STATISTICAL ANALYSIS PLAN (**SAP**) FOR:

The Odense-Oslo Meniscectomy versus Exercise (OMEX) study.

A randomized controlled trial for treatment of degenerative meniscus tears in middle-aged patients with a 2-year follow-up

OMEX steering group:

Nina Jullum Kise, MD, PhD-student[1,2]

Ewa M Roos, PhD, PT, Professor[3] (Principal investigator)

May Arna Risberg, PhD, PT, Professor[4]

Statistical advisor:

Jonas Ranstam, Professor[5]

[1]Orthopedic Department, Martina Hansens Hospital, Baerum, Norway

[2]Institute of clinical medicine, Faculty of medicine, University of Oslo, Norway

[3]Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

[4]Norwegian Research Center for Active Rehabilitation (NAR), Department of Orthopedics, Oslo University Hospital and Department of Sports Medicine, Norwegian School Sport Sciences

[5]Lund University and Skane University Hospital, Lund, Sweden

Table of contents

1. Study Synopsis……....3
2. Study Objectives and outcomes……..4
2.1. Primary Objective and Outcome……..4
2.2. Secondary Objectives and Outcomes……...……...6
2.3. Exploratory Objectives and Outcomes…….……...…….6
2.4. Descriptive Assessments.……...……..7
2.5. Economic Evaluation……...7
2.6. Radiologic Evaluation……..8
2.7. Specification of Endpoints……...……8
2.7.1. Primary Endpoint……..8
2.7.2. Secondary Endpoints……9
3. Study Design…….....10
3.1. Sample Size……..……..10
3.2. Randomization and Blinding…….…11
4. Study Population……...12
4.1. Subject Disposition……....12
5. Statistical Analysis……...….13
5.1. Primary Endpoint……...13
5.2. Secondary Endpoints…….....13
6. Major Protocol Deviations…….....14
7. Implementation of Analysis Plan……..14
8. Figures……...15
8.1. Figure 1. Flow chart / Study design……...15
8.2. Figure 2. Flow chart / Groups analyzed in ITT, PP and AT-analyses……..…….16
9. Tables……17
9.1. Table 1. Adverse Events.……17
9.2. Table 2. Baseline characteristics……18
9.3. Table 3. Treatment-related variables…….....20
9.4. Table 4. Outcome at 2 years……...21
10. References……...22
11. Appendix……Separate file
11.1. Appendix 1. PhD Study protocol…….....

1. Study Synopsis

Meniscal tears are common and arthroscopic meniscal surgery is among the most frequently performed procedures in orthopedic surgery[1,2]. Only one out of ten procedures includes repairs[2] and most operations comprise arthroscopic partial meniscal resections only.

75% of these procedures are performed in those above 35 years of age[3]. As many as 300 persons per 100 000 are going through arthroscopic partial meniscal resections annually[2,4,5]. A recent study reveals that the incidence of meniscal procedures performed in Denmark is doubled from year 2000 to 2011[6].

Most meniscal tears in particular in those 35 years and above, are degenerative tears and may be regarded as the first sign of osteoarthritis (OA)[7,8]. However, degenerative meniscal tears may be present in asymptomatic knees as well and degeneration in meniscal tissue could be looked upon as wrinkles with age and alternatives to surgical treatment should be considered[9].

A tear in a meniscus with degenerative changes is often associated with preexisting structural changes in the articular cartilage that may represent early-stage osteoarthritis[10]. The menisci have protective properties relative to the joint cartilage surfaces and removal of meniscal tissue leads to degeneration[11-13] that correlates with the amount of tissue removed[11]. The risk for OA is definitely increased in knees after meniscal resections[7,14].

Short- and long-terms studies have shown that exercise therapy improves function and activity level in patients with degenerative meniscal tears whether they have undergone meniscal resections or not[15-18]. Several studies have shown that arthroscopy is no better than sham surgery[19-21], and a recent meta-analysis concludes that there is moderate evidence to suggest that there is no benefit to arthroscopic meniscal debridement for degenerative meniscal tears in comparison with non-operative or sham treatments in middle-aged patients with mild or no concomitant osteoarthritis[22]. Only one recent study reveals that patients treated with meniscal resection combined with instructions for exercise therapy had significant lower pain level after 12 months than patients treated with instruction for exercise therapy only[23].

Considering the large and assumedly increasing number of patients undergoing surgery, the large number of studies that have shown that meniscal resection increases the risk for knee osteoarthritis, the documented favorable effects of exercise therapy, the scarce documented long-term effects of partial meniscal resection surgery, and the lack of knowledge upon natural development without surgery, there is need for more high quality long-term follow-up randomized controlled studies (RCTs) of the long term effect of exercise therapy versus partial meniscal resection for patients with degenerative meniscal tears.

In this RCT we randomized 140 middle-aged patients (35-60 years) with MRI-verified degenerative medial meniscal tears and no or early knee osteoarthritis (Kellgren-Lawrence (KL) grade of 0, 1 or 2[12]) to treatment with either arthroscopic partial meniscal resection or a 12-weeks exercise therapy program[24]. The primary outcome is change in knee function measured by $KOOS_4$, a composite of four of the five subscales in the knee specific questionnaire KOOS (Knee injury and Osteoarthritis Outcome Score)[25], from baseline to the 2-year follow-up (Figure 1).

2. Study Objectives and Outcomes

All methods used in this study are described in the PhD study protocol dated April 2012 (Appendix 1). All outcomes were obtained from all participants at baseline and at follow-ups at 3 and 12 months. The 2-year follow-up was performed as a survey included KOOS sent by post and includes all outcomes except the isokinetic muscle strength tests and the performance tests (Table 4). SF-36 was included at all follow-ups but will in this 2-year follow-up study be analyzed only for base line and main outcome at 2 years (see section 2.7).

The 24 months follow-up was finalized in December 2014.

2.1. Primary Objective and Outcome

The primary objective is to compare difference in change in knee function measured by $KOOS_4$ from baseline to the 2-year follow-up (including follow-ups at 3 months and 1 year) between patients randomized to exercise therapy or arthroscopic partial meniscal resections. The reason for choosing this primary endpoint is as follows:

The KOOS is a knee-specific instrument, developed to assess the patients' opinion about their knee and associated problems. The KOOS evaluates both short-term and long-term consequences of knee injury[26]. The KOOS subscales has previously been found reliable and are validated for patients with a range of knee injuries including those having knee arthroscopy and those with knee OA[26,27].

The KOOS holds 42 items in 5 separately scored subscales; Pain, other Symptoms, Function in daily living (ADL), Function in Sport and Recreation (Sport/Rec), and knee-related Quality of Life (QOL)[26]. Each item in KOOS is scored from 0-4 on a Likert scale. Subscale scores are given separately (see www.koos.nu for user's guide and scoring) ranging from 0 (worst) to 100 (best). In this study, each subscale except ADL will be calculated according to the instructions in the user's guide and the composite $KOOS_4$ will be calculated giving each subscale equally large impact on the $KOOS_4$ score using this formula:

$KOOS_4$ = (KOOS Pain + KOOS Symptoms + KOOS Sport/Rec + KOOS QOL)/4

An average KOOS score derived from any number of calculated subscales scores can be used as primary endpoint in an RCT, if defined a priori. However, since $KOOS_4$ has not been subjected to psychometric validation, they are intended for statistical purposes only and cannot be interpreted clinically. The individual KOOS subscales must be analyzed as secondary outcomes to enable clinical interpretation of the contributions of each subscale of the KOOS score.

One reason for choosing a composite KOOS score as the primary endpoint is to avoid issues with *multiplicity*. Another reason is to avoid bias between groups because *group belongingness* only may favor scores in some KOOS subscales. The third reason is that especially the KOOS subscale ADL has *low effect size* and is actually not optimal for measuring differences in knee function when analyzed in this patient population. The assumed group belongingness bias and effect size difficulties will be discussed below:

Group belongingness: Bias between the patient groups will certainly occur as a consequence of the different treatment interventions (exercises versus surgery). It is reasonable to suggest

that the context for each group at follow-ups is different due to experiences and knowledge the patients have acquired during either surgery or exercise therapy. This may have an effect on the change in scores. Gauffin et al has shown higher KOOS pain scores (less pain) after 12 months in patients that have gone through arthroscopy in addition to exercise, compared to patients that had exercise only[23]. This may in fact be a result from placebo effect of the surgery and the patients' expectations of good results[20,28].

In a previously published study by our group based on the 20 first subjects in the exercise therapy group of the OMEX trial, patients scored statistically and clinically significant higher in all KOOS subscales after 12 weeks exercise therapy[24]. The highest effect sizes (1.7 and 1.7) were seen in the subscales Sport/Rec and QOL, compared to the lower effect sizes in the subscales ADL, Pain and Symptoms (1.1, 1.5 and 0.5 respectively). This may actually reflect an unintended side effect of the exercise intervention: During the exercises, the patients have had frequent contact with the physiotherapist who assumedly have given instructive information and taught the patients that physical activity is not dangerous in spite of the experience of pain. Thereby, the patients have the courage to get more physical active and this may influence on the changes in knee related quality of life, as well, and changes in the subscales QOL and Sport/Rec.

Therefore, the group belongingness itself may have influence on the self-reported follow-up scores.

Effect sizes: In the above mentioned study by Stensrud et al, the smallest effect sizes after exercise therapy were found for the KOOS subscales Symptoms and ADL (0.5 and 1.1)[24]. This is analogue to Herrlin et al´s study of patients having arthroscopy and exercises compared to exercises only[16], where the smallest changes in scores were seen in the subscales Symptoms and ADL. Likewise, also in Gauffin et al´s study, the smallest changes were seen in the subscales Symptoms and ADL, and there were no significant differences between the treatment groups for the subscales Symptoms and ADL in the per-protocol analyses of mean scores at 12 months, but a significant difference in favor of the surgery group for the subscale Symptoms in the as-treated-analysis [23].

Therefore, the KOOS subscales Symptoms and ADL seems to have low effect sizes and only small, if any, differences between groups are expected.

Another argument for excluding the subscale ADL from the composite score: This patient population is middle-aged (35-60 years), the patients have no serious comorbidities (due to the inclusion and exclusion criteria) and it is reasonable to suspect that the patients at baseline actually not have serious problems with usual activities of daily life, even though they experience knee pain and functional limitations. Thereby, the KOOS subscale ADL could be considered less valid and leave less room for improvement in this population. On the other hand this is an early OA group, and their baseline score is about 50-60. I think this paragraph could be omitted since this the speculations are not substantiated by facts.

Therefore, we have chosen not to use a single subscale as the primary outcome of this RCT, but to use a composite KOOS score; the $KOOS_4$, without the subscale ADL.

2.2. Secondary Objectives and Outcomes

The secondary objectives are to compare change from baseline to the 2 years follow-up (including all follow-ups) between groups in a range of outcomes, except for isokinetic muscle strength and functional performance, for those the endpoint is 1-year follow-up.

Secondary outcomes differ from exploratory outcomes, because secondary outcomes will possibly enable clinical interpretations; not only be explanatory or hypothesis generating. Secondary outcomes give support to the primary outcome and will contribute to better interpretation and eventually more knowledge. Since secondary outcomes only are seen as supportive for the primary outcome, multiplicity is not considered to be a problem and not statistically corrected for.

These outcomes are:

1) The five subscales of KOOS in hierarchic sequence: a. QOL b. Sport/Rec. c. Pain d. Symptoms e. ADL.

2) Functional performance: a. The one-leg hop test b. The 6 meter timed hop test c. The knee-bending test

3) Isokinetic muscle strength: a. Quadriceps strength b. Hamstrings strength

4) SF-36: Physical component summary (PCS) and mental component summary (MCS)

5) Global knee rating (much better – better– unchanged – worse – much worse)

7) To analyze the number needed to treat (NNT), based on a 15% difference in change in KOOS between groups from baseline to the 2 years applied in recent studies in similar patient populations[29,30], (see section 2.7.1.).

8) The health economic utility instruments EQ-5D index including EQ-5D VAS is included in the baseline and follow-up surveys. This will not be a part of this 2-year follow-up, but will be published later in a separate paper (see section 2.5.).

2.3. Exploratory Objectives and Outcomes

The exploratory objectives are to compare change from baseline to the 2-year follow-up (including all follow-ups) between the groups in a range of outcomes. These outcomes will only be exploratory and/or hypothesis generating, which is why multiplicity is not considered to be a problem and not statistically corrected for.

These outcomes are:

1) Weight change in percent (from baseline to 3 and 12 months follow-up).

2) Change in physical activity level a. physical activity level b. intensity c. exercise sessions per week d. hours being physically active per week (based upon questions in baseline questionnaire: Kind of activity / number of sessions a week / total hours a week and questions in 2-years questionnaire: How often / how intense / how long sessions) (from baseline to 24

months follow-up).

3) Pain intensity on a 100 mm VAS with terminal descriptors of 'no pain' (0) and 'worst pain possible' (100) (from baseline to 3 and 12 months follow-up).

Further exploratory objectives may be added later on.

2.4. Descriptive Assessments

Baseline characteristics will be presented in a table (Table 2).

Furthermore, the following treatment-related variables will be presented descriptively:

1) Compliance with exercise during the 12 weeks is recorded from the patients exercise diary notes. The patients were instructed to complete two to three exercise sessions a week over twelve weeks (24-36 sessions). Compliance is assessed as the total number of exercise sessions completed out of the minimum of 24 sessions. *Excellent compliance* is defined as participation in 24 or more sessions (more than 100 %), *satisfactory compliance* is defined as 19-23 sessions (80-100 %) and *poor compliance* is defined as 18 or less sessions (less than 80 %). Less than 80 % compliance is defined as not completed treatment allocation.
2) Pain level during the 12 weeks exercise treatment is recorded from the patients exercise diary notes where they have reported pain on a visual analogue scale (VAS) during and after each training session.
3) Satisfaction with the treatment effect will be registered at each follow-up until 12 months on a five-point Likert scale (very dissatisfied, dissatisfied, neither satisfied nor dissatisfied, satisfied, very satisfied).
4) Patients crossing over from the exercise group to arthroscopy and patients in both groups going through additional surgeries during the 24 months follow-up period will be registered (Table 3).
5) Patients crossing over from the arthroscopy group to exercise, not have had postoperative active physiotherapy (exercises) more than 80 % of twice a week in 12 weeks (19 or more sessions)(equivalent to 80 % in the exercise group).
6) Adverse events (AE) and seriously adverse events (SAE) will be registered and categorized into index knee or sites other than index knee. The project workers will record any adverse events that the participant experiences or tells them about and at all follow-ups all participants are asked about potential adverse events. For all participants in both groups including all patients crossing over (see section 2.7.1.) to the other intervention a project worker will look through hospital records to register if any pre-defined perioperative or postoperative adverse events occurred, or adverse events related to the exercise therapy (Table 1).

2.5. Economic Evaluation

The EQ-5D will be applied in a health economic evaluation. This will not be a part of the 2-year follow-up of this trial but is planned for the 5-year follow-up and will be published in a separate paper.

2.6. Radiographic Evaluation

One of the inclusion criteria in this RCT is osteoarthritis evaluated according to KL in a fixed-flexion posterior-anterior radiography. This procedure will be repeated in the 5-year follow-up and results will be published in a separate paper. Radiography is not performed in the 2-yeas follow-up.

2.7. Specification of Endpoints

2.7.1. Primary Endpoint

The trial is designed as a superiority trial, i.e. we expect that one of the treatment groups will improve more than the other group in the primary outcome $KOOS_4$ from baseline to the primary endpoint after 2 years. Thereby, the primary endpoint is based upon the between-group difference in change in $KOOS_4$.

The primary analysis of the primary outcome ($KOOS_4$) will be analyzed in an intention-to-treat (ITT) analysis. It implies that for all patients randomized to one treatment, should be analyzed according to the treatment to which the patient was allocated, irrespective of whether they received this or some other treatment, or no treatment at all[31] (Figure 2). The purpose of this ITT principle is to preserve the theoretical basis for the validity of the statistical results, specifically by eliminating the possibility that patients with known or unknown prognostic factors are systematically selected to a treatment[32].

The *ITT population* is equal to all patients randomized to the two treatment arms, and the dataset is equal to the "all patients randomized set" (APRS).

Missing values due to patients´ absence from follow-ups or withdrawal from the study are not uncommon in clinical trials[31]. Several approaches are described for handling missing data in the ITT analysis, and among them "last observation carried forward" (LOCF) (the last observed outcome is regarded as the final outcome), multiple imputation (MI) (stochastic technique which depends on model-based imputation of multiple values for each missing observation) and mixed model (MM) analysis. A MM analysis includes all patients with at least one baseline or follow-up value, and includes both fixed and random factors. The MM method is shown to have higher statistical precision in analyzing data from an RCT, than LOCF or MI methods[31]. Therefore, MM method is chosen for the analysis of the primary outcome in this trial.

Treatment effect will be determined as difference between groups in change in the primary outcome $KOOS_4$ from baseline to the 2-year follow-up. Since KOOS contains the full and original version of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), it has been usual to apply a minimal important change (MIC) of 10 points, which has been demonstrated for WOMAC[26]. Recent studies in OA patient populations[29,30] have applied a MIC of 15%. However, percentage change from baseline is not recommended as an outcome in controlled trials, since it has low statistical power, is highly sensitive to changes in variance and fails to protect from bias in the case of baseline imbalance.

We acknowledge that MIC is dependent on context factors such as population, intervention, and time to follow-up[33]. In this trial, a possible bias may be a result of the different interventions (see section 2.1.) and this may affect the scores in the different KOOS subscales different in the two treatment groups, and thereby the change in $KOOS_4$ may differ in the two groups as a result of the bias alone. Therefore, in the between-group analyses, the MIC is dubious and of poor value. On the contrary, in the within-group analyses, it is still reasonable to assess a 10-point difference as MIC.

Superiority will be tested using the two-sided 95% confidence interval (CI) of the mean change in $KOOS_4$ between the two treatment groups.

2.7.2. Secondary Endpoints

Secondary endpoints will be analyzed for between- and within-group differences using ITT, PP (per protocol) and AT (as treated) analyses (see sections 2.7.1 and 5.2.)(Figure 2).

The *PP population* will be defined as the following:

In the arthroscopy group, included patients are: All patients randomized to arthroscopy that went through surgery, and not have had postoperative active physiotherapy (exercises) more than 80 % of twice a week in 12 weeks (19 or more sessions)(equivalent to 80 % in the exercise group), are included.

In the exercise group patients who participated in the program with at least 80 % compliance during the 12 weeks intervention period and did not cross over to the surgical treatment arm during the 2 years period and not have had any other surgery to the index knee (osteotomy, arthroplasty) within two years, are included.

This means that the following will be excluded from the PP analysis:

1) Those who were randomized to arthroscopy
    a. but never went through surgery
    b. but had postoperative exercise therapy supervised by physiotherapists (defined as 19 or more sessions)

2) Those who were randomized to exercise therapy supervised by physiotherapist
    a. but never went through exercise or participated in less than 19 sessions (80 % of the minimum of 24 exercise sessions)
    b. but crossed over to the surgical treatment arm within 2 years

The *AT population* will be defined as the following:

In the arthroscopy group, included patients are: All patients randomized to arthroscopy that went through surgery, and not have had postoperative active physiotherapy (exercises) more than 80 % of twice a week in 12 weeks (19 or more sessions)(equivalent to 80 % in the exercise group), are included. Additionally; Patients from the exercise group that within 2 years crossed over to surgery, are included in the arthroscopy group.

In the exercise group patients who participated in the program with at least 80 % compliance

during the 12 weeks intervention period and did not cross over to the surgical treatment arm during the 2 years period and not have had any other surgery to the index knee (osteotomy, arthroplasty) within two years, are included. Additionally; patients from the arthroscopy group that within two years crossed over to exercise (performed more than 80 % of twice a week in 12 weeks (19 or more sessions), are included in the exercise group.

The KOOS subscales Pain, Symptoms, ADL, Sport/Rec and QOL, EQ- 5D Index, EQ-5 VAS, weight and global knee function will be presented as mean (with 95% CI) for each treatment group. Between-group differences in change from baseline to 2 years will be statistically assessed. Each subscale of the KOOS will be presented graphically for its development over the 2 years period.

All issues during the trial found in the treatment records from the project physiotherapist, hospital records or the questionnaire from the follow-ups will be assessed to determine whether it represents an AE or not. AE will be presented in a table (see Table 1) and analyzed statistically by comparing actual numbers of serious AE (site other than index knee, index knee and all serious events) and non-serious AE (site other than index knee, index knee and all serious events).

AE associated with surgery will be given in Table 1 for the following groups:

1. Those receiving arthroscopic meniscal resection by randomization

2. All treatment dependent AE reported during and after surgery in those who had surgery even though randomized to the exercise therapy throughout the follow up period

AE associated with exercise therapy will be given in Table 1 for the following groups:

1. Those remaining in the 'Exercise therapy' group throughout the follow up period

2. All treatment dependent AE reported prior to the surgery in those who had surgery even though randomized to the exercise therapy

3. Those randomized to surgery but crossing over to exercise during the follow-up period

3. Study Design

3.1. Sample Size

Primary outcome for the RCT is the ITT- analysis of the between-group difference in change in $KOOS_4$ from baseline to the 2-year follow-up.

We wanted to detect a 10-points difference (Δ) with a standard deviation (SD) of 15, with a level of power (β) of 90% and a level of significance (α) of 0.05. The calculation was performed using this formula, where the f (α, β) is a constant (given by Stuart J. Pocock; "Clinical trials. A practical approach":

$2 \times SD^2 / \Delta^2 \times f(\alpha, \beta) = 2 \times 15^2 / 10^2 \times 10.5 = 47.25 \approx 48$ (patients in each group)

Estimated dropout rate at 2 years was 15%, thereby, the number of subjects (N) randomized was to be:

N – (15 % out of N) = 47.25
N = 55.6 ≈ 56 (patients in each group)

Thus, 112 patients randomized to two groups are necessary for the ITT-analysis.

3.2. Randomization and Blinding

The schedule for randomization was randomly generated using a computer before the initiation of the trial. To conceal the outcomes of the randomization, the allocation numbers were put in concealed, opaque C5 envelopes. An independent staff member prepared the envelopes. These were kept in a locked location accessible only by one research assistant. Following the informed consent and completion of the baseline measures, the envelopes were opened by the patients and the allocation was revealed.

The randomization was stratified for gender, to ensure equal gender distribution in the two groups. In both genders there were used block randomization with blocks of eight.

The randomization was not stratified for center, because initially the trial was planned to be finalized in one hospital only; the Oslo University Hospital (OUH). Due to less patient flow than suspected, the recruitment was after patient number 53 taken over by Martina Hansens Hospital (MHH). The recruitment started at OUH in October 2009 and continued after patient number 54 at MHH from May 2011 and finished (patient number 140) in September 2012.

All patients went through baseline and follow-up tests at NIMI (Norsk Idrettsmedisinsk Institutt), Ullevål, Oslo. Patients randomized to arthroscopy were treated in the respective hospital that they were recruited from, and patients randomized to exercise therapy went through the interventions at NIMI (patients recruited between patient number 1 and 83) or at Gnist Trening og Fysioterapi, Bærum, an external physiotherapist institute.

Equality in patient treatment in the two interventional hospitals and two physiotherapist institutes respectively, was considered extremely important, especially on account of the lack of randomization stratification for center. Similarity in the recruitment and treatment procedures was ensured by written procedures and protocols and head-to-head instructions and dialogues. Only one surgeon was involved in recruiting at OUS, and another one at MHH. Fifty-four out of the total 64 (84 %) surgical procedures were performed by these two surgeons. Equivalently, the exercise therapy was performed by one dedicated physiotherapist in each institute, following similar procedures and protocols.

The RCT is a single-blinded study. The outcome assessor was blinded to group allocation, was not involved in providing the interventions, and unaffiliated with the treatment sites. The participants, the project physiotherapist and the surgeons delivering part of the interventions could not be blinded.

The statistician performing the statistical analyses will be blinded to group allocation.

Järvinen et al have estimated that blinded interpretation of study results feasibly and

effectively might diminish interpretation bias[34]. Therefore, the writing committee of this study (see section 7) will, prior to breaking the code, conduct three interpretations of the results, one assuming that *Group I* was randomized to exercise, and the other assuming that Group I was randomized to arthroscopy, and *Group II* to the opposite intervention. In this superiority trial there are three possible results;

1) Group I is statistically superior to Group II
2) Group II is statistically superior to Group I
3) No statistically significant difference between the groups

Until the writing committee has agreed in writing on the clinical interpretation of the results the randomization code will not be broken, ensuring that bias during data interpretation is kept to a minimum.

4. Study Population

4.1. Subject Disposition

Study procedures, including recruitment strategies and inclusion and exclusion criteria, have been described in the PhD protocol (Appendix 1). Patients included in the trial were randomized to: A) Exercise therapy or B) Arthroscopic partial meniscal resection. All consecutive patients who came to the two inclusion sites on the days of the orthopedic surgeons outpatient clinic fulfilling the eligibility criteria were asked to enter the study. Patients fulfilling all eligibility criteria, but refusing to participate in the randomization, were registered.

Crossovers are a common problem in studies randomizing to surgical or non-surgical treatments[35,36]. In this study participants who experienced impairment of their symptoms or lack of improvement during the 2-year observation period, were reassessed at any time when necessary by the orthopedic surgeon who was responsible for the study. Indication for surgery (crossing over from exercise to arthroscopy) or re-operation (meniscal re-resection or other surgery in the index knee within 2 years) was based upon patient's history, clinical and radiological examinations.

Patients from the arthroscopy group that had postoperative active exercise therapy supervised by a physiotherapist more than 80 % of twice a week in 12 weeks (19 or more sessions)(equivalent to 80 % in the exercise group) were regarded as crossing over to the exercise group.

The frequency of crossovers from exercise to arthroscopy and vice versa, re-operations with meniscal re-resections or other surgery in the index knee within 2 years will be registered and reported (Table 3).

5. Statistical Analysis

5.1. Primary Endpoint

In this RCT, the primary outcome is the between-group difference in change in $KOOS_4$ from baseline to 2-year follow-up. p-values and 95% CI for the main outcome will be presented to assess superiority. An independent statistician will analyze the primary outcome.

Multiple regression analysis will be made using a mixed model (MM) ANOVA (analysis of variance) in all patients randomized data set (APRS). In this model, patient is the random factor and time points (baseline, 3, 12 and 24 months), treatment arm (exercise, arthroscopy) and recruitment site (OUH, MHH) are fixed factors. Gender, a randomization stratification factor will be encountered for to avoid variance. Furthermore interactions between the fixed factors will be included in the model.

To reduce the variance, results of continuous variables will be adjusted for baseline imbalance.

Preconditions for multiple regression analysis are independency, normal distribution and constant variance of the residuals. This will be checked in SPSS by plotting a histogram over standardized residuals, a normality plot for standardized residuals and a plot with standardized residuals contra standardized predictive value. SPSS will make a list of residuals larger then 3 and smaller than -3.

Different statistical analysis software will be used; SPSS, Stata, SAS.

The program code for the analysis of the primary outcome is in SAS:

```
PROC MIXED;
CLASS TREAT TIME CENTRE;
MODEL CHANGE= BASELINE*TIME TREAT*TIME CENTRE SEX TIME;
REPETED TIME / SUBJECT=PATIENT TYPE=UN;
RUN;
```

TREAT is the treat variable, TIME is nominal time since baseline in months, CENTER indicates center, CHANGE is change from baseline in $KOOS_4$, BASELINE is $KOOS_4$ at baseline, SEX is the patients gender, and PATIENT is the patient´s id.

5.2. Secondary Endpoints

Secondary outcomes are between- and within-groups comparisons of the change from baseline to the 2-year follow-up in all secondary endpoints except for the isokinetic muscle strength and functional performance data, where 1-year follow-up is the end point.

All outcomes will be checked for normality and statistical methods will be dependent on data distribution. When fulfilling the requirements for regression analyses, ANOVA analyses of the ITT populations will be conducted; otherwise non-parametrical tests will be used.

Finally, PP- and AT-analyses will be conducted in a range of secondary outcomes.

6. Major Protocol Deviations

In the Clinical Trial document (registration number NCT01002794) published in October 25. 2009 before the start of inclusion of patients the primary outcome is settled as "KOOS", not more specified. In the PhD-protocol (April 2012) the composite $KOOS_4$, (an averaged score of the four KOOS subscales: Pain, Symptoms, Sport and Recreation and QOL) was decided as the primary outcome for the RCT.

There is no major difference between this SAP and the PhD-protocol dated in April 2012, or the Clinical Trial document published in 2009 prior to study start.

There is one minor deviation from the PhD-protocol: It was planned to define patients in the exercise group scoring 44 or less in the QOL outcome KOOS as failures and thereby fulfill the criteria for crossover to the arthroscopy group. This definition was not used in the clinical situation. The indication for arthroscopy (or reoperations) was determined by the treating orthopedic surgeon based upon the patient's history, clinical findings and radiologic exams.

7. Implementation of Analysis Plan

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

The implementation of this SAP will be as follows:

1. A 'data collection form' will be outlined in a collaboration between the database manager, statistician and principal investigator.

2. The database manager will code each treatment arm into 'treatment I' and 'treatment II' and thus leaving all others blinded from treatment during the analyses.

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

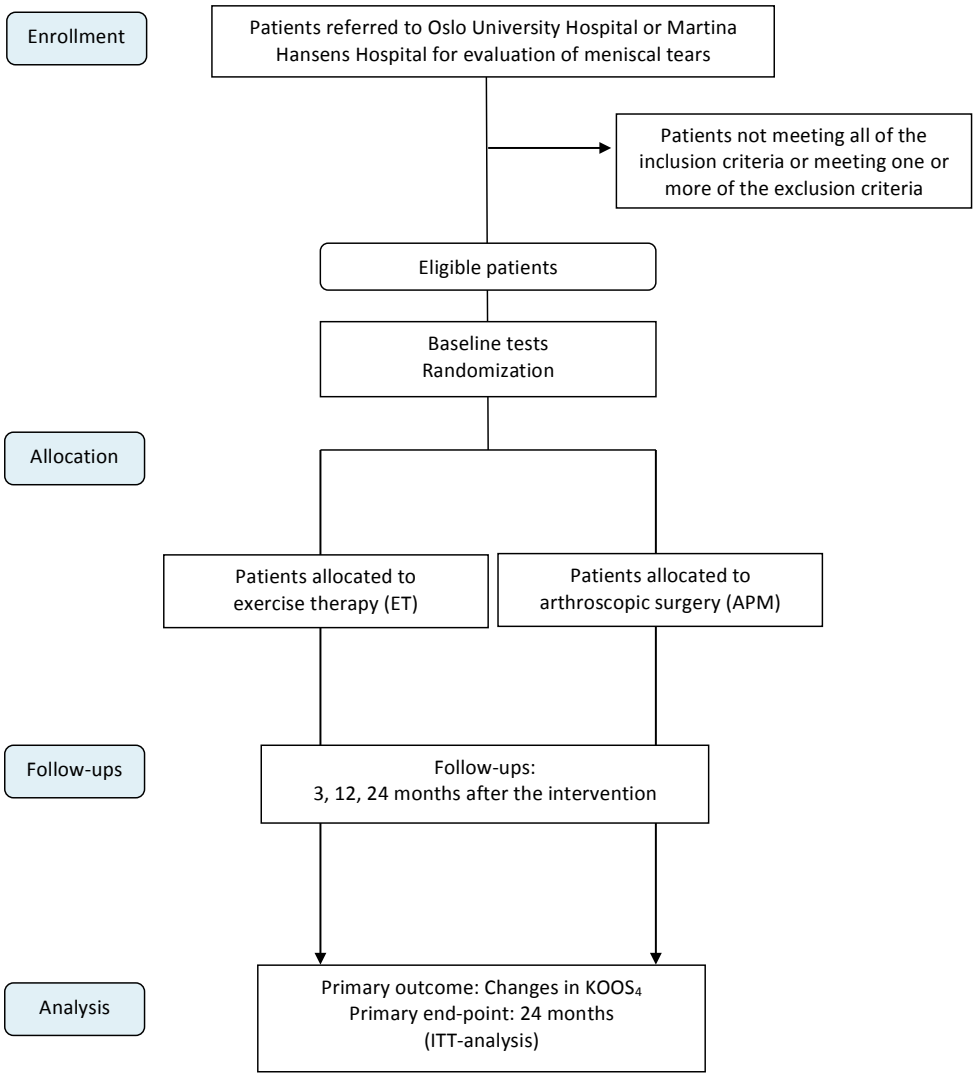
4. Primary, secondary and exploratory endpoint analyses will be made blinded from treatment allocation.

5. Results will be presented to the writing committee of the trial where any uncertainties will be clarified and blinded interpretations of the primary endpoint results will be conducted prior to breaking the allocation[34].

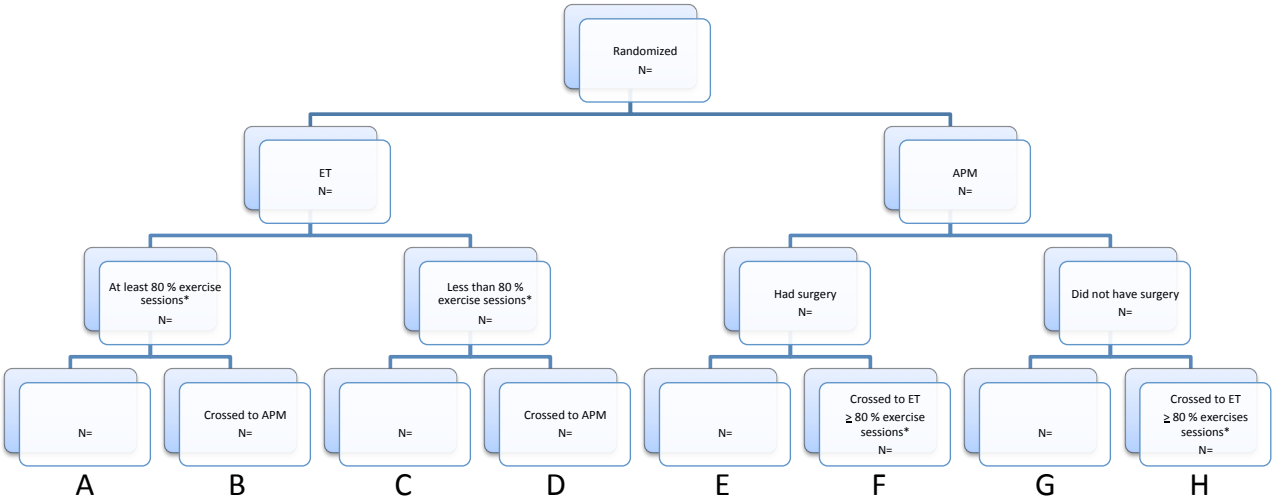
The writing committee consists of the OMEX steering group (including Nina Jullum Kise, MD, PhD-student, Ewa M Roos, PhD, PT, Professor (principal investigator), May Arna Risberg, PhD, PT, Professor (main supervisor for Nina Jullum Kise), Jonas Ranstam, Professor (Statistical advisor) and Silje Stensrud, PhD, PT.

8. Figures

8.1 Figure 1. Flow chart / Study design



8.2 Figure 2. Flow chart / Groups analyzed in ITT, PP and AT-analyses



ITT analysis:	ET-group = A+B+C+D	APM-group = E+F+G+H
PP analysis:	ET-group = A	APM-group = E
AT analysis:	ET-group = A+F+H	APM-group = E+B+D

*Supervised exercise therapy;
Less than vs. 80 % or more of the prescribed twice weekly sessions for 12 weeks (cut-off 19 sessions).

9. Tables

9.1. Table 1. Adverse events

Adverse events	Exercise therapy group	Arthroscopic surgery group	p-values
Seriously adverse events (SAE)			
Site other than the index knee, n (%)			
Cardiovascular			
Gastrointestinal			
Other			
Index knee, n (%)			
Pain			
Swelling			
Subjective instability			
Decreased range of motion			
Infection			
Deep vein thrombosis/pulmonary embolus			
Adverse events (AE)			
Site other than the index knee, n (%)			
Index knee, n (%)			

9.2. Table 2. Baseline characteristics

Baseline characteristics	Exercise therapy group	Arthroscopic surgery group
Men, n (%)		
Right knee, n (%)		
Age (years), mean (median)		
Weight, mean (median)		
Body mass index, mean (median)		
Smoking, n (%)		
Previous smoker n (%)		
Analgesics daily n (%)		
Education level, n (%)		
Primary school		
University level		
Activity level, mean (range)		
Level		
Sessions per week		
Hours per week		
Intensity		
Duration of symptoms, mean (median)		
2-12 months		
>12 months - 2 years		
> 2years		
Radiographic knee OA severity (Kellgren- Lawrence), n (%)		
Grade 0		
Grade 1		
Grade 2		
MRI; medial meniscus: degeneration grade, n (%)		
Grade 1		
Grade 2		
Grade 3		
Grade 4		
MRI; medial meniscus: localization of tear, n (%)		
Posterior horn		
Corpus		
Anterior horn		
MRI; medial meniscus: tear morphology, n (%)		
Non		
Horizontal		
Vertical		
Radial		
Longitudinal		
Root tear		
Complex		
Flap dislocated to medial recess		
Flap not dislocated		

MRI; medial meniscus: extrusion, n (%)		
None		
< 25 %		
25-49 %		
50-75 %		
> 75 %		
MRI; lateral meniscus: tear, n (%)		
KOOS scores, mean (median) (range)		
KOOS		
Pain		
Symptoms		
ADL		
Sport/Rec		
QOL		
Global knee rating, mean (median) (range)		
Much better		
Better		
Unchanged		
Worse		
Much worse		
SF-36, mean (median) (range)		
Physical component summary (PCS)		
Mental component summary (MCS)		
Muscle strength and performance, mean (median) (range)		
Hamstrings		
Quadriceps		
One leg hop test		
6 meter timed hop test		
Knee bending test		

9.3. Table 3. Treatment-related variables

Variable	Exercise therapy group	Arthroscopic surgery group	p-values
Exercise group; Compliance with exercise during the 12 weeks, n (%)			
Excellent; > 100			
Satisfactory; 100-80 %			
Poor; < 80 %			
Exercise group; Pain level during the 12 weeks exercise treatment, n (%)			
VAS 0-2			
VAS 3-7			
VAS 8-10			
Exercise group; Crossing over to arthroscopy, n (%)			
< 12 weeks			
3-12 months			
1-2 years			
Exercise group; Reoperations after crossing over, n (%)			
Meniscal reresection			
Other surgery / index knee			
Surgery other knee			
Arthroscopy group; Crossing over to exercise, n (%)			
< 12 weeks			
3-12 months			
1-2 years			
Arthroscopy group; Reoperations during 2 years, n (%)			
Meniscal reresection			
Other surgery / index knee			
Surgery other knee			
Patient satisfaction at follow-ups, n (%)			
Very dissatisfied			
Dissatisfied			
Neither satisfied nor dissatisfied			
Satisfied			
Very satisfied			

9.4. Tabell 4. Self-reported outcome at 1 and 2 years and muscle strength and performance at 1 year

Variables	Within-group difference		Between-group difference	p-values
	Exercise therapy group	Arthroscopic surgery group		
Primary endpoint: change in **$KOOS_4$** from baseline to 2 years (mean) (95% CI)				
Months follow-up (mean) (95% CI)				
Secondary endpoints:				
Mean change in KOOS subscale scores p-values (95% CI) (2 years)				
Pain				
Symptoms				
ADL				
Sport/Rec				
QOL				
Global knee rating, mean (median) (range) (1 year)				
Much better				
Better				
Unchanged				
Worse				
Much worse				
Mean change in other scores p-values (95% CI) (1 year)				
VAS				
BMI				
Self efficacy				
Mean change in activity level (95% CI) (2 years)				
Level				
Sessions per week				
Hours per week				
Intensity				
Mean change in muscle strength and performance (95% CI) (1 year)				
Hamstrings				
Quadriceps				
One-leg hop test				
6 meter timed hop test				
Knee bending test				
Mean change in SF-36 (95% CI) (2 years)				
Physical component summary (PCS)				
Mental component summary (MCS)				

10. References

1. Lyman S, Oh LS, Reinhardt KR, et al. Surgical decision making for arthroscopic partial meniscectomy in patients aged over 40 years. Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association 2012;28:492-501 e1.
2. Abrams GD, Frank RM, Gupta AK, Harris JD, McCormick FM, Cole BJ. Trends in meniscus repair and meniscectomy in the United States, 2005-2011. The American journal of sports medicine 2013;41:2333-9.
3. Thorlund JB, Hare KB, Lohmander LS. Large increase in arthroscopic meniscus surgery in the middle-aged and older population in Denmark from 2000 to 2011. Acta orthopaedica 2014;85:287-92.
4. Lohmander LS, Englund PM, Dahl LL, Roos EM. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. The American journal of sports medicine 2007;35:1756-69.
5. Reigstad O, Grimsgaard C. Complications in knee arthroscopy. Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA 2006;14:473-7.
6. Thorlund JB, Hare KB, Lohmander LS. Large increase in arthroscopic meniscus surgery in the middle-aged and older population in Denmark from 2000 to 2011. Acta orthopaedica 2014.
7. Englund M. The role of the meniscus in osteoarthritis genesis. The Medical clinics of North America 2009;93:37-43, x.
8. Englund M, Roos EM, Roos HP, Lohmander LS. Patient-relevant outcomes fourteen years after meniscectomy: influence of type of meniscal tear and size of resection. Rheumatology 2001;40:631-9.
9. Risberg MA. Degenerative meniscus tears should be looked upon as wrinkles with age--and should be treated accordingly. British journal of sports medicine 2014;48:741.
10. Noble J, Hamblen DL. The pathology of the degenerate meniscus lesion. The Journal of bone and joint surgery British volume 1975;57:180-6.
11. King D. The healing of semilunar cartilages. 1936. Clinical orthopaedics and related research 1990:4-7.
12. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Annals of the rheumatic diseases 1957;16:494-502.
13. Fairbank TJ. Knee joint changes after meniscectomy. The Journal of bone and joint surgery British volume 1948;30b:664-70.
14. Roos H, Lauren M, Adalberth T, Roos EM, Jonsson K, Lohmander LS. Knee osteoarthritis after meniscectomy: prevalence of radiographic changes after twenty-one years, compared with matched controls. Arthritis and rheumatism 1998;41:687-93.
15. Herrlin S, Hallander M, Wange P, Weidenhielm L, Werner S. Arthroscopic or conservative treatment of degenerative medial meniscal tears: a prospective randomised trial. Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA 2007;15:393-401.
16. Herrlin SV, Wange PO, Lapidus G, Hallander M, Werner S, Weidenhielm L. Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA 2013;21:358-64.
17. Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. The New England journal of medicine 2013;368:1675-84.

18. Yim JH, Seon JK, Song EK, et al. A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. The American journal of sports medicine 2013;41:1565-70.
19. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. The New England journal of medicine 2013;369:2515-24.
20. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. The New England journal of medicine 2002;347:81-8.
21. Kirkley A, Birmingham TB, Litchfield RB, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. The New England journal of medicine 2008;359:1097-107.
22. Khan M, Evaniew N, Bedi A, Ayeni OR, Bhandari M. Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 2014;186:1057-64.
23. Gauffin H, Tagesson S, Meunier A, Magnusson H, Kvist J. Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: a prospective, randomised, single-blinded study. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 2014.
24. Stensrud SR, M.A., Roos E. . A 12-Week Exercise Therapy Program in Middle Aged Patients With Degenerative Meniscus Tears: A Case Series With 1-Year-Follow-Up. JOSPT 2012.
25. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scandinavian journal of medicine & science in sports 1998;8:439-48.
26. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health and quality of life outcomes 2003;1:64.
27. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. The Journal of orthopaedic and sports physical therapy 1998;28:88-96.
28. Buchbinder R, Osborne RH, Ebeling PR, et al. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. The New England journal of medicine 2009;361:557-68.
29. Hurley MV, Walsh NE, Mitchell H, Nicholas J, Patel A. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. Arthritis care & research 2012;64:238-47.
30. Villadsen A, Overgaard S, Holsgaard-Larsen A, Christensen R, Roos EM. Immediate efficacy of neuromuscular exercise in patients with severe osteoarthritis of the hip or knee: a secondary analysis from a randomized controlled trial. The Journal of rheumatology 2014;41:1385-94.
31. Ranstam J, Turkiewicz A, Boonen S, Van Meirhaeghe J, Bastian L, Wardlaw D. Alternative analyses for handling incomplete follow-up in the intention-to-treat analysis: the randomized controlled trial of balloon kyphoplasty versus non-surgical care for vertebral compression fracture (FREE). BMC medical research methodology 2012;12:35.

32. Bubbar VK, Kreder HJ. The intention-to-treat principle: a primer for the orthopaedic surgeon. The Journal of bone and joint surgery American volume 2006;88:2097-9.
33. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta orthopaedica 2014;85:39-48.
34. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. Journal of clinical epidemiology 2014;67:769-72.
35. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. The New England journal of medicine 2008;358:794-810.
36. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS. A randomized trial of treatment for acute anterior cruciate ligament tears. The New England journal of medicine 2010;363:331-42.