully comparable. At the same we don't know the clinical relevance of small RC tears or SLAP lesions, which don't result in obvious muscle weakness and/or are not apparent in MRA. In the end, if this bias proves clinically relevant in our analysis, it will skew our results by favouring the ASD group in the pragmatic comparison (ASD vs. ET). Another concern related to the pragmatic comparison (ASD vs. ET) is that the progressive exercise therapy regimen carried out in the ET group is different from the postoperative rehabilitation carried out by patients in the ET group, for obvious reasons; surgically treated patients need time to recover from the initial surgical trauma. Furthermore, ASD patients are also subject to some degree of postoperative immobilization, extended sick leave, and modifications in pain medication and activities, all of which potentially have an effect on the outcome of treatment.

Another obvious concern related to our study design is the discrepant timing of the start of the actual treatment between the ET and the two surgical groups due to the time required to arrange the surgery. Acknowledging this, the two-year follow-up was chosen as our primary time point for assessing the benefits of treatment, as we assume that by this time the potential confounding effect of slightly different follow-up times should be diluted to a minimum. This is also the reason why we use data from the shorter-term follow-up visits (i.e.: visits performed at 3, 6 and 12 months after randomization) primarily to illustrate the trajectory of the treatment response in the three groups. Concerns over the varying time span from the randomization of the patients to the trial to the actual induction of treatment (due to delay in surgery) also applies to the CSAW trial[20]. To compensate for the waiting list effects, the CSAW investigators have chosen a slightly different strategy: Although the primary outcome assessment is performed at 6 months after randomization in CSAW trial, they have introduced additional follow-up assessments, referenced from surgery, for patients waiting for longer than 4 months for their surgery after

randomization. They have also set a secondary outcome measurement point at 1-year post randomization.

Author affiliations

Mika Paavola
Helsinki University Central Hospital,
Department of Orthopaedics and Traumatology,
Töölö hospital,
Topeliuksenkatu 5
P.O. Box 266, 00029 HUS,
Helsinki, Finland
Email: mika.paavola@hus.fi

Antti Malmivaara
National Institute for Health and Welfare
Center for Health and Social Economics
Mannerheimintie 166
00270 Helsinki
Finland
Email:antti.malmivaara@thl.fi

Simo Taimela
University of Helsinki,
Department of Orthopaedics and Traumatology,
Töölö hospital, Topeliuksenkatu 5,
00014 University of Helsinki, Helsinki, Finland
Email: simo.taimela@evalua.fi

Kari Kanto
Hatanpään sairaala, Tampere
Hatanpäänkatu 24, 33900 Tampere, Finland
Email: kari.kanto@fimnet.fi

Teppo L N Järvinen
University of Helsinki,
Department of Orthopaedics and Traumatology,
Töölö hospital, Topeliuksenkatu 5,
00014 University of Helsinki, Helsinki, Finland
Email: teppo.jarvinen@helsinki.fi

Acknowledgements

The FIMPACT investigators would like to thank Professor Jonas Ranstam (Lund University) for the statistical analysis plan and acknowledge Dr. Pekka Paavolainen and Professor Markku Järvinen for

their critical contributions and support on the early phases of the planning and execution of the trial. Virginia Mattila is acknowledged for her linguistic expertise and language revisions.

Contributions

Concept and design: Mika Paavola (MP), Antti Malmivaara (AM), Simo Taimela (ST) and Teppo Järvinen (TJ).

Drafting and critical revision of the article for important intellectual content: MP, AM, ST, TJ and Kari Kanto (KK).

Final approval of the article: MP, AM, ST, TJ, KK.

Ensuring the accuracy of the work: MP, AM, ST, TJ, KK.

Statistical expertise: Jonas Ranstam (JR).

Obtaining of funding: TJ and Markku Järvinen (MJ).

Funding

The FIMPACT trial was supported by the Sigrid Juselius Foundation, the Competitive Research Fund of Pirkanmaa and Helsinki University Central Hospital Districts, the Academy of Finland, and the Jane and Aatos Erkko Foundation. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of this protocol.

Figure 1: Flowchart of the trial: enrolment, assigned intervention and follow-up scheme.

Figure 2: Study design and interpretation of results.

Table 1: Baseline characteristics.

Table 2: A diagram outlining the follow-up scheme used.

Appendix 1: FIMPACT investigators.

Appendix 2: Statistical analysis plan (SAP).

Appendix 3: Blinded data interpretation plan.

REFERENCES

1. Paloneva J, Lepola V, Karppinen J, et al. Declining incidence of acromioplasty in Finland. *Acta Orthop* 2015;86(2):220-4. doi: 10.3109/17453674.2014.977703
2. Vitale MA, Arons RR, Hurwitz S, et al. The rising incidence of acromioplasty. *J Bone Joint Surg Am* 2010;92(9):1842-50. doi: 10.2106/JBJS.I.01003
3. Engebretsen K, Grotle M, Bautz-Holter E, et al. Radial extracorporeal shockwave treatment compared with supervised exercises in patients with subacromial pain syndrome: single blind randomised study. *Bmj* 2009;339:b3360. doi: 10.1136/bmj.b3360 [published Online First: 2009/09/17]
4. Diercks R, Bron C, Dorrestijn O, 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. doi: 10.3109/17453674.2014.920991
5. Oh LS, Wolf BR, Hall MP, et al. Indications for rotator cuff repair: a systematic review. *Clinical orthopaedics and related research* 2007;455:52-63. doi: 10.1097/BLO.0b013e31802fc175
6. Aurora A, McCarron J, Iannotti JP, et al. Commercially available extracellular matrix materials for rotator cuff repairs: state of the art and future trends. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al]* 2007;16(5 Suppl):S171-8. doi: 10.1016/j.jse.2007.03.008
7. Michener LA, McClure PW, Karduna AR. Anatomical and biomechanical mechanisms of subacromial impingement syndrome. *Clinical biomechanics* 2003;18(5):369-79.
8. Koester MC, George MS, Kuhn JE. Shoulder impingement syndrome. *The American journal of medicine* 2005;118(5):452-5. doi: 10.1016/j.amjmed.2005.01.040
9. Neer CS, 2nd. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *The Journal of bone and joint surgery American volume* 1972;54(1):41-50.
10. Yu E, Cil A, Harmsen WS, et al. Arthroscopy and the dramatic increase in frequency of anterior acromioplasty from 1980 to 2005: an epidemiologic study. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 2010;26(9 Suppl):S142-7. doi: 10.1016/j.arthro.2010.02.029
11. Judge A, Murphy RJ, Maxwell R, et al. 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. doi: 10.1302/0301-620X.96B1.32556
12. Dong W, Goost H, Lin XB, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. *Medicine (Baltimore)* 2015;94(10):e510. doi: 10.1097/MD.0000000000000510
13. Shi LL, Edwards TB. The role of acromioplasty for management of rotator cuff problems: where is the evidence? *Advances in orthopedics* 2012;2012:467571. doi: 10.1155/2012/467571
14. Brox JI, Staff PH, Ljunggren AE, et al. Arthroscopic surgery compared with supervised exercises in patients with rotator cuff disease (stage II impingement syndrome). *Bmj* 1993;307(6909):899-903. [published Online First: 1993/10/09]
15. Haahr JP, Ostergaard S, Dalsgaard J, et al. Exercises versus arthroscopic decompression in patients with subacromial impingement: a randomised, controlled study in 90 cases with a one year follow up. *Ann Rheum Dis* 2005;64(5):760-4. doi: 10.1136/ard.2004.021188 [published Online First: 2005/04/19]
16. Ketola S, Lehtinen J, Rousi T, et al. No evidence of long-term benefits of arthroscopicacromioplasty in the treatment of shoulder impingement syndrome: Five-year results of a randomised controlled trial. *Bone Joint Res* 2013;2(7):132-9. doi: 10.1302/2046-3758.27.2000163 [published Online First: 2013/07/10]
17. Ernst E, Resch KL. Concept of true and perceived placebo effects. *Bmj* 1995;311(7004):551-3. [published Online First: 1995/08/26]
18. Wartolowska K, Judge A, Hopewell S, et al. Use of placebo controls in the evaluation of surgery: systematic review. *Bmj* 2014;348:g3253. doi: 10.1136/bmj.g3253

19. Buchbinder R. Meniscectomy in patients with knee osteoarthritis and a meniscal tear? *N Engl J Med* 2013;368(18):1740-1. doi: 10.1056/NEJMe1302696 [published Online First: 2013/03/20]
20. Beard D, Rees J, Rombach I, et al. The CSAW Study (Can Shoulder Arthroscopy Work?) - a placebo-controlled surgical intervention trial assessing the clinical and cost effectiveness of arthroscopic subacromial decompression for shoulder pain: study protocol for a randomised controlled trial. *Trials* 2015;16:210. doi: 10.1186/s13063-015-0725-y
21. Ellman H. Arthroscopic subacromial decompression: analysis of one- to three-year results. *Arthroscopy* 1987;3(3):173-81. [published Online First: 1987/01/01]
22. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al]* 2009;18(6):927-32. doi: 10.1016/j.jse.2009.03.021
23. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clinical orthopaedics and related research* 1987(214):160-4.
24. Henseler JF, Kolk A, van der Zwaal P, et al. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al]* 2015;24(3):376-81. doi: 10.1016/j.jse.2014.07.003
25. Matsen FA, 3rd, Ziegler DW, DeBartolo SE. Patient self-assessment of health status and function in glenohumeral degenerative joint disease. *J Shoulder Elbow Surg* 1995;4(5):345-51. [published Online First: 1995/09/01]
26. Godfrey J, Hamman R, Lowenstein S, et al. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *J Shoulder Elbow Surg* 2007;16(3):260-7. doi: 10.1016/j.jse.2006.07.003 [published Online First: 2006/12/26]
27. Tashjian RZ, Deloach J, Green A, et al. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. *J Bone Joint Surg Am* 2010;92(2):296-303. doi: 10.2106/JBJS.H.01296 [published Online First: 2010/02/04]
28. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med* 2001;33(5):328-36. [published Online First: 2001/08/09]
29. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. *Qual Life Res* 2008;17(3):485-95. doi: 10.1007/s11136-008-9317-6 [published Online First: 2008/02/22]
30. Bowling A. Measuring health. A review of quality of life measurement scales. 3rd ed: Berkshire: Open University Press 2004.
31. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83. [published Online First: 1992/06/11]
32. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *Bmj* 2002;324(7351):1417. [published Online First: 2002/06/18]
33. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med* 2013;369(26):2515-24. doi: 10.1056/NEJMoa1305189 [published Online First: 2013/12/27]
34. Ranstam J. Multiple P-values and Bonferroni correction. *Osteoarthritis Cartilage* 2016;24(5):763-4. doi: 10.1016/j.joca.2016.01.008
35. Sun X, Ioannidis JP, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. *Jama* 2014;311(4):405-11. doi: 10.1001/jama.2013.285063
36. Neer CS, 2nd. Impingement lesions. *Clin Orthop Relat Res* 1983(173):70-7. [published Online First: 1983/03/01]
37. Dong W, Goost H, Lin XB, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. *Medicine (Baltimore)* 2015;94(10):e510. doi: 10.1097/MD.0000000000000510 [published Online First: 2015/03/12]

38. Ketola S, Lehtinen J, Rousi T, et al. Which patients do not recover from shoulder impingement syndrome, either with operative treatment or with nonoperative treatment? *Acta Orthop* 2015;86(6):641-6. doi: 10.3109/17453674.2015.1033309
39. Tangtrakulwanich B, Kapkird A. Analyses of possible risk factors for subacromial impingement syndrome. *World J Orthop* 2012;3(1):5-9. doi: 10.5312/wjo.v3.i1.5
40. Bigliani LU MD, April EW The morphology of the acromion and its relationship to rotator cuff tears. *Orthop Trans* 1986;10:228.
41. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *J Clin Epidemiol* 2014;67(7):769-72. doi: 10.1016/j.jclinepi.2013.11.011 [published Online First: 2014/02/25]
42. Skou ST, Roos EM, Laursen MB, et al. A Randomized, Controlled Trial of Total Knee Replacement. *N Engl J Med* 2015;373(17):1597-606. doi: 10.1056/NEJMoa1505467 [published Online First: 2015/10/22]
43. Investigators F, Bhandari M, Jeray KJ, et al. A Trial of Wound Irrigation in the Initial Management of Open Fracture Wounds. *N Engl J Med* 2015;373(27):2629-41. doi: 10.1056/NEJMoa1508502
44. Jonas WB, Crawford C, Colloca L, et al. To what extent are surgery and invasive procedures effective beyond a placebo response? A systematic review with meta-analysis of randomised, sham controlled trials. *BMJ Open* 2015;5(12):e009655. doi: 10.1136/bmjopen-2015-009655 [published Online First: 2015/12/15]
45. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving. *Bmj* 1999;319(7211):652-3. [published Online First: 1999/09/10]
46. Djulbegovic B, Paul A. From efficacy to effectiveness in the face of uncertainty: indication creep and prevention creep. *Jama* 2011;305(19):2005-6. doi: 10.1001/jama.2011.650
47. Jarvinen TL, Sievanen H, Kannus P, et al. The true cost of pharmacological disease prevention. *Bmj* 2011;342:d2175. doi: 10.1136/bmj.d2175
48. Donigan JA, Wolf BR. Arthroscopic subacromial decompression: acromioplasty versus bursectomy alone--does it really matter? A systematic review. *Iowa Orthop J* 2011;31:121-6. [published Online First: 2011/11/19]
49. Cates C, Karner C. Clinical importance cannot be ruled out using mean difference alone. *Bmj* 2015;351:h5496. doi: 10.1136/bmj.h5496
50. Guyatt GH, Juniper EF, Walter SD, et al. Interpreting treatment effects in randomised trials. *Bmj* 1998;316(7132):690-3. [published Online First: 1998/04/02]
51. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials* 2007;8:31. doi: 10.1186/1745-6215-8-31 [published Online First: 2007/10/27]
52. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg* 2009;18(6):927-32. doi: 10.1016/j.jse.2009.03.021 [published Online First: 2009/06/19]
53. Norman G, Monteiro S, Salama S. Sample size calculations: should the emperor's clothes be off the peg or made to measure? *BMJ* 2012;345:e5278. doi: 10.1136/bmj.e5278
54. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121(3):200-6.
55. Lenza M, Buchbinder R, Takwoingi Y, et al. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. *Cochrane Database Syst Rev* 2013;9:CD009020. doi: 10.1002/14651858.CD009020.pub2 [published Online First: 2013/09/26]

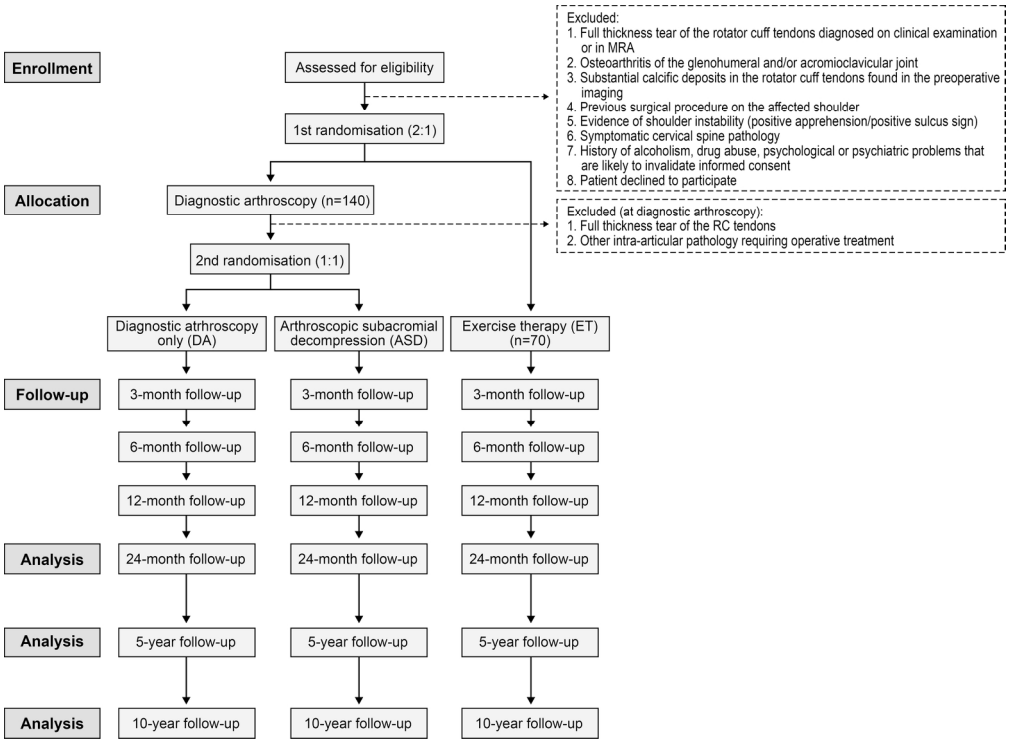
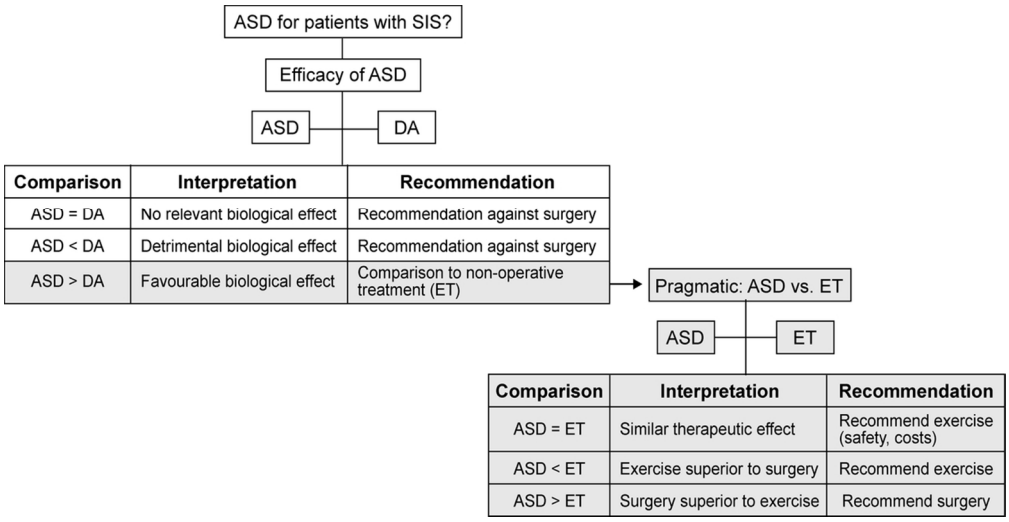


Figure 1.

184x134mm (300 x 300 DPI)



Comparison	Interpretation	Recommendation
ASD = DA	No relevant biological effect	Recommendation against surgery
ASD < DA	Detrimental biological effect	Recommendation against surgery
ASD > DA	Favourable biological effect	Comparison to non-operative treatment (ET)

Comparison	Interpretation	Recommendation
ASD = ET	Similar therapeutic effect	Recommend exercise (safety, costs)
ASD < ET	Exercise superior to surgery	Recommend exercise
ASD > ET	Surgery superior to exercise	Recommend surgery

Figure 2.

107x54mm (300 x 300 DPI)

FIMPACT INVESTIGATORS

Steering Committee: Mika Paavola (Chair, Helsinki University Central Hospital), Teppo Järvinen (Co-Chair, University of Helsinki and Helsinki University Central Hospital), Antti Malmivaara (National Institute for Health and Welfare), Simo Taimela (University of Helsinki)

Writing Committee: Teppo Järvinen (Co-Chair, University of Helsinki and Helsinki University Central Hospital), Simo Taimela (Co-Chair, University of Helsinki), Antti Malmivaara (National Institute for Health and Welfare), Mika Paavola (Helsinki University Central Hospital), Kari Kanto (Hatanpää City Hospital, Tampere).

Participating Clinical Sites: *Helsinki University Central Hospital/Jorvi Hospital*: Kalevi Hietaniemi, Juha Kalske, Vesa Lepola, Jyrki Salmenkivi, Sikri Tukiainen. – *Helsinki University Central Hospital/Herttoniemi Hospital*: Jarkko Pajarinen, Mikko Salmela, Vesa Savolainen, Ilkka Sinisaari. – *Hatanpää City Hospital, Tampere*: Timo Järvelä, Kari Kanto, Janne Lehtinen, Mikael Salmela.

FIMPACT Methods Center: Leena Caravitis, Sari Karesvuori, Pirjo Toivonen (Project management), Mathias Bäck (Data management), Ville Haapamäki (Imaging) (Helsinki University Central Hospital and University of Helsinki); Jari Inkinen (Physiotherapy) (Kunnon Klinikka, Tampere), Esa Läärä (Randomization) (University of Oulu), Harri Sintonen (QoL outcomes) (National Institute for Health and Welfare).

Statistical Analysis Plan (SAP) for:

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT), 2-year follow-up

Simo Taimela[1] and Teppo L N Järvinen[1]

Statistical advisor:

Jonas Ranstam

[1] Department of Orthopedics and Traumatology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

STUDY SYNOPSIS

Introduction: Arthroscopic subacromial decompression (ASD) is the most commonly performed surgical intervention for shoulder pain, yet evidence on its efficacy is limited. The rationale for the surgery rests on the tenet that symptom relief is achieved through removal of a bony acromial spur and the resulting decompression of the tendon passage. Acknowledging the potential placebo effect of surgery, the primary objective of this superiority trial is to compare the efficacy of ASD versus diagnostic arthroscopy (DA) in patients with shoulder impingement syndrome (SIS), where DA differs only by the lack of subacromial decompression. As a non-surgical treatment option, a third group of supervised progressive exercise therapy (ET) will allow for pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies.

Methods/Design: FIMPACT trial is an ongoing multicentre, three-group randomised controlled study with a primary objective of assessing the efficacy of the ASD vs. DA and a secondary objective of comparing ASD to exercise therapy (ET) in a pragmatic setting. We performed two-fold concealed allocation, first by randomizing patients to surgical (ASD or DA) or conservative (ET) treatment in 2:1 ratio and then those allocated to surgery further to ASD or DA in 1:1 ratio. Our two primary outcomes are pain at rest and arm activity assessed with visual analog scale (VAS), while the secondary outcomes are functional assessment (Constant score and Simple shoulder test), quality of life (15D and SF-36), patient satisfaction, proportions of responders and non-responders, reoperations/treatment conversions, all at 2 years post-randomization, as well as adverse effects and complications. We recruited a total of 210 patients from 3 tertiary referral centres. We will conduct the primary analysis on the intention-to-treat basis.

TRIAL REGISTRATION

ClinicalTrials.gov NCT00428870 (first registered January 29, 2007).

STUDY OBJECTIVES AND OUTCOMES

This statistical analysis plan (SAP) is accompanying the actual study protocol of the FIMPACT trial, a document that elaborates the methods used in detail. All outcomes were inquired from participants at baseline and follow-ups (6 and 24 months) and selected additional measures at 3 and 12 months (for details, see Table 1). The last patient reached the primary endpoint, the 24-month follow-up, in September 2015.

Table 1: Outcomes and follow-up time points

Assessment	Screening	Enrolment (Baseline)	Surgery	3 Months	6 Months	12 Months	24 Months	5 years	10 years
Screening form	X								
Informed consent		X							
Baseline characteristics form		X							
X-ray and MRI	X								X
Randomisation		X (1st)	X (2nd)						
Arthroscopic findings form			X						
Follow-up form*				X		X			
Clinical examination		X			X		X	X	X
Complications/adverse effects form**			(X)	(X)	(X)	(X)	(X)	(X)	(X)
VAS, at rest		X		X	X	X	X	X	X
VAS, at arm activity		X		X	X	X	X	X	X
Constant- Murley Score		X			X		X	X	X
Simple Shoulder Test (SST)		X			X		X	X	X
SF-36		X		X	X	X	X	X	X
15D		X		X	X	X	X	X	X
Return to work				X	X	X	X	X	X
Return to previous leisure activities				X	X	X	X	X	X
Responder analysis				X	X	X	X	X	X
Patients satisfaction to the treatment				X	X	X	X	X	X
Patients assessment of the treatment allocation				X					
Health resource utilization				X	X	X	X	X	X

* Letter/telephone interview

** If required

DESCRIPTIVE OUTCOMES

At screening, the participants filled out a questionnaire to record gender, age, hand dominance, weight, height, level of education (socioeconomic status), workload (type of work), physical activity level, sports discipline, subjective health, symptoms (onset, frequency, and severity), use of pain medications, prior treatments, expectations to treatment, generic health state, and disease-specific scores. To exclude patients with concomitant shoulder pathology (particularly rotator cuff rupture), magnetic resonance imaging with contrast (MRA) was acquired for each participant.

OBJECTIVES AND PRIMARY OUTCOME

The primary objective of this trial is to compare the efficacy of arthroscopic subacromial decompression (ASD) versus diagnostic arthroscopy (DA) in patients with SIS. The trial is designed as a superiority trial, i.e. we expected in the power calculation that the ASD will result in greater pain relief at 24-month follow-

up than DA (or ET). The 24-month follow-up was chosen as the primary endpoint, since this time point is a commonly held "minimal requirement" for any procedure in the field (orthopaedics) and most commonly used in the trials assessing the treatment of SIS.

The primary hypothesis: The primary hypothesis of our FIMPACT trial is that ASD is superior to DA in patients with SIS.

To enable pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies on SIS, a non-surgical (third) treatment option of supervised progressive exercise therapy (ET) is also included (ASD vs. ET).

Additional hypothesis: The relative benefits of ASD and ET will be assessed without a priori hypothesis on the superiority of one or the other.

As the primary outcome measure, a visual analogue scale (0-100) was used to measure the patient's perceived pain intensity at rest and at arm activity during the 24 hours preceding the assessment. We considered 15 as the minimal clinically important difference (MCID) for SIS.[1]

SECONDARY OUTCOMES

Our secondary outcome measures are listed below. These outcomes will only be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not considered to be a problem[2].

Constant-Murley score

Constant-Murley score (CS) is the most commonly used scoring system for evaluation of various disorders of the shoulder[3]. It consists of both objective (range of motion and strength) and subjective measurements (pain assessment, work load, and leisure time activities), which are summarized in a score between 0 and 100. A higher score indicates better shoulder function. The minimal detectable change (MDC) of the Constant score is 17 for patients with SIS[4]

In addition, as night pain is considered one of the hallmark symptoms in patients with SIS and our two primary outcome measures (patient's perceived pain intensity at rest and at arm activity in the last 24 hours) do not specifically address this issue, a specific question from the Constant-Murley score (unaffected sleep: "Yes" or "No") will be analysed separately.

SST

The simple shoulder test (SST) was developed to assess any impairment of the patient's activities of daily living[5]. The SST consists of 12 questions with yes (1) or no (0) response options. The maximum SST score is 12 indicating normal shoulder function, minimum score of 0 points refers severely diminished shoulder function. The SST has good reliability and responsiveness in patients with rotator cuff symptoms[6]. The MCID for the SST in rotator cuff disease is 2 points[7].

15D

The 15D instrument is a generic health-related quality of life (HRQoL) instrument comprising 15 dimensions[8]. For each dimension, the respondent must choose one of the five levels that best describes his/her state of health at that the moment (the best level being 1 and the worst level being 5). A set of utility or preference weights is used in an addition aggregate formula to generate a single index number, the utility or 15D score. The maximum 15D score is 1 (no problems on any dimension) and the minimum score is 0 (being dead). The responsiveness, reliability and validity of 15D have been thoroughly established, and this instrument has been used extensively in clinical and healthcare research[9 10].

SF-36

Short form or SF-36 is a generic HRQoL instrument to quantify the physical, functional, and psychological aspects of health-related quality of life. It consists of 36 questions in eight subscales that assess physical, functional, social, and psychological well-being[11]. Score ranges from 0 to 100, where a higher score is associated with better health. The physical and mental component summary scales (PCS and MCS, respectively) are then calculated as composites of the related subscales. SF-36 is one of most widely used measure of health-related quality of life[12].

Patient satisfaction and Responder analysis

We elicited patients' global assessment of satisfaction to the treatment with this question: "Are you satisfied with the treatment you have received?" We used a VAS scale ranging from 0 (completely disappointed) to 100 (completely satisfied).

Additionally, we elicited patient satisfaction to the treatment outcome with the following question at each follow-up time point (Table 1): "How satisfied are you with the outcome of your treatment?" on a 5-item scale. Participants who reported very satisfied or satisfied will be categorized as "Responders" and patients who responded very dissatisfied or dissatisfied as "Non-responders".

Return to previous leisure activities

Similarly, at each follow-up (Table 1), participants were asked to respond to the following question: "Have you been able to return to your previous leisure activities?" ("yes" or "no").

Patients' perception of operative treatment-group assignment

At the 3-month follow-up point, the patients in the two operative groups were asked to guess whether they had undergone ASD or DA.

Health resource utilization and costs

For the cost-effectiveness analysis, at each follow-up visit the participants were asked to fill in a questionnaire inquiring about the use of healthcare resources. The questionnaire contains a list of items of healthcare resources available and the participants were asked to fill in the number of visits per item during the recall period of each follow-up time point. The resource use will be calculated based on the number of visits times unit cost per item and expressed as mean costs by items of resource use, and the mean direct total health care resource costs. All costs will be discounted to the 2016 price level.

Time to return to work

Information about return to work was recorded at each follow-up time point (Table 1).

Complications and adverse effects

Complications directly related to the interventions were registered. The participants were also encouraged to contact the participating hospitals if any adverse effects occurred and contacts to the health care system were monitored at every follow-up visit. Potential adverse effects (AE) were categorized to serious adverse effects (SAE) and minor adverse effects (MAE) if the participants sought treatment. Death, cardiovascular or gastrointestinal effects, deep venous thrombosis, pulmonary embolism, systemic or local infection were categorised as SAEs and shoulder symptoms like pain, swelling and decreased range of motion were categorised as MAEs. The number and severity of complications and adverse effects will be assessed.

EXPLORATORY OUTCOMES

We have identified three potentially important effect modifying factors. We will perform subgroup analyses with the primary endpoint as the outcome and the direction of hypothesized effect described as below[14].

Duration of symptoms

We will compare the treatment effects stratified based on the duration of symptoms (those with < 6/12 months vs. those > 6/12 months). We hypothesize that subacromial decompression will work better in patients with duration of symptoms > 6 months than for patients with symptoms < 6 months.

Severity of symptoms

We will compare the treatment effects in patients with severe (VAS 70 or more), moderate (VAS 55 to 69), and mild (VAS less than 55) symptoms at baseline. We hypothesize that subacromial decompression will work better in patients with more severe (VAS 70 or more) than moderate (VAS 55 to 69) or mild (VAS less than 55) symptoms at baseline.

Acromial anatomy

We will compare the treatment effects in patients with flat (type I), curved (type II), or hooked (type III) acromion according to classification by Bigliani et al.[15] We hypothesize that subacromial decompression will work better in patients with hooked (type III) than curved (type II) or flat (type I) acromion at baseline.

STUDY DESIGN

Sample size

The sample size calculation was based on the two primary outcome measures, VAS at rest and at arm activity, at 24 months post randomization. FIMPACT trial was powered to detect a minimal clinically important improvement (MCII) in a VAS pain score (improvement of at least 15; assumed standard deviation 25) between ASD and DA (or ET). To achieve a somewhat unconventional (stringent) 90% study power and using a two-sided Type I error rate (5%), our trial requires 68 patients per study group to show clinically meaningful advantage of ASD over DA (or ET). Acknowledging the stringent power threshold, only 3% surplus was reserved for potential loss to follow up/crossovers (3%), and accordingly, the recruitment target was set at 70 patients per treatment group.

Randomization and blinding

To obtain three balanced study groups (of similar group size), we performed a two-fold, sequential randomization. In Phase I, the participants were randomized into non-surgical or surgical treatment with allocation ratio 1:2. In the Phase II, those allocated to surgical treatment were further randomized to ASD or DA with 1:1 ratio. An independent statistician with no clinical involvement in the execution of the trial prepared separate randomization lists for each study centre using a computer-generated algorithm. Randomization was carried out using sequentially numbered sealed opaque envelopes. The envelopes were kept in a secure, agreed location at each centre. To ensure concealment, block randomization was applied using blocks varying in size randomly, the block size known only by the statistician.

To initially enter a participant into the study (Phase I), an envelope containing the treatment assignment [non-surgical (ET) or surgery (ASD or DA), ratio 1:2] was opened during the baseline appointment. Participants randomized to ET started standardized physiotherapy within 2 weeks of the baseline appointment. Participants allocated to surgical treatment were scheduled for surgery aimed to be completed within 12 weeks of randomization.

At the day of surgery, an arthroscopic examination was first carried out to confirm the eligibility of the participant (to rule out full-thickness RC tear and other obvious intra-articular pathology). Research/staff nurse then completed the randomization procedure (Phase II) by opening an envelope containing the surgical treatment allocation (ASD or DA, ratio 1:1). The allocation was revealed to the surgeon by showing the paper, but not expressed verbally.

The full follow-up process is shown in figure 1. In brief, the participants filled in the above noted (mailed) outcome questionnaires at 3, 6, 12 and 24 months post randomization, in addition to which they were also assessed clinically at 6 and 24 months (and 5 and 10 years) post randomisation by a study physiotherapist unaware of treatment allocation, treatment given or possible unblinding. Outcome assessors were instructed not to inquire anything about prior treatment. Further, participants wore a t-shirt on all follow-up examinations. Data analysis will be done in a blinded manner by the study statistician (JR) not directly involved in the study.

STUDY POPULATION

Subject disposition

Study procedures, including recruitment strategies and inclusion and exclusion criteria, are presented in detail in the accompanying actual study protocol.

STATISTICAL ANALYSIS

Data will be analysed in a blinded manner. All p-values will be reported to 3 decimal places with those less than 0.001 reported as $p < 0.001$. The criterion for statistical significance will be set at alpha = 0.05.

Primary analysis

The primary analysis will be carried out according to the intention-to-treat (ITT) principle: participants are retained in the groups to which they were initially randomized. The primary comparison on the efficacy of ASD (ASD vs. DA) will be performed as a between-group comparison using a repeated measures mixed-effects model (RMMM). Study group and time of assessment (baseline, 3, 6, 12 and 24 months) will be included as fixed factors and patient as a random factor. The model will include interactions between study group and time of assessment. The baseline value will be included as a covariate. An unstructured covariance structure will be assumed. If the model cannot be fitted, compound symmetry will be assumed instead. The number of degrees of freedom will be assessed using Satterthwaite's method. The RMMM model will be used to quantify the treatment effect as the difference between the groups in pain scores (VAS) with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization. To safeguard against potential multiplicity bias[2], we will require a statistically significant treatment effect on both of our primary outcome variables, i.e., pain at rest and pain at activity (Table 2).

Table 2: Primary comparison ASD vs ET: Outcomes of the trial at 24 months follow-up.

	ASD ET	Improvement from baseline	Between-Group Difference in Improvement from Baseline
Primary outcomes		ASD ET	
VAS (rest)			
VAS (at arm activity)			
Secondary outcomes			
Constant-Murley Score			
SST			

SF-36			
15D			
Time to return to work			
Return to previous leisure activities			
Responder analysis			
Patients satisfaction to the treatment			
Patients assessment of the treatment allocation			
Complications and adverse effects			

Abbreviations: VAS, visual analogue scale; SST, Simple Shoulder Test; SF-36, Short form- 36

The same statistical model will also apply to the pragmatic comparison of the relative benefits of surgical vs. non-operative treatment strategies on SIS (ASD vs. ET) (Table 3).

Table 3. Secondary comparison ASD vs ET: Outcomes of the trial at 24 months follow-up.

	ASD ET	**Improvement from baseline**	**Between-Group Difference in Improvement from Baseline**
Primary outcomes		**ASD ET**	
VAS (rest)			
VAS (at arm activity)			
Secondary outcomes			
Constant-Murley Score			
SST			
SF-36			
15D			
Time to return to work			
Return to previous leisure activities			

Responder analysis			
Patients satisfaction to the treatment			
Patients assessment of the treatment allocation			
Complications and adverse effects			

Abbreviations: VAS, visual analogue scale; SST, Simple Shoulder Test; SF-36, Short form- 36

Secondary analyses

We will also use the RMMM model to analyse secondary outcomes (Table 2 and 3) where applicable. The results will be reported as the differences between the groups with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization.

Furthermore, instead of focusing only on the statistical significance of the mean differences between treatment groups in the VAS (i.e., the mean improvement from baseline to 24 months), we will also carry out "a responder analysis". In principle, this analysis allows physicians to inform a patient of his or her chance of experiencing a clinically meaningful improvement from the treatment, both in absolute terms and in comparison, to a control group. The difference between responders and non-responders can be considered the net-benefit of the treatment. One proposed means to carry out a responder analysis relies on the assessment of the proportion of patients reaching the patient-acceptable symptom state (PASS) and the patient-disappointing symptoms state (PDSS). As no universal consensus exists on either the PASS or the PDSS in the context of SIS, we chose to anchor our responder analysis to the patient's assessment of satisfaction with the shoulder treatment outcome: Patients reporting very satisfied or satisfied will be categorized as "Responders" and those reporting very dissatisfied or dissatisfied as "Non-responders". Given the obvious coarseness of this approach, we plan to evaluate the appropriate criteria for PASS and PDSS in more detail in the future, exploring the potential contribution of, e.g., arm pain at rest and at activity, shoulder function, and night pain.

Categorical variables, the rates of unblinding, reoperation, treatment conversion, complications and adverse effects will be analysed using logistic regression analysis or Poisson regression dependent on whether subjects with complications or (multiple) complications (per subject) are analysed.

These secondary analyses will be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not a problem[2].

Sensitivity analyses

The following two sensitivity analyses will be carried out: 1) per-protocol analysis, in which the above noted primary analyses will be carried out again with patients who received the interventions as allocated will be redone; 2) and potential effects due to the treatment providing centres.

As all the participants in the ASD group have received the critical therapeutic element (subacromial decompression), no treatment group conversion is possible in this group.

In the per-protocol comparison of the efficacy of ASD (ASD vs. DA), we define the DA per-protocol population as those participants who have not received ASD during the 24-month follow-up (who have not crossed over to ASD).

In the per-protocol comparison of the effectiveness of ASD (ASD vs. ET), we define the ET per-protocol population as those participants who have not received ASD during the 24-month follow-up (who have not crossed over to ASD).

INTERPRETATION OF RESULTS

To safeguard against potential risk of bias during interpretation, a method of "blinded data interpretation" will be used[17]. In brief, an independent statistician will provide the Steering/Writing committee of the FIMPACT trial with blinded results from the analyses with study groups labelled as group A, group B, and group C. This data will be presented to the Steering/Writing Committee, who will then contemplate on the interpretation of the results until a consensus is reached and agree in writing on all alternative interpretations of the findings. Once reaching a consensus, we will record the minutes of this meeting as a statement of interpretation document signed by all members of the Writing Committee. Only after reaching this common agreement will the data manager and independent statistician break the randomization code.

There was also variation in the actual execution of the follow-up assessments, particularly in the earlier time-points (3- and 6-month follow-up visits).

IMPLEMENTATION OF ANALYSIS PLAN

This SAP will be used as a work description for the statistician performing the analyses. All analyses will be performed by the same statistician and none of the investigators involved in this trial will perform any of the statistical analyses.

The implementation of the SAP will be as follows:

1. A 'data collection form' will be outlined in a collaboration between the database manager (Leena

Caravitis), statistician and principal investigators (Mika Paavola and Teppo Järvinen).

2. The database manager will code each treatment arm into ‘treatment A’, ‘treatment B’ and ‘treatment C’, thus leaving all others blinded to group assignment during the analyses.

3. Blinded data will be delivered to the statistician according to the ‘data collection form’.

4. Primary, secondary and exploratory endpoint analyses will be made blinded to group assignment.

5. Results will be presented to the trial Writing and Steering committee, any uncertainties will be clarified and blinded interpretations of the primary endpoint results will be conducted prior to unblinding of data.

REFERENCES

1. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2009;**18**(6):927-32.
2. Ranstam J. Multiple P-values and Bonferroni correction. Osteoarthritis Cartilage 2016;**24**(5):763-4.
3. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clinical orthopaedics and related research 1987(214):160-4.
4. Henseler JF, Kolk A, van der Zwaal P, et al. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2015;**24**(3):376-81.
5. Matsen FA, 3rd, Ziegler DW, DeBartolo SE. Patient self-assessment of health status and function in glenohumeral degenerative joint disease. J Shoulder Elbow Surg 1995;**4**(5):345-51.
6. Godfrey J, Hamman R, Lowenstein S, et al. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. J Shoulder Elbow Surg 2007;**16**(3):260-7.
7. Tashjian RZ, Deloach J, Green A, et al. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am 2010;**92**(2):296-303.
8. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001;**33**(5):328-36.
9. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res 2008;**17**(3):485-95.
10. Bowling A. *Measuring health. A review of quality of life measurement scales.* 3rd ed: Berkshire: Open University Press, 2004.
11. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;**30**(6):473-83.
12. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. Bmj 2002;**324**(7351):1417.
13. Hamilton DF, Lane JV, Gaston P, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. BMJ Open 2013;**3**(4).
14. Sun X, Ioannidis JP, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. Jama 2014;**311**(4):405-11.
15. Bigliani LU MD, April EW The morphology of the acromion and its relationship to rotator cuff tears. Orthop Trans 1986;**10**:228.
16. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc., 1987.
17. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol 2014;**67**(7):769-72.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Interpretation of Blinded Data, Statement of Interpretation

Background assumptions regarding our primary comparison (ASD vs. DA)

1) This superiority RCT is designed to address the true *efficacy* of arthroscopic subacromial decompression (ASD), i.e., can ASD theoretically work? Accordingly, we have chosen patients that – based on the existing literature – represent optimal responders to this index surgical procedure.
2) Conceding that the act of surgery *per se* produces a profound placebo response, a 'true' treatment effect is impossible to disentangle from the nonspecific (placebo or meaning) effects – such as the patients' or researchers' expectations of benefit – without a placebo comparison group.
3) The only difference between ASD and DA treatment groups is that the subacromial decompression, the critical therapeutic (surgical) element, has been carried out for patients in the ASD group.
    a. The critical therapeutic (surgical) element is the component of the surgical procedure that is believed to provide the therapeutic effect (here, subacromial decompression), being distinct from aspects of the procedures that are diagnostic or required to access the disease being treated (here, shoulder arthroscopy).
    b. Apart from the critical therapeutic element, the treatment of the ASD and DA groups is identical, i.e., all "placebo or meaning effect" related to the entire treatment and care is identical.
4) To be deemed effective, ASD should provide a statistically significant benefit over DA in both of the two primary outcomes, pain at rest and activity assessed with a visual analog scale (VAS), as determined by the mean VAS difference between the groups. This is to safeguard against potential multiplicity bias[2].

If ASD is found effective (see above), it should also provide a clinically relevant benefit over DA according to following rationale:

1) There is a proven benefit as follows: Mean VAS-difference between ASD and DA shall exceed the threshold for the minimal clinically important difference (MCID) in VAS. We will consider 15 as the threshold for the minimal clinically important difference (MCID).

AND,

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

2) There is NO proven harm. If there is a proven benefit of ASD but significantly higher proportion of patients show adverse effects, the amount of benefits will be discussed in relation to the frequency and seriousness of the adverse effects.

Statistical commitments:

a) I-T-T is the primary data analysis, but per-protocol analysis will also be carried out.
b) The pre-specified time point of primary interest is 24 months after randomisation.
c) In addition to the two primary outcome parameters, we will also take into account the number of treatment conversions and re-operations, the incidence and seriousness of adverse effects between the ASD and DA groups, and the responder analysis.

Based on these theoretical commitments, our interpretation of the findings will be as follows:

a) If ASD is found superior to DA, the critical therapeutic element of the ASD procedure (subacromial decompression) has a clinically relevant effect on patients with symptoms consistent with SIS.
b) If ASD is not found superior to DA, the critical therapeutic element of the ASD procedure (subacromial decompression) does not have a clinically relevant effect on patients with symptoms consistent with SIS. Considering our efficacy design (study participants are 'optimal responders to ASD' and the surgeons are highly experienced), such finding would imply that ASD does not work at all.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Background assumptions regarding our secondary (independent) comparison (ASD vs. ET)

1) This pragmatic comparison is designed to address whether arthroscopic subacromial decompression (ASD) followed by postoperative rehabilitation is superior to supervised progressive exercise therapy (ET). We recognize that in this pragmatic comparison (ASD vs. ET) the supervised progressive exercise therapy regimen carried out in the ET group is different from the postoperative rehabilitation carried out by patients in the ASD group. In addition, the timing of the start of the actual treatment between the ET and ASD groups was somewhat discrepant due to the time required to arrange the surgery. The ASD patients are also subject to some degree of postoperative immobilization, sick-leave, and modification of pain medication and activities, unlike the patients in the ET group, all of which may also have an effect on the treatment outcome. However, these concord with the current best practice recommendations and the two-year follow-up chosen as our primary time point should dilute the effects of somewhat discrepant timing of the interventions.
2) To be deemed effective, either ASD or ET should provide a statistically significant benefit over ET or ASD, respectively, in both of our two primary outcomes, pain at rest and activity assessed with a visual analog scale (VAS), as determined by the mean VAS difference between the two treatment groups. This is to safeguard against potential multiplicity bias[2].
3) The following concern (apparent confounding) needs to be taken into account in the interpretation. Despite the thorough *preoperative* screening, 10% (14/136) allocated to the two surgical groups had to be excluded because of pathology found after the 1st random allocation. Although this does not have any effect on our primary comparison (ASD vs. DA), the ET and ASD groups are not fully comparable. This discrepancy will possibly skew our results by favouring the ASD group.

Acknowledging all this, if ASD (or ET) is found effective (statistically significant difference in both primary outcomes), it should also provide a clinically relevant benefit over ET (or ASD) according to following rationale:

1) There is a proven benefit as follows: Mean VAS-difference between ASD and ET shall exceed the threshold for the minimal clinically important difference (MCID) in VAS. We will consider 15 as the threshold for the minimal clinically important difference (MCID).

AND,

2) There is NO proven harm. If there is a proven benefit of ASD (or ET) but significantly higher proportion of patients show adverse effects, the amount of benefits will be discussed in relation to the frequency and seriousness of the adverse effects.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Statistical commitments:

a) I-T-T is the primary data analysis, but per-protocol analysis will also be carried out.
b) The pre-specified time point of primary interest is 24 months after randomisation.
c) In addition to the two primary outcome parameters, we will also take into account the number of treatment conversions and re-operations, the incidence and seriousness of adverse effects between the ASD and ET groups, and the responder analysis.
d) Given the Background assumption 3.a) (above, the discrepant number of patients with a shoulder pathology other than SIS), we will carry out a worst-case analysis by creating a subgroup of the ET group by removing seven (an equal number of patients excluded from both surgical treatment arms due to pathology found after 1st randomization) worst-cases/highest VAS-pain scores at the primary analysis time-point (24 months). The number of removed cases is based on the assumption that the prevalence of shoulder pathology is identical in the randomized population, while the decision to remove the individual with the highest VAS-pain scores at the end of the study basis on the assumption that shoulder pathology is an effect-modifying factor, predicting poor outcome.

Based on these theoretical commitments, our interpretation of the findings will be as follows:

a) If ASD is found superior to ET in both the complete case and the sensitivity (subgroup) analyses, ASD is a more effective treatment option than ET for patients with SIS.
b) If ET is found superior to ASD in both the complete case and the sensitivity (subgroup) analyses, our results suggests that ET is a more effective treatment option than ASD for patients with subacromial pain syndrome.
c) If there are no statistically significant differences between ASD and ET, ASD and ET are equally effective.

Supplementary 1: Patients satisfaction to the treatment given

How satisfied are you with the outcome of your treatment? Mark the answer closest to your situation.

1. Very satisfied, my shouder has healed completely.
2. Satisfied, I have only minor, activity related symptoms. My shoulder is much better than before treatment.
3. Somewhat satisfied, i have only minor symptoms. My shoulder is better than before treatment.
4. Dissatisfied, my shoulder is the same as before treatment.
5. Very dissatisfied, my shoulder is worse than before treatment.

BMJ Open

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT): A protocol for a randomized trial comparing arthroscopic subacromial decompression and diagnostic arthroscopy (placebo control), with an exercise therapy control, in the treatment of shoulder impingement syndrome

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014087.R2
Article Type:	Protocol
Date Submitted by the Author:	13-Feb-2017
Complete List of Authors:	Paavola, Mika; Helsinki University Central Hospital, Orthopaedics and traumatology Malmivaara, Antti; National Institute for Health and Welfare, Centre for Health and Social Economics Taimela, Simo; University of Helsinki, Department of Orthopedics and Traumatology Kanto, Kari; Hatanpaan kantasairaala Jarvinen, Teppo; University of Helsinki, Department of Orthopaedics and Traumatology
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Evidence based practice, Rehabilitation medicine, Sports and exercise medicine, General practice / Family practice
Keywords:	Orthopaedic & trauma surgery < SURGERY, Clinical trials < THERAPEUTICS, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, REHABILITATION MEDICINE, SPORTS MEDICINE

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT): A protocol for a randomized trial comparing arthroscopic subacromial decompression and diagnostic arthroscopy (placebo control), with an exercise therapy control, in the treatment of shoulder impingement syndrome

Mika Paavola[1], Antti Malmivaara[2], Simo Taimela[1], Kari Kanto[3], and Teppo L N Järvinen[1], for the FIMPACT investigators*

[1] Department of Orthopedics and Traumatology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

[2] National Institute for Health and Welfare, Center for Health and Social Economics, Helsinki, Finland

[3] Department of Orthopedics and Traumatology, Hatanpää City Hospital, Tampere, Finland

Address of corresponding author:
Mika Paavola, MD, PhD
Department of Orthopedics and Traumatology
Helsinki University Hospital / Töölö Hospital
Topeliuksenkatu 5
P.O. Box 266, 00029 HUS,
Helsinki, Finland
Phone: +358-50-4272481
Email: mika.paavola@hus.fi

* FIMPACT investigators, see appendix 1.

ABSTRACT

Introduction: Arthroscopic subacromial decompression (ASD) is the most commonly performed surgical intervention for shoulder pain, yet evidence on its efficacy is limited. The rationale for the surgery rests on the tenet that symptom relief is achieved through decompression of the rotator cuff tendon passage. The primary objective of this superiority trial is to compare the efficacy of ASD versus diagnostic arthroscopy (DA) in patients with shoulder impingement syndrome (SIS), where DA differs only by the lack of subacromial decompression. A third group of supervised progressive exercise therapy (ET) will allow for pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies.

Methods and Analysis: FIMPACT trial is an ongoing multicentre, three-group randomised controlled study. We performed two-fold concealed allocation, first by randomizing patients to surgical (ASD or DA) or conservative (ET) treatment in 2:1 ratio and then those allocated to surgery further to ASD or DA in 1:1 ratio. Our two primary outcomes are pain at rest and at arm activity, assessed using visual analog scale (VAS). We will quantify the treatment effect as the difference between the groups in the change in the VAS scales with the associated 95% confidence interval (CI) at 24 months. Our secondary outcomes are functional assessment (Constant score and Simple shoulder test), quality of life (15D and SF-36), patient satisfaction, proportions of responders and non-responders, reoperations/treatment conversions, all at 2 years post-randomization, as well as adverse effects and complications. We recruited a total of 210 patients from 3 tertiary referral centres. We will conduct the primary analysis on the intention-to-treat basis.

Ethics and Dissemination: The study was approved by the institutional review board of the Pirkanmaa Hospital District and duly registered at ClinicalTrials.gov. The findings of this study will be disseminated widely through peer-reviewed publications and conference presentations.

Trial registration: ClinicalTrials.gov NCT00428870 (first registered January 29, 2007).

Keywords: Acromion; Acromioplasty; Arthroscopy; Impingement; Physiotherapy; Placebo; Sham; Shoulder; Syndrome; Randomised; Trial

Strengths of this study

- Efficacy design: Strict eligibility criteria
- Placebo-surgery controlled trial: Blinding of both the participants and the outcome assessors in the comparison between index surgery and control (placebo surgery)
- Inclusion of a non-surgical treatment option to allow a pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies

Limitations of this study

- Potential confounding due to participants' knowledge of the treatment delivered in our secondary comparison between surgical and non-operative treatments

INTRODUCTION

Subacromial decompression is one of the most frequently performed procedures in orthopaedics[1 2]. It is carried out to treat patients with shoulder pain attributed to "subacromial impingement syndrome" (SIS). Conventional wisdom dictates that SIS is caused by 'impingement' of the rotator cuff (RC) between the humeral head and the overlying acromion while lifting the arm. The appropriateness of this mechanistic explanation has been challenged lately where the generic label of "subacromial pain syndrome" (SAPS) is currently advocated[3 4]. The aim of the subacromial decompression procedure, typically carried out arthroscopically, is to decompress the RC tendon passage through the subacromial space through resection and smoothening of the hypertrophied or prominent anterolateral undersurface of the acromion. Management of shoulder pain has been estimated to account for 4.5 million visits annually to physicians in the USA alone[5], accounting for US$3 billion in costs each year[6]. Since 44-65% of all shoulder complains are related to SIS, it is estimated that annual direct medical costs of SIS are over $1 billion in the USA[7 8].

Since the introduction of subacromial decompression surgery in the early 1970s[9], the number of procedures has steadily increased across the entire western world. With the advent of arthroscopy, the number of these surgeries has increased dramatically -- 5-fold from the 1980s to 2005 in the US[10] and 700% between 2000 and 2010 in the UK[11]. Remarkably, there is a stark absence of evidence from high-quality controlled trials to support the existing practice of performing subacromial decompression for patients with SIS. Two recent systematic reviews concluded that subacromial decompression provides no superior benefits in terms of pain relief, function, or quality of life compared to non-surgical treatment[12 13]. There is even a placebo controlled trial to show the beneficial effect of exercise therapy over placebo physiotherapy[14]. However, the proponents of the procedure have argued that the evidence is skewed in favour of the therapeutic potential of surgery due to a significant cross-over (5-15%) from conservative treatment to surgery[14-16]. Although such concern is obviously warranted, it should also be recalled that surgeons' own perceptions on the success of any surgery might similarly be biased due to a considerable surgical placebo effect.

The outcome of any medical (surgical) intervention – particularly when treating primarily subjective symptoms – is a cumulative effect of three main elements: placebo effects, critical therapeutic (surgical) element, and non-specific effects, most importantly, the normal variation in the course of the disease and the regression-to-the-mean phenomenon[17 18]. Conceding that the act of surgery *per se* produces a profound placebo response, a 'true' treatment effect is impossible to disentangle from the nonspecific (placebo) effects – such as the patients' or researchers' expectations of benefit – without a placebo

comparison group[19]. The critical therapeutic element is the component of the surgical procedure that is believed to provide the therapeutic effect (here, subacromial decompression), which are distinct from aspects of the procedures that are diagnostic or required to access the disease being treated (here, shoulder arthroscopy).

To the best of our knowledge, there is only one other ongoing study aiming to assess the true efficacy of subacromial decompression surgery in patients with SIS using a placebo controlled study design. According to the published protocol of this CSAW trial[20], the investigators have chosen a very similar approach to that of our FIMPACT trial. In brief, the CSAW trial is a three-group pragmatic RCT comparing arthroscopic acromioplasty, active monitoring with specialist reassessment, and investigational shoulder arthroscopy only. CSAW aims for recruitment of 300 patients with SIS to assess the efficacy of the surgery against no surgery, the need for a specific component of the surgery (acromioplasty), and the quantification of the possible placebo effect. As readily apparent, the two trials (FIMPACT vs. CSAW) are very similar in design with the only notable differences being the primary outcome measure (Pain at rest and after activity vs. Oxford Shoulder Score, a score that assesses both pain and ADL impairment), the primary outcome assessment point (24 months vs. 6 months), and the intervention delivered for the third group (exercise therapy vs. active monitoring with specialist reassessment), respectively.

The primary hypothesis of our FIMPACT trial is that ASD is superior to DA in patients with SIS. In addition, we will perform a pragmatic comparison of surgical and non-surgical treatment options (ASD vs. ET). The relative benefits of ASD and ET will be assessed without a priori hypothesis of the superiority of one or the other.

MATERIALS AND METHODS

Overview of study design

FIMPACT trial is an ongoing multicentre, three group randomised controlled superiority study with a primary objective to assess the efficacy of the ASD vs. DA in patients diagnosed with SIS. Our design also enables the pragmatic comparison of surgical and non-surgical treatment strategies (ASD vs. ET) (Figure 1). To obtain three balanced study groups (of similar group size), we performed a two-fold, sequential randomization as follows: First, we randomized patients to surgical or conservative treatment in 2:1 ratio and then randomized those allocated to surgery to ASD or DA in 1:1 ratio. The initial patient screening for the trial began at one site (Tampere) in February 1, 2005 and was then expanded to two additional tertiary referral centres in March 2006 and December 2006 to improve recruitment and overall generalisability of the results. The recruitment was completed (all 210 required patients enrolled) in August 2013.

Ethical approval

Ethical approval was obtained on December 28, 2004 from the institutional review board (IRB) of the Pirkanmaa Hospital District (R04200). Local research and development approvals were gained for each recruiting centre.

Participant selection

We assessed for eligibility all patients complaining of subacromial shoulder pain to any of the participating clinics. These participants were screened according to the inclusion and exclusion criteria and a recruitment surgeon confirmed the clinical diagnosis of SIS. To qualify as a recruitment surgeon, all trial surgeons had to have experience of more than 500 shoulder arthroscopies before the start of the trial. Detailed clinical examination of the shoulder was performed on all referred patients to rule out possible instability, clinical signs of rotator cuff rupture, frozen shoulder or other causes of symptoms. Standard x-rays and MRI were obtained from all potential participants and assessed by both a musculoskeletal radiologist and an orthopaedic surgeon. For patients found eligible for this study (fulfilling indications for ASD), we obtained written informed consent and randomised them into non-operative or operative group (1:2) immediately after the baseline appointment. If patient had bilateral symptoms, only one shoulder was included in the study.

Eligibility criteria

We used specific eligibility criteria to ensure that recruited participants were only those with SIS. Accordingly, a standardized clinical examination was first performed, followed by a subacromial injection test. To exclude patients with concomitant pathology, particularly rotator cuff rupture, standard x-rays and magnetic resonance imaging with intra-articular contrast injection (MRA) were carried out on all potential participants.

Inclusion criteria

1) Adult men or women ages 35 to 65 years
2) Subacromial pain for greater than 3 months with no relief from non-operative means (physiotherapy, non-steroidal anti-inflammatory medication, corticosteroid injections, and rest)
3) Pain provoked by abduction and positive painful arc -sign
4) Positive impingement test (temporary relief of pain by subacromial injection of lidocaine)
5) Pain in at least 2 out of 3 of isometric tests (abduction 0° and 30° or external rotation)
6) Provision of informed consent from the participant
7) Ability to speak, understand and read in the language of the clinical site

Exclusion criteria

1) Full thickness tear of the rotator cuff tendons diagnosed on clinical examination (marked weakness in any of the examined muscles) or magnetic resonance imaging with intra-articular contrast (MRA)
2) Osteoarthritis of the glenohumeral and/or acromioclavicular joint diagnosed on clinical examination or on x-rays
3) Substantial calcific deposits in the rotator cuff tendons found in the preoperative imaging
4) Previous surgical procedure on the affected shoulder
5) Evidence of shoulder instability (positive apprehension/positive sulcus sign)
6) Symptomatic cervical spine pathology
7) History of alcoholism, drug abuse, psychological or psychiatric problems that are likely to invalidate informed consent
8) Patient declined to participate

Recruitment process

Consultant orthopaedic surgeons carried out eligibility screening among patients referred to the study centres through standard clinical practice for shoulder pain. Patients meeting the eligibility criteria were introduced to the study. If patients expressed interest in participating, written information about the study was provided and they were asked to opt in. If the interest continued, arrangements were made for obtaining required imaging (x-rays and MRA) and for a separate baseline appointment.

Informed consent

At the first appointment, all participants were introduced to the detailed written information about the study and asked to sign a written informed consent form. At the baseline appointment (arranged within 45 days of initial contact), baseline data was completed and participant's willingness to participate in the study was confirmed. This procedure ensured that all potential participants had a reflection period for consent of at least 48 hours before giving their final consent to participate. Particular attention was paid to ensure that the participants realized that on entering the study they may receive only diagnostic arthroscopy, in which case the subacromial decompression would not be performed. They were also informed that participation in the study is entirely voluntary and any decision they make would not affect their possible future care. In addition, every participant was informed of their right to withdraw from the trial whenever they desire without the need to supply any reason for such decision.

Baseline assessment

Baseline assessment included documentation of gender, birth date, education, employment, hand dominance, time from the onset of symptoms, recreational habits, and employment status. We asked participants to assess their general heath and usage of pain medication. Modalities of any prior conservative treatment were also recorded (Table 1).

Table 1: Baseline characteristics

	ASD	DA	ET
Age (years), mean (SD)			
Gender (female/male), n (%)			
Dominant hand affected, n (%)			
Social economic status/ work load			
Heavy manual labor (construction work etc.), n (%)			
Heavy manual labor (variable workload), n (%)			
Mostly manual labor including daily office work, n (%)			
Mostly office work with occasional manual assignments, n (%)			
Full-time office work, n (%)			

Unemployed, n (%)	
Pensioner/disability pensioner, n (%)	
Student, n (%)	
Homemaker/housewife/other, n (%)	
Subjective health	
Duration of symptoms (Months), mean (SD)	
Ability to work normally regardless of the shoulder symptoms? (yes/no), n (%)	
Recreational ability regardless of the shoulder symptoms? (yes/no), n (%)	
Prior treatments	
Rest, n (%)	
Pain medication, n (%)	
Topical pain medication, n (%)	
Corticosteroid injection, n (%)	
Ultrasound, laser or any other similar therapies, n (%)	
Physiotherapy including exercise therapy, n (%)	
Other, n (%)	
Generic health states	
15D	
SF-36	
Pain measurements/Shoulder scores	
Pain at rest (0-100 VAS scale), mean (SD)	
Pain during activity (0-100 VAS scale), mean (SD)	
Constant- Murley score (CM), mean (SD)	
The simple shoulder test (SST), mean (SD)	

Baseline clinical symptoms

The recruiting surgeon carried out a clinical history and a clinical examination related to shoulder pain. Shoulder complaints other than SIS, such as full-thickness rotator cuff tears, frozen shoulder, osteoarthritis of the acromioclavicular joint and instability were ruled out as much as clinical diagnosis allows.

Baseline imaging

Standard x-rays of the shoulder were obtained to assess possible glenohumeral or acromioclavicular osteoarthritis. A magnetic resonance image with intra-articular contrast medium (MRA) was also obtained to rule out any other intra- or extra-articular pathologies. A musculoskeletal radiologist and an orthopaedic surgeon assessed all the images.

Randomisation and concealment

We used a two-phase sequential randomization. In Phase I, the participants were randomized into non-surgical or surgical treatment with allocation ratio 1:2. In the Phase II, those allocated to surgical treatment were further randomized to ASD or DA with 1:1 ratio (Figure 1).

An independent statistician with no involvement in the execution of the trial prepared separate randomization lists for each study centre using a computer-generated algorithm. Randomization was carried out using sequentially numbered sealed opaque envelopes. The envelopes were kept in a secure, agreed location at each centre. To ensure concealment, block randomization was applied using blocks varying in size randomly, the block size known only by the statistician.

To initially enter a participant into the study (Phase I), an envelope containing the treatment assignment [non-surgical (ET) or surgery (ASD or DA), ratio 1:2] was opened during the baseline appointment. Participants randomized to ET started standardized physiotherapy within 2 weeks of the baseline appointment. Participants allocated to surgical treatment were scheduled for surgery aimed to be completed within 12 weeks of randomization.

At the day of surgery, a diagnostic arthroscopy was first carried out to confirm the eligibility of the participant (to rule out full-thickness RC tear and other obvious intra-articular pathology). Research/staff nurse then completed the randomization procedure (Phase II) by opening an envelope containing the surgical treatment allocation (ASD or DA, ratio 1:1). The allocation was revealed to the surgeon by showing the paper, but not expressed verbally.

Interventions

Diagnostic arthroscopy (DA)

All participants in the two operative groups first underwent arthroscopic examination of the shoulder with the use of standard posterior and lateral portals and a 4-mm arthroscope. To maintain concealment, the surgery was carried out under general anesthesia. The orthopaedic surgeon evaluated and graded possible intra-articular pathologic changes. The rotator cuff integrity was evaluated also from the subacromial space without performing routine bursectomy. If the integrity of the rotator cuff could not be assessed, bursal tissue was bluntly stretched with troachar or resected on the tendon side to allow visualisation. If arthroscopic examination revealed any unexpected pathology (such as capsular pathology, full-thickness rotator cuff tear, or osteoarthritis), the patient was treated according current clinical practice guidelines for the given pathology while under the same anesthesia. In such a case, the participant

was excluded from the trial. Patients with partial tears were included in the study, while patients with a full-thickness tear were excluded and rotator cuff repair was carried out.

After the arthroscopic examination of the glenohumeral joint and subacromial space, confirming the eligibility of the participant, the participants were randomly assigned to receive either ASD or DA only. If the patient was allocated to the DA group, the operation was terminated. To ensure concealment of the participants and the staff other than those in the operating theatre, the participants were kept in the operating theatre for the required time to perform subacromial decompression.

Arthroscopic subacromial decompression (ASD)

Debridement of the subacromial bursa was performed with a shaver and/or electrocoagulation, followed by the resection of the bony spurs and projecting anterolateral undersurface of the acromion by a shaver as described by Ellman[21].

Postoperative care

In both the ASD and the DA group, the postoperative rehabilitation was identical. All surgically treated participants received one visit to an independent physiotherapist for guidance and instructions for home exercises. Subsequent rehabilitation was carried out according to the standardized rehabilitation protocols of the participant centres. Since the initial rehabilitation after a surgery needs to be "tempered" due to joint irritation, the rehabilitation protocol of the operatively treated groups (ASD and DA) was not identical to the ET group.

Exercise therapy (ET)

In the exercise therapy (ET) group, supervised progressive physiotherapy was started within 2 weeks of randomization using a standardized protocol. The protocol was based on the same principles as the regimen shown effective for the treatment of SIS earlier[14], but was updated – with the help of the principal investigator of the original study[14] – to conform with the state-of-the-art exercise therapy for SIS. The regimen was based on daily home exercises, but also included 15 visits to an independent physiotherapist for guidance and monitoring of the progress, carried out approximately once a week. The aim of the supervised exercise treatment was to restore painless, normal mobility of the shoulder girdle, eliminate any capsular tightness and to increase the dynamic stability of the glenohumeral joint and the scapula.

Compliance to treatment allocation and possible crossover

Participants allocated to ET group were told at the time of giving consent that they would be allowed to consider crossing over to the ASD group if they didn't get adequate relief of symptoms (preferably no sooner than 6 months post randomization). Similarly, in the two surgical treatment groups, the participants were informed of the possibility of unblinding if debilitating symptoms persisted 6 months or more after operation. If the participant was allocated to DA group, ASD was then offered. If the participant had undergone ASD, he/she was offered extended physiotherapy. No pre-specified criteria were used for determining "inadequate relief of symptoms/debilitating symptoms", rather it was left to the participants and the study physicians to make the clinical judgment together.

Outcome measures

The outcomes used in this study and the timetable for follow-up assessments are summarised in Table 2.

Table 2: Outcomes and follow-up time points

Assessment	Screening	Enrolment (Baseline)	Surgery	3 Months	6 Months	12 Months	24 Months	5 years	10 years
Screening form	X								
Informed consent		X							
Baseline characteristics form		X							
X-ray and MRI	X								X
Randomisation		X (1st)	X (2nd)						
Arthroscopic findings form			X						
Follow-up form*				X		X			
Clinical examination		X			X		X	X	X
Complications/adverse effects form**			(X)	(X)	(X)	(X)	(X)	(X)	(X)
VAS, at rest		X		X	X	X	X	X	X
VAS, at arm activity		X		X	X	X	X	X	X
Constant- Murley Score		X			X		X	X	X
Simple Shoulder Test (SST)		X			X		X	X	X
SF-36		X		X	X	X	X	X	X
15D		X		X	X	X	X	X	X
Return to work				X	X	X	X	X	X
Return to previous leisure activities				X	X	X	X	X	X
Responder analysis				X	X	X	X	X	X
Patients satisfaction to the treatment				X	X	X	X	X	X
Patients assessment of the treatment allocation				X					
Health resource utilization				X	X	X	X	X	X

* Letter/telephone interview

** If required

Primary outcome measure

VAS

As the primary outcome measure, we used a visual analogue scale (VAS) to measure the patient's perceived pain intensity at rest and at arm activity during the 24 hours preceding the assessment. Shoulder pain was assessed on a 100 mm scale ranging from 0 (no pain) to 100 (extreme pain). We considered 15 as the minimal clinically important difference (MCID) for VAS[22].

Secondary outcome measures

Constant-Murley score

Constant-Murley score (CS) is the most commonly used scoring system for evaluation of various disorders of the shoulder[23]. It consists of both objective (range of motion and strength) and subjective measurements (pain assessment, work load, and leisure time activities), which are summarized in a score between 0 and 100. A higher score indicates better shoulder function. The minimal detectable change (MDC) of the Constant score is 17 for patients with SIS[24].

In addition, as night pain is considered one of the hallmark symptoms in patients with SIS and our two primary outcome measures (patient's perceived pain intensity at rest and at arm activity in the last 24 hours) do not specifically address this issue, a specific question from the Constant-Murley score (unaffected sleep: "Yes" or "No") will be analysed separately.

SST

The simple shoulder test (SST) was developed to assess any impairment of the patient's activities of daily living [25]. The SST consists of 12 questions with yes (1) or no (0) response options. The maximum SST score is 12 indicating normal shoulder function, minimum score of 0 points refers severely diminished shoulder function. The SST has good reliability and responsiveness in patients with rotator cuff symptoms[26]. The MCID for the SST in rotator cuff disease is 2 points[27].

15D

The 15D instrument is a generic health-related quality of life (HRQoL) instrument comprising 15 dimensions[28]. For each dimension, the respondent must choose one of the five levels that best describes his/her state of health at that moment (the best level being 1 and the worst level being 5). A set of utility or preference weights is used in an addition aggregate formula to generate a single index number, the utility or 15D score. The maximum 15D score is 1 (no problems on any dimension) and the minimum score is 0 (being dead). The responsiveness, reliability and validity of 15D have been thoroughly established, and this instrument has been used extensively in clinical and healthcare research[29 30].

SF-36

Short form or SF-36 is a generic HRQoL instrument to quantify the physical, functional, and psychological aspects of health-related quality of life. It consists of 36 questions in eight subscales that assess physical, functional, social, and psychological well-being[31]. Score ranges from 0 to 100, where a higher score is associated with better health. The physical and mental component summary scales (PCS and MCS, respectively) are then calculated as composites of the related subscales. SF-36 is one of most widely used measure of health-related quality of life[32].

Patient satisfaction and Responder analysis

We elicited patients' global assessment of satisfaction to the treatment with this question: "Are you satisfied with the treatment you have received?" We used a VAS scale ranging from 0 (completely disappointed) to 100 (completely satisfied).

Additionally, we elicited patient satisfaction to the treatment outcome with the following question at each follow-up time point (Table 2): "How satisfied are you with the outcome of your treatment?" on a 5-item scale. Participants who reported very satisfied or satisfied will be categorized as "Responders" and patients who responded very dissatisfied or dissatisfied as "Non-responders".

Return to previous leisure activities

Similarly, at each follow-up (Table 2), participants were asked to respond to the following question: "Have you been able to return to your previous leisure activities?" ("yes" or "no").

Patients' perception of operative treatment-group assignment

At the 3-month follow-up point, the patients in the two operative groups were asked to guess whether they had undergone ASD or DA.

Health resource utilization and costs

For the cost-effectiveness analysis, at each follow-up visit the participants were asked to fill in a questionnaire inquiring about the use of healthcare resources. The questionnaire contains a list of items of healthcare resources available and the participants were asked to fill in the number of visits per item during the recall period of each follow-up time point. The resource use will be calculated based on the number of visits times unit cost per item and expressed as mean costs by items of resource use, and the mean direct total health care resource costs. All costs will be discounted to the 2016 price level.

Time to return to work

Information about return to work was recorded at each follow-up time point (Table 2).

Complications and adverse effects

Complications directly related to the interventions were registered. The participants were also encouraged to contact the participating hospitals if any adverse effects occurred and contacts to the health care system were monitored at every follow-up visit. Potential adverse effects (AE) were categorized to serious adverse effects (SAE) and minor adverse effects (MAE) if the participants sought treatment. Death, cardiovascular or gastrointestinal effects, deep venous thrombosis, pulmonary embolism, systemic or local infection were categorised as SAEs and shoulder symptoms like pain, swelling and decreased range of motion were categorised as MAEs. The number and severity of complications and adverse effects will be assessed.

Follow-up

The full follow-up process is shown in figure 1. In brief, the participants filled in the above noted (mailed) outcome questionnaires at 3, 6, 12 and 24 months post randomization, in addition to which they were also assessed clinically at 6 and 24 months (and 5 and 10 years) post randomisation by a

study physiotherapist unaware of treatment allocation, treatment given or possible unblinding. Outcome assessors were instructed not to inquire anything about prior treatment. Further, participants wore a t-shirt on all follow-up examinations.

Adherence and loss to follow-up

Several procedures were implemented to limit loss to follow-up, including excluding individuals likely to pose suboptimal adherence to study follow-up, obtaining verified contact information from each consented participant, and having a local research nurse remind participants of upcoming follow-up/clinic visits. All attempts were made to make follow-up as convenient for the patients as possible. Participants were required to visit the outpatient clinic only at 6 months and 24 months (and 5 and 10 years) post randomisation, while the 3- and 12-month follow-ups were carried out using mailed questionnaires to minimize inconvenience to the participants. The follow-up visits had no more discomfort for the participant than the routine clinical shoulder examinations. The follow up schedule did not involve extra costs to the participants. Follow-up rate was monitored throughout the trial and patients who did not return follow-up questionnaires would receive reminder telephone calls. Using strategies highly similar to these in our previous placebo-surgery controlled trial[33], a 99% follow-up rate was achieved.

The number and proportion of individuals eligible for and compliant with each follow-up was documented. Individuals who died during the study (from causes unrelated to the study or procedure) will be tabulated. An analysis of the demographic and prognostic characteristics will be carried out between the individuals who withdrew and those who remained in the study. For continuous variables, parametric or non-parametric analysis of variance will be used. For categorical variables, $\chi 2$ or Fisher's exact test will be applied.

Missing items

We will use multiple imputation to handle missing data for those statistical analyses that cannot handle occasional missing values. All variables to be included in the final analyses will be included in the chained equations imputation model. The imputation algorithm, fully conditional specification (FCS), uses a specific univariate model for each variable and, for each specific imputed dataset, iteratively imputes each variable with missing values and uses the imputed values in the imputation of other variables.

Sample size

The sample size calculation was based on the two primary outcome measures, VAS at rest and at arm activity, at 24 months post randomization. FIMPACT trial was powered to detect a minimal clinically important improvement (MCII) in a VAS pain score (improvement of at least 15; assumed standard deviation 25) between ASD and DA (or ET). To achieve a somewhat unconventional (stringent) 90% study power and using a two-sided Type I error rate (5%), our trial requires 68 patients per study group to show clinically meaningful advantage of ASD over DA (or ET). Acknowledging the stringent power threshold, we reserved only 3% surplus for potential loss to follow up/crossovers (3%), and accordingly, we set the recruitment target at 70 patients per treatment group.

Safety analysis

There are no anticipated safety issues with the FIMPACT Study. Identically to our previous placebo-surgery controlled trial[33], an interim analysis, as requested by the ethics board, was carried out after the enrolment of 45 participants by an independent data and safety monitoring board (the National Institute for Health and Welfare) to ensure that the rates of complications or reoperations were within acceptable limits (within the normal rate of complications and/or reoperations related to shoulder arthroscopy). Since we found no marked discrepancy in our crude assessment of the incidence of complications/reoperations, no unsealing of group assignments (unblinding) was carried out. No other interim analysis was carried out.

Data management

Questionnaire forms on paper were the primary data collection tools for the study. Upon receipt of the questionnaire forms, a study nurse made a visual check of the responses and queried missing data when possible. Research assistants, blinded to the group allocation, stored the forms into an electronic database by double data entry to minimize typing errors. The researchers, blinded to the group allocation, perform a visual check of the data in the electronic database and then queried all missing, implausible, and inconsistent data. Patient records in the participating hospitals were used when collecting missing data or interpreting inconsistent or implausible data. The final analysis was performed on data transferred to the file "FIMPACT-full data_final", having been documented as meeting the cleaning and approval requirements of our independent statistician and after the finalisation and approval of the accompanying statistical analysis plan (SAP) document. Participant files will be maintained in storage (both in

electronic and paper format) at the coordinating centre for a period of 10 years after completion of the study (10 year follow-up visits).

STATISTICAL METHODS

Statistical Analysis plan (SAP)

A statistical analysis plan (SAP) is published along this protocol. An independent statistician who is unaware of the group assignments will perform all the analyses.

We will summarise the baseline characteristics of the participants by group, reported as a mean (standard deviation) or median (first quartile, third quartile) for continuous variables and count (percent) for categorical variables.

We will analyse the data in a blinded manner. All p-values will be reported to 3 decimal places with those less than 0.001 reported as $p < 0.001$. The criterion for statistical significance will be set at alpha = 0.05.

Primary analysis

We will carry out the primary analysis according to the intention-to-treat (ITT) principle: participants are retained in the groups to which they were initially randomized.

The primary comparison will be on the efficacy of ASD (ASD vs. DA). We will perform the primary comparison on the efficacy of ASD (ASD vs. DA) as a between-group comparison using a repeated measures mixed-effects model (RMMM). Study group and time of assessment (baseline, 3, 6, 12 and 24 months) will be included as fixed factors and patient as a random factor. The model will include interactions between study group and time of assessment. The baseline value will be included as a covariate. The RMMM model will be used to quantify the treatment effect as the difference between the groups in pain scores (VAS) with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization. To safeguard against potential multiplicity bias[34], we will require a statistically significant treatment effect on both of our primary outcome variables, i.e., pain at rest and pain at activity.

The same statistical model will also apply to the pragmatic comparison of the relative benefits of surgical vs. non-operative treatment strategies on SIS (ASD vs. ET).

Secondary analyses

We will also use the RMMM model to analyse secondary outcomes where applicable. The results will be reported as the differences between the groups with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization.

Categorical variables, reoperations or treatment conversions, and complications as well as adverse effects will be analysed using logistic regression analysis or Poisson regression dependent on whether subjects with complications or (multiple) complications (per subject) are analysed.

These secondary analyses will be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not a problem[2].

Sensitivity analyses

We will carry out the following sensitivity analyses: 1) per-protocol analyses, in which the above noted primary and secondary analyses will be carried out again with patients who received the interventions as allocated; 2) and potential effects due to the treatment providing centres.

Subgroup analyses and Hypothesized Effects

We have identified three important subgroups. We will perform these three subgroup analyses with the primary endpoint as the outcome and the direction of hypothesized effect described[35]:

1) Duration of symptoms – Neer originally suggested that ASD should be considered for patients with persistent symptoms despite over one year of conservative treatment[36]. Recent RCTs failing to find efficacy on ASD (vs. conservative treatment) have prompted arguments that ASD should be reserved to situations when long-term conservative treatment has failed[37]. Although a recent study specifically addressed this question and failed to support this hypothesis[38], we still intend to compare the treatment effects of participants stratified based on the duration of symptoms. Accordingly, we will compare those with symptoms less than 12 months to those with symptoms longer than 12 months. We hypothesize that subacromial decompression will work better in patients with duration of symptoms > 12 months than for patients with symptoms < 12 months.
2) Severity of symptoms - A subgroup analysis will also be conducted comparing the treatment effects in patients with severe (VAS 70 or more), moderate (VAS 55 to 69), and mild (VAS less than 55) symptoms at baseline. We hypothesize that subacromial decompression will work better in patients

with more severe (VAS 70 or more) than moderate (VAS 55 to 69) or mild (VAS less than 55) symptoms at baseline.

3) Acromial anatomy - A hook-type acromion has been suggested as an independent risk factor for subacromial impingement[39]. To assess the validity of this suggestion, a subgroup analysis will be conducted comparing the treatment effects in patients with flat (type I), curved (type II), or hooked (type III) acromion according to classification by Bigliani et al.[40] We hypothesize that subacromial decompression will work better in patients with hooked (type III) than curved (type II) or flat (type I) acromion at baseline.

Effect modifying and mediating factors

Multiple regression models were used to assess the potential effect modifying factors (e.g., age, gender, psychological well-being, mental health, occupational shoulder load, education level, and hand dominance) and effect mediating factors (e.g., absence of complications and adherence to rehabilitation) on pain, functional disability and quality of life. These analyses are supportive, explanatory and/or hypothesis generating.

Blinded data interpretation

To safeguard against potential risk of bias during interpretation, we will use our recently introduced method of "blinded data interpretation"[41]. So far, this method has been successfully applied to three previous trials[33 42 43]. In brief, an independent statistician will provide the Writing committee of the FIMPACT trial (authors of this protocol) with blinded results from the analyses with study groups labelled as group A, group B, and group C. The Writing Committee will then contemplate on the interpretation of the results until a consensus is reached and agree in writing on all alternative interpretations of the findings. Once reaching a consensus, we will record the minutes of this meeting as a statement of interpretation document signed by all members of the Writing Committee. Only after reaching this common agreement will the data manager and independent statistician break the randomization code.

DISCUSSION

In this protocol paper, we describe the execution of a randomised, placebo-surgery controlled trial for the assessment of the efficacy of arthroscopic subacromial decompression (ASD) in patients with sub-acromial impingement syndrome (SIS). Acknowledging the potential of surgery to produce powerful placebo effects[44], our primary comparator is diagnostic arthroscopy, differing from the ASD only by lacking the critical therapeutic element of the ASD (subacromial decompression). We will also conduct the pragmatic comparison of surgical and non-surgical treatment options of SIS by including a third group of progressive exercise therapy (ET) (Figure 1, ASD vs. ET).

Interpretations and generalizability

Our interpretation scheme primarily rests on the tenet that the minimum requirement for the clinical viability of ASD is that it needs to show superiority to DA - a therapeutically inert and thus a clinically non-viable option. To test this, we have chosen a classic *efficacy* or *"can it work" (proof-of-concept)* design[45-47]: The recruited participants are those who - according to current evidence - should have an "optimal response" to ASD and the participants and outcome assessors are blinded to the interventions given. This design should thus yield findings that are widely applicable to patients with characteristic clinical signs and symptoms of SIS. We will also compare ASD with non-operative treatment option for SIS, the progressive ET, in a more pragmatic comparison, which is confounded by the lack of blinding of the participants (Figure 2).

The generalizability of our primary (efficacy) comparison may be questioned as the patients are carefully selected (strict eligibility criteria) and treated by experienced shoulder surgeons. Nevertheless, the eligibility criteria are in agreement with the existing treatment guidelines on SIS [4]. The results should thus be applicable to the specific populations currently receiving treatment for their SIS. As for the skill-level of the surgeons, the index surgical procedure (ASD) is a relatively simple procedure and thus likely not very sensitive to individual surgeons' experience. For example, the amount of bone removed from the undersurface of the acromion seems to have at best a marginal effect on the outcome. Even bursectomy alone has been shown to produce the same therapeutic effect as standard acromioplasty[48].

Rationale for outcome assessment and statistical analysis

Traditionally, the assessment of the treatment effects of two or more interventions has relied primarily on the statistical significance of the mean differences of the intervention groups. However, as described in a recent paper[49], to truly assess the clinical relevance of a treatment, one also needs information about the distribution of individual responses. In essence, one needs to look at how many people on treatment and on comparator group(s) had a response at least as great as the minimum (clinically) important difference (MCID). Such individuals have been described as "responders," and this approach of comparing treatment groups as a "responder analysis"[50 51]. The authors[49] suggested that "*Clinical trials should specify in their protocol that they will report the distribution of results in individual participants as well as the mean difference. Researchers should publish plots of individual results and responder analyses in clinical trial reports.*" The FIMPACT trial adheres to this suggested approach. Accordingly, we will elaborate several relevant and often interrelated issues, such as the study power, the primary outcomes and their interpretation, the minimal clinically important difference (MCII), as well as the approach we have chosen for carrying out a responder analysis.

Study power

Traditionally the sample size is calculated based on the minimal clinically important difference or change (MCID or MCII), i.e., the smallest change in measurement that signifies an important/detectable improvement in a patient's symptom(s). MCII/-D is not a static value even for one outcome instrument, but rather can have different values when assessed with different methods or *in different patient populations*. We chose VAS at rest and during arm activity as our primary outcomes, because shoulder pain is the primary complaint of patients with SIS. The FIMPACT trial was powered to detect an improvement of at least 15 on a 0-100 VAS scale[52] between ASD and ET. This yielded a sample size estimate of 70 participants per group. To safeguard against lack of study power, we chose a statistical threshold of 90% over the more conventional 80%. In this context, Norman et al.[53] recently introduced a thought-provoking proposal arguing that a standard ('off-the-peg') sample size of 64 per group would be just as valid an estimate as one obtains by more traditional ('made-to-measure') sample size calculations[53]. Finally, although the statistical power is a vital step in the *planning phase* of any clinical trial, the actual quality of evidence (certainty in the obtained estimates) can only be appropriately assessed from the confidence intervals (CI) of the data obtained[54].

Responder analysis

As noted above, instead of focusing only on the statistical significance of the mean differences between treatment groups in the VAS (i.e., the mean improvement from baseline to 24 months), we will also carry out "a responder analysis". In principle, this analysis allows physicians to inform a patient of his or her chance of experiencing a clinically meaningful improvement from the treatment, both in absolute terms and in comparison, to a control group. The difference between responders and non-responders can be considered the net-benefit of the treatment. One proposed means to carry out a responder analysis relies on the assessment of the proportion of patients reaching the patient-acceptable symptom state (PASS) and the patient-disappointing symptoms state (PDSS). As no universal consensus exists on either the PASS or the PDSS in the context of SIS, we chose to anchor our responder analysis to the patient's assessment of satisfaction with the shoulder treatment outcome: Patients reporting very satisfied or satisfied will be categorized as "Responders" and those reporting very dissatisfied or dissatisfied as "Non-responders". Given the obvious coarseness of this approach, we plan to evaluate the appropriate criteria for PASS and PDSS in more detail in the future, exploring the potential contribution of, e.g., arm pain at rest and at activity, shoulder function, and night pain.

Ethics of placebo surgery

A recent systematic review of the use of surgical placebo shows that in more than half of these studies the treatment group that included critical surgical/therapeutic element had no greater effect than a placebo group [18]. The review also showed that risks of adverse effects were small and the placebo group was safer than the surgery under investigation. These findings make a compelling case for the use of surgical placebo controls when a placebo effect may be present. Regarding the ethics of surgical placebo controls, the authors of the review state "*Placebo controlled surgical trials raise important ethical concerns but are justified when there is a genuine equipoise; that is, a disagreement in the medical community about whether one treatment is superior to another, because standard treatment does not exist or its efficacy is questioned.*" They continue by concluding: "*Placebo controlled trials in surgery are as important as they are in medicine, and they are justified in the same way. They are powerful, feasible way of showing the efficacy of surgical procedures. They are necessary to protect the welfare of present and future patients as well as to conduct proper cost effectiveness analyses. Only then may publicly funded surgical interventions be distributed fairly and justly. Without such studies ineffective treatment may continue unchallenged.*" Our views regarding the ethics of using a surgical placebo group are perfectly aligned with these notions.

Limitations of the study

One possible confounder in our trial is that subacromial pain is also the hallmark symptom of a rotator cuff tear, although the latter patients usually also represent with muscle weakness. To exclude patients with a (clinically-relevant) rotator cuff tear, our eligibility screening included two preoperative assessments: (a) clinical exams targeted at finding obvious weakness of the rotator cuff muscles and (b) MRA, an imaging modality with a shown 92 specificity and 94 sensitivity for "full-thickness" RC tears[55]. In addition to these, we also carried out (c) a diagnostic arthroscopy in the ASD and DA groups prior to randomisation. Despite the thorough *preoperative* screening, 10% (14/136) of the participants allocated to the two surgical groups had to be excluded because of AC-arthrosis (n=1) or intra-articular pathology found at diagnostic arthroscopy (n=13). Although this does not have any effect on our primary comparison (ASD vs. DA), one could argue that the ET and operatively treated groups (ASD and DA) are not fully comparable. At the same we don't know the clinical relevance of small RC tears or SLAP lesions, which don't result in obvious muscle weakness and/or are not apparent in MRA. In the end, if this bias proves clinically relevant in our analysis, it will skew our results by favouring the ASD group in the pragmatic comparison (ASD vs. ET). Another concern related to the pragmatic comparison (ASD vs. ET) is that the progressive exercise therapy regimen carried out in the ET group is different from the postoperative rehabilitation carried out by patients in the ET group, for obvious reasons; surgically treated patients need time to recover from the initial surgical trauma. Furthermore, ASD patients are also subject to some degree of postoperative immobilization, extended sick leave, and modifications in pain medication and activities, all of which potentially have an effect on the outcome of treatment.

Another obvious concern related to our study design is the discrepant timing of the start of the actual treatment between the ET and the two surgical groups due to the time required to arrange the surgery. Acknowledging this, the two-year follow-up was chosen as our primary time point for assessing the benefits of treatment, as we assume that by this time the potential confounding effect of slightly different follow-up times should be diluted to a minimum. This is also the reason why we use data from the shorter-term follow-up visits (i.e.: visits performed at 3, 6 and 12 months after randomization) primarily to illustrate the trajectory of the treatment response in the three groups. Concerns over the varying time span from the randomization of the patients to the trial to the actual induction of treatment (due to delay in surgery) also applies to the CSAW trial[20]. To compensate for the waiting list effects, the CSAW investigators have chosen a slightly different strategy: Although the primary outcome assessment is performed at 6 months after randomization in CSAW trial, they have introduced additional follow-up assessments, referenced from surgery, for patients waiting for longer than 4 months for their surgery after

randomization. They have also set a secondary outcome measurement point at 1-year post randomization.

ETHICS AND DISSEMINATION

Ethics

FIMPACT trial is conducted in accordance with the principles of the Declaration of Helsinki. This trial has been approved by the institutional review board (IRB) of the Pirkanmaa Hospital District and each participating centre granted clinical trial authorisation prior to recruitment. The trial has been registered to ClinicalTrials.gov registry and any revisions about the protocol are documented in this registry. For each participant, informed consent is obtained prior to any study-related procedures.

Dissemination policy

We aim to produce high-impact publications of the results of the trial and present the findings to the clinicians who manage shoulder pain in the front line. The investigators will be involved in preparing drafts of the manuscripts, abstracts, press releases and any other publications arising from the trial. The final reporting will follow the Consolidated Standards of Reporting Trials (CONSORT) Statement guidelines. Authorship will be determined in accordance with the International Committee of Medical Journal Editors (ICMJE) guidelines and other contributors will be acknowledged. The funders will be acknowledged in all resulting publications. There is no intended use of professional writers.

Author affiliations

Mika Paavola
Helsinki University Central Hospital,
Department of Orthopaedics and Traumatology,
Töölö hospital,
Topeliuksenkatu 5
P.O. Box 266, 00029 HUS,
Helsinki, Finland
Email: mika.paavola@hus.fi

Antti Malmivaara
National Institute for Health and Welfare
Center for Health and Social Economics
Mannerheimintie 166
00270 Helsinki
Finland
Email:antti.malmivaara@thl.fi

Simo Taimela
University of Helsinki,
Department of Orthopaedics and Traumatology,
Töölö hospital, Topeliuksenkatu 5,
00014 University of Helsinki, Helsinki, Finland
Email: simo.taimela@helsinki.fi

Kari Kanto
Hatanpään sairaala, Tampere
Hatanpäänkatu 24, 33900 Tampere, Finland
Email: kari.kanto@fimnet.fi

Teppo L N Järvinen
University of Helsinki,
Department of Orthopaedics and Traumatology,
Töölö hospital, Topeliuksenkatu 5,
00014 University of Helsinki, Helsinki, Finland
Email: teppo.jarvinen@helsinki.fi

Acknowledgements

The FIMPACT investigators would like to thank Professor Jonas Ranstam (Lund University) for the statistical analysis plan and acknowledge Dr. Pekka Paavolainen and Professor Markku Järvinen for their critical contributions and support on the early phases of the planning and execution of the trial.

Contributions

Concept and design: Mika Paavola (MP), Antti Malmivaara (AM), Simo Taimela (ST) and Teppo Järvinen (TJ).

Drafting and critical revision of the article for important intellectual content: MP, AM, ST, TJ and Kari Kanto (KK).

Final approval of the article: MP, AM, ST, TJ, KK.

Ensuring the accuracy of the work: MP, AM, ST, TJ, KK.

Statistical expertise: Jonas Ranstam (JR).

Obtaining of funding: TJ and Markku Järvinen (MJ).

Primary Sponsor: MP.

Funding

The FIMPACT trial was supported by the Sigrid Juselius Foundation, the Competitive Research Fund of Pirkanmaa and Helsinki University Central Hospital Districts, the Academy of Finland, and the Jane and Aatos Erkko Foundation. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of this protocol.

Competing interests

Dr. Taimela reports personal fees from Evalua group of companies, personal fees from DBC group of companies, and personal fees from insurance companies, outside the submitted work. Dr. Kanto reports an honorarium for a lecture from Linvatec, outside the submitted work. Dr. Järvinen reports an honorarium for a lecture on osteoporosis from AMGEN (donated to AllTrials campaign). Authors not named here have disclosed no conflicts of interest.

Availability of data and materials

The principal investigator (MP) will have access to the full dataset. Given that the informed consent of the FIMPACT trial did not include a provision for data sharing (trial launched in 2005 when such policy was not endorsed), at present the dataset cannot be shared due to a potential breach of the Finnish Personal Data Act. We intend to rectify this situation by renewing the informed consents of the trial.

Consent for publication

Not applicable.

Date and version identifier

August 30, 2016:	Original protocol, (v1.0)	bmjopen-2016-014087
December 5, 2016:	Revised protocol, major revision (v2.0)	bmjopen-2016-014087.R1
February 13, 2017:	Re-revised protocol, minor revision (v2.1)	bmjopen-2016-014087.R2

All versions available at http://bmjopen.bmj.com

Figure 1: Flowchart of the trial: enrolment, assigned intervention and follow-up scheme.

Figure 2: Study design and interpretation of results.

Appendix 1: FIMPACT investigators.

Appendix 2: Statistical analysis plan (SAP).

Appendix 3: Blinded data interpretation plan.

Appendix 4: SPIRIT checklist.

Appendix 5: Informed consent form (in Finnish).

REFERENCES

1. Paloneva J, Lepola V, Karppinen J, et al. Declining incidence of acromioplasty in Finland. *Acta Orthop* 2015;86(2):220-4. doi: 10.3109/17453674.2014.977703
2. Vitale MA, Arons RR, Hurwitz S, et al. The rising incidence of acromioplasty. *J Bone Joint Surg Am* 2010;92(9):1842-50. doi: 10.2106/JBJS.I.01003
3. Engebretsen K, Grotle M, Bautz-Holter E, et al. Radial extracorporeal shockwave treatment compared with supervised exercises in patients with subacromial pain syndrome: single blind randomised study. *Bmj* 2009;339:b3360. doi: 10.1136/bmj.b3360 [published Online First: 2009/09/17]
4. Diercks R, Bron C, Dorrestijn O, 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. doi: 10.3109/17453674.2014.920991
5. Oh LS, Wolf BR, Hall MP, et al. Indications for rotator cuff repair: a systematic review. *Clinical orthopaedics and related research* 2007;455:52-63. doi: 10.1097/BLO.0b013e31802fc175
6. Aurora A, McCarron J, Iannotti JP, et al. Commercially available extracellular matrix materials for rotator cuff repairs: state of the art and future trends. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al]* 2007;16(5 Suppl):S171-8. doi: 10.1016/j.jse.2007.03.008
7. Michener LA, McClure PW, Karduna AR. Anatomical and biomechanical mechanisms of subacromial impingement syndrome. *Clinical biomechanics* 2003;18(5):369-79.
8. Koester MC, George MS, Kuhn JE. Shoulder impingement syndrome. *The American journal of medicine* 2005;118(5):452-5. doi: 10.1016/j.amjmed.2005.01.040
9. Neer CS, 2nd. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *The Journal of bone and joint surgery American volume* 1972;54(1):41-50.
10. Yu E, Cil A, Harmsen WS, et al. Arthroscopy and the dramatic increase in frequency of anterior acromioplasty from 1980 to 2005: an epidemiologic study. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 2010;26(9 Suppl):S142-7. doi: 10.1016/j.arthro.2010.02.029
11. Judge A, Murphy RJ, Maxwell R, et al. 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. doi: 10.1302/0301-620X.96B1.32556
12. Dong W, Goost H, Lin XB, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. *Medicine (Baltimore)* 2015;94(10):e510. doi: 10.1097/MD.0000000000000510
13. Shi LL, Edwards TB. The role of acromioplasty for management of rotator cuff problems: where is the evidence? *Advances in orthopedics* 2012;2012:467571. doi: 10.1155/2012/467571
14. Brox JI, Staff PH, Ljunggren AE, et al. Arthroscopic surgery compared with supervised exercises in patients with rotator cuff disease (stage II impingement syndrome). *Bmj* 1993;307(6909):899-903. [published Online First: 1993/10/09]
15. Haahr JP, Ostergaard S, Dalsgaard J, et al. Exercises versus arthroscopic decompression in patients with subacromial impingement: a randomised, controlled study in 90 cases with a one year follow up. *Ann Rheum Dis* 2005;64(5):760-4. doi: 10.1136/ard.2004.021188 [published Online First: 2005/04/19]
16. Ketola S, Lehtinen J, Rousi T, et al. No evidence of long-term benefits of arthroscopicacromioplasty in the treatment of shoulder impingement syndrome: Five-year results of a randomised controlled trial. *Bone Joint Res* 2013;2(7):132-9. doi: 10.1302/2046-3758.27.2000163 [published Online First: 2013/07/10]
17. Ernst E, Resch KL. Concept of true and perceived placebo effects. *Bmj* 1995;311(7004):551-3. [published Online First: 1995/08/26]
18. Wartolowska K, Judge A, Hopewell S, et al. Use of placebo controls in the evaluation of surgery: systematic review. *Bmj* 2014;348:g3253. doi: 10.1136/bmj.g3253

19. Buchbinder R. Meniscectomy in patients with knee osteoarthritis and a meniscal tear? *N Engl J Med* 2013;368(18):1740-1. doi: 10.1056/NEJMe1302696 [published Online First: 2013/03/20]
20. Beard D, Rees J, Rombach I, et al. The CSAW Study (Can Shoulder Arthroscopy Work?) - a placebo-controlled surgical intervention trial assessing the clinical and cost effectiveness of arthroscopic subacromial decompression for shoulder pain: study protocol for a randomised controlled trial. *Trials* 2015;16:210. doi: 10.1186/s13063-015-0725-y
21. Ellman H. Arthroscopic subacromial decompression: analysis of one- to three-year results. *Arthroscopy* 1987;3(3):173-81. [published Online First: 1987/01/01]
22. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al]* 2009;18(6):927-32. doi: 10.1016/j.jse.2009.03.021
23. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clinical orthopaedics and related research* 1987(214):160-4.
24. Henseler JF, Kolk A, van der Zwaal P, et al. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al]* 2015;24(3):376-81. doi: 10.1016/j.jse.2014.07.003
25. Matsen FA, 3rd, Ziegler DW, DeBartolo SE. Patient self-assessment of health status and function in glenohumeral degenerative joint disease. *J Shoulder Elbow Surg* 1995;4(5):345-51. [published Online First: 1995/09/01]
26. Godfrey J, Hamman R, Lowenstein S, et al. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *J Shoulder Elbow Surg* 2007;16(3):260-7. doi: 10.1016/j.jse.2006.07.003 [published Online First: 2006/12/26]
27. Tashjian RZ, Deloach J, Green A, et al. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. *J Bone Joint Surg Am* 2010;92(2):296-303. doi: 10.2106/JBJS.H.01296 [published Online First: 2010/02/04]
28. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med* 2001;33(5):328-36. [published Online First: 2001/08/09]
29. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. *Qual Life Res* 2008;17(3):485-95. doi: 10.1007/s11136-008-9317-6 [published Online First: 2008/02/22]
30. Bowling A. Measuring health. A review of quality of life measurement scales. 3rd ed: Berkshire: Open University Press 2004.
31. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83. [published Online First: 1992/06/11]
32. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *Bmj* 2002;324(7351):1417. [published Online First: 2002/06/18]
33. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med* 2013;369(26):2515-24. doi: 10.1056/NEJMoa1305189 [published Online First: 2013/12/27]
34. Ranstam J. Multiple P-values and Bonferroni correction. *Osteoarthritis Cartilage* 2016;24(5):763-4. doi: 10.1016/j.joca.2016.01.008
35. Sun X, Ioannidis JP, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. *Jama* 2014;311(4):405-11. doi: 10.1001/jama.2013.285063
36. Neer CS, 2nd. Impingement lesions. *Clin Orthop Relat Res* 1983(173):70-7. [published Online First: 1983/03/01]
37. Dong W, Goost H, Lin XB, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. *Medicine (Baltimore)* 2015;94(10):e510. doi: 10.1097/MD.0000000000000510 [published Online First: 2015/03/12]

38. Ketola S, Lehtinen J, Rousi T, et al. Which patients do not recover from shoulder impingement syndrome, either with operative treatment or with nonoperative treatment? *Acta Orthop* 2015;86(6):641-6. doi: 10.3109/17453674.2015.1033309
39. Tangtrakulwanich B, Kapkird A. Analyses of possible risk factors for subacromial impingement syndrome. *World J Orthop* 2012;3(1):5-9. doi: 10.5312/wjo.v3.i1.5
40. Bigliani LU MD, April EW The morphology of the acromion and its relationship to rotator cuff tears. *Orthop Trans* 1986;10:228.
41. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *J Clin Epidemiol* 2014;67(7):769-72. doi: 10.1016/j.jclinepi.2013.11.011 [published Online First: 2014/02/25]
42. Skou ST, Roos EM, Laursen MB, et al. A Randomized, Controlled Trial of Total Knee Replacement. *N Engl J Med* 2015;373(17):1597-606. doi: 10.1056/NEJMoa1505467 [published Online First: 2015/10/22]
43. Investigators F, Bhandari M, Jeray KJ, et al. A Trial of Wound Irrigation in the Initial Management of Open Fracture Wounds. *N Engl J Med* 2015;373(27):2629-41. doi: 10.1056/NEJMoa1508502
44. Jonas WB, Crawford C, Colloca L, et al. To what extent are surgery and invasive procedures effective beyond a placebo response? A systematic review with meta-analysis of randomised, sham controlled trials. *BMJ Open* 2015;5(12):e009655. doi: 10.1136/bmjopen-2015-009655 [published Online First: 2015/12/15]
45. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving. *Bmj* 1999;319(7211):652-3. [published Online First: 1999/09/10]
46. Djulbegovic B, Paul A. From efficacy to effectiveness in the face of uncertainty: indication creep and prevention creep. *Jama* 2011;305(19):2005-6. doi: 10.1001/jama.2011.650
47. Jarvinen TL, Sievanen H, Kannus P, et al. The true cost of pharmacological disease prevention. *Bmj* 2011;342:d2175. doi: 10.1136/bmj.d2175
48. Donigan JA, Wolf BR. Arthroscopic subacromial decompression: acromioplasty versus bursectomy alone--does it really matter? A systematic review. *Iowa Orthop J* 2011;31:121-6. [published Online First: 2011/11/19]
49. Cates C, Karner C. Clinical importance cannot be ruled out using mean difference alone. *Bmj* 2015;351:h5496. doi: 10.1136/bmj.h5496
50. Guyatt GH, Juniper EF, Walter SD, et al. Interpreting treatment effects in randomised trials. *Bmj* 1998;316(7132):690-3. [published Online First: 1998/04/02]
51. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials* 2007;8:31. doi: 10.1186/1745-6215-8-31 [published Online First: 2007/10/27]
52. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg* 2009;18(6):927-32. doi: 10.1016/j.jse.2009.03.021 [published Online First: 2009/06/19]
53. Norman G, Monteiro S, Salama S. Sample size calculations: should the emperor's clothes be off the peg or made to measure? *BMJ* 2012;345:e5278. doi: 10.1136/bmj.e5278
54. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121(3):200-6.
55. Lenza M, Buchbinder R, Takwoingi Y, et al. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. *Cochrane Database Syst Rev* 2013;9:CD009020. doi: 10.1002/14651858.CD009020.pub2 [published Online First: 2013/09/26]

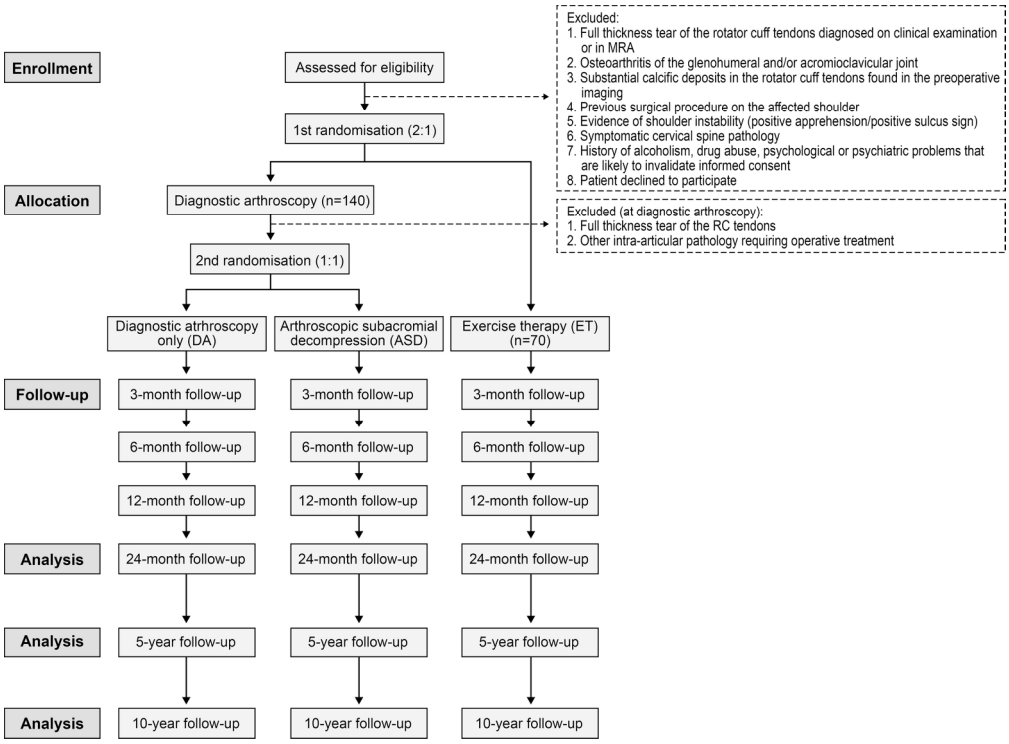
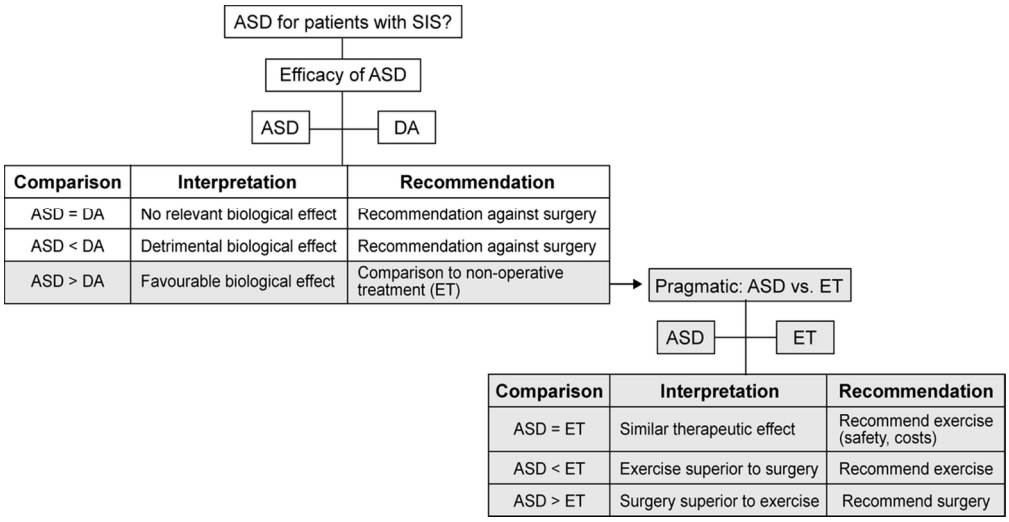


Figure 1: Flowchart of the trial: enrolment, assigned intervention and follow-up scheme.

184x134mm (300 x 300 DPI)



Comparison	Interpretation	Recommendation
ASD = DA	No relevant biological effect	Recommendation against surgery
ASD < DA	Detrimental biological effect	Recommendation against surgery
ASD > DA	Favourable biological effect	Comparison to non-operative treatment (ET)

Comparison	Interpretation	Recommendation
ASD = ET	Similar therapeutic effect	Recommend exercise (safety, costs)
ASD < ET	Exercise superior to surgery	Recommend exercise
ASD > ET	Surgery superior to exercise	Recommend surgery

Figure 2: Study design and interpretation of results.

107x54mm (300 x 300 DPI)

FIMPACT INVESTIGATORS

Steering Committee: Mika Paavola (Chair, Helsinki University Central Hospital), Teppo Järvinen (Co-Chair, University of Helsinki and Helsinki University Central Hospital), Antti Malmivaara (National Institute for Health and Welfare), Simo Taimela (University of Helsinki)

Writing Committee: Teppo Järvinen (Co-Chair, University of Helsinki and Helsinki University Central Hospital), Simo Taimela (Co-Chair, University of Helsinki), Antti Malmivaara (National Institute for Health and Welfare), Mika Paavola (Helsinki University Central Hospital), Kari Kanto (Hatanpää City Hospital, Tampere).

Participating Clinical Sites: *Helsinki University Central Hospital/Jorvi Hospital*: Kalevi Hietaniemi, Juha Kalske, Vesa Lepola, Jyrki Salmenkivi, Sikri Tukiainen. – *Helsinki University Central Hospital/Herttoniemi Hospital*: Jarkko Pajarinen, Mikko Salmela, Vesa Savolainen, Ilkka Sinisaari. – *Hatanpää City Hospital, Tampere*: Timo Järvelä, Kari Kanto, Janne Lehtinen, Mikael Salmela.

FIMPACT Methods Center: Leena Caravitis, Sari Karesvuori, Pirjo Toivonen (Project management), Mathias Bäck (Data management), Ville Haapamäki (Imaging) (Helsinki University Central Hospital and University of Helsinki); Jari Inkinen (Physiotherapy) (Kunnon Klinikka, Tampere), Esa Läärä (Randomization) (University of Oulu), Harri Sintonen (QoL outcomes) (National Institute for Health and Welfare).

Statistical Analysis Plan (SAP) for:

Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT), 2-year follow-up

Simo Taimela[1] and Teppo L N Järvinen[1]

Statistical advisor:

Jonas Ranstam

[1] Department of Orthopedics and Traumatology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

STUDY SYNOPSIS

Introduction: Arthroscopic subacromial decompression (ASD) is the most commonly performed surgical intervention for shoulder pain, yet evidence on its efficacy is limited. The rationale for the surgery rests on the tenet that symptom relief is achieved through removal of a bony acromial spur and the resulting decompression of the tendon passage. Acknowledging the potential placebo effect of surgery, the primary objective of this superiority trial is to compare the efficacy of ASD versus diagnostic arthroscopy (DA) in patients with shoulder impingement syndrome (SIS), where DA differs only by the lack of subacromial decompression. As a non-surgical treatment option, a third group of supervised progressive exercise therapy (ET) will allow for pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies.

Methods/Design: FIMPACT trial is an ongoing multicentre, three-group randomised controlled study with a primary objective of assessing the efficacy of the ASD vs. DA and a secondary objective of comparing ASD to exercise therapy (ET) in a pragmatic setting. We performed two-fold concealed allocation, first by randomizing patients to surgical (ASD or DA) or conservative (ET) treatment in 2:1 ratio and then those allocated to surgery further to ASD or DA in 1:1 ratio. Our two primary outcomes are pain at rest and arm activity assessed with visual analog scale (VAS), while the secondary outcomes are functional assessment (Constant score and Simple shoulder test), quality of life (15D and SF-36), patient satisfaction, proportions of responders and non-responders, reoperations/treatment conversions, all at 2 years post-randomization, as well as adverse effects and complications. We recruited a total of 210 patients from 3 tertiary referral centres. We will conduct the primary analysis on the intention-to-treat basis.

TRIAL REGISTRATION

ClinicalTrials.gov NCT00428870 (first registered January 29, 2007).

STUDY OBJECTIVES AND OUTCOMES

This statistical analysis plan (SAP) is accompanying the actual study protocol of the FIMPACT trial, a document that elaborates the methods used in detail. All outcomes were inquired from participants at baseline and follow-ups (6 and 24 months) and selected additional measures at 3 and 12 months (for details, see Table 1). The last patient reached the primary endpoint, the 24-month follow-up, in September 2015.

Table 1: Outcomes and follow-up time points

Assessment	Screening	Enrolment (Baseline)	Surgery	3 Months	6 Months	12 Months	24 Months	5 years	10 years
Screening form	X								
Informed consent		X							
Baseline characteristics form		X							
X-ray and MRI	X								X
Randomisation		X (1st)	X (2nd)						
Arthroscopic findings form			X						
Follow-up form*				X		X			
Clinical examination		X			X		X	X	X
Complications/adverse effects form**			(X)	(X)	(X)	(X)	(X)	(X)	(X)
VAS, at rest		X		X	X	X	X	X	X
VAS, at arm activity		X		X	X	X	X	X	X
Constant- Murley Score		X			X		X	X	X
Simple Shoulder Test (SST)		X			X		X	X	X
SF-36		X		X	X	X	X	X	X
15D		X		X	X	X	X	X	X
Return to work				X	X	X	X	X	X
Return to previous leisure activities				X	X	X	X	X	X
Responder analysis				X	X	X	X	X	X
Patients satisfaction to the treatment				X	X	X	X	X	X
Patients assessment of the treatment allocation				X					
Health resource utilization				X	X	X	X	X	X

* Letter/telephone interview

** If required

DESCRIPTIVE OUTCOMES

At screening, the participants filled out a questionnaire to record gender, age, hand dominance, weight, height, level of education (socioeconomic status), workload (type of work), physical activity level, sports discipline, subjective health, symptoms (onset, frequency, and severity), use of pain medications, prior treatments, expectations to treatment, generic health state, and disease-specific scores. To exclude patients with concomitant shoulder pathology (particularly rotator cuff rupture), magnetic resonance imaging with contrast (MRA) was acquired for each participant.

OBJECTIVES AND PRIMARY OUTCOME

The primary objective of this trial is to compare the efficacy of arthroscopic subacromial decompression (ASD) versus diagnostic arthroscopy (DA) in patients with SIS. The trial is designed as a superiority trial, i.e. we expected in the power calculation that the ASD will result in greater pain relief at 24-month

follow-up than DA (or ET). The 24-month follow-up was chosen as the primary endpoint, since this time point is a commonly held "minimal requirement" for any procedure in the field (orthopaedics) and most commonly used in the trials assessing the treatment of SIS.

The primary hypothesis: The primary hypothesis of our FIMPACT trial is that ASD is superior to DA in patients with SIS.

To enable pragmatic assessment of the relative benefits of surgical vs. non-operative treatment strategies on SIS, a non-surgical (third) treatment option of supervised progressive exercise therapy (ET) is also included (ASD vs. ET).

Additional hypothesis: The relative benefits of ASD and ET will be assessed without a priori hypothesis on the superiority of one or the other.

As the primary outcome measure, a visual analogue scale (0-100) was used to measure the patient's perceived pain intensity at rest and at arm activity during the 24 hours preceding the assessment. We considered 15 as the minimal clinically important difference (MCID) for SIS.[1]

SECONDARY OUTCOMES

Our secondary outcome measures are listed below. These outcomes will only be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not considered to be a problem[2].

Constant-Murley score

Constant-Murley score (CS) is the most commonly used scoring system for evaluation of various disorders of the shoulder[3]. It consists of both objective (range of motion and strength) and subjective measurements (pain assessment, work load, and leisure time activities), which are summarized in a score between 0 and 100. A higher score indicates better shoulder function. The minimal detectable change (MDC) of the Constant score is 17 for patients with SIS[4]

In addition, as night pain is considered one of the hallmark symptoms in patients with SIS and our two primary outcome measures (patient's perceived pain intensity at rest and at arm activity in the last 24 hours) do not specifically address this issue, a specific question from the Constant-Murley score (unaffected sleep: "Yes" or "No") will be analysed separately.

SST

The simple shoulder test (SST) was developed to assess any impairment of the patient's activities of daily living[5]. The SST consists of 12 questions with yes (1) or no (0) response options. The maximum SST score is 12 indicating normal shoulder function, minimum score of 0 points refers severely diminished shoulder function. The SST has good reliability and responsiveness in patients with rotator cuff symptoms[6]. The MCID for the SST in rotator cuff disease is 2 points[7].

15D

The 15D instrument is a generic health-related quality of life (HRQoL) instrument comprising 15 dimensions[8]. For each dimension, the respondent must choose one of the five levels that best describes his/her state of health at that the moment (the best level being 1 and the worst level being 5). A set of utility or preference weights is used in an addition aggregate formula to generate a single index number, the utility or 15D score. The maximum 15D score is 1 (no problems on any dimension) and the minimum score is 0 (being dead). The responsiveness, reliability and validity of 15D have been thoroughly established, and this instrument has been used extensively in clinical and healthcare research[9 10].

SF-36

Short form or SF-36 is a generic HRQoL instrument to quantify the physical, functional, and psychological aspects of health-related quality of life. It consists of 36 questions in eight subscales that assess physical, functional, social, and psychological well-being[11]. Score ranges from 0 to 100, where a higher score is associated with better health. The physical and mental component summary scales (PCS and MCS, respectively) are then calculated as composites of the related subscales. SF-36 is one of most widely used measure of health-related quality of life[12].

Patient satisfaction and Responder analysis

We elicited patients' global assessment of satisfaction to the treatment with this question: "Are you satisfied with the treatment you have received?" We used a VAS scale ranging from 0 (completely disappointed) to 100 (completely satisfied).

Additionally, we elicited patient satisfaction to the treatment outcome with the following question at each follow-up time point (Table 1): "How satisfied are you with the outcome of your treatment?" on a 5-item scale. Participants who reported very satisfied or satisfied will be categorized as "Responders"

and patients who responded very dissatisfied or dissatisfied as “Non-responders”.

Return to previous leisure activities

Similarly, at each follow-up (Table 1), participants were asked to respond to the following question: “Have you been able to return to your previous leisure activities?” (“yes” or “no”).

Patients’ perception of operative treatment-group assignment

At the 3-month follow-up point, the patients in the two operative groups were asked to guess whether they had undergone ASD or DA.

Health resource utilization and costs

For the cost-effectiveness analysis, at each follow-up visit the participants were asked to fill in a questionnaire inquiring about the use of healthcare resources. The questionnaire contains a list of items of healthcare resources available and the participants were asked to fill in the number of visits per item during the recall period of each follow-up time point. The resource use will be calculated based on the number of visits times unit cost per item and expressed as mean costs by items of resource use, and the mean direct total health care resource costs. All costs will be discounted to the 2016 price level.

Time to return to work

Information about return to work was recorded at each follow-up time point (Table 1).

Complications and adverse effects

Complications directly related to the interventions were registered. The participants were also encouraged to contact the participating hospitals if any adverse effects occurred and contacts to the health care system were monitored at every follow-up visit. Potential adverse effects (AE) were categorized to serious ad-verse effects (SAE) and minor adverse effects (MAE) if the participants sought treatment. Death, cardio-vascular or gastrointestinal effects, deep venous thrombosis, pulmonary embolism, systemic or local infection were categorised as SAEs and shoulder symptoms like pain, swelling and decreased range of motion were categorised as MAEs. The number and severity of complications and adverse effects will be assessed.

EXPLORATORY OUTCOMES

We have identified three potentially important effect modifying factors. We will perform subgroup analyses with the primary endpoint as the outcome and the direction of hypothesized effect described as below[14].

Duration of symptoms

We will compare the treatment effects stratified based on the duration of symptoms (those with < 6/12 months vs. those > 6/12 months). We hypothesize that subacromial decompression will work better in patients with duration of symptoms > 6 months than for patients with symptoms < 6 months.

Severity of symptoms

We will compare the treatment effects in patients with severe (VAS 70 or more), moderate (VAS 55 to 69), and mild (VAS less than 55) symptoms at baseline. We hypothesize that subacromial decompression will work better in patients with more severe (VAS 70 or more) than moderate (VAS 55 to 69) or mild (VAS less than 55) symptoms at baseline.

Acromial anatomy

We will compare the treatment effects in patients with flat (type I), curved (type II), or hooked (type III) acromion according to classification by Bigliani et al.[15] We hypothesize that subacromial decompression will work better in patients with hooked (type III) than curved (type II) or flat (type I) acromion at baseline.

STUDY DESIGN

Sample size

The sample size calculation was based on the two primary outcome measures, VAS at rest and at arm activity, at 24 months post randomization. FIMPACT trial was powered to detect a minimal clinically important improvement (MCII) in a VAS pain score (improvement of at least 15; assumed standard deviation 25) between ASD and DA (or ET). To achieve a somewhat unconventional (stringent) 90% study power and using a two-sided Type I error rate (5%), our trial requires 68 patients per study group

to show clinically meaningful advantage of ASD over DA (or ET). Acknowledging the stringent power threshold, only 3% surplus was reserved for potential loss to follow up/crossovers (3%), and accordingly, the recruitment target was set at 70 patients per treatment group.

Randomization and blinding

To obtain three balanced study groups (of similar group size), we performed a two-fold, sequential randomization. In Phase I, the participants were randomized into non-surgical or surgical treatment with allocation ratio 1:2. In the Phase II, those allocated to surgical treatment were further randomized to ASD or DA with 1:1 ratio. An independent statistician with no clinical involvement in the execution of the trial prepared separate randomization lists for each study centre using a computer-generated algorithm. Randomization was carried out using sequentially numbered sealed opaque envelopes. The envelopes were kept in a secure, agreed location at each centre. To ensure concealment, block randomization was applied using blocks varying in size randomly, the block size known only by the statistician.

To initially enter a participant into the study (Phase I), an envelope containing the treatment assignment [non-surgical (ET) or surgery (ASD or DA), ratio 1:2] was opened during the baseline appointment. Participants randomized to ET started standardized physiotherapy within 2 weeks of the baseline appointment. Participants allocated to surgical treatment were scheduled for surgery aimed to be completed within 12 weeks of randomization.

At the day of surgery, an arthroscopic examination was first carried out to confirm the eligibility of the participant (to rule out full-thickness RC tear and other obvious intra-articular pathology). Research/staff nurse then completed the randomization procedure (Phase II) by opening an envelope containing the surgical treatment allocation (ASD or DA, ratio 1:1). The allocation was revealed to the surgeon by showing the paper, but not expressed verbally.

The full follow-up process is shown in figure 1. In brief, the participants filled in the above noted (mailed) outcome questionnaires at 3, 6, 12 and 24 months post randomization, in addition to which they were also assessed clinically at 6 and 24 months (and 5 and 10 years) post randomisation by a study physiotherapist unaware of treatment allocation, treatment given or possible unblinding. Outcome assessors were instructed not to inquire anything about prior treatment. Further, participants wore a t-shirt on all follow-up examinations. Data analysis will be done in a blinded manner by the study statistician (JR) not directly involved in the study.

STUDY POPULATION

Subject disposition

Study procedures, including recruitment strategies and inclusion and exclusion criteria, are presented in detail in the accompanying actual study protocol.

STATISTICAL ANALYSIS

Data will be analysed in a blinded manner. All p-values will be reported to 3 decimal places with those less than 0.001 reported as $p < 0.001$. The criterion for statistical significance will be set at alpha = 0.05.

Primary analysis

The primary analysis will be carried out according to the intention-to-treat (ITT) principle: participants are retained in the groups to which they were initially randomized. The primary comparison on the efficacy of ASD (ASD vs. DA) will be performed as a between-group comparison using a repeated measures mixed-effects model (RMMM). Study group and time of assessment (baseline, 3, 6, 12 and 24 months) will be included as fixed factors and patient as a random factor. The model will include interactions between study group and time of assessment. The baseline value will be included as a covariate. An unstructured covariance structure will be assumed. If the model cannot be fitted, compound symmetry will be assumed instead. The number of degrees of freedom will be assessed using Satterthwaite's method. The RMMM model will be used to quantify the treatment effect as the difference between the groups in pain scores (VAS) with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization. To safeguard against potential multiplicity bias[2], we will require a statistically significant treatment effect on both of our primary outcome variables, i.e., pain at rest and pain at activity (Table 2).

Table 2: Primary comparison ASD vs ET: Outcomes of the trial at 24 months follow-up.

	ASD ET	Improvement from baseline	Between-Group Difference in Improvement from Baseline
Primary outcomes		**ASD ET**	
VAS (rest)			

VAS (at arm activity)			
Secondary outcomes			
Constant-Murley Score			
SST			
SF-36			
15D			
Time to return to work			
Return to previous leisure activities			
Responder analysis			
Patients satisfaction to the treatment			
Patients assessment of the treatment allocation			
Complications and adverse effects			

Abbreviations: VAS, visual analogue scale; SST, Simple Shoulder Test; SF-36, Short form- 36

The same statistical model will also apply to the pragmatic comparison of the relative benefits of surgical vs. non-operative treatment strategies on SIS (ASD vs. ET) (Table 3).

Table 3. Secondary comparison ASD vs ET: Outcomes of the trial at 24 months follow-up.

	ASD ET	Improvement from baseline	Between-Group Difference in Improvement from Baseline
Primary outcomes		**ASD ET**	
VAS (rest)			
VAS (at arm activity)			
Secondary outcomes			
Constant-Murley Score			
SST			

SF-36			
15D			
Time to return to work			
Return to previous leisure activities			
Responder analysis			
Patients satisfaction to the treatment			
Patients assessment of the treatment allocation			
Complications and adverse effects			

Abbreviations: VAS, visual analogue scale; SST, Simple Shoulder Test; SF-36, Short form- 36

Secondary analyses

We will also use the RMMM model to analyse secondary outcomes (Table 2 and 3) where applicable. The results will be reported as the differences between the groups with the associated 95% confidence interval (CI) and p-value at 24 months post-primary randomization.

Furthermore, instead of focusing only on the statistical significance of the mean differences between treatment groups in the VAS (i.e., the mean improvement from baseline to 24 months), we will also carry out "a responder analysis". In principle, this analysis allows physicians to inform a patient of his or her chance of experiencing a clinically meaningful improvement from the treatment, both in absolute terms and in comparison, to a control group. The difference between responders and non-responders can be considered the net-benefit of the treatment. One proposed means to carry out a responder analysis relies on the assessment of the proportion of patients reaching the patient-acceptable symptom state (PASS) and the patient-disappointing symptoms state (PDSS). As no universal consensus exists on either the PASS or the PDSS in the context of SIS, we chose to anchor our responder analysis to the patient's assessment of satisfaction with the shoulder treatment outcome: Patients reporting very satisfied or satisfied will be categorized as "Responders" and those reporting very dissatisfied or dissatisfied as "Non-responders". Given the obvious coarseness of this approach, we plan to evaluate the appropriate criteria for PASS and PDSS in more detail in the future, exploring the potential contribution of, e.g., arm pain at rest and at activity, shoulder function, and night pain.

Categorical variables, the rates of unblinding, reoperation, treatment conversion, complications and adverse effects will be analysed using logistic regression analysis or Poisson regression dependent on whether subjects with complications or (multiple) complications (per subject) are analysed.

These secondary analyses will be supportive, explanatory and/or hypothesis generating, which is why multiplicity is not a problem[2].

Sensitivity analyses

The following two sensitivity analyses will be carried out: 1) per-protocol analysis, in which the above noted primary analyses will be carried out again with patients who received the interventions as allocated will be redone; 2) and potential effects due to the treatment providing centres.

As all the participants in the ASD group have received the critical therapeutic element (subacromial decompression), no treatment group conversion is possible in this group.

In the per-protocol comparison of the efficacy of ASD (ASD vs. DA), we define the DA per-protocol population as those participants who have not received ASD during the 24-month follow-up (who have not crossed over to ASD).

In the per-protocol comparison of the effectiveness of ASD (ASD vs. ET), we define the ET per-protocol population as those participants who have not received ASD during the 24-month follow-up (who have not crossed over to ASD).

INTERPRETATION OF RESULTS

To safeguard against potential risk of bias during interpretation, a method of "blinded data interpretation" will be used[17]. In brief, an independent statistician will provide the Steering/Writing committee of the FIMPACT trial with blinded results from the analyses with study groups labelled as group A, group B, and group C. This data will be presented to the Steering/Writing Committee, who will then contemplate on the interpretation of the results until a consensus is reached and agree in writing on all alternative interpretations of the findings. Once reaching a consensus, we will record the minutes of this meeting as a statement of interpretation document signed by all members of the Writing Committee. Only after reaching this common agreement will the data manager and independent statistician break the randomization code.

There was also variation in the actual execution of the follow-up assessments, particularly in the earlier time-points (3- and 6-month follow-up visits).

IMPLEMENTATION OF ANALYSIS PLAN

This SAP will be used as a work description for the statistician performing the analyses. All analyses will be performed by the same statistician and none of the investigators involved in this trial will perform any of the statistical analyses.

The implementation of the SAP will be as follows:

1. A 'data collection form' will be outlined in a collaboration between the database manager (Leena Caravitis), statistician and principal investigators (Mika Paavola and Teppo Järvinen).

2. The database manager will code each treatment arm into 'treatment A', 'treatment B' and 'treatment C', thus leaving all others blinded to group assignment during the analyses.

3. Blinded data will be delivered to the statistician according to the 'data collection form'.

4. Primary, secondary and exploratory endpoint analyses will be made blinded to group assignment.

5. Results will be presented to the trial Writing and Steering committee, any uncertainties will be clarified and blinded interpretations of the primary endpoint results will be conducted prior to unblinding of data.

REFERENCES

1. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2009;**18**(6):927-32.
2. Ranstam J. Multiple P-values and Bonferroni correction. Osteoarthritis Cartilage 2016;**24**(5):763-4.
3. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clinical orthopaedics and related research 1987(214):160-4.
4. Henseler JF, Kolk A, van der Zwaal P, et al. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al] 2015;**24**(3):376-81.
5. Matsen FA, 3rd, Ziegler DW, DeBartolo SE. Patient self-assessment of health status and function in glenohumeral degenerative joint disease. J Shoulder Elbow Surg 1995;**4**(5):345-51.
6. Godfrey J, Hamman R, Lowenstein S, et al. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. J Shoulder Elbow Surg 2007;**16**(3):260-7.
7. Tashjian RZ, Deloach J, Green A, et al. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am 2010;**92**(2):296-303.
8. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001;**33**(5):328-36.
9. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res 2008;**17**(3):485-95.
10. Bowling A. *Measuring health. A review of quality of life measurement scales.* 3rd ed: Berkshire: Open University Press, 2004.
11. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;**30**(6):473-83.
12. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. Bmj 2002;**324**(7351):1417.
13. Hamilton DF, Lane JV, Gaston P, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. BMJ Open 2013;**3**(4).
14. Sun X, Ioannidis JP, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. Jama 2014;**311**(4):405-11.
15. Bigliani LU MD, April EW The morphology of the acromion and its relationship to rotator cuff tears. Orthop Trans 1986;**10**:228.
16. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc., 1987.
17. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol 2014;**67**(7):769-72.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Interpretation of Blinded Data, Statement of Interpretation

Background assumptions regarding our primary comparison (ASD vs. DA)

1) This superiority RCT is designed to address the true *efficacy* of arthroscopic subacromial decompression (ASD), i.e., can ASD theoretically work? Accordingly, we have chosen patients that – based on the existing literature – represent optimal responders to this index surgical procedure.
2) Conceding that the act of surgery *per se* produces a profound placebo response, a 'true' treatment effect is impossible to disentangle from the nonspecific (placebo or meaning) effects – such as the patients' or researchers' expectations of benefit – without a placebo comparison group.
3) The only difference between ASD and DA treatment groups is that the subacromial decompression, the critical therapeutic (surgical) element, has been carried out for patients in the ASD group.
    a. The critical therapeutic (surgical) element is the component of the surgical procedure that is believed to provide the therapeutic effect (here, subacromial decompression), being distinct from aspects of the procedures that are diagnostic or required to access the disease being treated (here, shoulder arthroscopy).
    b. Apart from the critical therapeutic element, the treatment of the ASD and DA groups is identical, i.e., all "placebo or meaning effect" related to the entire treatment and care is identical.
4) To be deemed effective, ASD should provide a statistically significant benefit over DA in both of the two primary outcomes, pain at rest and activity assessed with a visual analog scale (VAS), as determined by the mean VAS difference between the groups. This is to safeguard against potential multiplicity bias[2].

If ASD is found effective (see above), it should also provide a clinically relevant benefit over DA according to following rationale:

1) There is a proven benefit as follows: Mean VAS-difference between ASD and DA shall exceed the threshold for the minimal clinically important difference (MCID) in VAS. We will consider 15 as the threshold for the minimal clinically important difference (MCID).

AND,

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

2) There is NO proven harm. If there is a proven benefit of ASD but significantly higher proportion of patients show adverse effects, the amount of benefits will be discussed in relation to the frequency and seriousness of the adverse effects.

Statistical commitments:

a) I-T-T is the primary data analysis, but per-protocol analysis will also be carried out.
b) The pre-specified time point of primary interest is 24 months after randomisation.
c) In addition to the two primary outcome parameters, we will also take into account the number of treatment conversions and re-operations, the incidence and seriousness of adverse effects between the ASD and DA groups, and the responder analysis.

Based on these theoretical commitments, our interpretation of the findings will be as follows:

a) If ASD is found superior to DA, the critical therapeutic element of the ASD procedure (subacromial decompression) has a clinically relevant effect on patients with symptoms consistent with SIS.
b) If ASD is not found superior to DA, the critical therapeutic element of the ASD procedure (subacromial decompression) does not have a clinically relevant effect on patients with symptoms consistent with SIS. Considering our efficacy design (study participants are 'optimal responders to ASD' and the surgeons are highly experienced), such finding would imply that ASD does not work at all.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Background assumptions regarding our secondary (independent) comparison (ASD vs. ET)

1) This pragmatic comparison is designed to address whether arthroscopic subacromial decompression (ASD) followed by postoperative rehabilitation is superior to supervised progressive exercise therapy (ET). We recognize that in this pragmatic comparison (ASD vs. ET) the supervised progressive exercise therapy regimen carried out in the ET group is different from the postoperative rehabilitation carried out by patients in the ASD group. In addition, the timing of the start of the actual treatment between the ET and ASD groups was somewhat discrepant due to the time required to arrange the surgery. The ASD patients are also subject to some degree of postoperative immobilization, sick-leave, and modification of pain medication and activities, unlike the patients in the ET group, all of which may also have an effect on the treatment outcome. However, these concord with the current best practice recommendations and the two-year follow-up chosen as our primary time point should dilute the effects of somewhat discrepant timing of the interventions.
2) To be deemed effective, either ASD or ET should provide a statistically significant benefit over ET or ASD, respectively, in both of our two primary outcomes, pain at rest and activity assessed with a visual analog scale (VAS), as determined by the mean VAS difference between the two treatment groups. This is to safeguard against potential multiplicity bias[2].
3) The following concern (apparent confounding) needs to be taken into account in the interpretation. Despite the thorough *preoperative* screening, 10% (14/136) allocated to the two surgical groups had to be excluded because of pathology found after the 1st random allocation. Although this does not have any effect on our primary comparison (ASD vs. DA), the ET and ASD groups are not fully comparable. This discrepancy will possibly skew our results by favouring the ASD group.

Acknowledging all this, if ASD (or ET) is found effective (statistically significant difference in both primary outcomes), it should also provide a clinically relevant benefit over ET (or ASD) according to following rationale:

1) There is a proven benefit as follows: Mean VAS-difference between ASD and ET shall exceed the threshold for the minimal clinically important difference (MCID) in VAS. We will consider 15 as the threshold for the minimal clinically important difference (MCID).

AND,

2) There is NO proven harm. If there is a proven benefit of ASD (or ET) but significantly higher proportion of patients show adverse effects, the amount of benefits will be discussed in relation to the frequency and seriousness of the adverse effects.

FIMPACT
Finnish Shoulder Impingement Arthroscopy Controlled Trial

Statistical commitments:

a) I-T-T is the primary data analysis, but per-protocol analysis will also be carried out.
b) The pre-specified time point of primary interest is 24 months after randomisation.
c) In addition to the two primary outcome parameters, we will also take into account the number of treatment conversions and re-operations, the incidence and seriousness of adverse effects between the ASD and ET groups, and the responder analysis.
d) Given the Background assumption 3.a) (above, the discrepant number of patients with a shoulder pathology other than SIS), we will carry out a worst-case analysis by creating a subgroup of the ET group by removing seven (an equal number of patients excluded from both surgical treatment arms due to pathology found after 1st randomization) worst-cases/highest VAS-pain scores at the primary analysis time-point (24 months). The number of removed cases is based on the assumption that the prevalence of shoulder pathology is identical in the randomized population, while the decision to remove the individual with the highest VAS-pain scores at the end of the study basis on the assumption that shoulder pathology is an effect-modifying factor, predicting poor outcome.

Based on these theoretical commitments, our interpretation of the findings will be as follows:

a) If ASD is found superior to ET in both the complete case and the sensitivity (subgroup) analyses, ASD is a more effective treatment option than ET for patients with SIS.
b) If ET is found superior to ASD in both the complete case and the sensitivity (subgroup) analyses, our results suggests that ET is a more effective treatment option than ASD for patients with subacromial pain syndrome.
c) If there are no statistically significant differences between ASD and ET, ASD and ET are equally effective.

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description	Addressed on page number
Administrative information			
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry	2

	2b	All items from the World Health Organization Trial Registration Data Set	
		Primary Registry and Trial Identifying Number	2
		Date of Registration in Primary Registry	2
		Secondary Identifying Numbers	6
		Source(s) of Monetary or Material Support	28
		Primary Sponsor	28
		Secondary Sponsor(s)	N/A
		Contact for Public Queries	1
		Contact for Scientific Queries	1
		Public Title	1
		Scientific Title	1
		Countries of Recruitment	Appendix 1
		Health Condition(s) or Problem(s) Studied	2
		Intervention(s)	2
		Key Inclusion and Exclusion Criteria	7
		Study Type	2
		Date of First Enrollment	6
		Target Sample Size	6
		Recruitment Status	6
		Primary Outcome(s)	13
		Key Secondary Outcomes	13-15
Protocol version	3	Date and version identifier	29
Funding	4	Sources and types of financial, material, and other support	28
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors	28
	5b	Name and contact information for the trial sponsor	1
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	28

	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	28
Introduction			
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	4-5
	6b	Explanation for choice of comparators	4-5
Objectives	7	Specific objectives or hypotheses	5
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)	6
Methods: Participants, interventions, and outcomes			
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	6, Appendix 1
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	7
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	10-12
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)	12
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)	16
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	N/A

Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	13-15
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	12 (Table2)
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	17
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size	16
Methods: Assignment of interventions (for controlled trials)			
Allocation:			
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	10
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	10
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	10,8
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	9-10, 20-21
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	12
Methods: Data collection, management, and analysis			

Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	17
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	16
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	17
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	18
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	19-20
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	16
Methods: Monitoring			
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	17
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial	17
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	15
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	N/A
Ethics and dissemination			

Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval	6
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)	26
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	6,8
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	N/A
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	17
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site	28
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	28
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	N/A
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	26
	31b	Authorship eligibility guidelines and any intended use of professional writers	26
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	28
Appendices			
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates	Appendix 5

Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	N/A

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated.

AHDAS OLKAPÄÄ -OIREYHTYMÄ-TUTKIMUS
R04200
päiv. 16.11.2011

POTILASTIEDOTE

Vertaileva tutkimus leikkaushoidon ja ei-leikkaushoidon vaikuttavuudesta ahdas olkapää-oireyhtymän (ns. impingement-syndrooma) hoidossa Hatanpään, Jorvin ja Herttoniemen sairaaloissa.

Hyvä potilas,

Teillä on todettu ahdas olkapää-oireyhtymä. Vaivan syynä on ajateltu olevan olkaluun pään ja olkalisäkkeen välinen ahtaus, joka puolestaan arvioidaan johtuvan rappeutuman aiheuttamista ahtauttavista muutoksista ja olkapään lihasvoimien ja lihastasapainon heikentymisestä johtuvasta olkaluun siirtymisestä ylöspäin.

Ahdas olkapää–oireyhtymää voidaan hoitaa joko ei-leikkauksellisesti (kuntoutus) tai leikkauksella. Perinteisesti ei-leikkauksellinen hoito toteutetaan käyttäen lepoa, olkalisäkkeen alaisia kortikosteroidipistoksia, kylmähoitoa, tulehduskipulääkettä ja fysioterapiaa. Leikkaushoitona taas käytetään olkalisäkkeen osittaista poistoa joko avoleikkauksella tai tähystyksessä. Leikkaushoitoon yhdistetään yleensä olkanivelen tähystystutkimus.

Ahdas olkapää–oireyhtymän perimmäiset syyt ja kipua aiheuttavat mekanismit ovat huonosti tunnetut. Hoitolinjan on ratkaissut pikemmin vakiintuneet hoitokäytännöt kuin tutkimuksellisesti osoitettu tietämys vaivan hoidosta. Alkuvaiheessa vaivaa on hoidettu ei-leikkauksellisesti kuormituksen vähentämisellä (lepo, sairasloma) ja kivun lievityksellä lääkehoidoin. Tehokasta pitkäaikaista lihaskuntoharjoittelua on harvoin kokeiltu. Mikäli ei-leikkauksellinen hoito ei ole tehonnut 3 kuukauden kuluessa, on usein päädytty leikkauksella tehtävään olkalisäkkeen osittaiseen poistoon. Tähystystutkimuksen yhteydessä tehtävä olkalisäkkeen osittainen poisto on todettu vertailevissa tutkimuksissa vähintään avoleikkauksen veroiseksi, mutta leikkaushoidon tehoa ei ole verrattu kuntoutuksen tehoon. Lisäksi leikkaushoitoon liittyy aina leikkauskomplikaatioiden riski, joista leikkausalueen bakteeritulehdus on yleisin.

Olemme käynnistäneet tutkimuksen, jossa selvitämme ahdas olkapää-oireyhtymän hoitojen vaikuttavuutta ja hyödyllisyyttä. Teillä on tutkimukseemme soveltuva ahdas olkapää-oireyhtymä ja pyydämmekin Teitä osallistumaan tutkimukseemme. Teidät arvotaan tutkimuksessamme lihaskuntoutusryhmään, leikkauksellisesti tehtävään olkalisäkkeen osapoiston ryhmään tai leikkaukselliseen olkapään pelkkään tähystystutkimusryhmään. Hoidon alkamista edeltää olkapään magneettitutkimus, ja mikäli siinä havaitaan jokin muu olkapääkipuanne selvittävä vaiva kuin ahdas olkapää-oireyhtymä, Teitä ei voida ottaa tutkimukseemme. Luonnollisesti teille kuitenkin järjestetään tämän todetun vaivan hoito.

Mikäli arvonta osoittaa teidät ei-leikkaukselliseen hoitoryhmään, teille aloitetaan 3 kuukauden ohjattu lihaskuntoharjoittelu, joka toteutetaan fysioterapeutin kanssa toteutettavalla 15 harjoittelukäynnillä ja itsenäisesti toteutettavin harjoittein. Fysioterapiakäynnit ovat teille maksuttomia. Hoidon alkuun teille määrätään sairaslomaa mikäli tilanteenne sitä edellyttää.

Mikäli teidät arvotaan leikkaukselliseen olkalisäkkeen osapoistoryhmään, saatte leikkausajan olkanivelen tähystykseen ja tähystyksen yhteydessä tapahtuvaan olkalisäkkeen osittaiseen poistamiseen. Toimenpide suoritetaan yleisanestesiassa (ns. nukutus) ja kotiudutte ensimmäisenä leikkauksen jälkeisenä päivänä noin vuorokauden sairaalaseurannan jälkeen. Mikäli toimenpiteenne toteutetaan päiväkirurgisena hoitona, on mahdollista, että pääsette kotiin jo operaatiopäivänä. Toimenpiteen jälkeen operoitua yläraajaa tulee pitää kantositeessä noin vuorokauden ajan. Leikkaushoitoa seuraa 4-6 viikon sairausloma ja 6 viikon kuluttua toimenpiteestä on jälkitarkastus poliklinikalla.

Mikäli teidät sitä vastoin arvotaan olkapään tähystysryhmään, saatte ajan toimenpiteeseen, jossa teille tehdään ainoastaan olkanivelen tähystys, mutta ei olkalisäkkeen osapoistoa. Mikäli tähystyksessä kuitenkin havaitaan olkapäävaivaanne selvittäviä rakenteellisia muutoksia (esim. kiertäjäkalvosimen repeämä), ne hoidetaan vallitsevien periaatteiden mukaisesti samassa leikkauksessa. Toimenpide suoritetaan yleisanestesiassa (ns. nukutus) ja kotiudutte ensimmäisenä leikkauksen jälkeisenä päivänä noin vuorokauden sairaalaseurannan jälkeen. Mikäli toimenpiteenne toteutetaan päiväkirurgisena hoitona, on mahdollista, että pääsette kotiin jo operaatiopäivänä. Toimenpiteen jälkeen operoitua yläraajaa tulee pitää kantositeessä noin vuorokauden ajan. Leikkaushoitoa seuraa 4-6 viikon sairausloma ja 6 viikon kuluttua toimenpiteestä on jälkitarkastus poliklinikalla.

Mikäli teidät arvotaan jompaankumpaan hoitoryhmistä, jossa ei suoriteta olkalisäkkeen osapoistoa ja olette 6 kuukautta hoidon aloittamisesta edelleen tyytymätön ahdas olkapää-oireyhtymänne hoitotulokseen, Teillä on mahdollisuus päästä viipymättä olkalisäkkeen tähystykselliseen osapoistoleikkaukseen.

Kaikissa hoitoryhmissä paranemista seurataan säännöllisin maksuttomin käynnein hoidostanne vastaavan sairaalan poliklinikalla 6 kuukautta, sekä 2, **5 ja 10 vuotta hoidon aloittamisen jälkeen**. Joka käynnin yhteydessä Teille tehdään tilannettanne ja olkapäänne kuntoutumista kartoittava kysely ja toimintakykytesti. Lisäksi tilannettanne kartoitetaan puhelimitse ja kotiin postitettavin kyselyin 3 kuukautta ja 12 kuukautta hoidon aloittamisen jälkeen.
Mikäli ette ole tyytyväinen tutkimuksessa seurattavan olkanivelenne tilanteeseen tai Teille tulee kysyttävää tutkimuksesta, voitte ottaa yhteyttä tämän potilastiedotteen lopussa mainittuun tutkimuksen yhteyshenkilöön.

Tutkimukseen osallistuminen antaa tärkeää tietoa ahdas olkapää-oireyhtymän hoidosta ja tulokset tullaan julkaisemaan kansainvälisissä lääketieteen alan julkaisuissa. Tutkimukseen osallistuminen on vapaaehtoista, eikä siitä kieltäytyminen vaikuta nykyisiin tai tuleviin hoitoihinne. Halutessanne voitte keskeyttää osallistumisenne tutkimukseen milloin tahansa. Tutkimuksen osallistuminen ei aiheuta teille vaaraa ja tutkimukseen osallistuessanne kuulutte normaaliin potilasvakuutuksen piiriin. Tutkimuksen yhteydessä saatte palautetta parantumisesta suoraan tutkivilta lääkäreiltä.

Terveisin,

Terveisin,

Kari Kanto
Ortopedi, erikoislääkäri
Tampereen kaupunki
Hatanpään sairaala

Pirjo Toivonen
tutkimushoitaja
Hatanpään sairaala
puh. 03-565 72827
pirjo.h.toivonen@tampere.fi

Ahdas olkapää-oireyhtymän hoitoa käsittelevä tutkimus (ETL-koodi R04200)

Minua on pyydetty osallistumaan yllä mainittuun tieteelliseen tutkimukseen ja olen saanut sekä kirjallista että suullista tietoa tutkimuksesta ja mahdollisuuden esittää siitä tutkijoille kysymyksiä.

Suostun osallistumaan yllä mainittuun tutkimukseen.

Ymmärrän, että tutkimukseen osallistuminen on vapaaehtoista ja että minulla on oikeus kieltäytyä siitä sekä perua suostumukseni milloin tahansa syytä ilmoittamatta. Ymmärrän myös, että tiedot käsitellään luottamuksellisesti.

Ahdas olkapää-oireyhtymäni hoitoa käsitteleviä tietoja voi tutkimusta varten pyytää seuraavista paikoista:

- ☐ Kansaneläkelaitos (Kela)
- ☐ Terveyden- ja hyvinvoinnin laitos (THL)
- ☐ Muut sairaalat
- ☐ Muut hoitolaitokset (esim. yksityislääkäriasemat)

Paikka ja pvm

_____ ___.___. 20_____

_____
tutkittavan allekirjoitus

_____
nimenselvennys

_____
henkilötunnus tai syntymäaika

_____
osoite

Paikka ja pvm

_____ ___.___.20_____

_____
suostumuksen vastaanottajan allekirjoitus

_____
nimenselvennys

Supplementary 1: Patients satisfaction to the treatment given

How satisfied are you with the treatment given? Mark the answer closest to your situation.

1. Very satisfied, my shouder has healed completely.
2. Satisfied, I have only minor, activity related symptoms. My shoulder is much better than before treatment.
3. Somewhat satisfied, i have only minor symptoms. My shoulder is better than before treatment.
4. Dissatisfied, my shoulder is the same as before treatment.
5. Very dissatisfied, my shoulder is worse than before treatment.