TITLE PAGE

The six-meter timed hop test is a prognostic factor for outcomes in patients with meniscal tears treated with exercise therapy or arthroscopic partial meniscectomy – A secondary, exploratory analysis of the Odense-Oslo Meniscectomy versus Exercise (OMEX) trial

Nina Jullum Kise[1]
Ewa M. Roos[2]
Silje Stensrud[3]
Lars Engebretsen[4]
May Arna Risberg[5]

[1]Corresponding author. Orthopaedic Department, Martina Hansens Hospital, PO-box 823, N-1306 Sandvika, Norway. E-mail: nina.kise@mhh.no, ninakise@hotmail.com

[2]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. Email: eroos@health.sdu.dk

[3] OsloMet – Oslo Metropolitan University, PO-box 4 St.Olavs Plass, N-0130 Oslo, Norway. E-mail: silje.stensrud@hioa.no

[4]Oslo University Hospital and University of Oslo, Oslo, Norway. OSTRC, The Norwegian School of Sports Sciences, Oslo, Norway. E-mail: lars.engebretsen@medisin.uio.no

[5] Division of Orthopedic Surgery, Oslo University Hospital and Department of Sports Medicine, The Norwegian School of Sport Sciences, Oslo, Norway. E-mail: m.a.risberg@nih.no

ABSTRACT

Purpose: To identify prognostic factors for two-year patient-reported outcomes in middle-aged patients with degenerative meniscal tears treated with exercise therapy (ET) or arthroscopic partial meniscectomy (APM).

Methods: One hundred and seven patients, with mean age 49.6 (SD 6.2) years and BMI 25.7 (SD 3.7), were included in this analysis of data from the OMEX trial (www.clinicaltrials.gov NCT01002794). Linear and Poisson regression models were built to explore associations between potential prognostic factors (patient characteristics, knee function-related and disease-related factors) and two-year patient-reported outcomes: the Knee Injury and Osteoarthritis Outcome Score (KOOS) subscales Pain, Symptoms, ADL, Sport/Rec, QoL and 5-point Global Rating of Change scales for knee pain (GRC Pain) and function (GRC Function). Analyses were performed for the whole cohort and for the two treatment groups (n=55 and 52) with adjustments for age, sex, BMI and baseline KOOS.

Results: For the whole cohort, a one-second better baseline six-meter timed hop test result was associated with 3.0-7.1 points better two-year scores for all KOOS subscales (95% CIs 1.1-5.2 to 4.1-10.1 points). A 1.36-2.63s better test was associated with scores equivalent to previously calculated clinical relevant differences for each KOOS subscale. For the groups of patients treated with ET and APM respectively, 1.96-3.38s and 0.52-1.80s better tests were associated with clinical relevant differences.

For the whole cohort, a one-second better test was associated with 26% (95% CI 15-38%) and 22% (95% CI 11-34%) higher possibility for better or much better GRC Pain and Function scores. Patients treated with ET had 17% (95% CI 2-33%) increased possibility for better or much better GRC Pain score, and patients treated with APM had

66% (95% CI 32-109%) and 70% (95% CI 38-109%) increased possibility for better or much better GRC Pain and Function scores.

Conclusions: The six-meter timed hop test result was a significant prognostic factor for two-year patient-reported outcomes in middle-aged patients with degenerative meniscal tears, especially in those treated with APM.

KEY WORDS

Degenerative meniscal tears

Degenerative meniscal lesions

Middle-aged

Arthroscopic partial meniscectomy

Exercise therapy

Prognostic factors

Lower extremity performance

Six-meter timed hop test

LEVEL OF EVIDENCE

Level II

INTRODUCTION

During the last years, the incidence of arthroscopic treatment for degenerative knee diseases has declined in Sweden and Finland [24], and in Norway the number of meniscal resections decreased from 14.927 in 2013 to 7.979 in 2016 (data from the Norwegian Patient Registry). This trend might result from high quality randomized controlled trials (RCTs) [11, 12, 14-16, 20, 32, 33, 42] and systematic reviews [19, 37, 39] comparing patient-reported outcomes in middle-aged patients with degenerative meniscal tears, 1 to 5 years after treatment with arthroscopic partial meniscectomy (APM) or exercise therapy (ET) or sham surgery. Significant improvements in pain and function are seen following both APM and ET, but differences between treatment groups are minor [11, 12, 14-16, 20, 32, 33, 42].

Implementation of these research findings into clinical practice is challenging [23], and guidance for clinicians and patients for tailored interventions is limited. On an individual level, there will always be patients who get worse, those who do not change, and those who improve, regardless of type of treatment. Furthermore, prognostic factors for outcomes after APM and ET could differ. Clinicians would like to optimize clinical outcomes by identifying patients' characteristics, as well as clinical- and disease-related factors to guide treatment decisions. Stratified medicine and personalized medicine address these issues (https://www.mrc.ac.uk/research/initiatives/stratified-medicine/). Stratified medicine tailors care to subgroups of patients via key characteristics, disease, or biomarkers. A first step in applying these principles for patients with degenerative meniscal tears would be to identify baseline characteristics that affect the prognosis following treatment.

In the Odense-Oslo Meniscectomy versus Exercise (OMEX) trial (www.clinicaltrials.gov NCT01002794), no significant differences were found in Knee Injury and Osteoarthritis Outcome Score (KOOS) after 2 years in groups of patients treated with ET or APM [20]. In this secondary exploratory analysis from the same trial, the hypothesis was that baseline characteristics could predict patient reported outcomes and the main objective was to identify prognostic factors for two-year outcomes for the whole cohort as well as for those treated with ET and APM separately.

MATERIALS AND METHODS

This study was a hypothesis-generating, per protocol analysis of data from the OMEX trial, a prospective RCT of middle-aged patients with degenerative medial meniscal tears treated with either a 12-week ET program or APM without concomitant ET or physiotherapy [20]. The meniscal tears were defined as "degenerative" if symptoms arose during normal physical activities without significant trauma, e.g. during walking, running or squatting. The 35-60 year-old patients had MRI-verified unilateral medial meniscal tears, no ligament injuries or locked knees and radiographic osteoarthritis (OA) grade 0-2 according to Kellgren and Lawrence (KL) [17]. KL grade 2 was defined as a definite presence of an osteophyte and possible joint space narrowing [30].

The RCT included 140 patients and 126 (90%) completed the two-year follow up [20]. Excluded from this secondary analysis were 16 patients who had not completed their allocated interventions (10 did not carry out the ET program according to a priori defined criteria [20] and six did not go through APM because they refused or their indications changed), five patients with OA severity of KL grade 2 or 3, four patients

with incomplete baseline data, and eight patients with incomplete two-year data. Hence, this secondary analysis is based on the 107 patients with full datasets (referred to as the "whole cohort") and on the groups treated with ET (n=55) and APM (n=52), respectively (Figure 1). Post hoc analyses of baseline and outcome variables showed similar results in the groups of excluded versus included patients.

Treatment strategies

The ET program consisted of progressive neuromuscular and strength exercises over 12 weeks (2-3 sessions a week) and has previously been described in detail [35]. The APMs were performed as standard arthroscopies, with 30 degrees optics, standard portals, and lavage with Ringer acetate. Examination of joint cartilage, ligaments and menisci were followed by resection of unstable meniscal tissue. Postoperatively, the patients were mobilized with crutches for 3-4 days and given oral and written instructions for home exercises aimed at reducing symptoms and regaining normal function [20].

Two-year outcomes

Pain and knee function at the two-year follow up were measured with the five subscales of KOOS (Pain, other Symptoms, Activities of Daily Living (ADL), Sport and Recreation (Sport/Rec) and Quality of Life (QoL) and 5-point Global Rating of Change (GRC) scales for Pain and Function. The KOOS is reliable and valid [27, 28] and study- and subscale-specific clinical relevant differences for patients included in the OMEX RCT have been reported to be 7.4, 8.4, 4.1, 10.9 and 13.6 points for Pain, Symptoms, ADL, Sport/Rec and QoL, respectively [20]. The GRC questions were: 1) "With respect to your knee disorder, how would you describe your pain now compared to when you were included in the

study and started treatment?" (GRC Pain), and 2) "With respect to your knee disorder, how would you describe your knee function now compared to when you were included in the study and started treatment?" (GRC Function). The patients defined their pain and knee function from baseline to follow up as much worse, worse, unchanged, better, or much better. To identify prognostic factors for scoring pain and function at 2 years as at least better than baseline, the GRC scales were dichotomized with the cut-off between "unchanged" and "better".

Prognostic factors

Since a tear in a degenerative meniscus is regarded as the first sign of OA [10], it is reasonable to assume that patients with higher total load of known risk factors for OA may be at higher risk of deterioration of knee pain and function compared to patients with lower total load of risk factors. Known risk factors for OA are higher age [18], female sex [18], higher BMI [18], cigarette smoking [1], impaired lower extremity performance [40], lower knee extension strength [25], lower physical activity level [22], higher grade of meniscal degeneration [6, 10], more meniscal extrusion [3] and receiving APM [10].

Knee function-related factors

Baseline scores of the five KOOS subscales, lower extremity performance tests, quadriceps muscle strength and physical activity level were included as knee function-related prognostic factors.

Lower extremity performance was measured by reliable and valid single-leg tests: The one-leg hop test for distance, the six-meter timed hop test and the maximum number of

knee bends in 30 seconds test [7, 8, 29]. The test procedures have been described previously [34] (and the six-meter timed hop test is illustrated in Figure S1, Supplementary appendix). Quadriceps muscle strength was measured using an isokinetic dynamometer, testing at 60^0/second (Biodex 6000 System; Biodex Medical Systems Inc, Shirley, NY, US) [9].

Physical activity was measured at baseline and at 2 years using two different questionnaires. At baseline the patients were asked "How many times a week (mean) do you exercise?" (frequency) and "How many hours a week (mean) do you exercise?" (duration), both questions related to the last six months before inclusion in the trial. At 2 years, additional information for physical activity was included; the activity questionnaire from a large health survey, the Nord-Trøndelag Health Study (HUNT 1) [21]. The patients were asked to report frequency, duration and intensity of physical activity on a Likert scale (Table S1, Supplementary appendix). Each score was weighted by intensity level, and the product of the three scores gave the HUNT 1 activity index, ranging from 0 (lowest) to 15 (highest) (Table S1, Supplementary appendix) [21]. Additionally, the patients were asked to estimate their physical activity level retrospectively over the previous 6 months prior to inclusion using the HUNT 1 activity index. Correlation analyses of baseline and two-year data on physical activity were performed to test the validity of including the retrospectively collected HUNT 1 activity index as a baseline variable. With a cut-off for acceptable correlation set to coefficients larger than 0.40, we found acceptable correlation for frequency and duration (correlation coefficients were 0.43 and 0.42, respectively). Therefore, the retrospective data (HUNT 1 activity index) collected at 2 years were accepted as baseline activity level measures.

198

199 *Disease-related factors*

200 Meniscal pathology at baseline was assessed with MRI, using grade of degeneration and
201 amount of extrusion. Meniscal degeneration was graded (0-3b, higher is worse) [2].
202 Grade 0 represents healthy meniscal tissue, grade 1 tissue degeneration inside the
203 meniscus, grade 2 a tear not reaching the surface of the meniscus, grade 3 a tear
204 penetrating one (3a) or both (3b) surfaces of the meniscus. Meniscal extrusion was
205 evaluated on the coronal sequence image with the largest tibial spine volume and
206 defined as meniscal subluxation crossing a vertical line on the medial margin of the tibia
207 without osteophytes. Extrusion was given in per cent (width of extruded meniscal tissue
208 relative to the total width of the meniscus in the same image, higher is worse) [13]
209 (Figure S2, Supplementary appendix).

210

211 *Statistical analyses*

212 The statistical computation was performed using IBM SPSS Statistics version 25 (IBM
213 Corp. 2017, Armonk, NY, US) (descriptive statistics and multiple linear regression
214 analyses) and Stata v15 (Stata 2017, College Station, TX, US) (Poisson regression
215 analyses). Summary statistics were, if nothing else is stated, for continuous
216 variables presented in terms of mean, standard deviation, and the number of
217 observations (N), and for categorical variables in terms of frequencies and per cent.
218 Comparison between groups was not part of this study; hence, no statistical analysis was
219 performed to evaluate group differences (Table 1). General assessments of fulfilment of
220 the statistical methods' underlying assumptions were made. The estimation uncertainty
221 of regression parameter estimates is presented in terms of 95% confidence intervals
222 (95% CIs) and p-values <0.050 were considered statistically significant.

Multiple linear regression models were built for the five continuous KOOS subscale outcomes. For interpretation of the results of the linear regression analyses, the previously calculated clinical relevant differences of each KOOS subscale [20] were used to calculate the exact level of independent variables needed to achieve the actual KOOS values. Poisson regression models, including the Hubert-White estimator, were built for the two dichotomized GRC outcomes (Pain and Function). This estimator provides asymptotically consistent estimates of the covariance matrix for parameter estimates without any distributional assumptions and even when the assumed model underlying the parameter is incorrect [41]. The analyses were performed for the whole cohort and repeated for the respective treatment groups. Hence, 21 regression models were analysed in total. Due to the exploratory nature of this secondary study adjustments for multiplicity was not included because they are usually not considered meaningful [4].

Statistical models for confounding adjustment were based upon clinical experience and literature studies [1, 3, 5, 10, 18, 22, 25, 40]. Additionally, Diagnostic Acyclic Graphs (DAGs) were generated with DAGitty (http://www.dagitty.net) [38] (Figure S3, Supplementary appendix) in order to define the statistical models that best reduced confounding while avoiding both adjustment bias and collider stratification bias [31]. Hence, age, sex, BMI and baseline KOOS for each subscale were identified as potential confounders.

RESULTS

Descriptive data including demographics, prognostic factors and two-year outcomes for the whole cohort and for the two treatment groups separately are presented in Table 1.

Results of the regression analyses are presented in Table 2. The six-meter timed hop test at baseline associated with clinical relevant outcomes after 2 years are presented in Table 3.

For the whole cohort (n=107), a one-second better hop test result was associated with 3.0 to 7.1 points better two-year scores for all five KOOS subscales (95% CI ranging from 1.1-5.2 to 4.1-10.1 points, and 26% (95% CI 15-38%) and 22% (95% CI 11-34%) increased possibility for scoring better or much better for GRC Pain and Function, respectively (Table 2). A 1.36-2.63 s better test was associated with two-year scores equivalent to the previously calculated clinical relevant differences for each KOOS subscale (Table 3).

For patients treated with ET (n=55), a one-second better hop test result was associated with better KOOS Symptoms, Sport/Rec and QoL (2.5 to 5.6 points, 95% CI from 0.2-4.8 to 2.1-9.1 points) and 17% (95% CI 2-33%) higher risk for better or much better GRC Pain score (Table 2). A 1.96-3.38 s better hop test was associated with clinical relevant differences (Table 3).

For patients treated with APM (n=52), a one-second better hop test result was associated with better KOOS Pain, ADL, Sport/Rec and QoL (7.6 to 10.9 points, 95% CI from 2.9-12.2 to 4.9 -17.0) and 66% (95% CI 32-109%) and 70% (95% CI 38-109%) higher risk for better or much better GRC Pain and Function scores, respectively (Table 2). A 0.52-1.80 s better test was associated with clinical relevant differences (Table 3). Higher activity level measured with the HUNT 1 activity index was associated with 65%

(95% CI 8-153%) and 83% (95% CI 5-221%) increased possibility for better or much better GRC Pain and Function scores, respectively (Table 2).

DISCUSSION

The principal finding of this study was that better knee performance at baseline measured with the six-meter timed hop test was a significant prognostic factor for less knee pain and better knee function after 2 years. Patients treated with APM had almost four times higher possibility for scoring better or much better in GRC Pain than patients treated with ET (66% versus 17%), and smaller differences in hop test results at baseline were associated with clinical relevant KOOS results [20] at 2 years (0.52-1.80 s versus 1.96-3.38 s).

To our knowledge, the six-meter timed hop test has not previously been used for middle-aged patients with degenerative meniscal tears, and psychometric properties of the test for this patient group is unknown. Mean hop time has been reported to be 1.82-1.86 s (SD 0.17-0.22 s) for young, healthy males [29] and 2.3 s (SD 0.2 s) for anterior cruciate ligament (ACL)-reconstructed patients after rehabilitation [36]. Our cohort used more time (2.84 s) and had a larger coefficient of variation compared to these populations (54% vs. 9-12%) [29, 36]. In our study, those who had the highest performance (the best quartile) had a mean hop time of 1.77 s (SD 0.17 s), and the second, third and forth quartile had a mean hop time of 2.21 s (SD 0.14 s), 2.83 s (SD 0.23 s) and 4.62 s (SD 2.23 s), respectively (Table S2, Supplementary appendix). Thus, one out of four middle-aged patients with degenerative meniscal tears hopped better or as well as young healthy males, and about half of the patients hopped better than ACL-reconstructed patients. However, this study show that those with a hop time of the

timeframe of 2.39 to 6.85 s (the lower quartile) should be informed that they would have better prognoses for outcomes if they improve their knee performance. Another finding was that higher activity level at baseline was prognostic for outcomes in patients treated with APM, but not ET. Consequently, patients should be encouraged to increased physical activity level or either prior to or following surgery.

Worse meniscal degeneration (grade 3b) was shown to be a statistically significant prognostic factor for worse two-year KOOS QoL in patients treated with ET, but not in patients treated with APM. This must be interpreted carefully and the clinical importance of this finding is uncertain. First, only 4 patients (7.3%) in the ET group had meniscal degeneration grade 3b. Second, only one radiologist has performed the MRI scorings. Third, no between-group comparisons have been done, hence, the association between worse meniscal degeneration and worse QoL in patients treated with ET does not imply better QoL in patients treated with APM.

This study has some limitations. First, this study does not include radiographs appropriate for evaluation of varus-valgus alignment. Second, MRI-evaluation included degeneration grade (0-3b, lower is better)[2] and measurement of meniscal extrusion [13]. More extensive classification systems as e.g. WORMS [26] might have strengthened our study. Third, in the subgroup analyses of GRC Pain and Function, reduced samples, especially for the APM-group, might have led to spurious results reflected in the wide 95% CIs. Forth, the external validity of this study might be weakened by the fact that these participants were highly educated, had higher activity levels than a younger (20-39 years old) Norwegian county population [21], were also only slightly over weighted (BMI 25.7) and only 7.5% reported daily smoking. Assumingly, these patients might

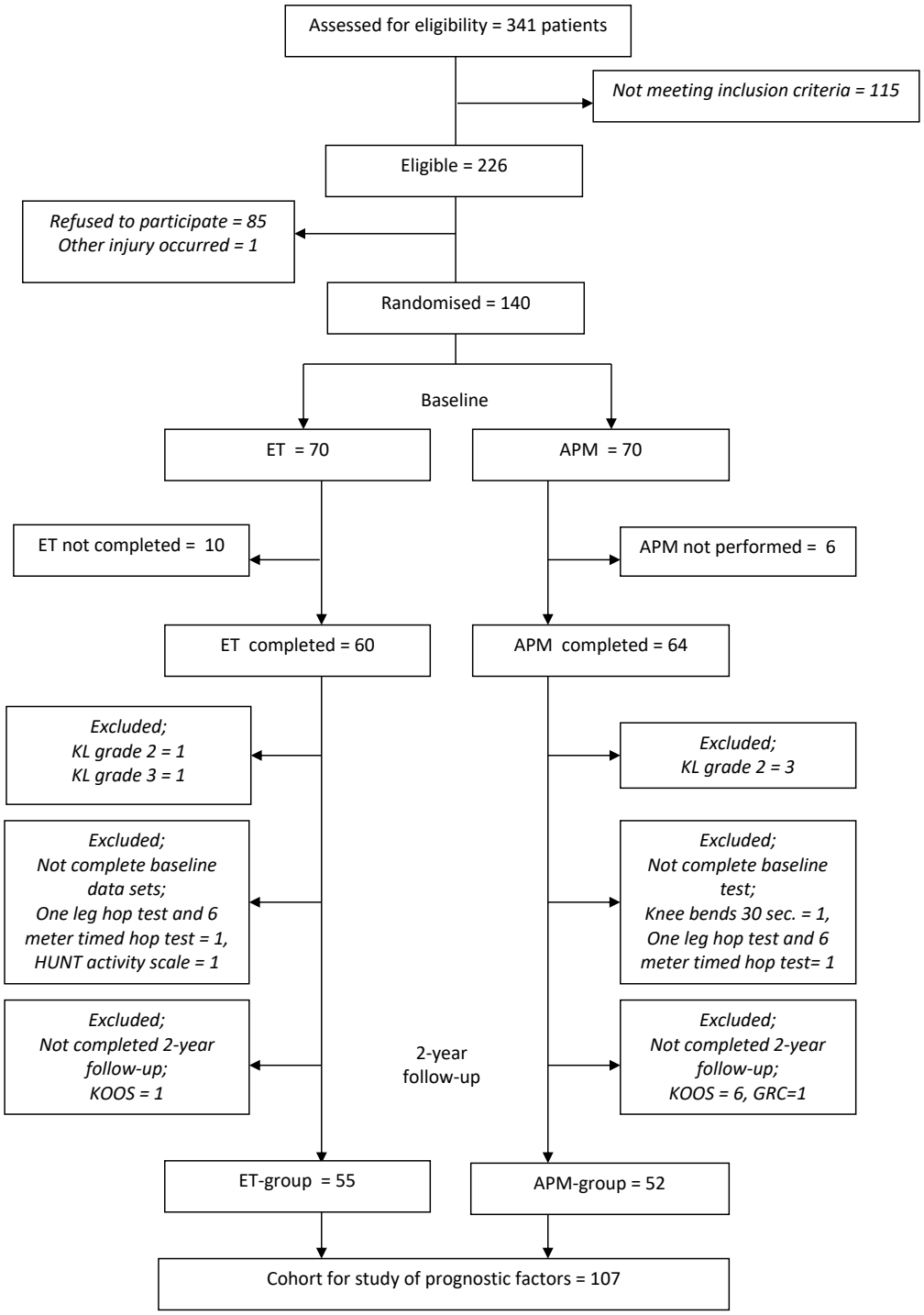
have been more prone to accept being included in a scientific trial including ET than less educated and less fit individuals.

To our knowledge, this is the first study to examine prognostic factors for patient-reported outcomes of pain and function in middle-aged individuals with degenerative meniscal tears without radiological knee OA. A degenerative meniscal tear is an early sign of knee OA [10] and identification of prognostic factors for outcomes available before OA is established might be a supplement to other risk-related factors such as high BMI [18], weak quadriceps muscle strength [25] and poor knee function [40]. This low-cost and quickly performed test is easily implemented in a clinical setting, and may give valuable information on future risk of worse knee pain and impaired knee function.

Conclusion

In patients with degenerative meniscal tears, a better six-meter timed hop test result at baseline was a significant prognostic factor for better patient-reported knee function after 2 years, especially in those treated with APM.

340 **Figure 1. Flow chart**



341

342 **Table 1. Participant characteristics**
343

		The whole cohort (n=107)	Patients treated with ET[1] (n=55)	Patients treated with APM[2] (n=52)
Demographics				
Gender, men (n(%))		65 (60.7)	34 (61.8)	31 (59.6)
Age, years (mean (SD))		49.6 (6.2)	50.1 (6.1)	49.2 (6.3)
BMI[3], kg/m^2 (mean (SD))		25.7 (3.7)	25.7 (4.0)	25.7 (3.5)
Smokers, (n(%))		8 (7.5)	1 (1.8)	7 (13.5)
Use analgetics daily, (n(%))		2 (1.9)	2 (3.6)	0 (0.0)
HUNT 1 activity index[4] (mean (SD))		3.4 (2.7)	2.6 (2.3)	4.2 (2.8)
Primary school only (n(%))		5 (4.7)	3 (5.5)	2 (3.8)
≥3-year higher education (n(%))		87 (81.3)	46 (83.6)	41 (78.8)
KL grade[5] (n (%))	KL = 0	83 (77.6)	42 (76.4)	41 (78.8)
	KL = 1	24 (22.4)	13 (23.6)	11 (21.2)
Prognostic factors				
Disease-related				
Meniscal degeneration grade[6] (n (%))	Grade 0	0 (0.0)	0 (0.0)	0 (0.0)
	Grade 1	1 (0.9)	0(0.0)	1 (1.9)
	Grade 2	9 (8.4)	5 (9.1)	4 (7.7)
	Grade 3a	86 (80.4)	46 (83.6)	40 (76.9)
	Grade 3b	11 (10.3)	4 (7.3)	7 (13.5)

Extrusion degree[7], % (mean (SD))		17.9 (18.9)	18.3 (17.7)	17.5 (20.3)
Treatment-related				
Intervention (n (%))	ET APM	55 (51.4) 52 (48.6)		
Patient-reported and patient-related				
KOOS[8] subscales				
Pain[9], points (mean (SD))		66.2 (17.1)	65.1 (19.2)	67.4 (14.8)
Symptoms[9], points (mean (SD))		74.9 (15.9)	71.1 (16.5)	79.0 (14.3)
ADL[9], points (mean (SD))		78.6 (18.0)	76.6 (19.6)	80.7 (16.1)
Sport/Rec[9], (mean (SD))		47.4 (24.0)	46.6 (24.4)	48.2 (23.8)
QoL[9], points (mean (SD))		44.0 (16.1)	41.9 (16.3)	46.3 (15.7)
Knee performance and thigh strength				
One-leg hop test for distance[9], cm (mean (SD))		79.9 (33.5)	78.2 (33.1)	81.7 (34.3)
Six-meter timed hop test[10], s (mean (SD))		2.84 (1.54)	3.00 (1.75)	2.70 (1.28)
Maximum knee bends in 30 s test[9] (mean (SD))		29.0 (10.5)	28.8 (10.5)	29.1 (10.7)

Peak torque knee extension Nm[9], (mean (SD))		159.3 (48.5)	157.2 (45.4)	161.5 (52.0)
Two-year outcomes				
KOOS subscales				
Pain[9], points (mean (SD))		88.2 (16.7)	86.9 (17.1)	89.5 (16.4)
Symptoms[9], points (mean (SD))		88.9 (13.9)	87.6 (13.1)	90.3 (14.6)
ADL[9], points (mean (SD))		92.6 (15.4)	91.5 (14.2)	93.9 (16.5)
Sport/Rec[9], points (mean (SD))		77.9 (23.8)	75.4 (23.9)	80.7 (23.7)
QoL[9], points (mean (SD))		76.4 (20.0)	72.8 (21.1)	80.2 (18.1)
5-point GRC[11] scales for knee function and pain at 2 years compared to baseline				
Knee Pain, n (%)	Much better	58 (54.2)	23 (41.8)	35 (67.3)
	Better	34 (31.8)	22 (40.0)	12 (23.1)
	Unchanged	8 (7.5)	7 (12.7)	1 (1.9)
	Worse	4 (3.7)	3 (5.5)	1 (1.9)
	Much worse	3 (2.8)	0 (0.0)	3 (5.8)
Knee Function, n (%)	Much better	56 (52.3)	24 (43.6)	32 (61.5)
	Better	34 (31.8)	17 (30.9)	17 (32.7)
	Unchanged	12 (11.2)	11 (20.0)	1 (1.9)
	Worse	4 (3.7)	3 (5.5)	1 (1.9)
	Much worse	1 (0.9)	0 (0)	1 (1.9)

[1]ET; Exercise therapy

[2]APM; Arthroscopic partial meniscectomy

[3]BMI; Body Mass Index, kg/m^2

[4]HUNT 1 activity index, range 0-15, higher is better. Questions were asked retrospectively at the two-year follow-up.

[5]According to Kellgren and Lawrence, grade 0-4, lower is better

[6]Grade 0-3b, lower is better

[7]Meniscal extrusion given in per cent evaluated on the coronal sequence image with the largest tibial spine volume, defined as meniscal subluxation crossing a vertical line on the medial margin of tibia without osteophytes, lower is better (Figure S2, Supplementary appendix)

[8]KOOS; the Knee Injury and Osteoarthritis Outcome Score

[9]Higher is better

[10]Lower is better

[11]GRC; the Global Rating of Change scale

Table 2. Statistically significant prognostic factors at baseline for outcomes after 2 years.

Two-year outcomes	Baseline prognostic factors	The whole cohort (n=107) Estimate	95% CI	p-value	Patients treated with ET[1] (n=55) Estimate	95% CI	p-value	Patients treated with APM[2] (n=52) Estimate	95% CI	p-value
KOOS[3]		Points[4]			Points[4]			Points[4]		
Pain	Six-meter timed hop test[5]	3.9	1.2 to 6.2	0.001				8.0	4.0 to 11.9	<0.001
Symptoms	Six-meter timed hop test[5]	3.4	1.4 to 5.3	0.001	2.5	0.2 to 4.8	0.032			
	Maximum knee bends in 30 s test[6]							0.6	0.2 to 1.1	0.003
ADL	Six-meter timed hop test[5]	3.0	1.1 to 5.2	0.003				7.8	4.0 to 11.7	<0.001
Sport/Rec	Six-meter timed hop test[5]	7.1	4.1 to 10.1	<0.001	5.6	2.1 to 9.1	0.002	10.9	4.9 to 17.0	0.001
QoL	Six-meter timed hop test[5]	5.2	2.5 to 7.9	<0.001	4.0	0.6 to 7.4	0.021	7.6	2.9 to 12.2	0.002
	Meniscal deg. grade 3b[7]				-22.7	-42.7 to -2.6	0.028			
GRC[8]		IRR[9]			IRR[9]			IRR[9]		
Pain	Six-meter timed hop test[5]	1.26	1.15 to 1.38	<0.001	1.17	1.02 to 1.33	0.021	1.66	1.32 to 2.09	<0.001
	HUNT[10]							1.65	1.08 to 2.53	0.021

Function	Six-meter timed hop test[5]	1.22	1.11 to 1.34	<0.001				1.70	1.38 to 2.09	<0.001
	HUNT[10]							1.83	1.05 to 3.21	0.033

[1]ET; Exercise therapy

[2]APM; Arthroscopic partial meniscectomy

[3]KOOS; the Knee Injury and Osteoarthritis Outcome Score

[4]KOOS points. One less second in six-meter timed hop test or higher number of knee bends gives the estimated number of KOOS points

[5]Seconds, lower is better

[6]Number, higher is better

[7]Meniscal degeneration, grade 1-3b, higher is worse

[8]GRC; the Global Rating of Change scale

[9]IRR = incidence rate ratio. One less second in six-meter timed hop test or one better point on HUNT 1 activity index gives the estimated incidence risk ratio to score better or much better

[10]HUNT 1 activity index, range 0-15, higher is better. Questions were asked retrospectively at the two-year follow-up

Table 3. The six-meter timed hop test results at baseline and associated clinical relevant better outcomes after 2 years

Two-year outcomes	The whole cohort (n=107)	Patients treated with ET[1] (n=55)	Patients treated with APM[2] (n=52)
KOOS[3]	Clinical relevant differences	Number of seconds better six-meter timed hop test associated with KOOS scores equivalent to the clinical relevant KOOS scores	
Pain	7.4 points	1.88 s	0.93 s
Symptoms	8.4 points	2.50 s	3.32 s
ADL	4.1 points	1.36 s	0.52 s
Sport/Rec	10.9 points	1.53 s	1.96 s	1.00 s
QoL, points	13.6 points	2.63 s	3.38 s	1.80 s
GRC[4]		Per cent higher possibility for scoring better or much better associated with one-second better six-meter timed hop test	
Pain		26%	17%	66%
Function		22%		70%

[1]ET; Exercise therapy

[2]APM; Arthroscopic partial meniscectomy

[3]KOOS; the Knee Injury and Osteoarthritis Outcome Score

[4]GRC; the Global Rating of Change scale

COMPETING INTERESTS

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that they have no support from any company for the submitted work; no relationships with any company that might have an interest in the submitted work in the previous 3 years; their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and they have no non-financial interests that may be relevant to the submitted work.

AUTHORS´ CONTRIBUTIONS

Ewa M. Roos, May Arna Risberg (MAR) and Silje Stensrud (SS) developed the concept and design for the RCT this study cohort is extracted from. SS, Lars Engebretsen and Nina Jullum Kise (NJK) collected the data. NJK and MAR developed the idea for this secondary study of data from the RCT. NJK wrote the manuscript. NJK and independent statistician Jonas Ranstam did the statistical analyses. All authors had full access to all of the data including statistical reports and tables in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

ACKNOWLEDGEMENTS

The authors acknowledge the patients in this study for their participation. Further, they acknowledge research coordinators Kristin Bølstad and Emilie Jul-Larsen for the organisation of the participants, physiotherapists Marte Lund, Karin Rydevik and Christian Vilming for assistance with data collection and radiologist Rana Tariq for reading the MRI scans. Nanna Kurtze is acknowledged for giving permission to copy her previously published table illustrating the calculation of the HUNT 1 activity index. The

authors also acknowledge the Norwegian Sports Medicine Clinic (NIMI), Oslo, Norway, for supporting the Norwegian Research Center for Active Rehabilitation (NAR) with rehabilitation facilities and research staff. NAR is a collaboration between the Norwegian School of Sports Sciences, Department of Orthopaedic Surgery, Oslo University Hospital, and NIMI. Finally, the authors acknowledge the Department of Orthopaedic Surgery, Oslo University Hospital, and the Department of Orthopaedic Surgery, Martina Hansens Hospital, Bærum, for accessibility to the outpatient and surgical clinics.

The OMEX RCT was funded by Sophies Minde Ortopedi AS, The Swedish Rheumatism Association, the Swedish Scientific Council; the Region of Southern Denmark, the Danish Rheumatism Association and the Health Region of South-East Norway. The researchers were independent of the funder.

REFERENCES

1. Amin S, Niu J, Guermazi A, Grigoryan M, Hunter DJ, Clancy M, et al. (2007) Cigarette smoking and the risk for cartilage loss and knee pain in men with knee osteoarthritis. Ann Rheum Dis 66:18-22
2. Anderson AF, Irrgang JJ, Dunn W, Beaufils P, Cohen M, Cole BJ, et al. (2011) Interobserver reliability of the International Society of Arthroscopy, Knee Surgery and Orthopaedic Sports Medicine (ISAKOS) classification of meniscal tears. Am J Sports Med 39:926-932
3. Badlani JT, Borrero C, Golla S, Harner CD, Irrgang JJ (2013) The effects of meniscus injury on the development of knee osteoarthritis: data from the osteoarthritis initiative. Am J Sports Med 41:1238-1244
4. Bender R, Lange S (2001) Adjusting for multiple testing--when and how? J Clin Epidemiol 54:343-349
5. Bloecker K, Guermazi A, Wirth W, Benichou O, Kwoh CK, Hunter DJ, et al. (2013) Tibial coverage, meniscus position, size and damage in knees discordant for joint space narrowing - data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 21:419-427
6. Bloecker K, Guermazi A, Wirth W, Benichou O, Kwoh CK, Hunter DJ, et al. (2013) Tibial coverage, meniscus position, size and damage in knees discordant for joint space narrowing - data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 21:419-427
7. Bolgla LA, Keskula DR (1997) Reliability of lower extremity functional performance tests. J Orthop Sports Phys Ther 26:138-142
8. Bremander AB, Dahl LL, Roos EM (2007) Validity and reliability of functional performance tests in meniscectomized patients with or without knee osteoarthritis. Scand J Med Sci Sports 17:120-127
9. Drouin JM, Valovich-mcLeod TC, Shultz SJ, Gansneder BM, Perrin DH (2004) Reliability and validity of the Biodex system 3 pro isokinetic dynamometer velocity, torque and position measurements. Eur J Appl Physiol 91:22-29
10. Englund M, Roos EM, Lohmander LS (2003) Impact of type of meniscal tear on radiographic and symptomatic knee osteoarthritis: a sixteen-year followup of meniscectomy with matched controls. Arthritis Rheum 48:2178-2187
11. Gauffin H, Sonesson S, Meunier A, Magnusson H, Kvist J (2017) Knee Arthroscopic Surgery in Middle-Aged Patients With Meniscal Symptoms: A 3-Year Follow-up of a Prospective, Randomized Study. Am J Sports Med 45:2077-2084
12. Gauffin H, Tagesson S, Meunier A, Magnusson H, Kvist J (2014) Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: a prospective, randomised, single-blinded study. Osteoarthritis Cartilage 22:1808-1816
13. Ha JK, Jang HW, Jung JE, Cho SI, Kim JG (2014) Clinical and radiologic outcomes after meniscus allograft transplantation at 1-year and 4-year follow-up. Arthroscopy 30:1424-1429
14. Herrlin S, Hallander M, Wange P, Weidenhielm L, Werner S (2007) Arthroscopic or conservative treatment of degenerative medial meniscal tears: a prospective randomised trial. Knee Surg Sports Traumatol Arthrosc 15:393-401
15. Herrlin SV, Wange PO, Lapidus G, Hallander M, Werner S, Weidenhielm L (2013) Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. Knee Surg Sports Traumatol Arthrosc 21:358-364
16. Katz JN, Brophy RH, Chaisson CE, de Chaves L, Cole BJ, Dahm DL, et al. (2013) Surgery versus physical therapy for a meniscal tear and osteoarthritis. N Engl J Med 368:1675-1684
17. Kellgren JH, Lawrence JS (1957) Radiological assessment of osteo-arthrosis. Ann Rheum Dis 16:494-502
18. Kerkhof HJ, Bierma-Zeinstra SM, Arden NK, Metrustry S, Castano-Betancourt M, Hart DJ, et al. (2014) Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. Ann Rheum Dis 73:2116-2121
19. Khan M, Evaniew N, Bedi A, Ayeni OR, Bhandari M (2014) Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. CMJ 186:1057-1064
20. Kise NJ, Risberg MA, Stensrud S, Ranstam J, Engebretsen L, Roos EM (2016) Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. BMJ 354:i3740
21. Kurtze N, Rangul V, Hustvedt BE, Flanders WD (2008) Reliability and validity of self-reported physical activity in the Nord-Trondelag Health Study: HUNT 1. Scand J Public Health 36:52-61
22. Lin W, Alizai H, Joseph GB, Srikhum W, Nevitt MC, Lynch JA, et al. (2013) Physical activity in relation to knee cartilage T2 progression measured with 3 T MRI over a period of 4 years: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 21:1558-1566
23. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. (2014) Biomedical research: increasing value, reducing waste. Lancet 383:101-104
24. Mattila VM, Sihvonen R, Paloneva J, Fellander-Tsai L (2016) Changes in rates of arthroscopy due to degenerative knee disease and traumatic meniscal tears in Finland and Sweden. Acta Orthop 87:5-11

25. Oiestad BE, Juhl CB, Eitzen I, Thorlund JB (2015) Knee extensor muscle weakness is a risk factor for development of knee osteoarthritis. A systematic review and meta-analysis. Osteoarthritis Cartilage 23:171-177
26. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, et al. (2004) Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage 12:177-190
27. Roos EM, Lohmander LS (2003) The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 1:64
28. Roos EM, Roos HP, Ekdahl C, Lohmander LS (1998) Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports 8:439-448
29. Ross MD, Langford B, Whelan PJ (2002) Test-retest reliability of 4 single-leg horizontal hop tests. J Strength Cond Res 16:617-622
30. Schiphof D, Boers M, Bierma-Zeinstra SM (2008) Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. Ann Rheum Dis 67:1034-1036
31. Shrier I, Platt RW (2008) Reducing bias through directed acyclic graphs. BMC Med Res Methodol 8:70
32. Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H, et al. (2018) Arthroscopic partial meniscectomy versus placebo surgery for a degenerative meniscus tear: a 2-year follow-up of the randomised controlled trial. Ann Rheum Dis 77:188-195
33. Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H, et al. (2013) Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. N Engl J Med 369:2515-2524
34. Stensrud S, Risberg MA, Roos EM (2014) Knee function and knee muscle strength in middle-aged patients with degenerative meniscal tears eligible for arthroscopic partial meniscectomy. Br J Sports Med 48:784-788
35. Stensrud S, Roos EM, Risberg MA (2012) A 12-week exercise therapy program in middle-aged patients with degenerative meniscus tears: a case series with 1-year follow-up. J Orthop Sports Phys Ther 42:919-931
36. Sueyoshi T, Nakahata A, Emoto G, Yuasa T (2017) Single-Leg Hop Test Performance and Isokinetic Knee Strength After Anterior Cruciate Ligament Reconstruction in Athletes. Orthop J Sports Med 5:2325967117739811
37. Swart NM, van Oudenaarde K, Reijnierse M, Nelissen RG, Verhaar JA, Bierma-Zeinstra SM, et al. (2016) Effectiveness of exercise therapy for meniscal lesions in adults: A systematic review and meta-analysis. J Sci Med Sport 19:990-998
38. Textor J, Hardt J, Knuppel S (2011) DAGitty: a graphical tool for analyzing causal diagrams. Epidemiology 22:745
39. Thorlund JB, Juhl CB, Roos EM, Lohmander LS (2015) Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms. Br J Sports Med 49:1229-1235
40. Thorstensson CA (2004) Reduced functional performance in the lower extremity predicted radiographic knee osteoarthritis five years later. Annals of the Rheumatic Diseases 63:402-407
41. White H (1982) Maximum likelihood estimation of misspecified models. Econometrica 50:1-25
42. Yim JH, Seon JK, Song EK, Choi JI, Kim MC, Lee KB, et al. (2013) A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. Am J Sports Med 41:1565-1570

SUPPLEMENTARY APPENDIX

Table S1. The HUNT 1 activity index. (Questions about exercise in HUNT 1 [1]).

EXERCISE*	
By exercise we mean, for example, walking, skiing, swimming or training/sport	
How frequently do you exercise?	
(Give an average)	
☐ Never	(0)
☐ Less than once a week	(.5)
☐ Once a week	(1)
☐ 2-3 times a week	(2.5)
☐ Almost every day	(5)
If you do such exercise as frequently as once or more times a week: How hard do you push yourself?	
(Give an average)	
☐ I take it easy without breaking into a sweat or losing my breath	(1)
☐ I push myself so hard that I lose my breath and break into a sweat	(2)
☐ I push myself to near-exhaustion	(3)
How long does each session last?	
(Give an average)	
☐ Less than 15 minutes	(.10)
☐ 16-30 minutes	(.38)
☐ 30 minutes to 1 hour	(.75)
☐ More than 1 hour	(1.0)

Table legend:

*Numbers in parentheses indicate score used for each response when calculating the summary index.

The HUNT 1 activity index is a product of the frequency, intensity and duration scales (range 0-15, higher is better).

The table is copied with permission from the first author [1].

546 **Table S2. Baseline six-meter timed hop test in quartiles**

	The whole cohort (n=107)	Patients treated with ET[1] (n=55)	Patients treated with APM[2] (n=52)
Quartile	Time, sec[3] mean (SD)	Time, sec mean (SD)	Time, sec mean (SD)
1 (superior 25%)	1.77 (0.17)	1.80 (0.16)	1.75 (0.19)
2	2.21 (0.14)	2.30 (0.18)	2.15 (0.13)
3	2.83 (0.23)	2.95 (0.18)	2.67 (0.28)
4 (inferior 25%)	4.62 (2.23)	5.00 (2.65)	4.22 (1.74)

547

548 Table legend:

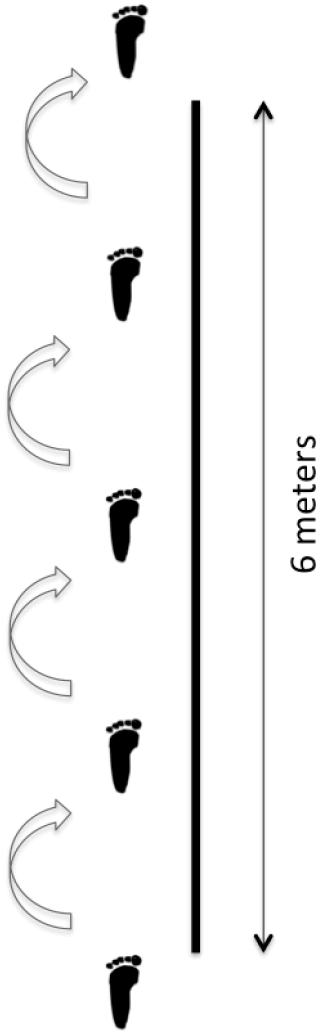
549 [1]Exercise Therapy

550 [2]Arthroscopic Partial Meniscectomy

551 [3]seconds

552

553 **Figure S1. Diagrammatic presentation of the six-meter timed hop test**



554
555

556 Figure legend:

557 The patient hopped the six-meter distance as quickly as possible and the time (in

558 seconds) was recorded using an ordinary stopwatch. One practice trial was followed by

559 two test trials, and the best out of the two was recorded [2, 3].

560

561 **Figure S2. Meniscal extrusion**

562

563 Figure legend:

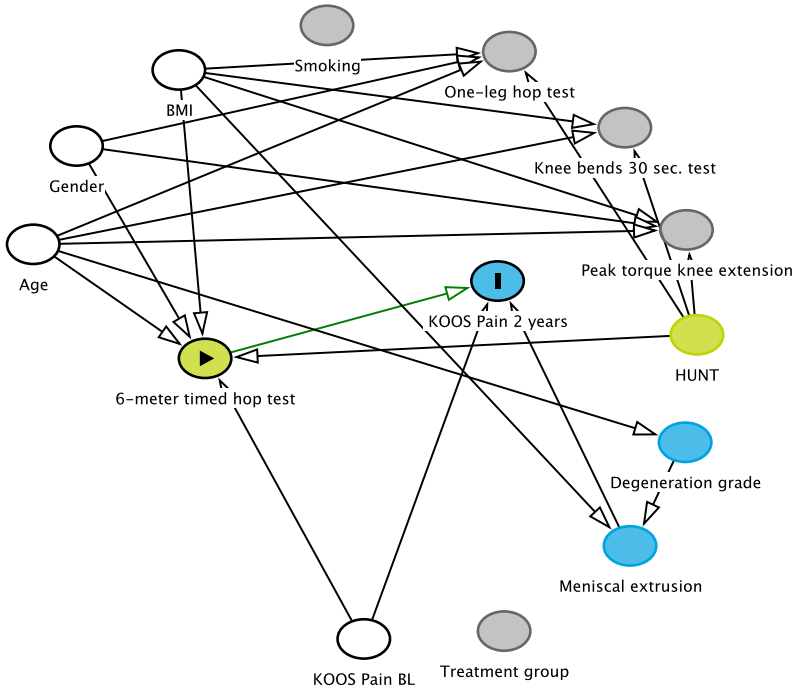
564 Meniscal extrusion measured on the coronal sequence image with the largest tibial spine

565 volume: a/b x 100 = extrusion in per cent.

566

567 **Figure S3. Diagnostic acyclic graph**

568



569

570 Figure legend:

571 An example of a DAG [4] with KOOS Pain at two years as the dependent variable. The

572 final regression model should be adjusted for the potential confounders KOOS Pain at

573 baseline (BL), age, gender and BMI (body mass index). (The figure is made using this

574 web-site http://www.dagitty.net)

575

576

577 REFERENCES

578

579 1. Kurtze N, Rangul V, Hustvedt BE, Flanders WD (2008) Reliability and validity of self-reported
580 physical activity in the Nord-Trondelag Health Study: HUNT 1. Scand J Public Health 36:52-61
581 2. Ross MD, Langford B, Whelan PJ (2002) Test-retest reliability of 4 single-leg horizontal hop tests. J
582 Strength Cond Res 16:617-622
583 3. Stensrud S, Roos EM, Risberg MA (2012) A 12-week exercise therapy program in middle-aged patients
584 with degenerative meniscus tears: a case series with 1-year follow-up. J Orthop Sports Phys Ther
585 42:919-931
586 4. Textor J, Hardt J, Knuppel S (2011) DAGitty: a graphical tool for analyzing causal diagrams.
587 Epidemiology 22:745

588

589