From Department of Clinical Sciences Danderyd Hospital
Karolinska Institutet, Stockholm, Sweden

PROXIMAL HAMSTRING AVULSIONS:

Surgery or not?

Elsa Pihl

Karolinska Institutet

Stockholm 2024

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2024

ISBN 978-91-8017-139-7

Cover illustration: Andreas Johansson

PROXIMAL HAMSTRING AVULSIONS:
Surgery or not?
Thesis for Doctoral Degree (Ph.D.)

By

Elsa Pihl

The thesis will be defended in public at Danderyds sjukhus, föreläsningssal Aulan, Entrévägen 2, Målpunkt K, plan 3, Stockholm, 15th of March 2024 at 9:am

Principal Supervisor:
MD PhD Carl Johan Hedbeck
Karolinska Institutet
Department of Clinical Sciences,
Danderyd Hospital

Co-supervisor(s):
Professor Kenneth B Jonsson
Uppsala University
Department of Surgical sciences

Professor Olof Sköldenberg
Karolinska Institutet
Department of Clinical Sciences,
Danderyd Hospital

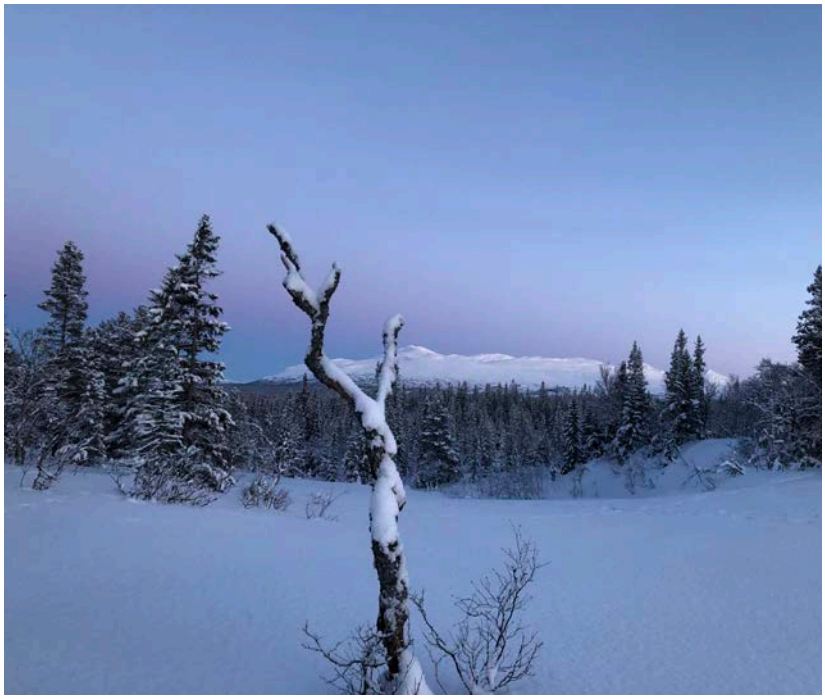
Opponent:
Professor Gino Kerkhoffs
University of Amsterdam Medical Centers
Department of Orthopedic Surgery and Sports Medicine
Division of Surgical Specialism

Examination Board:
Associate Professor Rikard Wedin
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Martin Hägglund
Linköpings University
Department of Health, Medicine, and Caring Sciences
Division of Prevention, Rehabilitation and Community Medicine

Associate Professor
Hanna Björnsson Hallgren
Linköping University
Department of Biomedical and Clinical Sciences
Division of Surgery, Orhopedics and Oncology

Till min älskade familj Gustav, Klara och Eva...



2021-02-10

...och till alla de kalla vinterdagarna i februari.

Popular science summary of the thesis

The focus of this thesis is proximal hamstring avulsion injury. The injury occurs when the individual suddenly splits, e.g., slipping while cross-country or water-skiing. The hamstring muscles try to resist the force of stretching when slipping, and the load can make the proximal hamstring tendons avulse from their origin. The injury is most prevalent in relatively active middle-aged people and is painful. In the acute phase, it is difficult to elongate the stride, crutches are often needed, and it is often painful to sit since there is a haematoma around the buttocks. After several days, a large bruise is likely to be present on the back of the thigh, and some patients experience neurological symptoms in their lower leg.

There are two treatment options: surgical treatment, suturing the tendons back to their footprint at the ischial tuberosity followed by a long period of rehabilitation; or nonsurgical treatment involving a similar long period of rehabilitation. The aim of this thesis is to determine which treatment is preferable for this injury and what to expect in terms of outcome.

The study population in the first study was comprised of patients who were treated for proximal hamstring avulsions at Danderyd Hospital between 2007 and 2013. In 2015, we performed a follow-up study with two different questionnaires, one designed to assess lower limb impairments and one designed to assess proximal hamstring avulsions. We were not able to observe any difference between the patients treated with surgery and those not treated with surgery, but the surgically treated patients reported that they were more physically active at the time of follow-up. However, since the choice of treatment was made long before the study started, we cannot conclude that the two treatment options had similar results. It could also be that the patients received the optimal treatment based on their injury type, physical activity level and medical status.

In the second study, we evaluated recovery after proximal hamstring avulsion at least two years after injury. Recovery was assessed with different questionnaires, as well as with strength and functionality tests. From 2018 to 2019, we examined the clinical data of patients who were injured between 2007 and 2016. The most common questionnaires for assessing the conditions of patients with proximal hamstring avulsions were used, but the scores from the questionnaires did not

reflect the function of the injured leg, which was assessed with functional tests. Running was the most frequently reported activity that was limited by the injury.

To properly assess recovery after proximal hamstring avulsions, different types of assessments are needed.

In the third study, we examined the quality of the hamstring muscles using MRI at follow-up. We examined the MRI images of the same group of patients at the same time points as in study number two. We found that the hamstring muscles of the injured leg decreased in volume and showed greater fatty infiltration than the muscles of the uninjured leg. This change in muscle quality had a weak but significant association with decreased muscle strength in the injured leg compared to the uninjured leg. At follow-up, we concluded that fatty infiltration and muscle atrophy are likely to occur as a result of proximal hamstring avulsions, and muscle quality impairment is weakly correlated with muscle weakness in the injured leg.

Patients from eight hospitals in Sweden and two hospitals in Norway participated in the fourth study. We randomized middle-aged patients who suffered from proximal hamstring avulsions to receive either surgical treatment or nonsurgical treatment. Thereafter, the patients were followed for two years with questionnaires, functional tests and MRI. In cases where either the doctor or patient had a strong preference for one of the treatment options, the patient was invited to join an observational cohort and undergo the same protocol for follow-up as the randomized patient. With the randomized design, we were able to determine whether the nonsurgically treated patients were in a worse condition at follow-up than the surgically treated patients. We did not find any difference in the patient-reported outcomes between the treatment options. The conclusion was that nonsurgical treatment was not worse than surgical treatment and should therefore be recommended as the treatment of choice for the middle-aged patients.

Abstract

In paper 1, a retrospective cohort study of patients treated for proximal hamstring avulsions at Danderyd Hospital from 2007–2013 was conducted. The study was performed in 2015. The primary outcome was the subjective patient reported outcome Lower Extremity Functional Scale (LEFS), and the exposure was surgical treatment. We hypothesized that there would be no difference in the LEFS, between the two treatments at this long-term follow-up. The results showed similar LEFS scores in the surgically treated 74 (SD±12) and nonsurgically treated 72 (SD±16) patients, which were also true after adjusting for confounders. There were some differences between the groups, with surgically treated patients having more severe injuries and reporting more hours of physical activity at the follow-up than nonsurgically treated patients. To obtain an evidence-based treatment decision for proximal hamstring avulsions, studies at higher scientific levels are needed.

Paper 2 consists of a cross-sectional cohort study. Patients treated for proximal hamstring avulsions at Danderyd Hospital between 2007 and 2016 were included. The study was performed from 2018-2019. The main outcomes were the correlation among subjective, patient-reported outcome measurements (PROMS), the Perth Hamstring Assessment Tool (PHAT) score and the Lower Extremity Functional Scale (LEFS) score and their correlation. The secondary outcomes were the correlation of PROMs with objective performance-based tests. Moreover, we explored which activity patients perceived to be most limited several years after injury. We hypothesized that there would be good correlations between the PHAT and LEFS scores and at least moderate correlations between these questionnaires and performance-based tests. We found strong correlations between the PHAT and LEFS scores ($r=0.832$, $p<0.001$). The LEFS was more appropriately aligned with the performance-based tests than the PHAT. Of all the physical performance tests performed at follow-up, only the isokinetic test could discriminate between injured and uninjured legs. Patients most frequently reported activity limitation was running. Since the PHAT, LEFS and physical performance-based tests seem to assess different aspects of recovery, both subjective and objective outcome measures are recommended to be used for follow-up after proximal hamstring avulsion.

In paper 3, we performed a cross-sectional study on the same cohort of patients in the same setting as in Study 2. The primary outcomes were hamstring muscle

volume and fatty infiltration at least 2 years after injury. The secondary outcome was the correlation of these parameters with isokinetic muscle strength. The conditions of the injured and uninjured legs were compared. We hypothesized that the injured leg would have greater fatty infiltration and atrophy than the uninjured leg at follow-up and that these findings would correlate with muscle weakness. We found that, on average, the hamstring muscle volume was reduced by 9% (SD±11%) compared to that of the uninjured leg. Fatty infiltration was significantly more severe in the injured hamstrings than in the uninjured hamstrings ($p<0.001$). Reduced muscle volume and increased fatty infiltration were significantly weakly correlated with isokinetic strength test results ($r=0.357-494$, $p< 0.001-0.013$). At follow-up, we concluded that fatty infiltration and muscle atrophy are likely to occur as a result of proximal hamstring avulsions, and muscle quality impairment is weakly correlated with muscle weakness in the injured leg.

In Study 4, a randomized, noninferiority, multicentre, preference-tolerated clinical trial was performed. Patients from ten study sites in Sweden (8) and Norway (2) participated. Patients were eligible for inclusion if they had an acute (within 4 weeks) proximal hamstring injury and were aged 30 to 70 years. Patients were randomly selected to undergo surgical or nonsurgical treatment. Surgical treatment included reinsertion of the tendons followed by rehabilitation, and nonsurgical treatment included rehabilitation only. If the patients and doctors could not reach a consensus on treatment, the patients were invited to join an observational follow-up cohort. The primary outcome was the PHAT score at two years post treatment. A noninferiority margin of 10 points on the PHAT was set for the lower limit of the two-sided 95% CI. The secondary outcomes consisted of the LEFS score, physical performance-based test results and muscle quality analysis results on MRI.

We enrolled 119 patients in the randomized trial and 97 in the observational cohort. According to the intention-to-treat analysis, the mean PHAT scores were similar, with mean (±SD) scores of 80.4(±19.3) and 77.7(±20.0) in the surgical and nonsurgical groups, respectively. The prespecified inferiority limit was not crossed (mean difference, -2.1; (95%CI -9.3 to 5.1) $p =0.017$ for noninferiority). According to the inverse probability weighted analysis of both cohorts combined, the mean difference in the PHAT score was -2.6 (95%CI, -7.9 to 2.8). Analyses of secondary outcomes including the mean LEFS score

difference of -2.1 (95%CI, -5.7 to 1.5) supported noninferiority.

The conclusion was that patients with proximal hamstring avulsions who were treated nonsurgically do not have worse PHAT scores than patients who were treated surgically; therefore, the treatment of choice for middle-aged patients should be nonsurgical treatment.

List of scientific papers

I. Pihl E, Skoldenberg O, Nasell H, Jonhagen S, Kelly-Pettersson P, Hedbeck CJ. Patient-reported outcomes after surgical and non-surgical treatment of proximal hamstring avulsions in middle-aged patients. BMJ Open Sport Exerc Med. 2019;5(1):e000511**.** DOI: 10.1136/bmjsem-2019-000511

II. Pihl E, Jonsson K .B., Berglöf M, Brodin N, Sköldenberg O, Hedbeck CJ. Exploring the Perth Hamstring Assessment Tool and Lower Extremity Functional Scale in a Proximal Hamstring Avulsion Cohort: A cross-sectional Study. American J Sports Med. 2021;49(7):1732–1740. DOI: 10.1177/03635465211008568

III. Pihl E, Skorpil M, Sköldenberg O, Hedbeck CJ, Jonsson K. B. At mid- to long-term follow-up after proximal hamstring tendon avulsion; there was greater fatty infiltration, muscle atrophy and strength deficit in the hamstring muscles of the injured leg than in the uninjured legh. J Orthop Surg Res 2023 Vol. 18(1):114 DOI: 10.1186/s13018-023-03582-2

IV. Pihl E, Laszlo S, Rosenlund AM, Kristofferssen MH, Schilcher J, Hedbeck CJ, Skorpil M, Micoli C, Eklund M, Sköldenberg O, Frihagen F, Jonsson K. B. Operative versus non-operative treatment of proximal hamstring avulsions.

Scientific papers not included in thesis

I. Pihl E, Kristoffersen MH, Rosenlund AM, Laszlo S, Berglof M, Ribom E, Eriksson K, Frihagen F, Mattila V.M., Schilcher J, Eklund M, Snellman G, Skorpil M, Sköldenberg O, Hedbeck CJ, Jonsson K. B. The proximal hamstring avulsion clinical trial (PHACT)-a randomised controlled non-inferiority trial of operative versus non-operative treatment of proximal hamstrings avulsions: study protocol. BMJ Open 2019 Vol. 9(9):e031607. DOI: 10.1136/bmjopen-2019-031607

II. Laszlo S, Nilsson N, Pihl E, Mattila V.M, Schilcher J. Sköldenberg O, Frihagen F and Jonsson K.B. Proximal Hamstring Tendon Avulsions: A Survey of Orthopaedic Surgeons' Current Practices in the Nordic Countries. Sports Med Open 2022 Vol. 8(1):49. DOI: 10.1186/s40798-022-00439-6

Contents

1 Literature review 3
1.1 Anatomy 3
1.2 Epidemiology 5
1.3 Injury mechanism 6
1.4 Diagnosis 6
1.4.1 Clinical findings 6
1.4.2 Magnetic resonance imaging (MRI) 7
1.5 Treatment 9
1.5.1 Surgical treatment 10
1.5.2 Nonsurgical treatment 10
1.6 Adverse events and complications following proximal hamstring avulsions. 12
1.7 Outcomes 12
1.7.1 Important properties of outcome measures 13
1.7.2 Measures for evaluating the properties of instruments 15
2 Research aim 17
2.1 Specific aims 17
2.1.1 Study 1 17
2.1.2 Study 2 17
2.1.3 Study 3 17
2.1.4 Study 4 17
3 Materials and methods 19
3.1 Study designs 19
3.1.1 Cohort study 19
3.1.2 Randomized controlled trial 19
3.2 Instruments relevant to the thesis 21
3.2.1 Patient reported outcome measures (PROM) 21
3.2.2 Physical Performance-based Tests 25
3.2.3 Magnetic resonance imaging (MRI) 29
3.3 Statistical methods relevant to the thesis 31
3.3.1 Parametric test 31
3.3.2 Non-parametric tests 31
3.3.3 Intraclass correlation coefficient (ICC) 31
3.3.4 Spearman correlation coefficient 32
3.3.5 Linear regression 32

3.3.6 Imputation ... 32
3.3.7 Propensity score ... 32
3.3.8 Intention to treat (ITT) analysis ... 33
3.3.9 Per-protocol (PP) analysis ... 33
3.3.10 As-treated (AS) analysis ... 33
3.4 Patients and methods ... 35
3.4.1 Study 1 ... 35
3.4.2 Study 2 ... 36
3.4.3 Study 3 ... 37
3.4.4 Study 4 ... 38
3.5 Ethical considerations ... 41
3.5.1 Ethical considerations of study 1 ... 41
3.5.2 Ethical considerations for Studies 2 and 3 ... 41
3.5.3 Ethical considerations of Study 4 ... 42
4 Results ... 43
4.1 Study 1 ... 44
4.1.1 Participants and baseline characteristics ... 44
4.1.2 Main results ... 44
4.2 Study 2 ... 46
4.2.1 Participants and baseline characteristics ... 46
4.2.2 Main results ... 46
4.2.3 Secondary outcomes ... 47
4.3 Study 3 ... 48
4.3.1 Participants and baseline characteristics ... 48
4.3.2 Main results ... 48
4.3.3 Secondary outcomes ... 49
4.4 Study 4 ... 51
4.4.1 Participants and baseline characteristics ... 51
4.4.2 Main results ... 52
4.4.3 Secondary outcomes ... 54
4.4.4 Adverse events ... 55
5 Discussion ... 57
5.1 General discussion and methodological considerations ... 57
5.2 Discussion of Study 1 ... 61
5.2.1 Strengths ... 61
5.2.2 Limitations ... 62
5.3 Discussion of Study 2 ... 62

5.3.1 Strengths ... 62
5.3.2 Limitations ... 63
5.4 Discussion of Study 3 ... 63
5.4.1 Strengths ... 63
5.4.2 Limitations ... 64
5.5 Discussion of Study 4 ... 65
5.5.1 Strengths ... 65
5.5.2 Limitations ... 65
6 Conclusions ... 67
6.1 Study 1 ... 67
6.2 Study 2 ... 67
6.3 Study 3 ... 67
6.4 Study 4 ... 67
7 Points of perspective ... 69
8 Populärvetenskaplig sammanfattning på svenska ... 71
9 Acknowledgements ... 75
10 References ... 79
11 Appendices ... 87
11.1 Perth Hamstring Assessment Tool ... 87
11.2 Lower Extremity Functional Scale ... 88
11.3 Patient-Specific Functional Scale ... 89

List of abbreviations

AS	As treated
CI	Confidence Interval
ICC	Intraclass Correlation Coefficient
ICF	International Classification of Functioning, Disability and Health
ITT	Intention to treat
LEFS	Lower Extremity Functional Scale
LSI	Limb Symmetry Index
MCID	Minimal Clinical Important Difference
MDC	Minimal Detectable Change
MET	Metabolic Equivalent of Task
MRI	Magnetic Resonance Imaging
N	Newton
OR	Odds Ratio
P	P value
PHA	Proximal Hamstring Avulsion
PHAT	Perth Hamstring Assessment Tool
PHIQ	Proximal Hamstring Injury Questionnaire
PP	Per protocol
PROM	Patient Reported Outcome Measure
PSFS	Patient Specific Functional Scale
RCT	Randomized Controlled Trial
SD	Standard Deviation
SEM	Standard Error of Measurement

Introduction

In the human body, there are intricate interactions between contracting muscles and lengthening muscles. When muscles contract, they pull on tendons. This generates a load on the tendon origins at different joints and makes the joints and thereby the extremities move.

A typical tendon avulsions occurs when the contraction force from the muscle exceeds the strength of the tendon-muscle junction, the tendon itself or the tendon-bone junction.

The specific focus of this thesis is tendon avulsion of proximal hamstring origin, proximal hamstring avulsions (PHA's), and the two different treatment options and their outcomes.

However, what is a good result after treatment? When studying the recovery of a condition that interferes with a person's daily life, such as proximal hamstring avulsions, the focus should be on what is important for the affected person in terms of activities of daily living and recreation. Within the research field of proximal hamstring avulsions there are until now no qualitative study aiming at exploring what activities these patients have problems in performing, nor any consensus on what the 'gold standard' for assessing outcome is.

The first paper on proximal hamstring avulsions was published in 1988 and listed in PubMed, and the authors described two rare cases of proximal hamstring avulsions without any visible fractures on X-ray. The patients'history included sudden onset of pain in the buttocks, difficulty weight bearing, palpable defects at the buttocks and weakness in knee flexion. After surgical repair, function was restored[1]. The first study of the thesis was conducted 27 years later, in 2015. At that time, research on such subject matter was still limited, with only 52 papers published on this subject in PubMed. There were no prospective studies but there were case reports and case series of surgically treated patients. Since then, the knowledge base for proximal hamstring avulsions has expanded substantially, and 149 papers have been published up to late 2023 (Figure 1).

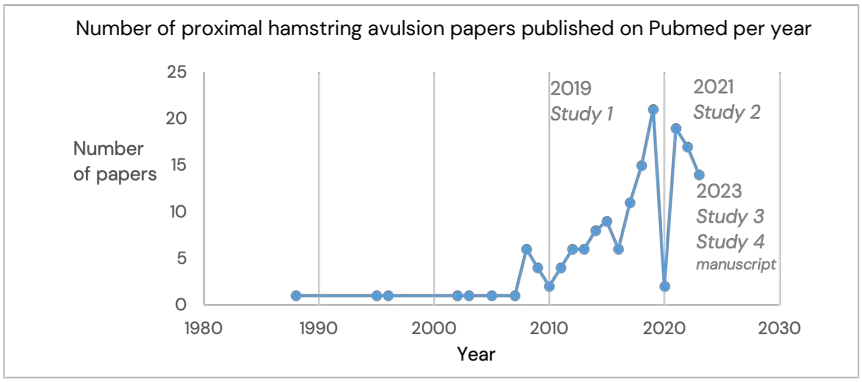


Figure 1: Evolution of the field of proximal hamstring avulsions until October 2023, as indicated by the number of papers retrieved from PubMed when searching for literature on proximal hamstring avulsions. Combined with the timing of the studies performed for this thesis.

The impression after reading papers discussing current clinical practices and the most cited literature review was that surgical treatment is increasingly becoming the treatment of choice for proximal hamstring avulsions[2-6].

However, there is potential publication bias because of the relatively small number of nonsurgically treated patients studied. In the frequently-cited review by Bodendorfer et al., only 3.6% of the patients were nonsurgically treated[4]. Furthermore, almost all studies comparing treatment options are based on retrospective data, which can lead to bias by indication[7, 8]. The need for studies with higher level of evidence is frequently argued.

To my knowledge, only one prospective cohort study, with the inherent problem of bias by indication, compared the two treatment options and found no difference in patient-reported outcomes[9].

The question of which treatment that is preferable for what patients and what the expected outcomes after treatment are remains unanswered.

1 Literature review

1.1 Anatomy

The main action of the hamstring muscle complex is to extend the hip and to flex the knee[10]. The complex consists of the biceps femoris long and short head, semitendinosus, and semimembranosus muscles (Figure 2). All except the biceps femoris short head attach to the ischial tuberosity, where the long head of the biceps femoris and semitendinosus meet to share a common conjoint tendon. Sato et al. reported that some of the fibres of the biceps tendon extend to the sacrotuberous ligament[11]. The short head of the biceps femoris originates from the middle third of the linea aspera; i.e., the short head of the biceps femoris does not cross the hip joint but only the knee joint[11]. Studies of human cadavers have shown that the semitendinosus region partly consists of muscle fibres, whereas the biceps femoris region of the conjoint tendon consists of only tendon fibres[11, 12]. The semimembranosus has its own tendon attached to the ischial tuberosity, proximally, laterally and anteriorly to the common origin of the other two hamstring muscles[11, 12](Figure 3).

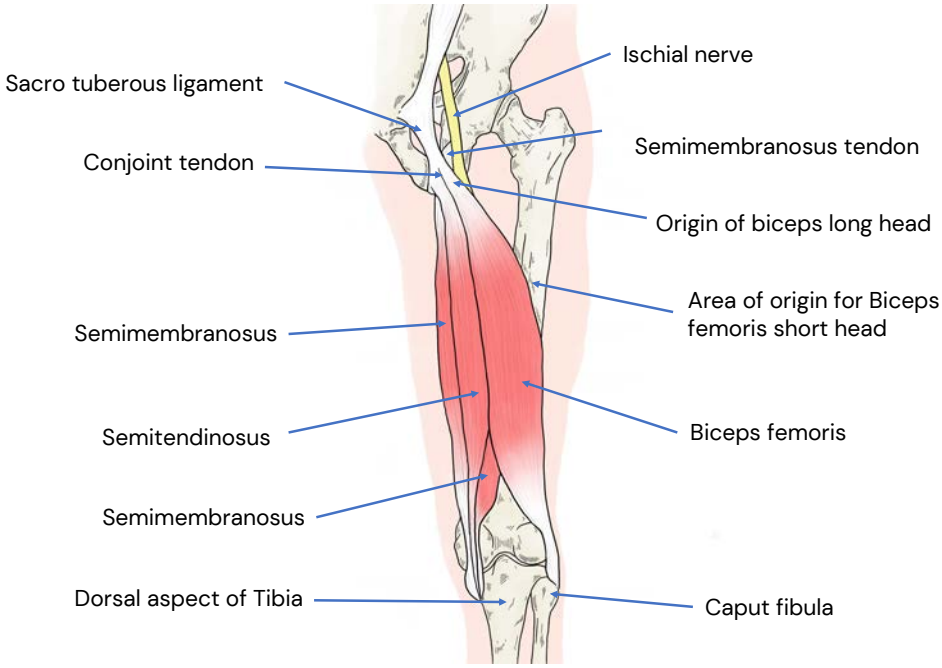


Figure 2: Schematic image of the anatomy of the hamstring muscles, dorsal view.

The insertion of the biceps femoris is located at the head of the fibula, and the insertion of the other two muscles are at the proximal medial tibia.

The nerve supply comes from the sciatic and tibial nerves and originates from the L5 and S1 nerve roots[10, 11]. The long head of the biceps femoris is supplied by the tibial nerve, and the short head is supplied by the fibular nerve. The semitendinosus and semimembranosus muscles are supplied by the tibial nerve[13].

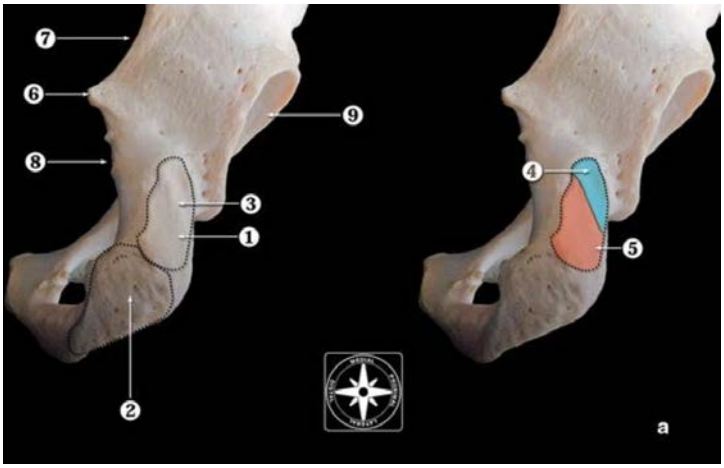


Figure 3: Image retrieved from van der Made et al 2015[12], presenting the attachment of the proximal hamstring tendons to the footprint. Posterior view of the right coxal bone showing the ischial tuberosity, which can be divided into two regions. 1 Upper region. 2 Lower region. 3 Vertical ridge, which divides the upper region in two facets. 4 Lateral facet, for insertion of the tendon of the semimembranosus muscle. 5 Medial facet, for insertion of the conjoint tendon of the long head of the biceps femoris and semitendinosus muscle. 6 Sciatic spine. 7 Greater sciatic notch. 8 Lesser sciatic notch. 9 Acetabulum. Printed with permission from Springer Nature license number 5654650644792.

1.2 Epidemiology

Hamstring muscle injuries account for roughly 26-39% of all sport injuries in elite athletes[3, 14]; however, proximal hamstring avulsions are relatively uncommon, and only approximately 9% of all major hamstring muscle injuries are proximal hamstring avulsions[15].

However, the proximal hamstring avulsions predominantly affect middle-aged patients who are nonelite athletes[5, 16]. According to a review from 2017, only 20 of 772 the selected patients were stratified as elite athletes[4]. Both sexes are equally represented, but females tend to be injured at an older age and more often sustain their injury during daily life activities than males, who are more likely to be injured during sporting activities[5, 16] (Figure 4).

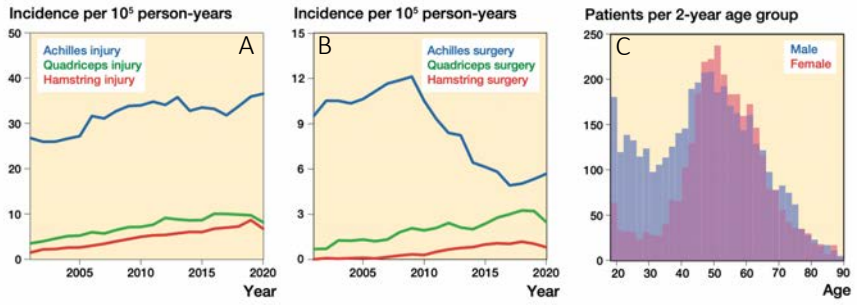


Figure 4: Images of the incidence of injury, surgical treatment and number of patients injured retrieved from Laszlo S and Jonsson KB 2023[5] A, Incidence per 10^5 person-years for Achilles, quadriceps, and hamstring injuries (ICD-10 S86.0, S76.1, and S76.3). B, Incidence of surgical treatment for the respective injuries in all Swedish adults aged between 18 and 90 years between 2001 and 2020. C Age and sex distribution of patients with hamstring injuries. Although hamstring injuries are relatively uncommon, they are being more frequently diagnosed, and surgery for such injuries is being more frequently performed. Females seem to be injured at a slightly older age. Printed with permission from ACTA Orthopaedica.

Additionally, adolescents can suffer from a similar injury, but this injury is often associated with bony avulsion from the ischial tuberosity and can be regarded as a different type of injury[17, 18].

The focus of this thesis is the vast majority of patients suffering from proximal hamstring avulsions, the middle-aged.

1.3 Injury mechanism

A proximal hamstring avulsion occurs when there is an eccentric load and when the tendon and muscle lengthen while the hamstring muscle contracts to resist the load. This typically occurs rapidly, during accidental splitting, e.g., while slipping on cross-country skis, slipping on a slippery floor, or sprinting (Figure 5)[16, 18, 19].

Figure 5: Percentage of activities performed by patients at the time of injury. The numbers were retrieved from Irger et al. 2019[16].

1.4 Diagnosis

The diagnosis is based on patient history, clinical examination, X-ray findings (absence of skeletal lesions) and MRI.

1.4.1 Clinical findings

As described in the first published paper on proximal hamstring avulsions, patients often present after an accidental split, complaining of severe and acute pain in the buttocks; additionally, it is common for patients to be unable to bear weight on the affected leg[1].

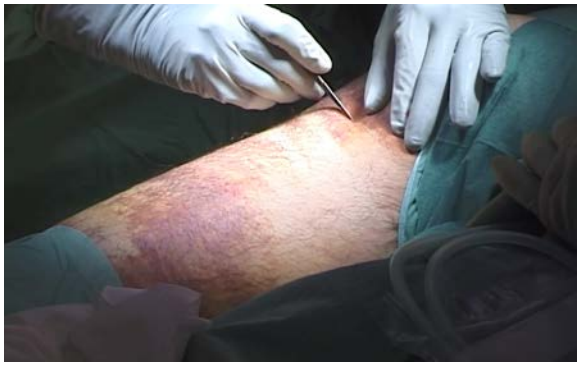


Figure 6: Bruising on the back of the thigh. Courtesy of M. D. Ph. D. Sven Jönhagen.

A palpable defect can sometimes be found distal to the ischial tuberosity, and after several days, spectacular bruising can be observed on the back of the thigh (Figure 6)[20-22].

Strength examination reveals weakness during both knee flexion and hip extension (Figure 7)[23]. The lower leg should be assessed for nerve symptoms since symptoms such as numbness, tingling, and foot drop can occur[20].

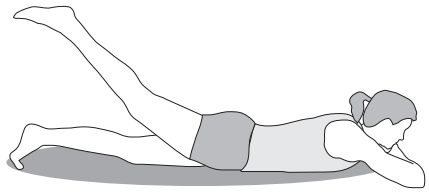


Figure 7: Negative hip extension test. Patients can extend their hips in the prone position. Described by Jönhagen et al. 2009 at Annual National Sports Medicine Meeting in Umeå, Sweden.

1.4.2 Magnetic resonance imaging (MRI)

MRI is the gold standard to confirm the diagnosis[24].

Wood et al. described the typical findings on proton density fat suppressed sequences in axial, coronal, and sagittal planes. MRI images show oedema and tendon avulsion with discontinuity between the tendon edge and ischial bone[19]. The classification was as follows. 1) Bone avulsion, 2) musculotendinous, 3) incomplete, 4) complete (no retraction), and 5) complete (retraction).

However, at present, there is no established measurement procedure available for quantifying the degree of tendon retraction. A survey among radiologists from 2017, published 2023, showed variability in tendon retraction

measurements because of the differences in choosing a proximal landmark for measurements and difficulties in locating the proximal tendon stump[25].

Recently, van der Made et al. presented a standardized assessment for proximal free tendon discontinuity using the 'dropped ice cream sign' and a standardized measurement procedure for acute tendon retraction for proximal full-thickness free hamstring tendon injury on MRI[24]. (Figure 8 and Figure 9).

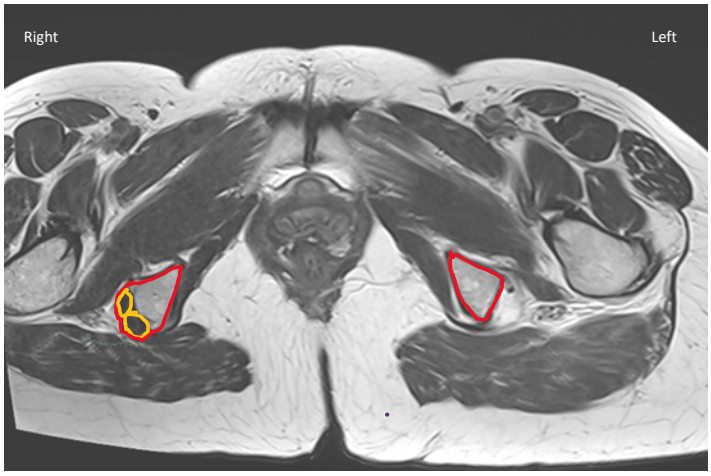


Figure 8: T2-weighted axial image. Right ischial tuberosity showing "two yellow scoops of ice cream" at the bone. The conjoint tendon is shown medially, and the semimembranosus tendon is shown laterally. The left ischial tuberosity showing "no scoop of ice cream and only cone (bone) left", both conjoint and semimembranosus tendons are avulsed.

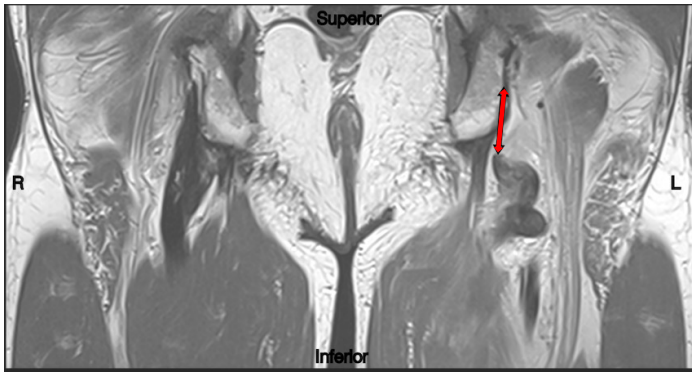


Figure 9: Coronal image and PD weighted tendons of the right leg are attached to the ischial tuberosity. The tendons of the left leg are avulsed and retracted the distance of the arrow. The footprint is located in the most cranial half of the ischial tuberosity.

Direct retraction was proposed as a standardized procedure for measuring tendon retraction, with an ICC of 0.88[24].

There are several injury patterns, which include both conjoint and semimembranosus tendon avulsion, or avulsion of the conjoint tendon or semimembranosus tendon only. The most common injury pattern described is complete tendon avulsion with both the conjoint tendon and the semimembranosus tendons avulsed[19, 24]; the distribution of injuries are presented in Table 1.

Table 1

Wood type	Distribution % of injury types	Injury pattern described by van der Made et al. 2022	Distribution % of injury types
1) Bony avulsion	1	**Conjoint tendon only** (Wood type 3)	3
2) Musculotendinous	1	**Injuries of conjoint tendon with retraction** (Wood type 3 and 5)	80
3) Incomplete	10	**Semimembranosus tendon only** (Wood type 3)	7
4) Complete no retraction	22	**Injuries of semimembranosus tendon with retraction** (Wood type 3 and 5)	88
5) Complete with retraction	66	**Injuries of both tendons** (Wood type 3 and 5)	90

Table 1: Distribution of different types of proximal hamstring injuries described by Wood et al 2008 and van der Made et al 2022[9, 19]. Van der Made et al reported that injuries of only conjoint tendon constitute of 3% of all injuries, and conjoint tendon injuries with retraction constitute 80% of all injuries. Injuries to both tendons with and without retraction constitute of 90% of all injuries. In conclusion, most injuries are complete (van der Made injuries of both tendons: 90%) and retracted (Wood type 5: 66%).

1.5 Treatment

There are two treatment options for proximal hamstring avulsions: surgical treatment involving reinsertion of the tendons to their origin at the ischial tuberosity followed by rehabilitation or nonsurgical treatment involving rehabilitation only.

1.5.1 Surgical treatment

In surgical treatment, the tendons are reattached to their footprint at the ischial tuberosity. Surgical treatment is, by orthopaedic surgeons, thought to be advantageous in that it restores the biomechanics of the muscle by improving muscle quality, reducing atrophy and improving function[3, 26].

The ischial tuberosity can be approached through a transverse incision in the gluteal crease or a longitudinal incision in the proximal thigh. The skin is incised, and then dissection is performed, taking care to avoid the posterior femoral cutaneous nerve. The sciatic nerve needs to be identified or at least considered before proceeding to avoid iatrogenic injury. After the gluteus maximus has been retracted, a longitudinal incision is made in the muscle fascia. If the paratenon is intact, it needs to be incised to mobilize the hamstring tendons. The footprint at the ischial tuberosity is made bare by debridement. Suture anchors are used to reinsert the tendons, and depending on the trademark, they are hammered or screwed into the bone. Different techniques are used to ensure that the sutures go through the tendons and are pushed towards the bone and anchors. The wound is closed layer by layer[27, 28].

1.5.2 Nonsurgical treatment

Nonsurgical treatment such as rehabilitation assisted by a physiotherapist is not as well documented or studied as the surgical procedure. There are several different rehabilitation protocols published and used[8, 9, 20]. Most of them recommend the patient to start rehabilitation as soon as possible after diagnosis[7, 9, 29]. The Askling protocol was modified and studied by Leger st Jean et al.[30] The original Askling protocol is the most common protocol used in the Stockholm area. However, this approach was constructed as a postoperative rehabilitation protocol and not as a nonsurgical treatment protocol. In theory, it should be possible to use the same rehabilitation protocol for both treatment options.

A modification of the Askling protocol is presented in (Figure 10).

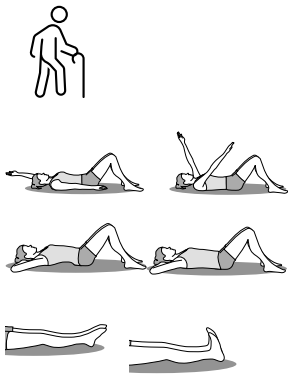
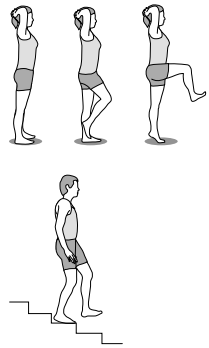
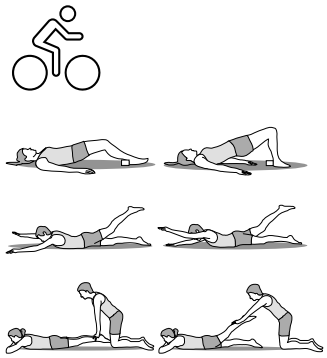
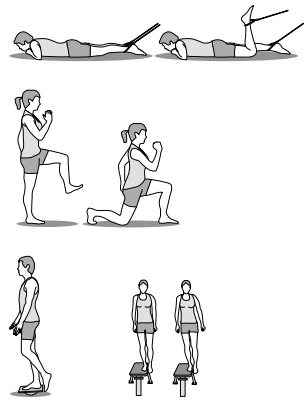
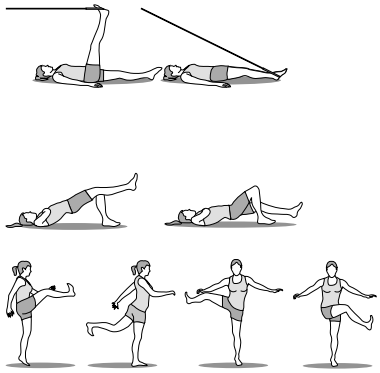
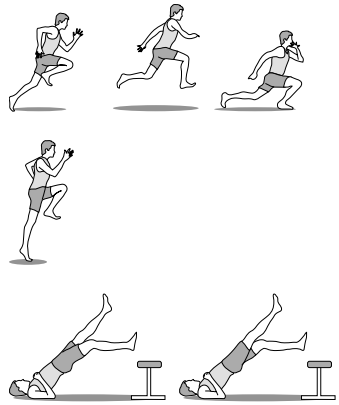
Week 1-2 	Full weight bearing walk with short strides and using crutches. No brace is used. Easy exercises for core and lower leg.	**Week 3-4** 	Release of crutches. Passive mobility exercises. Start of easy exercises on the injured leg. Focus should be on range of motion exercises.
Week 5-6 	If the wound is healed pool training can start. Stationary cycling can also be introduced. Hamstring strength exercises can be introduced.	**Week 7** 	Single leg exercises and hamstring curls can be started. Focus should be on proprioception.
Week 8 	Eccentric hamstring strength exercises are introduced. Dynamic mobility exercises. More preinjury activity like exercises.	**After 8 weeks, often much later** 	It is not recommended to return to preinjury activity level before functions like jumping and running can be performed without pain, stiffness, or the feeling of insecurity

Figure 10: Brief summary of the modified version of the Askling rehabilitation protocol, which has been used by most of the patients in this thesis.

1.6 Adverse events and complications following proximal hamstring avulsions.

In a review from 2017, the incidence of adverse events after surgery was found to be as high as 23%[4], and infection and neurological complications were the most common, followed by rerupture and thromboembolic complications. In studies of nonsurgically treated cohorts, adverse events were not described. However, it is known that some nonsurgically treated patients undergo "late surgery" due to neurological symptoms and weakness in the hamstring muscles[31]. Recent papers have presented neurological symptoms from both surgically (10-40%) and nonsurgically (33%-40%) treated cohorts[8, 9]. Additionally, problems with prolonged sitting have been reported after both treatments[29].

1.7 Outcomes

There are many different outcome measures used for follow-up after proximal hamstring avulsions[32, 33]. The outcome measures can be grouped into subjective; patient-reported outcome measures (PROMs), surveys, and objective; physical performance-based tests (including functional and strength tests), and imaging outcomes (mostly MRI).

In a recent review, of outcomes used after proximal hamstring avulsions, as many as thirty different outcome measures were found in the selected articles. The most frequently used were isokinetic strength testing, questions regarding return to sport and different questionnaires[33]. However, the authors concluded that there is no determined 'gold standard' of outcome measures after proximal hamstring avulsions[33].

The PROMs can be used to assess various aspects of outcomes. For example, there are general questions regarding quality of life, condition-specific questions and questions regarding pain on ordinal or continuous scales. Several different PROMs have been used for follow-up after proximal hamstring avulsions.

At the time the studies were performed for this thesis, the Perth Hamstring Assessment Tool (PHAT) was the only scale that had been validated in terms of construct and also found to yield reliable measurements for use after proximal hamstring avulsions[34]. However, more recently, the Sidney Hamstring Origin Rupture Evaluation (SHORE) was developed, and the validity of its construct was investigated through comparison with the PHAT for proximal hamstring avulsion

patients[35]. Other PROMS frequently used are the LEFS[36], the 12-item Short Form Survey (SF-12)[37], the Single Assessment Numeric Evaluation (SANE)[38]; additionally different uses of the visual analogue scale (VAS) and the Tegner Activity Score (TAS) are used[39].

The most commonly used performance-based tests are isokinetic strength measurements and single leg-hop tests[33].

Few studies have evaluated imaging outcomes such as fatty infiltration and assessment of tendon attachment[9, 26, 40]. To my knowledge, there is no report of muscle volume being assessed at follow-up, prior to the studies of this thesis.

1.7.1 Important properties of outcome measures

When evaluating patients' self-reported outcomes, it is important to know whether the instrument used assesses problems and concerns that are actually important to patients and whether it is free from different types of measurement errors. These concepts have been previously described in detail[41]. In Table *2*, different subgroups of reliability and validity and which of these concepts that are used in the studies of this thesis are presented.

Table 2

The different aspects of reliability and validity that are explored in the studies of the thesis		
Reliability		
Test-retest reliability	The ability of an instrument to give the same measure when used at different time points when the condition has not changed.	Assessed in study 2 using the ICC. PROM assessment and Physical performance-based tests were performed twice.
Intrarater reliability	The ability of an instrument to give the same score by the same person when used on different time points.	Assessed in study 3 using the ICC. Goutallier grading assessment and volume assessment on MRI examination were performed twice.
Measurement error	The systematic and random measurement error that is attributed to the true changes in construct to be measured.	Assessed in study 2 using the SEM measurement of PROMs.
Validity		
Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured.	Assessed in study 2 by counting the activities mentioned in the PSFS and their overlap with the items of the PHAT and LEFS.
Construct validity	The degree to which the scores of an instrument are consistent with the hypotheses that it relates to other scores or instrument and measures the construct to be measured.	Assessed in study 2 by comparing PROMs and physical performance-based tests results using Spearman correlation coefficient.
Cross cultural validity	The degree to which a translated instrument adequately aligns with the items of the original version of the instrument.	Assessed in studies 1, 2 and 4. Not measured but PROMs are translated and tested by the research group prior to the studies.
Criterion validity	The degree to which the scores of an instrument are an adequate reflection of a 'gold standard'.	Assessed in studies 2 and 3 by using the spearman correlation coefficient. Comparing isokinetic strength test results with other functional tests results and comparing isokinetic strength test results and MRI findings.

Table 2: Aspects of reliability and validity with definitions from Mokkink et al.[41] and which aspects are used in the thesis. This framework was originally designed for patient-reported outcomes; however, to obtain a comprehensive view of the measurements used in the thesis, also the objective outcome measures are included in the table.

1.7.2 Measures for evaluating the properties of instruments.

1.7.2.1 The standard error of measurement (SEM),

SEM is an estimate of the measurement error. It estimates how repeated measures are spread around the "true" score. The SEM is related to reliability, and a lower precision results in higher SEM values[42].

1.7.2.2 The ceiling effect

A ceiling effect occurs when a participant scores the highest possible score of an instrument. If more than 15% of participants are scoring maximum score, the instrument is regarded as not challenging enough for the assessed cohort. In such cases, the content validity and responsiveness of the instrument are limited[42].

1.7.2.3 Minimal detectable change (MDC)

MDC is defined as the minimal change that falls outside the measurement error in the score of an instrument[43].

1.7.2.4 Minimal clinically important differences (MCID)

MCIDs are patient-derived scores that reflect changes in a clinical intervention that are meaningful for the patient[44]

2 Research aim

The overall aim of this thesis is to provide novel and evidence-based data to help guide physicians in making treatment decisions for patients with proximal hamstring avulsion.

2.1 Specific aims

2.1.1 Study 1

To study the difference in patient-reported outcomes after proximal hamstring avulsions between surgically and nonsurgically treated patients.

2.1.2 Study 2

The purpose of this study was to explore commonly used subjective and objective outcome measures for the follow-up after proximal hamstring avulsions. Moreover, to investigate the most common activity limitations patients report at follow-up.

2.1.3 Study 3

To assess muscle fatty infiltration, volume, and their correlation with strength after proximal hamstring avulsions. Furthermore, to assess tendon attachment at follow-up.

2.1.4 Study 4

To study whether nonsurgical treatment is noninferior to surgical treatment measured with patient-reported outcome.

3 Materials and methods

An overview of the study designs used in the thesis is presented in Figure 11.

3.1 Study designs

3.1.1 Cohort study

In a cohort study a population is followed longitudinally over time; one part of the cohort is exposed to a condition or a treatment, and the other is unexposed. The exposure is the condition or treatment the researcher aims to explore, together with its relevant outcomes. At follow-up the outcome variables are compared between the exposed and unexposed groups[45].

3.1.1.1 Retrospective cohort study

In a typical medical retrospective cohort study, the data on exposure and confounders are collected from the patients' medical records, that already existed, when the study started. Data collection is therefore limited to the existing sources of data in the records. At follow-up, the outcomes are compared between the exposed and unexposed groups.

3.1.1.2 Cross-sectional cohort studies

According to the cross-sectional study design the outcome is measured in the study population at one time point. This design cannot provide any answers on causality since patient outcomes are not studied temporally.

3.1.2 Randomized controlled trial

The randomized controlled trial (RCT) has been regarded as the 'gold standard' of trials[45], as the randomization between exposures reduces the risk of bias associated with observational cohort studies. The patients in the study population are randomly assigned to either the exposed or unexposed group. At follow-up, the outcomes are compared between the exposed and unexposed groups, and conclusions about causality can be drawn. One advantage of the RCT design is that confounders are randomly distributed throughout both groups, which provides the opportunity to refine the treatment effect. One disadvantage of RCTs is that to obtain a homogenous group for study, the inclusion criteria are often strict, and it can be difficult to recruit patients from the source population. This approach often leads to high internal validity and low

external validity. However, a multicentre design is often associated with better external validity, but the internal validity is often affected by such a design.

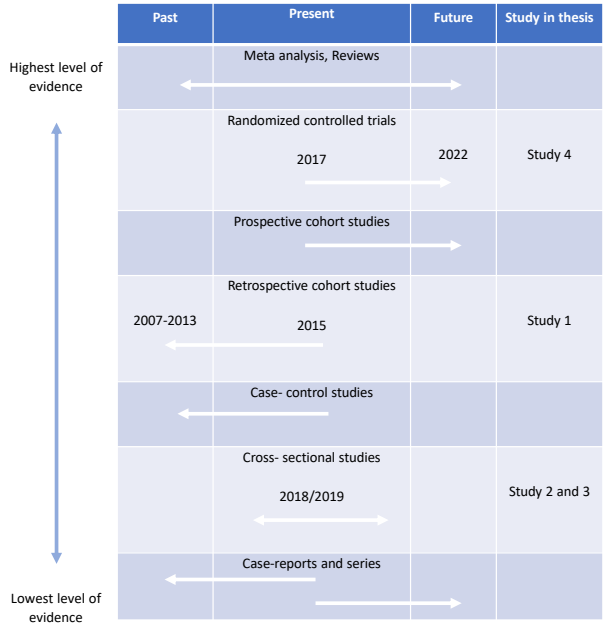


Figure 11: The hierarchy and relation to the temporality of different study designs and the study designs used in the studies of this thesis. The synthesis is made of design and hierarchy as described by Rothman 2012[45].

3.1.2.1 Partially randomized patient preference trials

In a partially randomized patient preference trial, patients who strongly preferred one of the treatment options, declined randomization, or received the treatment of their choice are followed in a parallel cohort with the same protocol as the randomized patients. The advantage of this design is that external validity is guaranteed since patients who do not consent to be randomized can also be followed, and studied if they are different from the ones randomized. The size of the study population often becomes larger in studies with this design, compared to classical RCT design[46].

3.1.2.2 Noninferiority trial

The aim of noninferiority studies is to demonstrate that the investigated treatment is not worse than the comparative treatment by more than a predetermined margin. The noninferiority margin can be determined in clinical studies on the basis of clinical knowledge and former studies on the chosen outcome measure. A noninferiority design is appropriate when, e.g., the cost and

risk are expected to be greater for the comparator than for the investigated treatment[47]. If the difference between the treatment option is not greater than the noninferiority margin, the treatment studied is considered noninferior to the comparator. The investigated treatment can be recommended because of lower costs and risks and no expectation of worse outcomes than the comparative treatment (Figure 12).

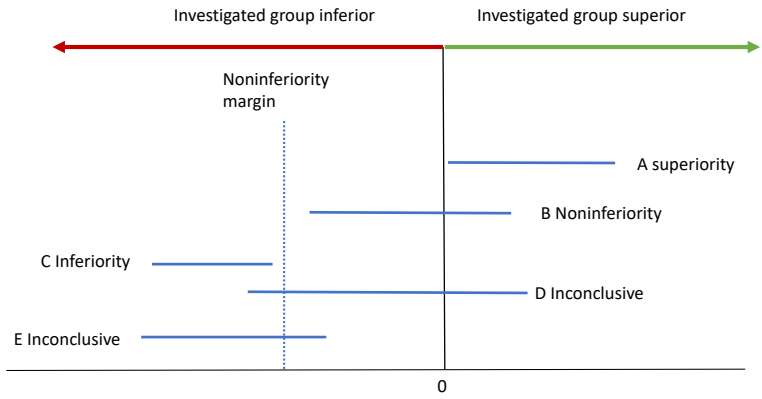


Figure 12: An image showing the strategy by which the results from noninferiority studies can be interpreted, as described by Groenwold et al.[47]. The blue lines represent the confidence intervals (CIs) for differences in outcomes between the investigated and control groups. A) A CI above 0 indicates that the investigated group is superior. B) A CI not below the noninferiority margin – the experimental variable is noninferior. C) The entire CI is below the noninferiority margin, so the experimental variable is inferior. D) Parts of the CI crosses the noninferiority margin, so the result is inconclusive. E) Parts of the CI crosses the noninferiority margin, so the result is inconclusive.

3.2 Instruments relevant to the thesis

3.2.1 Patient reported outcome measures (PROM)

3.2.1.1 Perth Hamstring Assessment Tool (PHAT)

The Perth Hamstring Assessment Tool (PHAT)[34] specifically assesses the proximal hamstring avulsion injury, and the maximum score is 100 points and minimum 0 points; the higher the score, the better the function. It consists of four domains: 1) pain during three different activities, 2) assessing the maximum amount of time the patient can perform three different activities, 3) current activity level and 4) level of local tenderness at buttocks.

The construct of PHAT has been validated for middle-aged, surgically treated proximal hamstring avulsion patients[34]. The PHAT was the only instrument that

has been validated for this population, during the time when the studies of this thesis were performed.

The measurement properties at one year after injury have been shown to be as follows: the mean score 74.1-76.7 (SD ±16.0 - ±22.5), the minimal detectable change (MDC) 16.4 points, the standard error of measurement (SEM) 5.9, the intraclass correlation coefficient (ICC) 0.84 and the ceiling effect 5-10%[28, 34]. The PHAT was translated to Swedish by our research group[48] using the guidelines of Guillemin et al.[49]. The original PHAT questionnaire is presented in the appendecis.

3.2.1.2 Lower Extremity Functional Scale (LEFS)

The LEFS is a generic instrument used for assessing injuries in the lower extremity[36]. It consists of 20 items related to different activities, and the patient is asked to score whether they have extreme difficulty or are unable to perform the activity; (0), quite a bit of difficulty (1), has moderate difficulty (2), has a little bit of difficulty (3), or has no difficulty (4). The maximum score is 80, which indicates that the patient has no difficulty performing the activities in the LEFS and minimum score is 0.

The LEFS was not specifically validated for assessing the ability of patients with proximal hamstring avulsions when the studies of this thesis was performed but was frequently used[4, 32, 33]. However, the LEFS is a responsive tool with good construct validity for assessing functional status during recovery in individuals with lower extremity musculoskeletal conditions[36].

The measurement properties of the LEFS in proximal hamstring avulsion patients at a minimum one-year follow-up has been presented to be: mean 72.77 (SD ±6.55) for patients who were surgically treated and 69.53 (SD ±4.04) for those who were nonsurgically treated[4]. In a review of the utility of the LEFS in cohorts of patients with lower limb injuries, further measures of responsiveness and reliability were presented: a minimal detectable change (MDC) of 8.1-15.3, a minimal clinically important difference (MCID) of 9 points, a minimal detectable, and good test-retest reliability (ICC 0.85-0.99)[50]. The ceiling effect at a mean of 28 months after injury in proximal hamstring patient has been presented to be 16%[28]. The LEFS was translated to Swedish by our research group[48] using the guidelines of Guillemin et al.[49]. The LEFS questionnaire is presented in the appendecis.

3.2.1.3 Proximal Hamstring Injury Questionnaire (PHIQ)

The Proximal Hamstring Injury Questionnaire (PHIQ) is a survey questionnaire developed for proximal hamstring avulsion patients[51]. The items does not provide a score but are rather questions concerning different aspects of recovery. The questionnaire includes items that are related to pain during different activities, the estimated grade of recovery and strength, the physical activity level at follow-up and perceived neurological symptoms.

No formal validity or reliability study has been performed for the PHIQ, but Green et al. studied overlapping questions between different questionnaires used for the follow-up of proximal hamstring avulsions and found that questions about neurological symptoms and self-reported recovery and strength were not included in the PHAT nor in the LEFS[32]. Some of these questions from the PHIQ are used in studies 1, 2 and 4 of this thesis. The PHIQ was translated to Swedish by our research group[48] using the guidelines of Guillemin et al.[49].

3.2.1.4 Patient specific functional scale (PSFS)

The PFSF is a questionnaire that is used to assess the most relevant problems experienced by the individual patient. Patient preference perspectives on subjective outcomes are provided[52, 53]. It is a short interview questionnaire in which the clinician/physiotherapists ask the patient to state three activities that they are unable to perform or are experiencing difficulty doing. The patient then indicates, on a scale, how difficult the activity is, where 0 represents being unable to perform the activity and 10 represents being able to perform the activity as before (minimum score 0, maximum 30).

The Swedish version of the PSFS has been validated in studies of the upper extremities, and the scale has good construct validity but has floor and ceiling effects[54, 55]. To my knowledge, the PFSF questionnaire was not used prior to this thesis to assess the recovery of proximal hamstring avulsion patients. Initially, the PSFS was produced to monitor the progress of rehabilitation, but in this thesis, we only used it at follow-up. The PSFS questionnaire is presented in the appendix.

3.2.1.5 International Physical Activity Questionnaire Short Form (IPAQ-short)

The IPAQ-short is a questionnaire that assess physical activity[56]. It consists of three types of different activities levels (walking, moderate and vigorous) undertaken in the four domains (leisure, domestic, work and transport related

physical activity). In the IPAQ short the total number of Metabolic Equivalent of Task (MET) minutes in activity are calculated as follows;

(minutes walking x 3.3 x number of walking days during 1 week)+(minutes in moderate activity level x 4 x days during 1 week with any moderate activity)+(minutes in vigorous-intensity activity x 8.0 x number of days with any vigorous-intensity activity).

It has acceptable measurement properties with fair to moderate criterion validity and high test-retest reliability[56]. I have not found any paper in which the researchers use the IPAQ-short for the assessment of proximal hamstring avulsion patients.

3.2.1.6 International Classification of Functioning, Disability and Health (ICF)

The International Classification of Functioning, Disability and Health, known more commonly as the ICF, is a classification of health and health-related domains. The ICF is the World Health Organization (WHO) framework for measuring health and disability at both the individual and population levels[57]. This framework was used in Study 2 to assess overlapping activities among the PSFS, PHAT and LEFS.

3.2.2 Physical Performance-based Tests

3.2.2.1 Single leg-hop and triple-hop for distance assessment

The aims of the single-leg hop and triple hop for distance tests are to assess function of an injured limb at follow-up. In the single-leg hop test, patients are instructed to start jumping and land, as far and as balanced and controlled as possible, on the same leg. In the triple hop, the patient is asked to jump three consecutive times on the same leg and land as controlled and balanced as possible. The total jumping distance from toe to toe or toe to heel is measured in metres and centimetres (Figure 13).

These hop tests were initially created and used for follow-up assessment of patients with anterior cruciate ligament injuries[58, 59]. The measurement properties from the first studies of these tests were inconsistent, the content validity was good but the construct validity was poor[58, 59]. Later, there are different reports of construct validity that is good, and test-retest reliability has been confirmed by high ICC values; however, other papers have reported conflicting conclusions regarding the measurement properties of hop tests[60, 61].

These tests have also been performed in many previous studies evaluating the post-treatment condition of proximal hamstring avulsion patients[7, 21, 29, 62]. By summarizing the results from those papers, the results of the hop tests in terms of construct and content validity appear to be inconsistent also in this patient population[7, 21, 29, 62].

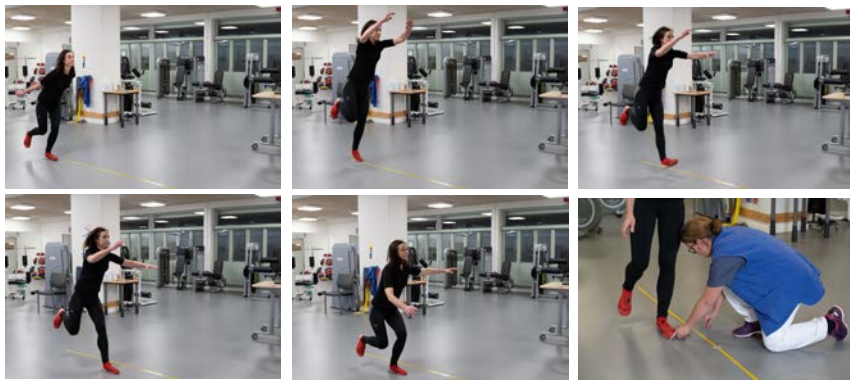


Figure 13: Single-leg hop test. Jump from one point to the farthest point possible on one leg. The distance was measured in centimeters from the start line to the point of the balanced landing. Single-leg hop tests were utilized in Studies 2 and 4 of the thesis. In these studies, the landing was measured at the heel not at the toe, as indicated in the photo.

3.2.2.2 Single step test

The single step test is a test in which the patient imitates the complex movement of walking up or down stairs and was initially constructed for the evaluation of knee arthritis patients[63]. In this test, patients are instructed to stand on the tested leg on a block (15 cm) while wearing an orthosis to keep the knee fully extended on the nontested leg. The patient is instructed to perform 20 squats as fast as possible, and the total time in seconds is recorded (Figure 14).

The interrater reliability has been shown to be excellent, the content and construct validity moderate, and the responsiveness good[63, 64].

To my knowledge, there are no other cohorts of patients with proximal hamstring avulsions that have been assessed with this test. Since the other functional tests used in the thesis are the more demanding hop tests, which were initially developed for younger anterior cruciate ligament patients, we believe that the single step test, developed for older knee arthritis patients, could be suitable as a complementary functional test for middle-aged proximal hamstring avulsion patients.

Figure 14: A modified version of the single step test is performed and used in studies 2 and 4 of this thesis. An orthosis is placed on the nontested knee in extension. The foot of the tested leg was placed on a box, and the time in seconds needed to perform the 20 squats was recorded. The height of the box was 15 cm (in the photo, the box was 20 cm) when this test was utilized in the studies of this thesis.

3.2.2.3 Isokinetic strength

Isokinetic strength is assessed with computer-based dynamometers. The dynamometer can be set to different torques, and ability of the patient to resist and contract against the torque of the machine is collected in Newtons (N) or the total workload in Joules (J) (Figure 15).

Isokinetic strength measurements have been considered the 'gold standard' for muscle strength measurements[65, 66]. For anatomical reasons, it is not possible to isolate hamstring muscles with such dynamometers, as agonists can contribute to both knee flexion and hip extension[67], which makes the content validity difficult to assess without using e.g., EMG.

However, such machines have been shown to have high interrater reliability in general[65, 66], and when testing the hamstrings of healthy individuals, the reliability of the ICCs and SEMs are good[68].

Isokinetic dynamometers have also been said to be the most effective assessment for evaluating strength after proximal hamstring avulsions[33, 69]. The content validity in terms of the peak torque of the injured leg compared to the uninjured leg in either knee flexion or hip extension after proximal hamstring avulsion has been reported to be 85%-91%[7, 21, 70, 71]

Figure 15: Isokinetic strength test measured with a Biodex Pro System 4 dynamometer (Biodex Medical Systems, Inc.). Patients can be positioned following a predetermined protocol (A), allowing for individual adjustment of, e.g., range of motion. When the dynamometer extends the leg, the patient should try to resist this force (C), and when the dynamometer allows flexion, the patient should try to contract against the by the dynamometer applied resistance (E). Different speeds of the dynamometer can be applied, and the peak torque (N) or total workload joule (J) can be recorded. The isokinetic strength test was used in Studies 2 and 3.

3.2.2.4 Isometric strength

For the assessment of isometric strength, handheld dynamometers can be used (Figure 16). Dynamometers can be positioned and used in many different ways. The dynamometer gives the maximum isometric force in Newtons (N).

Such dynamometers have previously been used for evaluating strength in proximal hamstring avulsion patients and for follow-up after treatment of many other conditions, and the reliability of these devices has been good (ICC 0.8-0.91), with varying results for hamstring muscle strength[7, 9, 62, 72].

Figure 16: Isometric strength tested with a handheld dynamometer (microFET 2; Hoggan Health Industries). Handheld dynamometers can be used in many different settings. In Studies 2 and 4, the patient was placed in the supine position with the heel on a 15 cm box and the knee in 15 degrees of flexion. The patient was instructed to apply a maximum contracting force at the dynamometer for 1-3 seconds. The maximum force in N was recorded.

3.2.3 Magnetic resonance imaging (MRI)

3.2.3.1 T1 weighted images

T1-weighted images reveals excellent contrast between muscle and fat tissue, fat bright and muscle dark. The advantage of T1-weighted sequences is that they are feasible and standard for all MRI scans. However, these methods cannot provide computer-based quantitative measurements of fatty infiltration, but fatty infiltration can be quantified with visual assessment on T1-weighted sequences[73].

3.2.3.2 Muscle volume via the slice-by-slice technique

The gold standard for muscle volume measurements has been stated to be slice-by-slice segmentation, which is time-consuming[74]. The cross-sectional area (CSA) is outlined for every segment in the image (Figure 17), and the software then adds the area of all the segments to calculate the muscle volume[74]. It is often performed on T1-weighted images[73]. To my knowledge there is no prior study that has assessed the muscle volume after proximal hamstring avulsions.

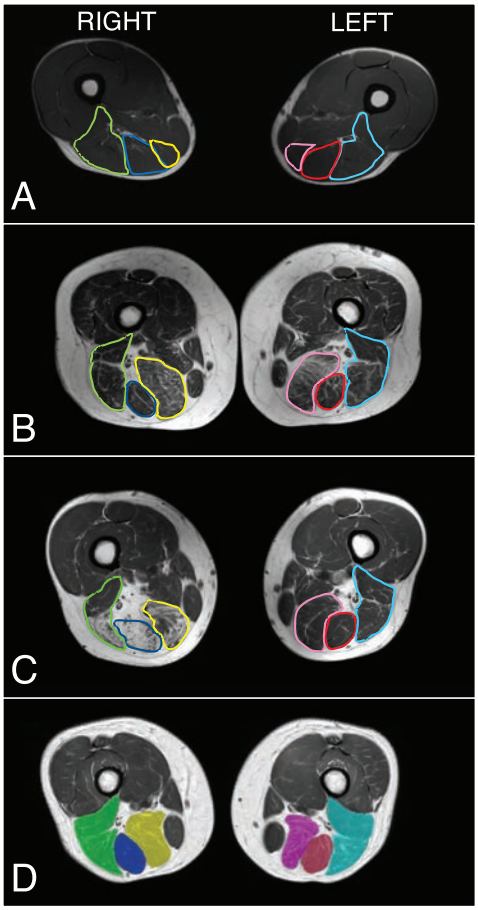
3.2.3.3 The Goutallier Classification of Fatty Infiltration

Fatty infiltration can be assessed according to the Goutallier grading system[75]. This classification was developed for assessing rotator cuff muscles using CT but was later used with T1-weighted MRI images[76]. In this system, cross-sectional images are graded on an ordinal scale as follows: 0) normal muscle, 1) some fatty streaks, 2) fat < muscle, 3) fat = muscle and 4) fat > muscle (Figure 17).

This classification system has not been validated for follow-up after proximal hamstring avulsions but has been previously used to examine patients after this injury[9, 26].

3.2.3.4 Dixon method for assessment of fatty infiltration and muscle volume

The Dixon method is used for quantifying fat and water with MRI[77]. The technique involves chemical shifts between water and fat. It delivers different images including the fat fraction, i.e. the percentage of fat in a voxel[73]. Using the Dixon method in combination with a semiautomatic segmentation method muscle volume and fat fraction can be quantified [78, 79].



Patient A
Goutallier 0 in all muscles of the segment.

Patient B
Right leg:
Goutallier 1 in BF and SemiT
Goutallier 3 in SemiM
Left leg:
Goutallier 2 in BF and SemiT
Goutallier 3 in SemiM

Patient C
Right leg:
Goutallier 0 in BF short head
Goutallier 4 in all other muscles
Left leg:
Goutallier 1 in all muscles of the segment.

Patient D
Muscles filled for volume quantification

BF Dx | BF Sin
SemiT Dx | SemiT Sin
SemiM Dx | SemiM Sin

Figure 17: Panels A, B and C show different grades of Goutallier classification in the different hamstring muscles. Panel D shows one cross-sectional area assessment of all hamstring muscles. To determine the muscle volume via the slice-by-slice technique, all slices of the muscle must be outlined. This process is very time consuming in long muscles, such as the hamstrings. These two methods were used in Study 3.

3.2.3.5 Tendon attachment

Tendon attachment after proximal hamstring avulsion has been assessed and reported in some papers. Chahal et al. reported that all examined patients had tendons attached at follow-up, and van der Made et al. reported that 95% of the surgically treated patients and 52% of the nonsurgically treated patients had tendons in continuity with the ischial tuberosity at follow-up[9, 26].

3.3 Statistical methods relevant to the thesis

In all studies of the thesis, a p value < 0.05 was considered to indicate statistical significance in all analyses, and the two-sided confidence intervals were set to 95%. The cut-off values for the strength of correlations were r=±0≤0.3 (negligible), r=±0.3<0.5 (low), r=±.5≤0.7 (moderate) and r=±0.7<1 (high)[80].

3.3.1 Parametric test

3.3.1.1 Student's t test

The student's test can be used to compare two groups when data are normally distributed. It tests the probability of the two samples coming from the same population and having the same mean value[81].

3.3.2 Non-parametric tests

3.3.2.1 Kruskal Wallis test

The Kruskal Wallis test is a nonparametric test testing if the samples have the same distribution, and is an extension of the two-sample Wilcoxon test[82].

3.3.2.2 Fisher's exakt test

The Fisher's exakt test is used to compare frequency distribution between two independent groups when the sample size is relatively small, but it can also be utilized comparing large samples[82].

3.3.2.3 Wilcoxon signed-rank test

The Wilcoxon singed-rank test computes whether there is a difference in the median score between two paired populations[82].

3.3.3 Intraclass correlation coefficient (ICC)

The intraclass correlation coefficient (ICC), can be used to compute interrater reliability estimates and is used as reliability parameter for continuous measurements. The ICC provides both the degree of correlation and agreement between measurements since it is an estimate of the interindividual variation divided by the entire variation in the sample and is expressed as a ratio of 0-1[42].

3.3.4 Spearman correlation coefficient

The Spearman correlation coefficient can be used to estimate the strength of an association between two nonnormally distributed variables. It uses the statistical dependence between the rankings of two variables. It can vary from -1 to 1.

3.3.5 Linear regression

Linear regression is used to investigate the relationship between the outcome variable (y) and some predictor variables (x), when assuming linearity in their relationship. Based on this assumptions, the variable y can be predicted by the combinations of the x variables and the regression coefficients[81].

3.3.6 Imputation

Imputation is a tool used to handle missing data; it can be used in different settings. Two of them are presented below.

3.3.6.1 Predictive mean matching (PMM)

When data are missing in an individual instrument e.g., if one answer of a questionnaire is missing, the data can be imputed using PMM. In the PMM all other items not missing within the same questionnaire are used to predict missing items.

3.3.6.2 Multiple Imputation by Chained Equations (MICE)

The MICE uses more complex methods of imputing missing data. For example, when an entire follow-up event is missing. The MICE imputation protocol is based on variables for predicting outcome of the missing event variable. The imputation is performed multiple times and considers the variability of the prediction for the missing data.

3.3.7 Propensity score

A propensity score gives the conditional probability to being exposed or unexposed given a certain set of variables. It is normally used in cohort studies to adjust for the selection bias in different treatment arms. Certain confounders are added in the model to get the probability of being exposed or unexposed, based on these covariates. Thereafter, the propensity scores obtained as a weight for each patient are used in the analysis of association of the exposure and outcome. The weight is based on how likely the patient is to be part of the exposed or unexposed group, i.e., how likely it is to be in the exposed group based on the confounders in the model. To even out this probability (weight) of

being exposed, given the covariates, the inverse probability (1/the probability weight) can be used in the model.

3.3.8 Intention to treat (ITT) analysis

The intention to treat analysis (ITT), is the 'gold standard' analysis method for RCT's. In a classical RCT design, a conservative estimate of the intervention effect is provided. The analysis consists of all participants who provides consent and are randomized into a study irrespective of the treatment received. The advantage with this method is that due to randomization the groups should not be affected by systematic selection or confounding bias. Since all patients will not follow their randomized treatment arm, some might not be compliant, some might cross over, etc., the ITT analysis will push towards the null effect.

However, owing to the noninferiority design this analysis may provide an anticonservative estimate. Since the two groups tends to be more similar in the ITT analysis it is more likely to give the result that the investigated group is not noninferior to the comparator.

3.3.9 Per-protocol (PP) analysis

The per-protocol (PP) analysis includes only those participants who completed the protocol for the treatment that they were originally allocated to by the randomization. The participants who crossover or drop out are excluded from this analysis. Per-protocol analyses are biased since they rely on post randomization events; such events can exclude participants from the analysis (e.g., a frequent treatment complication leading to many dropouts in only one of the allocation arms), which may lead to biased results. To adjust for the bias introduced by post randomization events usually inverse probability weighting using a propensity score model is used.

3.3.10 As-treated (AS) analysis

In the as-treated (AS) analysis, participants are analysed according to the treatment that they received, despite treatment allocation in the randomization procedure. This as-treated population is therefore based on post randomization events; which may lead to biased results; e.g., if more patients from treatment A than B crossover because of side effects, the information about the side effects might be lost in this analysis.

Table 3

Overview of the studies included in thesis				
	Study 1	**Study 2**	**Study 3**	**Study 4**
Design	Retrospective cohort study	Cross-sectional cohort study	Cross-sectional cohort study	International multi-centre, preference tolerant, noninferiority, randomized controlled trial
Setting	Orthopaedic department at Danderyd Hospital 2015	Orthopaedic department at Danderyd Hospital 2018-2019	Orthopaedic department at Danderyd Hospital 2018-2019	8 orthopaedic departments in Sweden and 2 in Norway 2017-2022
Participants	Patients diagnosed or treated for PHA 2007-2013	Patients diagnosed or treated for PHA 2007-2016	Patients diagnosed or treated for PHA 2007-2016	Patients 30-70 years with acute PHA 2017-2020
Intervention	Not applicable	Not applicable	Not applicable	Nonsurgical treatment
Exposure	Surgical treatment	Not applicable	Injured leg	Nonsurgical treatment
Control group	Nonsurgical treated	Not applicable	Noninjured leg	Surgical treatment
Data sources	Patient charts, questionnaires	Patient's charts, questionnaires, physical performance-based tests	Patient's charts, MRI, Biodex	Patient's charts, questionnaires, physical performance-based tests, MRI
Main outcomes	LEFS, PHIQ	LEFS, PHAT, PSFS, single leg-hop tests, single step test, isometric and isokinetic strength	Muscle volume, fatty infiltration, isokinetic strength, tendon attachment	PHAT, LEFS, IPAQ-SF, single-leg hop, single step test, isometric strength, fatty infiltration, muscle volume
Main statistical methods	Fischers exakt test, student t test, linear regression	ICC, spearman correlation coefficient, student t test, Kruskal Wallis test	ICC, spearman correlation, student t test, Wilcoxon signed ranked test	ITT, PP, AS, RCT+OBS with propensity score
Number in analysis (op/nonop)	47 (33/14)	50 (37/13)	48 (36/12)	Randomized 119 (58/61) OBS Cohort 97 (44/53)

3.4 Patients and methods

3.4.1 Study 1

3.4.1.1 Design and setting

A retrospective cohort study conducted at Danderyd Hospital in 2015.

3.4.1.2 Participants

Patients both surgically and nonsurgically treated at Danderyd Hospital were assessed for eligibility if they had been diagnosed or treated for proximal hamstring avulsions between 2007 and 2013 at the department. Patients who had previous other surgical procedures on the same leg after the proximal hamstring avulsion but before the follow-up, an MRI showing no tendon retraction or bony avulsion, and a previous proximal hamstring injury on the same leg were excluded.

3.4.1.3 Outcomes

The primary outcome was the LEFS, and the secondary outcomes were some of the questions from the PHIQ-questionnaire.

3.4.1.4 Data collection

Baseline characteristics were collected from patient records. At follow-up the questionnaires were posted to patients and the completed forms were sent back with prepaid envelopes.

3.4.1.5 Exposure

The exposure was surgical treatment. The outcomes were contrasted between surgical and nonsurgical treatment groups.

3.4.1.6 Statistics

The Fischer's exakt test was used for comparisons of categorical variables, and Student's t test was used for continuous variables. A linear regression analysis was performed to adjust for confounders that could affect the primary outcome. The regression model included the exposure variable and confounders (surgery/nonsurgery group, age, sex, American Society of Anaesthesiologists (ASA) classification and tendon retraction ≥2 cm). No power analysis was performed prior to the study. The statistical analyses were performed using SPSS V.23 for Mac.

3.4.2 Study 2

3.4.2.1 Design and setting

A cross-sectional cohort study conducted at Danderyd Hospital, 2018-2019.

3.4.2.2 Participants

Patients diagnosed or treated for proximal hamstring avulsion at Danderyd Hospital between year 2007 and 2016 were screened for eligibility from the surgical planning system or from the radiology administrative system, with a radiology report describing at least the conjoint tendon or the semimembranosus tendon avulsed from the ischial tuberosity. Patients were excluded if they had bilateral proximal hamstring avulsion, other major lower limb extremity injuries, diseases with sequelae unrelated to the proximal hamstring avulsion, or a severe medical condition at the time of follow-up. Patients were included despite treatment allocation.

3.4.2.3 Outcomes

The primary outcomes of Study 2 were the PHAT and LEFS scores and their correlation. The secondary outcomes were physical performance-based test results (isokinetic strength, isometric strength, single leg hop and triple hop for distance test and single step test) and correlations with the PHAT and LEFS scores and questions from the Patient Specific Functional Scale (PSFS) questionnaire.

3.4.2.4 Data collection

Medical records, including patient characteristics, injury mechanism, MRI data at diagnosis, time to treatment, treatment modality and adverse events, were reviewed. For reliability analysis, patients were invited to attend two study visits, complete the questionnaires, and complete the physical performance-based tests at both visits. At each visit, the PHAT and LEFS were collected. Thereafter, an experienced physiotherapist collected the PSFS data and then conducted the physical performance-based tests.

3.4.2.5 Statistics

To test test-retest reliability, the ICC was used, and the measurement error was assessed with the SEM. The Spearman correlation coefficient was calculated separately for correlations of all variables with the two PROMs used. Activities mentioned by one or more patients in PSFS from both visits were collected. The

different activities mentioned in the PSFS and the activities within the PHAT and LEFS were coded with the ICF, and the number of activities with corresponding ICFs between the PHAT, LEFS, and PSFS were counted. The Limb Symmetry Index (LSI) (test result injured leg/test result uninjured leg) was calculated for the physical performance-based tests.

Student's t test was used to compare the means, of injured versus uninjured legs, for normally distributed variables, and the Kruskal-Wallis test was used for nonnormally distributed variables. A complete case analysis was used in all the analyses.

In the power calculation performed prior to the study, which was based on Spearman correlations between the PHAT and LEFS as low as r=0.5, and with a p-value (2-tailed) of .025 and a power of 90%, a total sample size of 44 patients was needed. All the statistical analysis were performed using IBM SPSS Statistics for Mac Version 25.

3.4.3 Study 3

3.4.3.1 Design and setting

A cross-sectional cohort study conducted at Danderyd Hospital, 2018-2019.

3.4.3.2 Participants

Study 3 was performed on the same participants and on the same occasion as Study 2. Patients from Study 2 were excluded if they could not or did not want to undergo MRI.

3.4.3.3 Outcomes

In Study 3, the main outcomes were the fatty infiltration and muscle atrophy in the injured leg compared to the uninjured leg. The secondary outcomes were the correlation of fatty infiltration with muscle atrophy and the correlation with isokinetic strength. Furthermore, if the tendons were attached to their footprint at follow-up.

3.4.3.4 Data collection

Muscle quality (fatty infiltration and atrophy) was assessed via MRI (Philips Ingenia 3 T system) with a slice thickness of 5 mm in T1-weighted images. Fatty infiltration was classified according to the Goutallier grading system[75] using the open-source software Horos© version 3.3.6. A modification to the Goutallier

classification was performed by visually inspecting and grading each (biceps femoris, semitendinosus and semimembranosus) entire hamstring muscle separately and then adding the score of all three muscles to compute the modified Goutallier score of the entire hamstring muscle complex. To obtain test-retest reliability measures, the classification was performed twice for every muscle, once in June 2020 and once in September 2021. When there was a discrepancy between the two assessments the final decision of the Goutallier grade was made by a consultant in radiology.

Muscle volume was measured in cm^3 by slice-by-slice segmentation[74]. The open-source software ITK snap, version 3.8.0 (itksnap.org), was used for segmentation[83]. For reliability measure, the semitendinosus muscle was outlined twice for every second patient once in June 2020 and once in September 2021. Agreement was defined as repeated measures within 1 cm^3.

Tendon attachment was assessed on PD-weighted SPAIR sequencies. The attachment was grouped into total attachment (no high signal), partial attachment (some high signal), and no attachment (no continuity).

The isokinetic strength data were obtained from Study 2.

3.4.3.5 Statistics

Test-retest reliability for imaging outcome measures was calculated using the ICC. For comparisons of the means of the injured vs. uninjured legs, the paired t test were used for normally distributed variables, and the Wilcoxon signed rank test was for nonnormally distributed variables. Correlations were analysed with the Spearman correlation coefficient. A complete case analysis was used in all the analyses. No power calculation was performed prior to the study.

All the statistical analyses were performed using IBM SPSS Statistics for Mac Version 28.

3.4.4 Study 4

3.4.4.1 Design and setting

Study 4 was an international multicenter noninferiority preference tolerant randomized controlled trial. Ten study sites participated in the study, eight in Sweden (Danderyd Hospital, Uppsala Akademiska University Hospital, Linköping University Hospital, Umeå University Hospital, Skåne University Hospital, Södersjukhuset, Östersund Hospital, Örebro University Hospital) and two in

Norway (Oslo University Hospital and Haulekand University Hospital). The study was conducted between 2017 and 2022.

3.4.4.2 Participants

Patients were eligible if they were diagnosed or referred to the orthopaedic department at 1 of the 10 hospitals within four weeks after the injury, being at least moderately active, were between 30 and 70 years of age, had clinical examination and MRI result supporting proximal hamstring avulsion with at least two tendons avulsed from the ischial tuberosity.

Patients were excluded if they had moderate or severe diabetes; liver, pulmonary, kidney, psychiatric or heart disease; severe obesity (Body Mass Index (BMI) >35), or alcohol or substance abuse that significantly increased the risk of complications after surgical treatment. Additionally, high energy injury injuries or combinations of injuries affecting the lower extremity with sequelae, e.g., car accidents, were excluded.

After providing written informed consent, the study patients were randomized, by the assigned centralized study nurse, to either undergo surgical reinsertion of the tendons to the footprint followed by rehabilitation or receive nonsurgical treatment with rehabilitation. All patients followed the same standardized rehabilitation protocol.

3.4.4.3 Observational cohort

Patients who were eligible for inclusion in the trial, but where the patient or treating physicians' equipoise to treatment could not be reached, were invited to participate in the preference tolerant parallel observational follow-up cohort with identical treatment options and follow-up.

3.4.4.4 Outcome

The primary outcome measure was the PHAT score at 24 months. With the noninferiority design, the primary outcome must be a predefined noninferiority margin. We decided to use half of the SD of PHAT which at the time was presented to be ~16–21[34, 84] and which was also less than the presented MDC of PHAT 16.4[34]. With this information about the PHAT we decided the noninferiority margin to be, with a lower limit of the two-sided 95% confidence interval at -10 PHAT points between the surgically and nonsurgically treated groups.

The secondary outcomes were the PHAT, LEFS, IPAQ short, and physical performance-based test results at 24 months and how these outcomes improved during the follow-up. Additionally, questions regarding e.g., return to sport and adverse events were included as secondary outcomes during the study, together with MRI findings at the 24-month follow-up.

3.4.4.5 Data collection

The data were collected at the time of injury and 3 months, 6 months, 12 months and 24 months after injury at each study site (Figure *26*).

3.4.4.6 Statistics

We estimated that 60 patients in each randomized group was needed to achieve 74–90% power (depending on assumed standard deviation of the PHAT score) for demonstrating noninferiority when assuming a dropout rate of 10%, a standard deviation of the PHAT score in the range of 16-20 and using a 2.5% one-sided alpha. The PHAT at 24 months was contrasted between the treatment options with a linear regression model adjusted for age, sex, and degree of tendon retraction. If the lower boundary for the two-sided 95% confidence interval in the absolute difference of the PHAT score of nonsurgically and surgically treated groups was greater than -10 points, the nonsurgical treatment would be deemed to be noninferior. For secondary outcomes, two-sided 95% confidence intervals were reported.

The REDCap software[85] randomization tool was used to randomize the patients. Allocation tables with a random block size (2-6), stratified by study site were created by the statistician assigned to the study. Participants, physicians, and physical therapists were not blinded. Investigators who performed the statistical analysis were blinded during the analyses, and concealment was broken when the analysis was complete.

The primary analysis was the intention to treat (ITT) according to the randomized treatment allocation. Per-protocol (PP) and As-treated (AS) analyses were also performed. Patients were considered as treatment crossovers if the randomly assigned treatment was changed within 3 months.

To analyze the randomized cohort and observational cohorts together, propensity scores based on age, sex, study site, IPAQ-SF, and the degree of tendon retraction were calculated.

Great effort was taken to avoid missing data to ensure that the level of missing data and loss to follow-up were minimal. Missing data on individual instrument (e.g., PHAT) questions was imputed with predictive mean marching. Missed follow-up visits at specific follow-up timepoints were handled using multiple imputation by chained equations (MICE).

The global COVID-19 pandemic occurred during the execution of this trial and resulted in rescheduled and cancelled follow-up visits. Therefore, a special timeframe was used for every visit. Since the dates of PROM reporting, MRI, and physiotherapy visits differed, the data of the primary endpoint (PHAT) were used to set the time of the visit.

The analyses were performed using the software RStudio, R version 4.2.3 (2023-03-15 ucrt).

3.5 Ethical considerations

All four studies were approved by the appropriate ethical board and conducted according to the guidelines of the Helsinki Declaration. All patients in the four studies provided their written informed consent prior to participation.

3.5.1 Ethical considerations of study 1

There is always a violation of integrity when performing a medical record review. However, the extent of this violation was regarded as relatively small. Participation was voluntary, and all patients provided written informed consent before the review was performed. Ethical approval reference number 2015/622-31 EPN Stockholm.

3.5.2 Ethical considerations for Studies 2 and 3

Since this profound assessment of patients with proximal hamstring avulsions was not performed prior to these studies, the time spent, potential risk of reinjury during physical performance-based tests and potential violation of unexpected MRI findings were considered relatively small compared to the expected new findings and improved understanding of outcomes after treatment with proximal hamstring avulsions. Ethical approval reference number 2018/1260-31 EPN Stockholm.

3.5.3 Ethical considerations of Study 4

Surgical vs. nonsurgical treatment were two options considered acceptable in the catchment area of the study (Sweden and Norway). This approach made it ethically acceptable to randomize patients to one of the treatments.

This study contributes novel evidence-based data on how to treat future patients with proximal hamstring avulsions; this makes the potential risks for the patients who participated in the study worthwhile with respect to the new findings and more scientifically based for guiding treatment decisions for future patients. Ethical approval was granted by the Ethical Committee of Uppsala University DNR: 2017-170 with amendment 2019 with DNR 2019-00186 and the Norwegian Regional Ethical Committee (REC: 2017/1911).

4 Results

The baseline characteristics of the patients included in the studies are presented in Table *4*.

Table 4

Baseline characteristics of patients in thesis*					
		Study 1	**Study 2**	**Study 3**	**Study 4**
Sex, female, n (%)		28 (60)	24 (48)	22(46)	129 (60)
Age at injury, years, mean (SD)		51 (9)	50.9 (9.8)	50.8(9.7)	53 (8.5)
Activity at time of injury					
	Sporting injury, n (%)	28 (60)	31 (62)	30 (63)	117 (54)
	Vehicle accident, n (%)		2 (4)	2 (4)	9 (4)
	Slip or fall, n (%)	16 (34)	11 (22)	10 (21)	68 (31)
	Other, n (%)	3 (6)	6 (12)	6 (12)	22(10)
MRI Finding at diagnosis					
	Incomplete, n (%)	9 (19)	8 (16)	8 (17)	4 (2)
	Complete <2cm, n (%)	3 (6)	5 (10)	5 (10)	205 (95)
	Complete >2cm, n (%)	35 (74)	37 (74)	35 (73)	
	Missing**				7 (3)
Time to treatment, days, mean (SD)		19 (15)	17.4 (13.7)	17.6 (14.1)	14 (6.8)
Treatment					
	Surgical, n (%)	33 (70)	37 (74)	36 (75)	102 (47)
	Non-surgical, n (%)	14 (30)	13 (26)	12 (25)	114 (53)
Follow up time, years, mean (SD)		3.9 (1.4)	5.5 (2.7)	5.6 (2.7)	2.0 (0.5)

*% are rounded and might not add up to 100%
** we were unable to retrieve images from study site

4.1 Study 1

4.1.1 Participants and baseline characteristics

After screening, a total of 47 patients were enrolled in the study; 33 were treated surgically, and 14 were treated nonsurgically (Figure 18). The mean age at the time of injury was 51 (SD±9) years; all except one (age 23 years) were between 34 and 68 years of age.

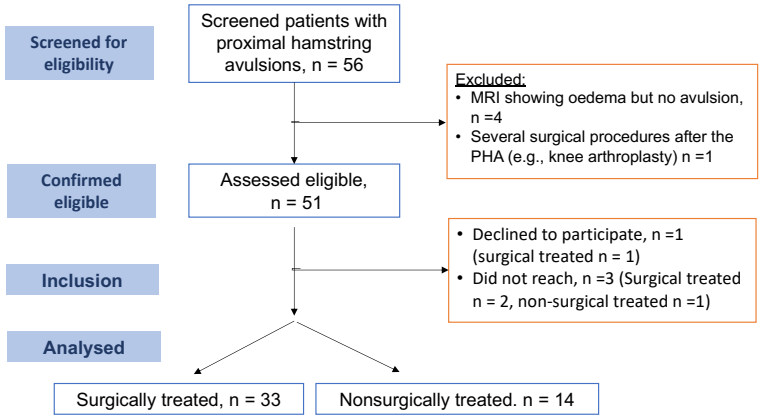


Figure 18: Flowchart of Study 1.

The baseline characteristics are presented in Table *4*. The only difference measured was that patients treated surgically had a greater proportion of tendons with a retraction ≥2 cm (p=0.025) than those not treated surgically. The surgeries were performed by four different surgeons. The rehabilitation protocol mostly used in the Stockholm area was the Askling protocol[20], and modifications of it. The patients were referred to different physiotherapists in the area for rehabilitation.

The medical records were screened for adverse events, and only surgically treated patients were found to have such events recorded. One patient suffered a postoperative pulmonary thrombosis, one patient had a wound infection, and one patient suffered from severe persistent pain postoperatively.

4.1.2 Main results

We could not find any difference in the LEFS score, the primary outcome, between the two treatments. The LEFS score was 72 (SD±16) in the nonsurgically treated group and 74 (SD±12) in the surgically treated group (p=0.80) (Figure 19).

When adjusting for confounders in a linear regression model, only increasing age was associated with a lower LEFS-score (B -0,5 [95% CI -0.9 to -0.0], p=0.037).

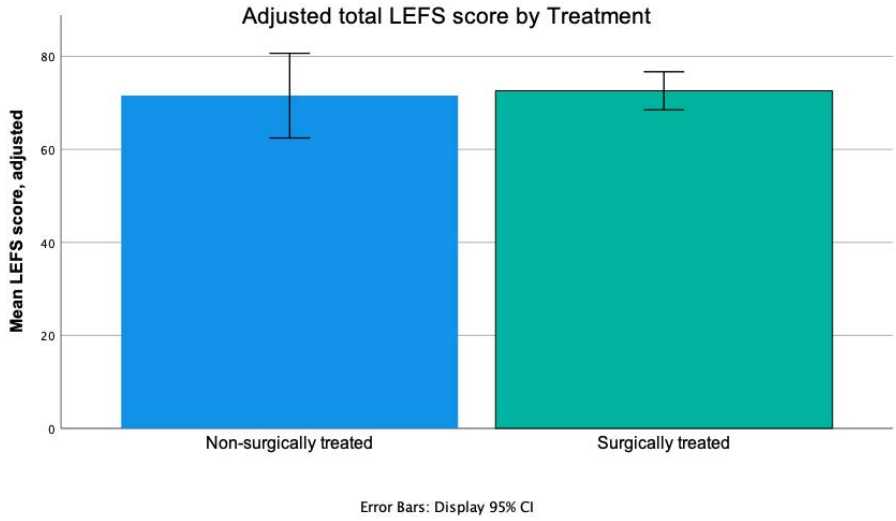


Figure 19: Adjusted mean total LEFS score at follow-up in the two treatment allocations. The regression model included age, sex, ASA classification and tendon retraction ≥2 cm.

The only between-group difference in outcome measures was that patients treated with surgery reported that they spent more time being physically active than the nonsurgically treated (p=0.02).

4.2 Study 2

4.2.1 Participants and baseline characteristics

We screened 103 patients and included 50 patients in the study (Figure 20). The baseline characteristics are presented in Table *4*. The mean age at follow-up was 57.2 years (±SD 10.2), 86% were satisfied with their treatment, 56% had returned to their preinjury activity level and, 20% experienced numbness or tingling in their affected leg at follow-up.

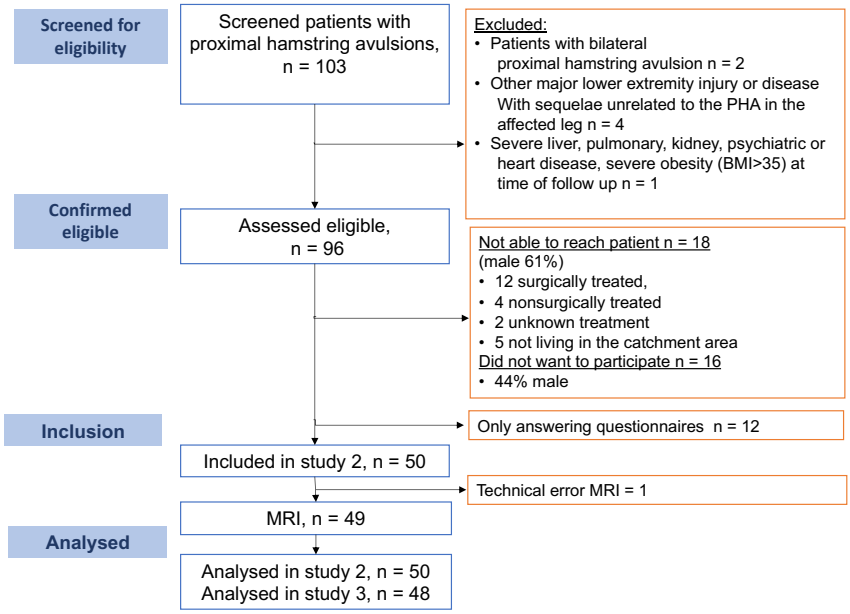


Figure 20: Flowchart of patients of Studies 2 and 3.

4.2.2 Main results

The mean PHAT score was 79.9 (±SD17.3), with 20% (10/50) scoring maximum score, and the mean LEFS score was 72.2 (±SD8.2), with 26.5% (13/49) scoring maximum score. The test-retest reliability measure of both the PHAT and LEFS was high (ICC=0.9).

In the questions from the PSFS questionnaire 28 patients (56%) mentioned a total of 31 activities they found difficult to perform due to their earlier proximal hamstring avulsion. The most common activity limitation was different types of running (mentioned by 16 patients) (Figure 21).

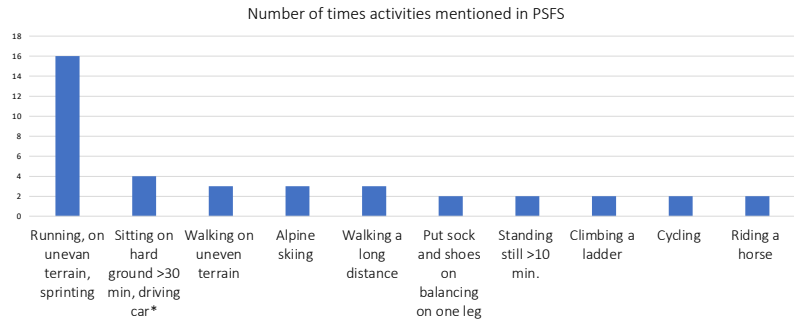


Figure 21: Number of activities mentioned in the PSFS questionnaire mentioned by at least one patient on at least one occasion. *Sitting on hard ground, 3 patients; driving car, 1 patient.

4.2.3 Secondary outcomes

4.2.3.1 Physical Performance-based Tests

Only the isokinetic strength test could discriminate between the injured and uninjured leg, and the Limb Symmetry Index (LSI) was 0.89, (95% CI=0.83 to 0.95) at follow-up. All the other tests used did not significantly differ in terms of performance results between the legs.

4.2.3.2 Correlations

The correlation between the PHAT and LEFS was strong (r=0.832, p<0.001) (Figure 22). There were weak statistically significant correlations of 3/12 performance-based outcome measures and the PHAT score, whereas the LEFS score was weakly but significantly correlated to 7/12 performance-based tests.

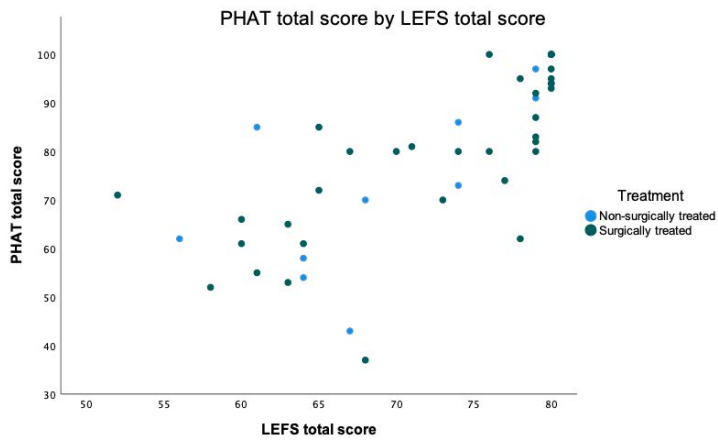


Figure 22: Scatterplot of the correlation between PHAT and LEFS scores. One dot can represent more than one subject scoring the same PHAT and LEFS.

4.3 Study 3

4.3.1 Participants and baseline characteristics

Forty-eight patients from the cohort provided data for the final analysis in Study 3 (Figure 20). The baseline characteristics are presented in Table *4*.

4.3.2 Main results

4.3.2.1 Fatty infiltration

We found a significantly greater fatty infiltration ($p<0.001$) in the injured hamstring muscles than in the uninjured hamstring muscles (Figure 23). The modified Goutallier classification exhibited high intrarater reproducibility, with an ICC of 0.82-0.91 for every hamstring muscle separatly.

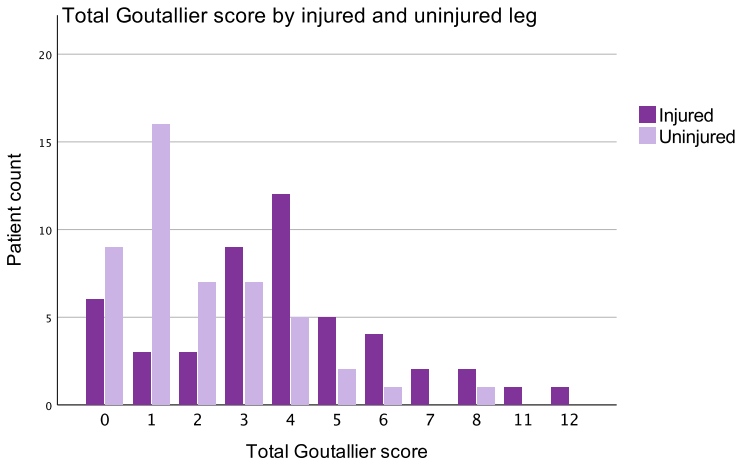


Figure 23: The distribution of modified Goutallier scores for injured and uninjured legs. More patients had higher scores (greater fatty infiltration) in the injured leg than in the uninjured leg.

4.3.2.2 Muscle volume

The mean total volume of the injured hamstring muscles was 1354 cm^3 (±SD 405), and that of the uninjured hamstrings muscle was 1506 cm^3 (±SD 480). The mean volume difference was -151.2 cm^3 (95% CI -215.0 to -87.4) and this was coherent with a mean deficit of 9% (±SD 11%) in the injured leg compared to the uninjured leg. The semimembranosus muscle had the greatest atrophy of the three hamstring muscles (Figure 24). The reproducibility of the muscle volume assessment was confirmed by an intrarater ICC of 0.99.

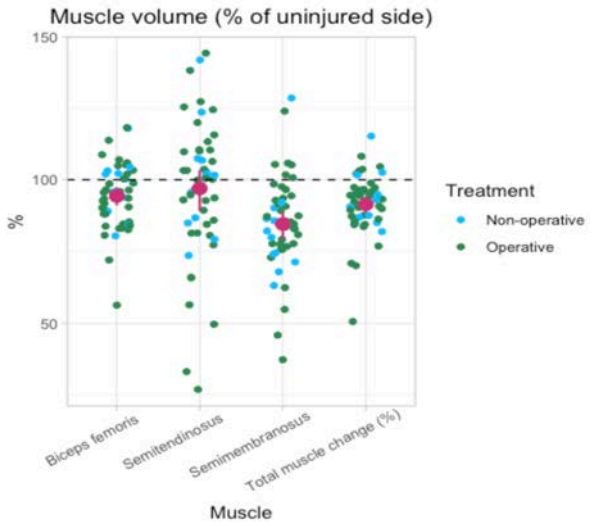


Figure 24: Distribution of hamstring muscle volume measured with the slice-by-slice technique in patients (surgically and nonsurgically treated) as a percentage of the volume of the hamstring muscles on the uninjured side. The total muscle volume and the three different muscles were measured separately. The purple dots represent the mean muscle volume loss and 95% CI.

4.3.3 Secondary outcomes

4.3.3.1 Tendon attachment

There were 29/48 of the patients that had their tendons completely attached to the origin, 13/48 patients had tendons partly attached, and 6/48 had tendons completely detached from the origin at follow-up (Figure 25).

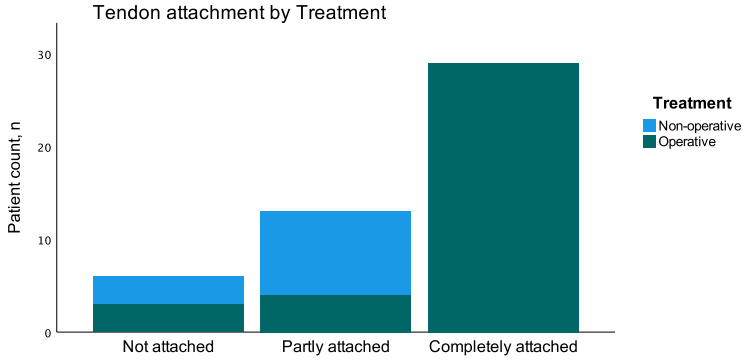


Figure 25: Tendon attachment at follow-up by treatment. The assessment was performed by one consultant in radiology on PD-weighted SPAIR images. The attachments were grouped into total attachment (no high signal), partial attachment (some high signal), and no attachment (no continuity).

4.3.3.2 Isokinetic strength

Forty-two patients were able to perform the peak torque test. The mean peak torque of the injured hamstrings was 60 Nm (±SD 24), whereas that of the uninjured hamstrings was 66 Nm (±SD 23) of the uninjured hamstrings. The mean deficit was 9% (±SD 22%, $p<0.008$).

4.3.3.3 Correlations

The correlations between muscle volume deficit of the injured and uninjured legs and the difference in total Goutallier score between injured and uninjured hamstrings and peak torque were low but statistically significant ($r=0.357-0.494$, $p<0.001-0.05$).

4.4 Study 4

4.4.1 Participants and baseline characteristics

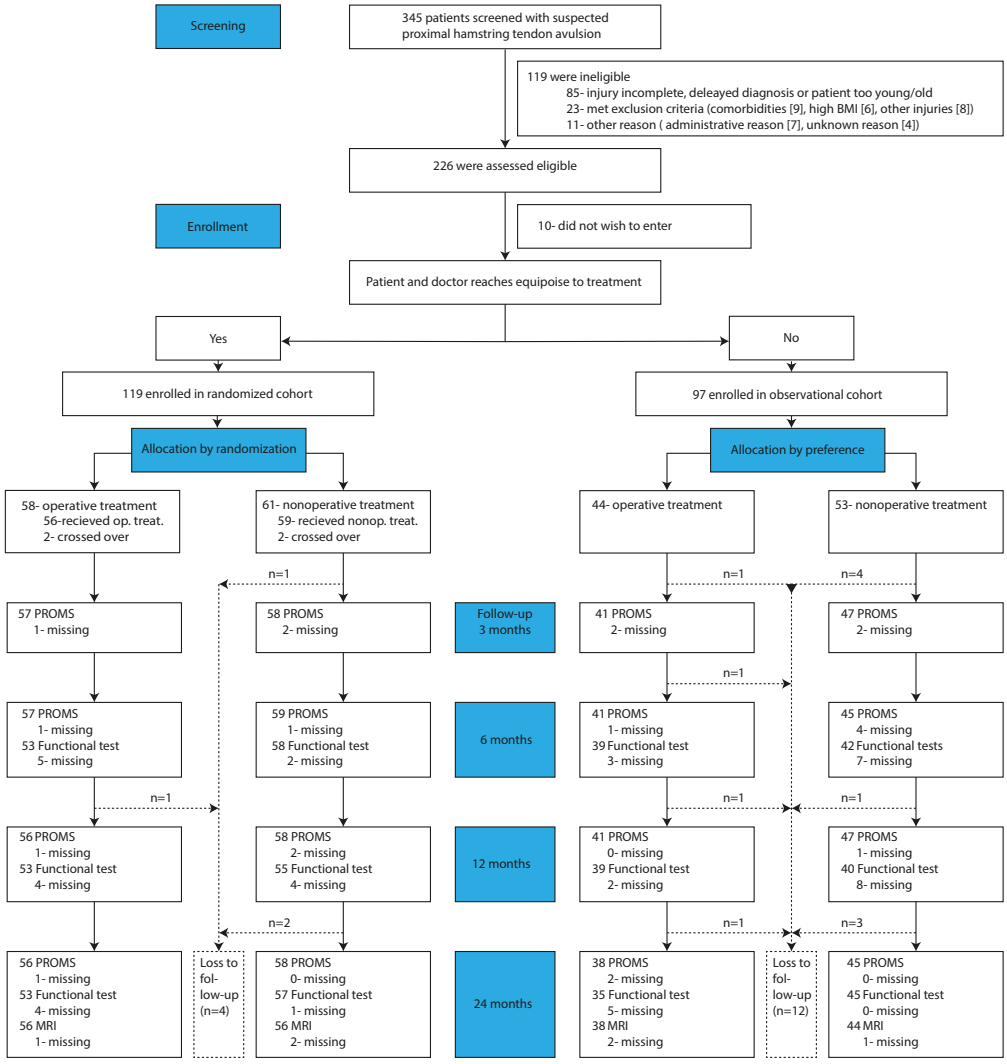


Figure 26: Flowchart of patients in Study 4. MRI denotes magnetic resonance imaging and PROMS is patient reported outcome measurements. Patients who were referred to one of the study sites with a suspected proximal hamstrings avulsion were screened. Inclusion criteria included age, time since injury and MRI verification of a complete avulsion of at least two tendons. The most common reasons for ineligibility were delayed diagnosis and incomplete tendon avulsions. Patients were enrolled in either the randomized cohort or in a parallel observational cohort. In the randomized cohort, two patients from each treatment group crossed over.

Patients were enrolled between October 2017 and July 2020, and the last patient was followed up in September 2022. In the randomized trial, 58 patients were randomized to surgery and 61 to nonsurgical treatment. There were two crossovers in each treatment arm. In the observational cohort, 44 patients were treated surgically and 53 were nonsurgically treated. At 24 months, 97% of the patients provided data for the analysis of the RCT, and 88% of the participants provided data for the observational cohort (Figure 26).

The mean age of the RCT study population was 53.2 (SD8.5) years. The sex distribution in the combined cohort (RCT + observational cohort) was 60% female and 40% male.

4.4.2 Main results

According to the primary intention-to-treat analyses, the difference between the nonsurgically treated group and surgically treated group was -2.1 points (95% CI -9.3 to 5.1; p=0.017 for noninferiority of the nonsurgical treatment (Figure 27).

The nonsurgical treatment was considered noninferior because the lower bound of the two-sided 95% confidence interval for the differences between the mean PHAT scores did not include the predefined noninferiority margin of -10 points. This was also true in the sensitivity analyses of PP and AS, as well as when using a propensity score adjusted combined analysis of the randomized and observational cohorts (Table 5).

The mean PHAT score decreased at the three-month visit in both the surgically and nonsurgically treated groups and gradually recovered at the 24-month visit (Figure 28).

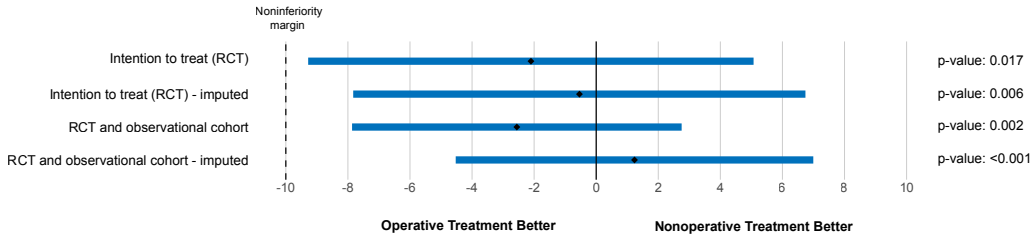


Figure 27: Intention-to-treat and imputed results of the primary outcome (difference in total PHAT score between operative and nonoperative treatment) for the randomized cohort and the randomized cohort combined with the observational cohort. Black diamonds indicate the absolute differences in PHAT score between the nonoperative and the operative groups, and horizontal blue bars indicate 95% confidence intervals. The intention-to-treat analysis included all randomized participants with follow-up completed at 24 months. We performed a prespecified model-based multiple imputation analysis in which PHAT outcomes for participants who underwent randomization but did not complete the 24 months follow-up visit were imputed based on treatment arm, age, gender, and PHAT outcomes at previous follow-up visits. In addition, the combined randomized and observational cohort were analyzed using inverse probability weighting, with a weight of 0.5 for the randomized participants. If the lower boundary of the two-sided 95% confidence interval for the difference in PHAT score (nonoperative group minus operative group) was greater than -10 points (dashed vertical line), nonoperative treatment was deemed to be noninferior. P values are for noninferiority of nonoperative treatment compared to operative treatment.

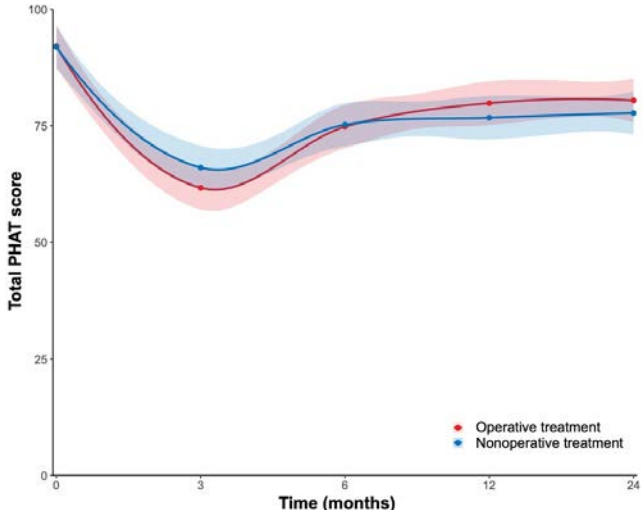


Figure 28: Graph shows the PHAT score of the operative (red) and nonoperative treatment (blue) groups at baseline, 3, 6, 12 and 24 months of follow-up. The mean PHAT score in the randomized cohort (ITT population) is represented by the dots at each time point. Red and blue Lines show a smoothed curve fit to the time points with 95% confidence bands. X-axis is on 2-log scale.

Table 5

Analyses of PHAT at 24 months						
Outcome	**Surgically treated**		**Nonsurgically treated**			
	No. Of Patients (Missing)	*mean (SD)*	*No. Of Patients (Missing)*	*mean (SD)*	Difference (95% CI)	P-value (noninferiority)
Primary outcome: Total PHAT score in RCT population						
ITT	56 (2)	80.4 (19.3)	58 (3)	77.7 (20.0)	-2.1 (-9.3, 5.1)	0.017
ITT - with imputation	58 (0)	79.2 (20.4)	61 (0)	78.0 (19.8)	-0.5 (-7.8, 6.7)	0.006
PP	54 (2)	79.9 (19.5)	57 (2)	78.5 (19.4)	-1.2 (-8.6, 6.2)	0.009
PP – with imputation	56 (0)	78.5 (20.8)	59 (0)	78.5 (19.3)	0.4 (-7.2, 7.9)	0.004
AS	55 (3)	79.1 (20.2)	59 (2)	79.0 (19.2)	-0.01 (-8.1, 8.1)	0.005
AS- with imputation	58(0)	78.0 (21.1)	61 (0)	79.1 (19.2)	1.0 (-7.0, 9.0)	0.004
Combined analyses RCT + observational cohort						
RCT + OBS	94 (8)	80.7 (18.4)	103 (11)	79.6 (18.4)	-2.6 (-7.9, 2.8)	0.002
RCT + OBS - with imputation	102 (0)	80.3 (18.8)	114 (0)	79.8 (18.5)	1.2 (-4.5, 7)	<0.001

The intention-to-treat analysis included all randomized participants with follow-up completed at 24 months. PHAT outcomes for participants who underwent randomization but did not complete the 24-month follow-up visit were imputed based on treatment arm, age, sex, and PHAT outcomes at follow-up visits at 3, 6, and 12 months. In addition, the randomized and observational populations were combined using inverse probability weighting, with a weight of 0.5 for the randomized participants. P values are for noninferiority of the nonsurgical treatment strategy to surgical treatment.

4.4.3 Secondary outcomes

Most of the secondary outcomes (LEFS score, IPAQ score and performance-based tests) did not differ between treatment groups at the 24-month visit. However, muscle quality was significantly affected in the nonsurgical treated group, in which the muscle volume of the injured hamstrings were lower than in the in the injured hamstrings of the surgically treated group. Additionally, in the nonsurgically treated group the percentage of fat were higher in the injured than in the noninjured hamstrings compared to the hamstrings of the surgically treated group. than those in the nonsurgical group. There was also a difference in

the number of patients who reported that they had fully returned to sports (59% in the surgically treated group compared to 40% in the nonsurgically treated group; OR=0.43; CI=0.20 to 0.69).

4.4.4 Adverse events

When analysing the RCT and observational cohorts together, 20 (9.2%) patients suffered some adverse event, 14 (14%) in the surgically treated group and six (5%) in the nonsurgically treated group. Of these, nine were classified as serious. Six serious adverse events were identified in the surgically treated group and consisted of one myocardial infarction, one deep infection, two deep venous thromboses and two suspected iatrogenic nerve injuries. In the nonsurgically treated patients, three serious adverse events were identified; two patients had ischialgia treated with neurolysis and one patient had a fall due to weakness and sustained a concussion.

5 Discussion

5.1 General discussion and methodological considerations

The aim of this thesis was to determine whether surgical or nonsurgical treatment yields superior outcomes for middle-aged patients who suffer proximal hamstring avulsions. It also seeks to establish a scientifically sound evaluation approach relevant to these patients. Both Studies 1 and 4, with compared treatment options, failed to demonstrate any significant difference in their primary outcomes. Most importantly, in the randomized controlled trial of Study 4, nonsurgical treatment was determined to be noninferior to surgical treatment for the primary patient-reported outcome (PHAT) at 2 years of follow-up.

Most patients included in the studies involved in this thesis were middle-aged and sustained an acute injury, including complete avulsion and retraction of all tendons, with treatment initiated within four weeks after injury. Several papers detailing characteristics of proximal hamstring avulsion patients[5, 9, 16, 19], suggest that the cohorts examined within the studies included in this thesis represent the typical profile of individuals affected by proximal hamstring avulsion. Consequently, the thesis holds high external validity.

Studies 2 and 3 demonstrate that the hamstrings muscles decline in quality and strength following proximal hamstring avulsions but that the degeneration of the injured hamstrings were only weakly correlated to patient-reported outcomes. In the analyses of surgically treated patients, we could not demonstrate a definitive protective effect of surgical repair on muscle degeneration. However, given the cross-sectional cohort design of the studies, a scientifically sound comparison between the treatment alternatives could not be made. The weak correlations between the different measurements demonstrate that evaluation of treatment of proximal hamstring avulsions is complex and is best done by a combination of subjective and objective measurements, which has also been concluded by others reviewing the use of outcome measures after proximal hamstring avulsions[33]. Interestingly, we found that running was the most frequent activity limitation mentioned by the patients. Additionally, in Study 4 the most omitted questions in the PHAT and LEFS questionnaires were the ones that included running. Future developments in this research field should involve qualitative investigations into patients perceived limitations and the design of construct-validated scores.

Even though the thesis was unable to show any major differences in patient-reported outcome measures between the treatment options, self-reported hours of physical training and return to sport rates at follow-up were in favour of surgical treatment in Studies 1 and 4. Additionally, differences were found in objective outcomes such as tendon healing in Study 3, and fatty infiltration and muscle volume in Study 4. There are several possible explanations for this discrepancy. Patients may adjust their demands to accommodate muscle weakness resulting from the injury. Additionally, other muscles might compensate for the loss of hamstring power. Specifically, the gastrocnemius muscle, sartorius muscle, and gracilis muscle contribute to knee flexion, while for hip extension, the gluteal muscles, piriformis muscle, and the adductor muscles are involved[13].

However, several methodological issues that need to be considered when interpreting the results.

First, the choice and quality of the outcome measures used will greatly influence the answers to the research questions posed, and it is important that the results of clinical studies are relevant to patients of interest. To my knowledge, there has been no qualitative study that has investigated the most important problems perceived by patients with proximal hamstring avulsions. Such a study could assist the research community in developing more specific outcome measures with respect to content and criterion validity that encompass the most relevant problems for proximal hamstring avulsion patients. In Study 2, we tried to better understand patients' perceived activity limitations, and if they were within the scope of, for proximal hamstring avulsions, the most frequently used questionnaires. We found that the LEFS covered more of these activities than the PHAT. However, the PHAT covers the pain dimension that the LEFS does not encompass, and pain is likely important to patients. Additionally, in the subgroup analysis of Study 2, the patients not satisfied with treatment scored a substantially lower PHAT score compared to the patients satisfied with treatment, mean difference 15.2 (95% CI, 1.4 to 29). This difference did not reach statical significance, there were only seven patients not satisfied with treatment. However, difference between satisfied and unsatisfied patients was not present in the LEFS score, mean difference 1.7 (95% CI, −5.2 to 8.6).

Furthermore, characteristics that are important when using outcome measures are the reliability and validity properties for the condition of interest. One of the

challenges of the research field on proximal hamstring avulsions is the disparate nature of such available measures. During the execution of Study 1, there was no subjective outcome measure presented with validity properties for this injury. The LEFS score was the most commonly used outcome measure and was therefore chosen as the primary outcome. Subsequently, when Studies 2-4 were conducted, the condition-specific Proximal Hamstring Assessment Tool (PHAT)[34] was validated for its construct, internal consistency, and sensitivity to change properties and was therefore chosen as the primary outcome in Study 4.

Since 2020, a new condition specific questionnaire, the Sydney Hamstring Origin Rupture Evaluation (SHORE)[35], has been validated for construct validity, internal consistency and reliability measures such as reproducibility. As it is fairly new, it has not yet been widely used in proximal hamstring cohorts and its future role as an outcome measure will have to be determined.

Recently Green et al. presented a study comparing the construct of the up until now most frequently used patient-reported outcome measures after proximal hamstring avulsions. The authors concluded that the PHAT, LEFS and 12-item Short Form Survey (SF -12) could be the most important measures for follow-up assessment after this injury[32]. However, this study investigated only the PROMs and not objective outcome measures. This thesis significantly extends the validity and reliability properties of both the LEFS, PHAT and also objective outcome measures for proximal hamstring avulsion patients.

For example, Study 2 revealed that both the LEFS and PHAT exhibited ceiling effects within proximal hamstring avulsion populations, a finding corroborated by others[28]. Nevertheless, strong reliability was observed for both the PHAT and LEFS which was also consistent with findings of other studies[34, 36, 50]. Additionally, the construct validity of the PHAT and LEFS was assessed through correlations between each other, which were found to be similar to previously presented corelations[28]. Furthermore, the content validity was assessed with the number of overlapping questions of the PHAT and LEFS and the activity limitations stated in the PSFS.

Additionally, the choice of a measure to assess everyday physical activity is problematic. It can either be done with questionnaires with the problem of discrepancy of self-estimated physical activity and actual physical activity performed, or by accelerometers which are expensive. With the limited budget,

we decided to collect information about physical activity via questionnaires. In Study 1, we tried to identify this entity with questions from the PHIQ questionnaire, i.e., total hours of physical training reported at follow-up visit. In Study 4, we used the IPAQ-SF questionnaire and one additional question on return to sport. None of these methods are perfect. There are many other different scores that are used to collect questionnaire-based information about physical activity[86]. The reason we chose the IPAQ-SF was that it is short, well documented, and relatively easy to fill in[56], although it has the problem of overestimating the physical activity level[87].

When designing Studies 2, 3, and 4, we conducted a thorough literature review to identify objective outcomes, such as performance-based tests, suitable for proximal hamstring avulsion patients. With no 'gold standard' for evaluating physical function, we decided to use several tests that had previously been used by others[7, 21, 26, 33, 88]. Additionally, we selected a test, the single step test[63]; we deemed this test appropriate for this middle-aged population, because it is less demanding than the more frequently used hop tests. None of these tests had previously been validated for use in the studied patient population.

We tried to assess the construct validity of the PHAT and LEFS by comparison with physical performance-based tests. Since the performance-based tests and the PROMS had at the most weak correlations, our conclusion, as well as that of others, is that the utilization of both subjective and objective outcome is needed, to cover different aspects of recovery[33].

Second, a major methodological concern, important not only in Study 1 but also in Study 3, is the problem of bias by indication. When patients are allocated to treatment according to indications outlined in a departmental treatment regime or based on the surgeon's preference, this affects the internal validity of the study and may contribute to a failure to find any true difference. This effect could work in several directions. For example, if the indication for surgery includes severe injury and the severity of injury affects the outcome negatively, then the indication would favour nonsurgical treatment in the comparison. Conversely, if young age is an indication for surgery and young age affects outcome positively then the results would be biased in favour of surgical treatment. An unlikely yet plausible explanation for not finding a difference between groups in Study 1 is that proficient surgeons selected the optimal

treatment for individual patients, i.e., have perfect indications. This scenario suggests that both treatments might yield comparable outcomes when administered to appropriately selected patients. Due to the cross-sectional design of Study 3, we did not aim to compare the treatment options and therefore did not suffer as much from the bias by indication. Instead, we decided to present the distribution of the outcomes by treatment options in a subgroup analysis.

Third, the conduct of a randomized controlled trial (RCT) presents the highest standard of scientific evidence in clinical trials. However, when investigating vastly different treatment approaches, such as surgical and nonsurgical methods, several challenges arise. One primary challenge involves mitigating selection bias and ensuring unbiased information delivery to patients by the recruiting personnel. This becomes particularly complex when the surgeon favours one treatment option, hindering the doctor and patient from reaching a consensus in the given case. In Study 4, this bias was observed through the varying abilities of different study sites to randomize patients versus enrolling them in the observational cohort. Additionally, patients themselves may possess strong preferences for a particular treatment, potentially refusing randomization. To address these potential biases, an observational cohort was generated for patients for whom consensus could not be reached. The aim of this approach is to lessen selection bias and enhance external validity by enrolling a broader spectrum of patients' representative of the source population. Hence, in Study 4, we adopted a patient preference design, incorporating an observational cohort to address these complexities.

5.2 Discussion of Study 1

In Study 1 we were unable to identify any difference between surgically and nonsurgically treated patients in terms of the patient-related outcome LEFS. However, the total hours of physical training were significantly greater in the surgically treated group at the time of follow-up.

5.2.1 Strengths

We managed to include 90.4% of the eligible patients. The choice of the LEFS score as the primary outcome measure was decided on the basis of being one for the lower limb injuries validated scores. Additionally, prior studies using it as the primary outcome[21, 26, 31] and one of them showing a statistically significant (but weak) correlation between the LEFS score and objective

outcome measures such as the single-leg hop test and isokinetic strength test, theoretically favouring the potential construct validity of the LEFS score for proximal hamstring avulsion patients[21].

5.2.2 Limitations

The greatest limitation of Study 1 was the bias by indication, the treatment decision was based on the orthopaedic surgeon's experience and local guidelines at the department. Some of the data from the study support this bias, with more severe injuries allocated to surgery. Furthermore, the surgically treated patients reported being more physically active at the follow-up. It could also be that the study design was too poor or the sample size too low to detect any difference. A formal power calculation was not performed prior to the study. With these biases and confounders an evidence-based conclusion on preferred treatment options could not be drawn, but it served as a motivator for conducting a study at a higher scientific level, study 4.

5.3 Discussion of Study 2

In Study 2, we found strong correlations between the PHAT and LEFS scores, but their correlations with physical performance-based tests results were weak. The isokinetic strength test was the only test that could discriminate between the injured and uninjured leg. Furthermore, running was the most mentioned activity that patients perceived as limited.

5.3.1 Strengths

With the described difficulty of choosing suitable outcome variables for the follow-up after proximal hamstring avulsions, high reliability measures (ICC's and SEM's) indicate that the outcome measures chosen were manageable. Furthermore, that the study was well standardized and accurate.

We demonstrated strong correlations between the PHAT and LEFS scores (r=0.83), which was supported by the findings of other studies (r = 0.68 – 0.81) [28, 34]. Additionally, the finding that only the isokinetic strength test was able to discriminate between the legs is supported by the findings of a relatively recent review concluding that isokinetic strength is the method of reference for assessing physical function after treatment for proximal hamstring avulsion[69].

Using the open-ended questions from the PSFS questionnaire at least two years after injury we presented novel data on patient perceived limitations, with

running being the most commonly activity limitation mentioned. This information is something future proximal hamstring avulsion patients should be informed of when they are in consultation with a physician.

5.3.2 Limitations

One explanation for the lack of validated subjective and objective outcome measures after proximal hamstring avulsions is that it is difficult to verify the construct validity of such outcomes. For example, objective outcomes should maybe be validated in terms of construct towards hamstring muscle strength, but isolating the hamstring muscles in a dynamic test is impossible, since there are agonists that work in both knee flexion and hip extension[67]. Therefore, we ultimately used many different scores and tests to perform objective assessment as much as possible in our proximal hamstring avulsion cohort, most of which were previously used by others [7, 8, 21, 26, 33, 34, 88]. However, we were aware of that none of the tests were perfect for hamstring muscle evaluation. Moreover, most of the previously used tests, such as the hop test, are produced for younger cohorts with anterior cruciate ligament injury[58, 59] and might not be optimal for the older proximal hamstring avulsion patients, in general. This reasoning is strengthened by the secondary outcome data of Study 4, in which approximately 25% of the data from the hop tests were missing, whereas only 15% of the data from single step test were missing.

5.4 Discussion of Study 3

In Study 3, we demonstrated that patients have more extensive fatty infiltration, muscle atrophy and loss of isokinetic strength in the injured than in the uninjured hamstring muscles several years after injury. When analysing only the surgically treated patients, the analysis generated similar findings, indicating that surgery cannot completely protect hamstring muscles from degeneration. Muscle quality impairment was significantly but weakly correlated with strength deficit in the injured leg.

5.4.1 Strengths

The measurements used to assess fatty infiltration of the muscles and muscle volume exhibited high reliability, implying that the MRI images were carefully assessed.

Our findings of a reduction in muscle volume and increase in fatty infiltration in the injured leg compared to uninjured leg, regardless of treatment, are supported

by studies of other tendon avulsion injuries[89]. However, they are opposing, the often mentioned by orthopaedic surgeons rational for treating tendon avulsions with surgery. Theoretically, surgery should restore the biomechanics, resulting in less muscle degeneration and less loss of strength compared to if not restored and the tendon is left to heal with some scarring of the closest tissue.

We demonstrated weak correlations (r = 0.357-0.494 ($p<0.001$-0.05)) between muscle volume deficit, fatty infiltration and strength deficit in the injured leg compared to the uninjured leg. In the literature similar findings have been presented after Achilles tendon rupture repair with muscle atrophy (11-13%), greater fatty infiltration (12-18%) and correlation to strength deficit in the affected leg at follow-up[89].

We also present novel data on nonsurgically treated patients and the healing of their proximal hamstring tendons. None of these patients had a completely healed tendon in our study. A previous study presenting data on tendon healing after proximal hamstring avulsion reported fewer tendons in continuity but that the formation of a neotendon was present in many cases[9]. In our cohort, this was observed in only one patient.

5.4.2 Limitations

The problem with the measurement of muscle volume by the slice-by slice technique is that the entire muscle volume is assessed, regardless of whether the muscle is infiltrated by fat or not. A muscle with almost only fatty degeneration can theoretically still have a large muscle volume, as seen on panel C of Figure 17.

With respect to the fatty infiltration assessments used, there are two major limitations. First, the Goutallier classification involves an individual assessment rather than a quantitative measure. The reason for choosing such a method was that when conducting the study, our MRI software did not allow us to establish an MRI protocol with the possibility for a quantitative measurement, such as Dixon[77]. Second, we decided to modify the Goutallier score, to obtain an understanding of the fatty degeneration of the entire muscle, and not only one slice of the muscle. I have not found any previous paper describing this approach to classify fatty infiltration and it has not been explored in regard to psychometric properties, which is of course another limitation of the study.

5.5 Discussion of Study 4

We present results that could change clinical practice for the treatment of acute proximal hamstring avulsions in middle-aged patients. With the first international, multicentre, patient preference, noninferiority randomized study, we found that nonsurgical treatment was noninferior to surgical treatment ($p<0.017$) for acute proximal hamstring avulsions in patients aged 30-70 years.

5.5.1 Strengths

The noninferiority of nonsurgical treatment was strengthened by the results being robust in sensitivity analyses and in most of the secondary outcomes. These findings are supported by recent studies with lower levels of evidence also presenting similar scores for both the PHAT and LEFS despite treatment[7-9, 23].

To prove the external validity of the results different approaches were used. The potential selection bias that is common in RCTs was partly addressed with the preference tolerant design. The results of the analysis were robust when the RCT and observational cohorts were analysed together. There were also few patients who were lost to follow-up, with 93% of the entire study population followed up to the 24-month visit.

Considering the potential placebo effect of surgery, one would expect that surgery would be associated with better patient-reported outcomes, which strengthens our findings that nonsurgical treatment is noninferior.

The only between-group differences found in our secondary outcomes, were a lower proportion of the nonsurgically treated patients who had fully returned to sports at follow up (OR 0.43 (0.20 to 0.93)) and a difference in muscle degeneration parameters assessed with MRI. The greater number of patients who returned to sports among the surgically treated patients is probably important for those prone to proximal hamstring avulsion, which comprises high-performance amateur athletes, and should be kept in mind in treatment decision making.

5.5.2 Limitations

As previously described the lack of different validated outcome measures after proximal hamstring avulsions is a great limitation for all the studies of the thesis, including Study 4.

The reason for choosing the noninferiority design was to investigate whether the nonsurgical treatment was inferior to surgical treatment, making the potential risks and costs worth considering for the increasingly used[5] but not evidence-based surgical treatment. However, the noninferiority design has limitations. First, the choice of the noninferiority margin is critical[47, 90]. Therefore, much effort was made to determine the noninferiority margin. Consensus meetings by experienced physicians and statisticians within the field were held, and final margin was set to be less than the, at that time, published MDC (16.4) and SD (±16.0- ±22.5) of the primary outcome, PHAT[28, 34]. Interpreting the results, one must keep in mind that if we had chosen a smaller margin, with the need for a larger sample size, our results might have been different. However, LEFS scores with similar SDs (±8- ±19) and MDC (8-15) [28, 48, 50] were also inferior in our study, which strengthens the noninferiority of nonsurgical treatment and supports our conclusion.

Several facts affect the internal validity of this multicentre study. We allowed for different surgical techniques and many different surgeons participated. This might cause the surgical treatment to differ between study sites and surgeons. Additionally, the rehabilitation was performed in many different clinics, and although a standardized rehabilitation protocol was used, it is difficult to ensure equality between physiotherapists and clinics. On the other hand, the external validity of such a design is strengthened.

6 Conclusions

6.1 Study 1

We found no difference in the LEFS score between treatment allocations. However, a greater proportion of surgically treated patients had more severe injuries than nonsurgically treated patients, and at follow-up, they reported being more physically active than nonsurgically treated patients.

6.2 Study 2

To obtain as broad perspective of recovery as possible, it is suggested that both subjective, patient-reported outcome measures, and objective, performance-based tests, are used for the follow-up after treatment for proximal hamstring avulsions. In addition, running seems to be an activity limitation of importance for these patients.

6.3 Study 3

At the mid- to long-term follow-up assessments, after proximal hamstring avulsions, the fatty infiltration is increased, and the muscle volume is reduced in the injured leg compared to the uninjured leg. These findings are, perhaps contra-intuitive, only weakly correlated with muscle strength loss of the injured leg.

6.4 Study 4

According to patient-reported outcome, nonsurgical treatment is not inferior to surgical treatment and should be the treatment of choice for most middle-aged patients with acute proximal hamstring avulsions.

In summary, the answer to the question of the title of the thesis is not.

7 Points of perspective

Summarizing Study 1, it became clear that to fill the knowledge gap of how, in an evidence-based manner, treat proximal hamstring avulsions, a prospective randomized design was needed.

Since then, one study that with a prospective design has evaluated what treatment that is preferable, without finding any difference in patient-reported outcomes[9]. There are also papers presenting the current state-of-the-art in treatment for proximal hamstring avulsions[2, 3]. Because of the lack of randomized level 1 studies the purpose of Study 4 was to use a randomized study design to fill the knowledge gap and present what treatment that is preferable in an evidence-based sound manner.

However, there are still many questions to answer.

First Study 2 noted the difficulty in choosing the "best" outcome measures since the subjective and objective outcome measures mostly used did not have any strong correlation, and seemed to cover different aspects of recovery. To better understand patients' perceptions of sequelae after proximal hamstring avulsions, a qualitative study design analysing the most important sequelae that patients suffer from after such injury is suggested. This could also contribute to constructing a comprehensive and patient-relevant outcome measure that might include both subjective and objective findings.

Until then, according to the findings of Study 2 and other studies, the preferred outcome measures for the follow-up assessment after proximal hamstring avulsions are PHAT, LEFS, SF-12 and isokinetic strength test[32, 33].

Second, in Study 3 the question was raised whether muscle degeneration was already present at the time of injury; such finding have been considered prognostic factors, e.g., after rotator cuff tears[91]. Prognostic studies that could predict outcome based on clinical and radiological data at admission for injury could better customize treatment for future patients with proximal hamstring avulsions.

Third, with the evidence-based data of Study 4, we suggest that there is no obvious advantage for surgical treatment after proximal hamstring avulsions for middle-aged patients and which might lead to the that current clinical practice change. However, there are still questions that remain: Are there subgroups of

patients who can exploit surgery, e.g., high-performance amateur athletes, who have greater demands for their physical function? Is there one surgical procedure that is preferable? Could the nonsurgical treatment improve? Is it possible to repeat our results with another RCT? And probably many more.

8 Populärvetenskaplig sammanfattning på svenska

Min avhandling handlar om hur patienter som drabbats av att bakre lårmuskelns senfästen slitits av från deras ursprung på sittbenet, kan behandlas och följas upp. Skadan inträffar vid en hastig spagatrörelse som till exempel när man halkar på längdskidor. Det innebär att bakre lårmusklerna utsätts för en kraft som vill sträcka ut musklerna trots att de försöker hålla emot. När kraften överstiger senornas hållfasthet lossnar de från sitt ursprung. Vanligen drabbar denna mycket smärtsamma skada relativt aktiva personer i medelåldern. I akutskedet blir det svårt att gå, ta ut steglängden och att sitta. Efter några dagar får många ett stort blåmärke på baksidan av låret. Vissa får påverkan på nervfunktionen i underbenet och foten.

Det finns två behandlingsalternativ. Det ena är att med en operation sy tillbaka senorna till sittbenet. Vilket innebär att skruvar försedda med trådar borras eller hamras in i sittbenet. Trådarna förs sedan genom senorna och ”dras hem” mot skruvarna. Därefter följer en lång period av rehabilitering. Det andra alternativet är att påbörja rehabilitering utan föregående operation. Denna avhandling försöker ta reda på vilken behandling som ger bäst resultat och hur patienterna mår när man följer upp dem minst två år efter skadan.

I den första studien följde vi år 2015 upp patienter som behandlats på Danderyds sjukhus åren 2007-2013. Vi jämförde hur de skattade sin självupplevda funktion med hjälp av frågeformulär. Det gick inte att uppmäta någon skillnad mellan de som behandlats med operation och efterföljande rehabilitering, och de som behandlats utan operation med endast rehabilitering. Eftersom valet av behandlingsmetod för respektive patient hade gjorts långt innan studien startade så är det svårt att avgöra om behandlingarna verkligen är likvärdiga. Den uteblivna skillnaden kan lika väl ha berott på att patienterna hade fått optimal behandling utifrån skadans allvarlighetsgrad, den enskildes aktivitetsnivå och den medicinska risken kopplad till kirurgi.

I den andra studien undersökte vi hur patienterna, både opererade och icke opererade, själva skattade sin återhämtning i frågeformulär men även hur de presterade i olika styrke- och funktionstester av benen. Här jämfördes det skadade med det oskadade benet. Vi undersökte vidare hur dessa olika typer av subjektiva och objektiva mått samvarierade med varandra. Ett annat formulär där patienterna själva fick ange vilka aktiviteter de tyckte var svåra att genomföra på grund av skadan användes också för att utvärdera resultatet. Patienterna

undersöktes 2018/2019 och de hade drabbats av skadan och behandlats på Danderyds sjukhus mellan 2007 och 2016. Frågeformulären som vi använde hade hög samvariation, men resultaten från formulären avspeglades inte i hur patienterna presterade i de olika styrke- och funktionstesterna av benet. Löpning var den aktivitet som flest patienter uppgav som begränsade på grund av skadan. Vår slutsats blev att för att göra en helhetsbedömning av en patient som drabbats av avslitning av bakre lårmuskelns senfästen behövs uppföljning, både med frågeformulär och objektiva mått såsom styrke- och funktionstester av benet.

I den tredje studien undersöktes hur bakre lårmuskelns kvalitet ser ut på en magnetkameraundersökning minst två år efter skadan och om utseendet i form av fettomvandling och minskad muskelvolym avspeglas i minskad muskelstyrka. Samma grupp av patienter undersöktes vid samma tillfälle som i studie två, d.v.s. både patienter som opererats och som inte opererats. Majoritet av patienterna hade tappat i muskelvolym och fått större fettomvandling av sina bakre lårmuskler i det skadade benet jämfört med det friska benet. Denna förändring i muskelns utseende samvarierade med en minskad styrka i muskeln. När den undergrupp av patienter som opererats för sin skada undersöktes separat fick vi samma resultat. Vår tolkning blev att efter en avslitning av bakre lårmuskelns senfästen tappar patienterna i muskelvolym, muskeln blir fettomvandlad och styrkan i benet försämras jämfört med det friska benet. Det verkar gälla trots att senorna sytts tillbaka till sitt ursprung med operation.

I den fjärde delstudien deltog 10 olika sjukhus i Sverige och Norge. Medelålders patienter som inkom till något av de deltagande sjukhusen med avslitning av bakre lårmuskelns senfästen lottades mellan kirurgisk och icke-kirurgisk behandling. Därefter följdes patienterna i två år med subjektiva och objektiva utfallsmått. De patienter som hade en stark önskan om en viss behandling eller om den enskilde ortopeden bedömde att en viss behandling var klart mer motiverad och därför inte gick att lotta, erbjöds samma uppföljning inom ramen för studien. Här ville vi på ett vetenskapligt korrekt sätt undersöka om den icke-kirurgiska behandlingen skulle vara underlägsen den kirurgiska behandlingen och i så fall motivera den potentiellt ökade risken och kostnaden med kirurgi. Vid uppföljningen efter två år kunde vi inte finna någon skillnad i av patienterna självskattat utfall mellan de två behandlingarna. Slutsatsen blev att den icke-kirurgiska behandlingen inte ger ett sämre patientrapporterat utfall än den

kirurgiska behandlingen och därför i första hand bör erbjudas till medelålders patienter efter skada av bakre lårmuskelns senfästen.

Sammanfattningsvis bidrar min avhandling till nya vetenskapliga insikter i att avslitning av bakre lårmuskelns senfäste, hos medelålders patienter, primärt skall behandlas icke-kirurgiskt samt vad en patient kan förvänta sig för självupplevt och funktionellt utfall efter skadan.

9 Acknowledgements

Jag vill tacka alla er som gjort denna avhandling möjlig att genomföra. Alla patienter som deltagit i studierna, kollegorna som arbetat med sjukvård när jag har forskat, och alla ni andra som på olika sätt stöttat mig under doktorandtiden. Ett särskilt tack till några av er:

Carl Johan Hedbeck tack för att du har lagt ner så stor möda på att hjälpa mig att få till den fina helheten som till slut blev min avhandling, och för att du då och då kommit med uppmuntrande ord när det har känts motigt.

Kenneth Jonsson för allt arbete du har lagt ner på de tre senare av mina delarbeten, framför allt det enorma fjärde delarbetet. Tack för dina genomtänkta och ofta kluriga synpunkter. De har utmanat och utvecklat min analytiska förmåga.

Olof Sköldenberg utan dig hade ”att forska” varit något helt annat på vår klinik. Du är generös, har tilltro till människor och har skapat en miljö där vi doktorander får tid och har möjlighet att lära oss att forska.

Mikael Skorpil tack för att du alltid återkommer med ett snabbt men också genomtänkt svar och för att du så generöst delat med dig av din tid och kunskap. Du har hela tiden ingjutit i mig att jag har förmågan att ro ”det hela” i land.

Nina Brodin även du har generöst delat med dig av dina kunskaper och din tid när jag behövt råd och uppmuntran. Din skarpsynthet har bidragit enormt till min utveckling och till att jag sakta börjat förstå hur patienternas perspektiv kan mätas och tolkas.

Gabriella Bröms med hjälp av dig som mentor och dina klocka synpunkter lyckades jag komma in på forskarskolan för kliniker inom epidemiologi. Utan den utbildningen hade vägen till färdig avhandling varit väldigt många år längre och kanske hade jag aldrig kommit fram till målet.

Sven Jönhagen du introducerade mig till ortopedi och idrottsmedicin på friidrottsbanan för länge sedan. Under de senaste åren, då jag har varit doktorand, har du varit min chef och du är även medförfattare till mitt första delarbete. Tack för att du följt och hela tiden har stöttat mig på den här långa resan.

Klinikens forskningssköterskor och personal: **Sofia Grindberg** utan din positiva energi och entusiasm hade det fjärde delarbetet aldrig blivit genomfört**, Paula Kelley-Pettersson** du har alltid tid för en fråga, kommer alltid med ett bra svar och är medförfattare till det första delarbetet, **Marie Ax** tack för all administrativ hjälp med delarbete 2 och 3. **Monica Öhlin** för administrativ hjälp i slutspurten av avhandlingen. **Saga Grindberg** för hjälp med datamonitorering i delarbete 4.

Mida Berglöf tack för din hjälp med utformning av protokoll och all datainsamling till delarbete 2 och 4. Jag vill också tacka **Sofie Mannerson, William Persson och Johan Fyrlund** som hjälpt till med mätningar till delarbete 4. **Alicia Lindström** och **Isabelle Berggren** för hjälp vid fotograferingen av de funktionella testerna.

Hans Nåsell du tog initiativ till det första delarbetet som lade hela grunden till avhandlingen, tack.

Per Widholm tack för att du tog dig tid att lära mig radiologiska mätmetoder och för din hjälp med radiologisk analys i delstudie 4.

Martin Magnéli tack för uppmuntran och för hjälp med grafisk presentation av data till delarbete 3.

Björn Salomonsson du har när möjlighet givits försökt att bredda min kunskap kring patientrapporterade utfallsmått vilket jag verkligen uppskattar.

Jag vill tacka medförfattarna till det fjärde delarbetet **Sofia Laszlo**, **Anne-Mari Rosenlund**, **Jörg Schilcher, Frede Frihagen.** Ett särskilt tack till **Chiara Micoli** och **Martin Eklund** för ert tålamod och hjälp med att försöka sätta mig in i olika statistiska metoder och till **Målfrid Holen Kristoferssen** för din positiva energi**.**

Jag vill tacka de som arbetar på MR avdelningen på Danderyds sjukhus: **Anders von Heijne** för att du har varit positivt inställd till att låta undersöka alla ”mina” hamstringspatienter, **Massi Farokhi** och **Alber Grau** som hjälp till att genomföra en stor mängd MR undersökningar.

Anderas Johansson tack för att du gjorde den fina illustrationen ”Storsnasen och skidåkarna” till framsidan och anatomiskissen i avhandlingen.

Jag vill också rikta ett tack till de **kvinnliga ortopeder** som stakat ut vägen framför mig.

Några har ställt upp extra mycket när livet varit svårt vilket har gjort att jag kunnat resa mig upp och börja om: **Jesper Hermodsson**, **Calle Sandell** och **Mikaela Koller** tack för alla härliga små och stora äventyr som vi gjort tillsammans eller på tu man hand, och för att ni delat de fina kalla vinterdagarna i februari med mig. **Hilde Wahl** för all din omtänksamhet och värme. **Clara Luthman** du väjer inte för att diskutera vare sig stort eller smått. **Maria Herrman** och **Lotten von Sydow** för att ni alltid finns där och alltid vill ses.

Klara Sondén våra liv har på ett mystiskt sätt följts åt sedan barnsben, genom skolår och läkarutbildning och även efter det du har funnits där i mitt livs olika skeenden, ett stort tack för det.

Lisa och **Göran** tack för all hjälp med familjen när det behövts under de senaste åren och för den fina kontakten ni har med våra barn.

Mamma och **Pappa** ni har gett mig frihet och trygghet att våga ta mig an livets olika utmaningar. Tack för att ni stöttar och hjälper mig och min familj. Utan er och **Maria** och **Pärs** hjälp hade det blivit mycket svårare att komma i mål med det här projektet. Ett särskilt tack till Maria för hjälp med korrekturläsning.

Slutligen vill jag rikta ett tack till min familj, denna bok är till er. **Gustav** du försöker enträget få mig att se det positiva i olika situationer även om jag ibland tycker att det är svårt. Du ger mig också frihet att göra det jag behöver och vill för att må bra och för att utvecklas. Jag beundrar dig för din förmåga att se det fina i det lilla, och jag hoppas det är något som du lär **Klara** och **Eva**, våra fina barn, som ser upp till sin älskade pappa så mycket. Tack.

10 References

1. Ishikawa, K., K. Kai, and H. Mizuta, *Avulsion of the hamstring muscles from the ischial tuberosity. A report of two cases.* Clin Orthop Relat Res, 1988(232): p. 153-5.
2. Chang, J.S., et al., *Management of hamstring injuries: current concepts review.* Bone Joint J, 2020. **102-b**(10): p. 1281-1288.
3. Plastow, R., et al., *London International Consensus and Delphi study on hamstring injuries part 2: operative management.* Br J Sports Med, 2023. **57**(5): p. 266-277.
4. Bodendorfer, B.M., et al., *Outcomes After Operative and Nonoperative Treatment of Proximal Hamstring Avulsions: A Systematic Review and Meta-analysis.* Am J Sports Med, 2017: p. 363546517732526.
5. Laszlo, S. and K.B. Jonsson, *Increasing incidence of surgically treated hamstring injuries: a nationwide registry study in Sweden between 2001 and 2020.* Acta Orthop, 2023. **94**: p. 336-341.
6. van der Made, A.D., et al., *Outcome After Surgical Repair of Proximal Hamstring Avulsions: A Systematic Review.* Am J Sports Med, 2015. **43**(11): p. 2841-51.
7. Shambaugh, B.C., et al., *A Comparison of Nonoperative and Operative Treatment of Complete Proximal Hamstring Ruptures.* Orthop J Sports Med, 2017. **5**(11): p. 2325967117738551.
8. Kanakamedala, A.C., et al., *No difference in clinical outcomes between operative and nonoperative management of minimally retracted proximal hamstring ruptures.* Knee Surg Sports Traumatol Arthrosc, 2023. **31**(7): p. 2739-2745.
9. van der Made, A.D., et al., *Proximal hamstring tendon avulsions: comparable clinical outcomes of operative and non-operative treatment at 1-year follow-up using a shared decision-making model.* Br J Sports Med, 2022.
10. Hoppenfeld S, d.P., Buckleym R, *Surgical Exposures in Orthopaedics. The Anatomic Approach*. Fourth edition ed. 2009: Wolters Kluwer. Lippincott Williams & Wilkins.
11. Sato, K., et al., *Anatomical study of the proximal origin of hamstring muscles.* J Orthop Sci, 2012. **17**(5): p. 614-8.
12. van der Made, A.D., et al., *The hamstring muscle complex.* Knee Surg Sports Traumatol Arthrosc, 2015. **23**(7): p. 2115-22.
13. Feneis H, D.W., *Anatomisk bildordbok*. 5th ed. Vol. 1. 2005: Liber.

14. Silvers-Granelli, H.J., et al., *Hamstring muscle injury in the athlete: state of the art*. J isakos, 2021. **6**(3): p. 170-181.

15. Koulouris, G. and D. Connell, *Evaluation of the hamstring muscle complex following acute injury*. Skeletal Radiol, 2003. **32**(10): p. 582-9.

16. Irger, M., et al., *Proximal hamstring tendon avulsion injuries occur predominately in middle-aged patients with distinct gender differences: epidemiologic analysis of 263 surgically treated cases*. Knee Surg Sports Traumatol Arthrosc, 2019.

17. Sikka, R.S., G.B. Fetzer, and D.A. Fischer, *Ischial apophyseal avulsions: proximal hamstring repair with bony fragment excision*. J Pediatr Orthop, 2013. **33**(8): p. e72-6.

18. Sarimo, J., et al., *Complete proximal hamstring avulsions: a series of 41 patients with operative treatment*. Am J Sports Med, 2008. **36**(6): p. 1110-5.

19. Wood, D.G., et al., *Avulsion of the proximal hamstring origin*. J Bone Joint Surg Am, 2008. **90**(11): p. 2365-74.

20. Askling, C.M., et al., *Total proximal hamstring ruptures: clinical and MRI aspects including guidelines for postoperative rehabilitation*. Knee Surg Sports Traumatol Arthrosc, 2013. **21**(3): p. 515-33.

21. Skaara, H.E., et al., *Self-reported and performance-based functional outcomes after surgical repair of proximal hamstring avulsions*. Am J Sports Med, 2013. **41**(11): p. 2577-84.

22. Subbu, R., H. Benjamin-Laing, and F. Haddad, *Timing of surgery for complete proximal hamstring avulsion injuries: successful clinical outcomes at 6 weeks, 6 months, and after 6 months of injury*. Am J Sports Med, 2015. **43**(2): p. 385-91.

23. Pihl, E., et al., *Patient-reported outcomes after surgical and non-surgical treatment of proximal hamstring avulsions in middle-aged patients*. BMJ Open Sport Exerc Med, 2019. **5**(1): p. e000511.

24. van der Made, A.D., et al., *Good Interrater Reliability for Standardized MRI Assessment of Tendon Discontinuity and Tendon Retraction in Acute Proximal Full-Thickness Hamstring Tendon Injury*. Am J Sports Med, 2021. **49**(9): p. 2475-2481.

25. Bloom, D.A., et al., *Variability of MRI reporting in proximal hamstring avulsion injuries: Are musculoskeletal radiologists and orthopedic surgeons utilizing similar landmarks?* Clin Imaging, 2023. **93**: p. 46-51.

26. Chahal, J., et al., *Clinical and magnetic resonance imaging outcomes after surgical repair of complete proximal hamstring ruptures: does the tendon heal?* Am J Sports Med, 2012. **40**(10): p. 2325-30.

27. Blakeney, W.G., et al., *A prospective evaluation of proximal hamstring tendon avulsions: improved functional outcomes following surgical repair*. Knee Surg Sports Traumatol Arthrosc, 2017. **25**(6): p. 1943-1950.

28. Best, R., et al., *Functional impairment after successful surgical reconstruction for proximal hamstring avulsion*. Int Orthop, 2019. **43**(10): p. 2341-2347.

29. Piposar, J.R., et al., *High-Grade Partial and Retracted (<2 cm) Proximal Hamstring Ruptures: Nonsurgical Treatment Revisited*. Orthop J Sports Med, 2017. **5**(2): p. 2325967117692507.

30. Leger-St-Jean, B., et al., *Accelerated rehabilitation results in good outcomes following acute repair of proximal hamstring ruptures*. Knee Surg Sports Traumatol Arthrosc, 2019. **27**(10): p. 3121-3124.

31. Cohen, S.B., et al., *Functional results and outcomes after repair of proximal hamstring avulsions*. Am J Sports Med, 2012. **40**(9): p. 2092-8.

32. Green, J.S., et al., *Comparative Analysis of Patient-Reported Outcome Measures for Proximal Hamstring Injuries: A Systematic Review*. Orthop J Sports Med, 2022. **10**(7): p. 23259671221104758.

33. Reza, T., et al., *Systematic Review of Outcome Measures Used After Proximal Hamstring Repair*. Orthop J Sports Med, 2021. **9**(5): p. 23259671211005101.

34. Blakeney, W.G., et al., *Proximal hamstring tendon avulsion surgery: evaluation of the Perth Hamstring Assessment Tool*. Knee Surg Sports Traumatol Arthrosc, 2017. **25**(6): p. 1936-1942.

35. French, S.R., et al., *Validation of the Sydney Hamstring Origin Rupture Evaluation (SHORE)*. Bone Joint J, 2020. **102-b**(3): p. 388-393.

36. Binkley, J.M., et al., *The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network*. Phys Ther, 1999. **79**(4): p. 371-83.

37. Ware, J., Jr., M. Kosinski, and S.D. Keller, *A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity*. Med Care, 1996. **34**(3): p. 220-33.

38. Williams, G.N., et al., *Comparison of the Single Assessment Numeric Evaluation method and two shoulder rating scales. Outcomes measures after shoulder surgery*. Am J Sports Med, 1999. **27**(2): p. 214-21.

39. Tegner, Y. and J. Lysholm, *Rating systems in the evaluation of knee ligament injuries*. Clin Orthop Relat Res, 1985(198): p. 43-9.

40. Lefevre, N., et al., *Returning to sports after surgical repair of acute proximal hamstring ruptures.* Knee Surg Sports Traumatol Arthrosc, 2013. **21**(3): p. 534-9.

41. Mokkink, L.B., et al., *The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument.* Braz J Phys Ther, 2016. **20**(2): p. 105-13.

42. Terwee, C.B., et al., *Quality criteria were proposed for measurement properties of health status questionnaires.* J Clin Epidemiol, 2007. **60**(1): p. 34-42.

43. Kovacs, F.M., et al., *Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain.* BMC Musculoskelet Disord, 2008. **9**: p. 43.

44. Cook, C.E., *Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense.* J Man Manip Ther, 2008. **16**(4): p. E82-3.

45. Rothman, *Epidemiology : an introduction (2nd ed.) 1. Oxford University Press. (2012).*

46. Wasmann, K.A., et al., *Partially randomised patient preference trials as an alternative design to randomised controlled trials: systematic review and meta-analyses.* BMJ Open, 2019. **9**(10): p. e031151.

47. Groenwold, R.H.H., E.R. Heerdink, and O.H. Klungel, *Challenges in the Design, Registration, and Reporting of Noninferiority Trials.* JAMA Psychiatry, 2022. **79**(3): p. 191-192.

48. Pihl, E., et al., *Exploring the Perth Hamstring Assessment Tool and Lower Extremity Functional Scale in a Proximal Hamstring Avulsion Cohort: A Cross-sectional Study.* Am J Sports Med, 2021. **49**(7): p. 1732-1740.

49. Guillemin, F., C. Bombardier, and D. Beaton, *Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines.* J Clin Epidemiol, 1993. **46**(12): p. 1417-32.

50. Mehta, S.P., et al., *Measurement Properties of the Lower Extremity Functional Scale: A Systematic Review.* J Orthop Sports Phys Ther, 2016. **46**(3): p. 200-16.

51. Sallay, P.I., et al., *Subjective and functional outcomes following surgical repair of complete ruptures of the proximal hamstring complex.* Orthopedics, 2008. **31**(11): p. 1092.

52. Chatman, A.B., et al., *The Patient-Specific Functional Scale: measurement properties in patients with knee dysfunction.* Phys Ther, 1997. **77**(8): p. 820-9.

53. Westaway, M.D., P.W. Stratford, and J.M. Binkley, *The patient-specific functional scale: validation of its use in persons with neck dysfunction.* J Orthop Sports Phys Ther, 1998. **27**(5): p. 331-8.

54. Rosengren, J. and N. Brodin, *Validity and reliability of the Swedish version of the Patient Specific Functional Scale in patients treated surgically for carpometacarpal joint osteoarthritis.* J Hand Ther, 2013. **26**(1): p. 53-60; quiz 61.

55. Mannberg Bäckman, S., et al., *Validity and sensitivity to change of the Patient Specific Functional Scale used during rehabilitation following proximal humeral fracture.* Disabil Rehabil, 2016. **38**(5): p. 487-92.

56. Craig, C.L., et al., *International physical activity questionnaire: 12-country reliability and validity.* Med Sci Sports Exerc, 2003. **35**(8): p. 1381-95.

57. WHO. *International Classification of Functioning, Disability and Health (ICF).* 2020; Available from: https://www.who.int/classifications/icf/en/.

58. Noyes, F.R., S.D. Barber, and R.E. Mangine, *Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture.* Am J Sports Med, 1991. **19**(5): p. 513-8.

59. Tegner, Y., et al., *A performance test to monitor rehabilitation and evaluate anterior cruciate ligament injuries.* Am J Sports Med, 1986. **14**(2): p. 156-9.

60. Davies, W.T., G.D. Myer, and P.J. Read, *Is It Time We Better Understood the Tests We are Using for Return to Sport Decision Making Following ACL Reconstruction? A Critical Review of the Hop Tests.* Sports Med, 2020. **50**(3): p. 485-495.

61. Hegedus, E.J., et al., *Clinician-friendly lower extremity physical performance measures in athletes: a systematic review of measurement properties and correlation with injury, part 1. The tests for knee function including the hop tests.* Br J Sports Med, 2015. **49**(10): p. 642-8.

62. Hofmann, K.J., et al., *Complete Avulsion of the Proximal Hamstring Insertion: Functional Outcomes After Nonsurgical Treatment.* J Bone Joint Surg Am, 2014. **96**(12): p. 1022-1025.

63. Marmon, A.R., et al., *Single-step test for unilateral limb ability following total knee arthroplasty.* J Orthop Sports Phys Ther, 2013. **43**(2): p. 66-73.

64. Loudon, J.K., et al., *Intrarater Reliability of Functional Performance Tests for Subjects With Patellofemoral Pain Syndrome.* J Athl Train, 2002. **37**(3): p. 256-261.

65. Lund, H., et al., *Learning effect of isokinetic measurements in healthy subjects, and reliability and comparability of Biodex and Lido dynamometers.* Clin Physiol Funct Imaging, 2005. **25**(2): p. 75-82.

66. de Araujo Ribeiro Alvares, J.B., et al., *Inter-machine reliability of the Biodex and Cybex isokinetic dynamometers for knee flexor/extensor isometric, concentric and eccentric tests*. Phys Ther Sport, 2015. **16**(1): p. 59-65.

67. Reichard, L.B., et al., *Testing knee extension and flexion strength at different ranges of motion: an isokinetic and electromyographic study*. Eur J Appl Physiol, 2005. **95**(4): p. 371-6.

68. Pincivero, D.M., S.M. Lephart, and R.A. Karunakara, *Reliability and precision of isokinetic strength and muscular endurance for the quadriceps and hamstrings*. Int J Sports Med, 1997. **18**(2): p. 113-7.

69. Fouasson-Chailloux, A., et al., *Strength assessment after proximal hamstring rupture: A critical review and analysis*. Clin Biomech (Bristol, Avon), 2020. **72**: p. 44-51.

70. Klingele, K.E. and P.I. Sallay, *Surgical repair of complete proximal hamstring tendon rupture*. Am J Sports Med, 2002. **30**(5): p. 742-7.

71. Konan, S. and F. Haddad, *Successful return to high level sports following early surgical repair of complete tears of the proximal hamstring tendons*. Int Orthop, 2010. **34**(1): p. 119-23.

72. van der Made, A.D., et al., *Assessment of Isometric Knee Flexor Strength Using Hand-Held Dynamometry in High-Level Rugby Players Is Intertester Reliable*. Clin J Sport Med, 2019.

73. Engelke, K., et al., *Magnetic resonance imaging techniques for the quantitative analysis of skeletal muscle: State of the art*. J Orthop Translat, 2023. **42**: p. 57-72.

74. Pons, C., et al., *Quantifying skeletal muscle volume and shape in humans using MRI: A systematic review of validity and reliability*. PLoS One, 2018. **13**(11): p. e0207847.

75. Goutallier, D., et al., *Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan*. Clin Orthop Relat Res, 1994(304): p. 78-83.

76. Fuchs, B., et al., *Fatty degeneration of the muscles of the rotator cuff: assessment by computed tomography versus magnetic resonance imaging*. J Shoulder Elbow Surg, 1999. **8**(6): p. 599-605.

77. Dixon, W.T., *Simple proton spectroscopic imaging*. Radiology, 1984. **153**(1): p. 189-94.

78. Karlsson, A., et al., *Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI*. J Magn Reson Imaging, 2015. **41**(6): p. 1558-69.

79. Widholm, P., et al., *Quantitative muscle analysis in facioscapulohumeral muscular dystrophy using whole-body fat-referenced MRI: Protocol*

development, multicenter feasibility, and repeatability. Muscle Nerve, 2022. **66**(2): p. 183-192.

80. Mukaka, M.M., *Statistics corner: A guide to appropriate use of correlation coefficient in medical research*. Malawi Med J, 2012. **24**(3): p. 69-71.

81. Harris, M. and G. Taylor, *Medical Statistics Made Easy*. 3 ed. 2014, Banbury: Scion Publishing.

82. Bowers, D., *Medical Statistics from Scratch: An Introduction for Health Professionals*. 4 ed. 2019, Newark: Wiley.

83. Yushkevich, P.A., et al., *User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability*. Neuroimage, 2006. **31**(3): p. 1116-28.

84. Best, R., et al., *[Surgical Refixation after Proximal Hamstring Tendon Avulsion Injuries: Does the Time of Surgery Influence Functional Outcomes?]*. Sportverletz Sportschaden, 2017. **31**(3): p. 160-166.

85. Harris, P.A., et al., *Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support*. J Biomed Inform, 2009. **42**(2): p. 377-81.

86. Terwee, C.B., et al., *Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties*. Osteoarthritis Cartilage, 2011. **19**(6): p. 620-33.

87. Lee, P.H., et al., *Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review*. Int J Behav Nutr Phys Act, 2011. **8**: p. 115.

88. Birmingham, P., et al., *Functional outcome after repair of proximal hamstring avulsions*. J Bone Joint Surg Am, 2011. **93**(19): p. 1819-26.

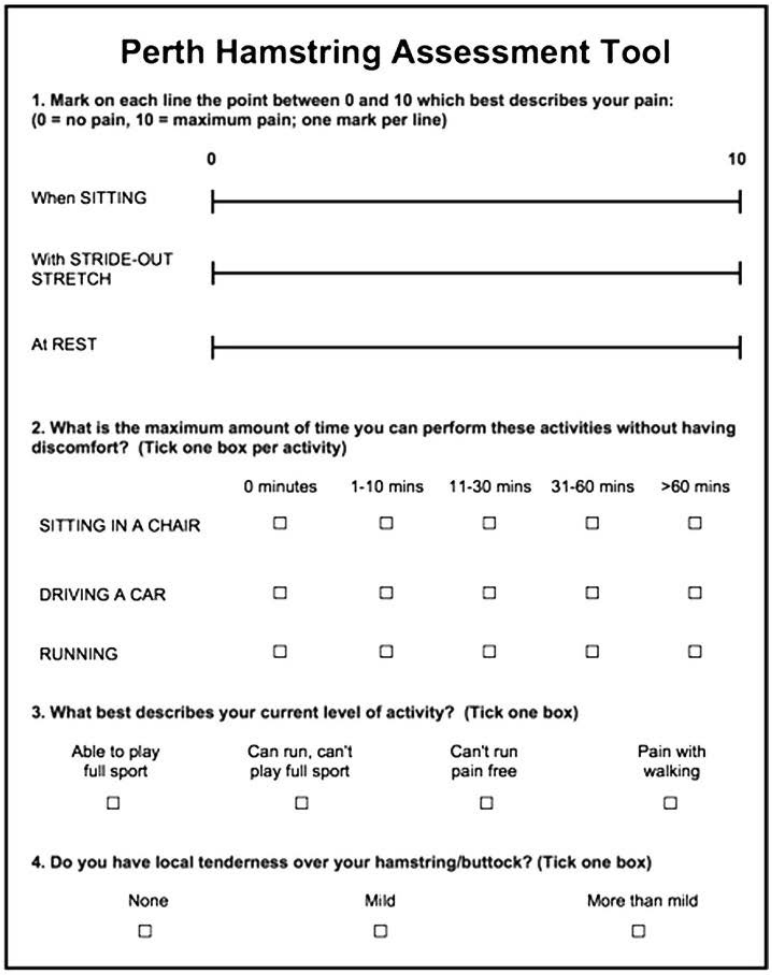
89. Heikkinen, J., et al., *Tendon Length, Calf Muscle Atrophy, and Strength Deficit After Acute Achilles Tendon Rupture: Long-Term Follow-up of Patients in a Previous Study*. J Bone Joint Surg Am, 2017. **99**(18): p. 1509-1515.

90. Mauri, L. and R.B. D'Agostino, Sr., *Challenges in the Design and Interpretation of Noninferiority Trials*. N Engl J Med, 2017. **377**(14): p. 1357-1367.

91. Aagaard, K.E., et al., *Factors associated with healing failure after early repair of acute, trauma-related rotator cuff tears*. J Shoulder Elbow Surg, 2023. **32**(10): p. 2074-2081.

11 Appendices

11.1 Perth Hamstring Assessment Tool



Perth Hamstring Assessment Tool

1. Mark on each line the point between 0 and 10 which best describes your pain:
(0 = no pain, 10 = maximum pain; one mark per line)

	0	10
When SITTING	⊢———	———⊣
With STRIDE-OUT STRETCH	⊢———	———⊣
At REST	⊢———	———⊣

2. What is the maximum amount of time you can perform these activities without having discomfort? (Tick one box per activity)

	0 minutes	1-10 mins	11-30 mins	31-60 mins	>60 mins
SITTING IN A CHAIR	☐	☐	☐	☐	☐
DRIVING A CAR	☐	☐	☐	☐	☐
RUNNING	☐	☐	☐	☐	☐

3. What best describes your current level of activity? (Tick one box)

Able to play full sport	Can run, can't play full sport	Can't run pain free	Pain with walking
☐	☐	☐	☐

4. Do you have local tenderness over your hamstring/buttock? (Tick one box)

None	Mild	More than mild
☐	☐	☐

FIGURE 2 – PHAT Scoring System:
Question 1: (10 – response) for each section. Ie Pain of 9 scores 1 point
Question 2: 0 mins=0, 1-10 mins=5, 11-30 mins=8, 31-60 mins=12, >60 mins=15
Question 3: Full sport=15, Can run=10, Can't run pain free=5, Pain with walking=0
Question 4: None=10, Mild=5, More than mild=0

11.2 Lower Extremity Functional Scale

Lower Extremity Functional Scale

We are interested in knowing whether you are having any difficulty at all with the activities listed below because of your lower limb problem for which you are currently seeking attention. Please provide an answer for **each** activity.

Today, do you or would you have any difficulty at all with:

(Circle one number on each line)

Activities	Extreme Difficulty or Unable to Perform Activity	Quite a Bit of Difficulty	Moderate Difficulty	A Little Bit of Difficulty	No Difficulty
a. Any of your usual work, housework, or school activities.	0	1	2	3	4
b. Your usual hobbies, recreational or sporting activities.	0	1	2	3	4
c. Getting into or out of the bath.	0	1	2	3	4
d. Walking between rooms.	0	1	2	3	4
e. Putting on your shoes or socks.	0	1	2	3	4
f. Squatting.	0	1	2	3	4
g. Lifting an object, like a bag of groceries from the floor.	0	1	2	3	4
h. Performing light activities around your home.	0	1	2	3	4
i. Performing heavy activities around your home.	0	1	2	3	4
j. Getting into or out of a car.	0	1	2	3	4
k. Walking 2 blocks.	0	1	2	3	4
l. Walking a mile.	0	1	2	3	4
m. Going up or down 10 stairs (about 1 flight of stairs).	0	1	2	3	4
n. Standing for 1 hour.	0	1	2	3	4
o. Sitting for 1 hour.	0	1	2	3	4
p. Running on even ground.	0	1	2	3	4
q. Running on uneven ground.	0	1	2	3	4
r. Making sharp turns while running fast.	0	1	2	3	4
s. Hopping.	0	1	2	3	4
t. Rolling over in bed.	0	1	2	3	4
Column Totals:					

SCORE: _____/80

Error (single measure): ±5 scale points
MDC: 9 scale points
MCID: 9 scale points

11.3 Patient-Specific Functional Scale

CLINICIAN TO READ AND FILL IN BELOW: Complete at the end of the history and prior to physical examination.

Initial assessment:

I am going to ask you to identify up to three important activities that you are unable to do or are having difficulty with as a result of your _____ problem. Today, are there any activities that you are unable to do or having difficulty with because of your _____ problem? (Clinician: show scale to patient and have the patient rate each activity.)

Follow-up assessments:

When I assessed you on (state previous assessment date), you told me that you had difficulty with (read all activities from list at a time). Today, do you still have difficulty with: (read and have patient score each item in the list)?

PATIENT-SPECIFIC ACTIVITY SCORING SCHEME (Point to one number):

0 1 2 3 4 5 6 7 8 9 10

Unable to perform activity.

Able to perform activity at same level as before injury or problem.

(Date and Score)

Activity	Initial					
1						
2						
3						
4						
5						
Additional						
Additional						

Figure.
Patient-specific functional scale.