

STATISTICAL ANALYSIS PLAN for the Achilles trial

Administrative information:

Sponsor name	Akershus University Hospital
Sponsor address	Sykehusveien 25, 1478 Lørenskog, Norway
EudraCT number / REC no	2012/530
Trial title	Treatment results after Achilles tendon rupture: A multi-center, randomized, clinical trial comparing conservative treatment, open repair and minimal invasive surgery.
Trial ID	The Achilles tendon project, nr 2739005
Trial registration number	NCT01785264

SAP and protocol version:

SAP version and date:	This SAP is version 1, dated 24. September 2020
Protocol version	This document has been written based on information contained in the study protocol version 1.1, dated 29. January 2013

SAP revision history:

Protocol version	SAP version	Section number changed	Description and reason for change	Date changed
1.0	1.0	NA	First edition of SAP.	24 Sep 2020

STATISTICAL ANALYSIS PLAN for the Achilles trial

ABBREVIATIONS

ASA	Acetylsalicylic acid
ATRS	Achilles tendon Total Rupture Score
BMI	Body Mass Index
CCS	Complete Case Set
CI	Confidence Interval
CMJ	Counter-Movement Jump
CT	Conservative Treatment
FAS	Full Analysis Set
LSI	Limb Symmetry Index
MCS	Mental Component Summary
MIS	Minimal Invasive Surgery
OR	Open Repair
PCS	Physical Component Summary
PF	Physical Function
PPS	Per Protocol Set
PROM	Patient Reported Outcome Measure
RCT	Randomized Controlled Trial
SD	Standard Deviation
SF-36	Short Form 36

STATISTICAL ANALYSIS PLAN for the Achilles trial

TABLE OF CONTENTS

1	INTRODUCTION	4
1.1	Background and rationale	4
1.2	Trial Objectives	4
2	TRIAL METHODS	4
2.1	Trial Design	4
2.2	Randomisation	4
2.3	Sample size	4
2.4	Statistical Framework	5
2.5	Statistical Interim Analyses and Stopping Guidance	6
2.6	Timing of Final Analysis	6
2.7	Timing of Outcome Assessments	6
3	STATISTICAL PRINCIPLES	7
3.1	Confidence Intervals and p-values	7
3.2	Adherence and Protocol Deviations	7
3.3	Analysis Populations	8
4	TRIAL POPULATION	8
4.1	Screening Data, Eligibility and Recruitment	8
4.2	Withdrawal/Follow-up	8
4.3	Baseline Patient Characteristics	9
5	ANALYSIS	9
5.1	Outcome Definitions	9
5.2	Analysis Methods	12
6	SAFETY ANALYSES	17
6.1	Adverse Events	17
6.2	Clinical Laboratory Parameters	17
6.3	Vital Signs	17
7	STATISTICAL SOFTWARE	17
8	REFERENCES	18
8.1	Literature References	18
8.2	Reference to Data Management Plan	18

1 Introduction

1.1 Background and rationale

The management of acute Achilles tendon ruptures has been the subject of a long-standing debate. Contemporary rehabilitation protocols emphasizing early weight bearing and mobilization have furthermore blurred differences in treatment results. Open repair (OR) may reduce the risk of re-rupture, but at the expense of surgical complications such as delay in wound-healing and infection. This has led to the development of percutaneous and minimal invasive surgical techniques (MIS). The demonstration of significant differences comparing OR, MIS and conservative treatment (CT), requires a large sample, and although the previously reported differences have lacked statistical significance, they may be of substantial clinical value. We have therefore conducted a three-arm RCT comparing CT, OR and MIS that is sufficiently large to be able to demonstrate significant differences.

1.2 Trial Objectives

1.2.1 Primary Objective

The primary objective is to assess if one of the three treatments is superior to at least one of the others measured by the Achilles tendon Total Rupture Score (ATRS) [2, 3] score at 1-2 years post rupture.

1.2.2 Secondary Objectives

The secondary objectives of this study are to assess if there are differences between the three treatment groups with regards to

- The ATRS at 3 and 6 months.
- The Short form 36 health survey (SF-36) [4, 5] at 6 months and 1-2 years.
- The physical test results in the Muscledab Measurement system [6] at 6 months and 1-2 years.
- Re-rupture rate between the three treatment groups at 1 year.

2 Trial Methods

2.1 Trial Design

The Achilles study is designed as a randomized, controlled, multicentre, single-country, comparative study.

2.2 Randomisation

Eligible patients were allocated in a 1:1:1 ratio between CT, OR and MIS. Treatment allocation was done by block randomization, stratified by hospital, with 6, 9 or 12 patients in each block. Details of allocation sequence generation was provided in a separate document unavailable to those who enrolled patients or assigned treatment.

2.3 Sample size

The sample size was based on the primary endpoint of ATRS at 1-2 years. The calculation was based on a one-way ANOVA analysis, assuming a standard deviation of 20 in each treatment group. A total sample size (study completers) of 480 (160 in each group) was deemed necessary to achieve 80%

STATISTICAL ANALYSIS PLAN for the Achilles trial

power to detect a difference of ATRS of 7 between the groups in two-sided tests at 5% significance level. To compensate for withdrawals/loss to follow-up, a total enrolment of 530 patients was targeted.

2.4 Statistical Framework

2.4.1 Hypothesis Test

This trial will employ a hierarchical testing procedure as follows:

1. First, a test of overall effect of treatment at 1-2 years is performed. The null hypothesis is that there is no difference in the change in ATRS score from baseline to 1-2 years between the three study arms. The alternative hypothesis is that there is a difference between at least two of the three treatment arms. The test will be performed at the (two sided) 5% significance level.
2. If the overall test shows that there is a significant difference on the 5% level, then tests of a difference in the change from baseline to 1-2 years ATRS score will be performed pairwise between the treatment arms. The null hypotheses in these tests will be that there is no difference between the treatment arms. The alternative hypotheses are that there is a difference between the two treatment arms being compared. These tests will be performed at the (two sided) 5% significance level.

The hierarchical testing procedure abides by the closed testing procedure, allowing a 5% significance level to be used in each test, while at the same time maintaining a family-wise error rate at 5%.

All other tests than described in step 1 and 2 above, will be regarded as supportive or exploratory.

Note that if the test in step 1 shows that there is no significant difference between treatment groups, a pairwise comparison between treatment groups will still be reported, but these are to be considered as secondary analyses.

2.4.2 Decision Rule

This protocol is designed to address a single primary endpoint, in a hierarchical fashion as described above. A difference in the effect of the treatment arms will be claimed if null hypothesis in step 1 of the hierarchical testing procedure outlined in Section 2.4.1 is rejected. That is, the two-sided p-value is less than 5%.

If step one of the testing procedure in step 1 in Section 2.4.1 shows significance, then tests outlined in step 2 of Section 2.4.1 will be performed.

1. Superiority of the OR group over the MIS group will be claimed if the two-sided p-value in the test comparing the change from baseline to 1-2 years ATRS score is less than 5%, and if the effect goes in favour of the OR group. Superiority of the MIS group will be claimed if the effect goes in favour of this group.
2. Superiority of the OR group over the CT group will be claimed if the two-sided p-value in the test comparing the change from baseline to 1-2 years ATRS score is less than 5%, and if the

STATISTICAL ANALYSIS PLAN for the Achilles trial

effect goes in favour of the OR group. Superiority of the CT group will be claimed if the effect goes in favour of this group.

3. Superiority of the MIS group over the CT group will be claimed if the two-sided p-value in the test comparing the change from baseline to 1-2 years ATRS score is less than 5%, and if the effect goes in favour of the MIS group. Superiority of the CT group will be claimed if the effect goes in favour of this group.

2.5 Statistical Interim Analyses and Stopping Guidance

There will be no interim analyses in this trial.

2.6 Timing of Final Analysis

The main analysis is planned when all patients have concluded a minimum of 334 days of follow up, all data up to two years have been entered, verified and validated and the primary database has been locked.

2.7 Timing of Outcome Assessments

For all clinically planned measures, visits should occur within a window of the scheduled visit. Visits outside visit window is regarded a protocol deviation. The target day and visits window are defined in the protocol as:

Visit Label	Target Day	Definition (Day window)
Diagnosis	-1	Prior to Day 0
V1. Baseline	Day 0 (Randomization)	Day 0
V2. Surgery/Casting	Day 1	< 8 days from injury
V2. 2 weeks	14 days from casting	Target day \pm 7 days
V3. 4 weeks	28 days from casting	Target day \pm 7 days
V4. 8 weeks	56 days from casting	Target day \pm 14 days
V5. 3 months	91 days from casting	Target day \pm 30 days
V6. 6 months	182 days from casting	Target day \pm 30 days
Last study visit*	364 days from casting	334 to 728

*The patients are physically tested by a physiotherapist at 1-2 years. Most often they are asked to complete the PROMs prior to the testing, either by link send by e-mail, on site on a tablet, or on paper when a tablet is not present or online. Some patients did not show up for the 1-2 years testing but did complete the PROMs. Some patients showed up for testing at 1-2 years but did not complete the questionnaires. The last study visit is defined as the last of either the 1-2 years test day or completion of the 1-2 years PROMs.

STATISTICAL ANALYSIS PLAN for the Achilles trial

For analysis and tabulation purposes, we define study time points as

Time Point Label	Target Day	Definition (Day window)
TP1. Baseline	Day 0 (Randomisation)	Information up to randomisation + 14 days
TP2. Month 3	91	Days 61 to 121
TP3. Month 6	182	Days 151 to 212
TP4. Year 1-2	364	Days 334 to 728

3 Statistical Principles

3.1 Confidence Intervals and p-values

All calculated p-values will be two-sided and compared to a 5% significance level. If a p-value is less than 0.05, the null hypothesis in the test will be discarded. Efficacy estimates for the comparison of specific treatment arms will be presented with two-sided 95% confidence intervals. This trial utilizes a hierarchical testing procedure for the primary endpoint (see Section 2.4.1), which maintains the type I error rate at 5%.

3.2 Adherence and Protocol Deviations

3.2.1 Adherence to Allocated Treatment

The number and proportion of patients that received the intervention they were randomized to will be presented.

3.2.2 Protocol Deviations

The following are pre-defined major protocol deviations regarded to affect the efficacy of the intervention:

- Entering the trial when the eligibility criteria should have prevented trial entry
 - Lack of sufficient skills in Norwegian language to understand the questionnaires.
 - Outside the age criteria (18-60 years).
 - Misdiagnosed.
 - Earlier injury to one of the Achilles tendons.
- Received other intervention than allocated to.
- Lack of technical competence for randomized surgical method.

The number (and percentage) of patients with major protocol deviations will be summarised by treatment group with details of type of deviation provided. All randomized patients will be used as the denominator to calculate the percentages. No formal statistical testing will be undertaken.

3.3 Analysis Populations

We define the following patient population in this trial.

- **All randomized patients:** All patients that have been randomized regardless if they actually received treatment or not.
- **Full analysis set (FAS):** All patients that are randomized, received treatment, and where ATRS was measured at least once post baseline (i.e. at 3 months, 6 months or 1-2 years). Patients are allocated to the treatment they were randomized to.
- **Per protocol set (PPS):** Similar to the FAS, but patients are allocated to the treatment they actually received.
- **Complete Case set (CCS):** The subset of patients in the FAS that has ATRS measurements at all follow-up visits. Patients are allocated to the treatment they were randomized to.

The FAS will be used for the primary analysis, while the PPS and the CCS will be used for sensitivity analyses.

4 Trial Population

4.1 Screening Data, Eligibility and Recruitment

The total number of screened patients and reasons for not entering the trial will be summarised and tabulated.

A CONSORT flow diagram (appendix A) will be used to summarise the number of patients who were:

- assessed for eligibility at injury
- eligible at injury
- eligible and randomised
- eligible but not randomised*
- received the randomised allocation
- did not receive the randomised allocation*
- lost to follow-up*
- randomised and included in the primary analysis
- randomised and excluded from the primary analysis*

*reasons will be provided.

4.2 Withdrawal/Follow-up

The status of eligible and randomised patients at trial end will be tabulated by treatment group according to whether they

- completed intervention, but not assessments.
- completed assessments, but not intervention.
- withdrew consent.
- lost to follow-up.

STATISTICAL ANALYSIS PLAN for the Achilles trial

- excluded due to delay from injury to surgery (>7 days).
- excluded due to protocol deviations after randomisation but occurring prior to surgery.
- Unable to measure the primary endpoint due to:
 - comorbidity that compromise rehabilitation or testing.
 - death during follow up.
 - re-rupture on the contralateral side during follow up.
 - other surgery on the lower extremities during follow up.

Time from randomisation to treatment discontinuation and time from randomisation to withdrawal/lost to follow-up will be presented graphically using a CONSORT flow diagram.

4.3 Baseline Patient Characteristics

The patient demographics and baseline characteristics to be summarised include study centre, age in years, gender, Body mass index (BMI), injured side (left/right), education, smoking status, ASA1/ASA2 and baseline ATRS.

Patient demographics and baseline characteristics will be summarised by randomised treatment arm and overall using descriptive statistics (N, mean, standard deviation, median) for continuous variables, and number and percentages of patients for categorical variables. Any clinical important imbalance between the treatment groups will be noted.

5 Analysis

5.1 Outcome Definitions

5.1.1 General Definitions and Derived Variables

5.1.1.1 Body Mass Index

Body Mass Index (BMI) = Body weight in kilograms divided by the square of the height in meters.

5.1.1.2 Achilles tendon Total Rupture Score

The acute Achilles tendon Total Rupture Score (ATRS) was developed in Sweden and published in 2007 (2). The ATRS is a patient reported outcome measure (PROM) designed to assess outcomes in patients with ruptures of the Achilles tendon. It contains ten questions, and each question is answered on an 11-point Likert scale ranging from 0 to 10. The total score is calculated by summing the individual Likert items. A score of 100 represents the absence of symptoms, whereas a score of 0 represents severe symptoms. The Norwegian version of the ATRS was validated, reliability tested and published in 2017/2018 (3). The Norwegian adaption of the ATRS demonstrates acceptable validity and reliability for use in the Norwegian population to assess clinical outcomes in patients with Achilles tendon ruptures.

If any of the 10 questions lacks an answer, this trial will follow the ATRS manual (8). Most importantly, if one or more of the 10 questions lacks an answer, the total ATRS score is treated as a missing value.

5.1.1.3 Short Form 36

STATISTICAL ANALYSIS PLAN for the Achilles trial

The Short Form 36 (SF-36) is a self-assessment health status questionnaire composed of 36 questions sorted into eight multi-item scales. The SF-36 also provides two summarized measures represented by the physical component summary (PCS) and the mental component summary (MCS). The validity and reliability of the Norwegian translation of SF-36 have been found to be satisfactory (4, 5).

5.1.2 Primary Outcome Definition

ATRS was assessed at baseline, 3 months, 6 months and 1-2 years post injury. The primary outcome is the change in the ATRS score from baseline to 1-2 years post injury. The last follow-up was planned 12 months post-injury, preferably at the 12 months test visit at one of the study physiotherapists. Patients that did not conduct the test battery or of some other reason did not complete the questionnaires at 12 months were contacted by phone and asked to answer the PROMs electronically via an encrypted link to the questionnaires sent by e-mail. In order to minimize loss to follow up the patients were allowed to answer the questionnaires up to 24 months post injury. This time extension was not done for the last included patients who did not answer the last set of PROM's or did not meet for the 12 months testing as the registration of completed forms and testing of patients was stopped 12 months after the last patient was included. An earlier publication reported only minor improvements occurring between the 1- and 2-year evaluations (7).

5.1.3 Secondary Outcomes Definitions

5.1.3.1 SF-36

The 36-Item Short Form Health Survey (SF-36) was assessed at baseline, 6 months and 1-2 years post injury. The PCS and MCS is composed of five and three subscales, respectively. The two SF-36 component summaries PCS and MCS in addition to the subscale physical function (PF) will be compared between the three treatment arms. As the ATRS and SF-36 are presented simultaneously, the time extension of up to 24 months post injury also applies to 12-month assessment of the SF-36 questionnaire. The difference in scores from baseline will be assessed at the 6 months and 1-2 years visits.

5.1.3.2 Re-rupture

Re-rupture is here defined as a total rupture of the same Achilles tendon as treated in the trial within the 12-month visit. Re-rupture will be treated as a dichotomous outcome. The diagnosis is done clinically without any need of imaging techniques. Re-ruptures were continuously reported to the principal investigator of the study. Patients suffering from either a re-rupture of the same Achilles tendon or a new rupture of the contralateral Achilles tendon were excluded from further testing and questionnaires.

5.1.3.3 ATRS score at intermittent visits

The change in ATRS scores from baseline to the intermittent visits (3 and 6 months) are regarded as continuous secondary endpoints.

5.1.3.4 Muscle lab measurements

The following 6 muscle lab measurement endpoint will be considered.

STATISTICAL ANALYSIS PLAN for the Achilles trial

1. Heel rise height (cm): For description see 3.
2. Hopping: A continuously rhythmical jump similar to skipping. The patients performed 25 jumps. The average air flight and floor contact times were documented, and the plyometric quotient (flight time/contact time) was used for data analysis.
3. Drop counter-movement jump (drop CMJ): The patients started by standing on one leg on a 20-cm-high wooden box. They were instructed to “fall” down onto the floor and, directly on landing, perform a maximum vertical one-legged jump. The maximum jumping height in centimeters was used for data analysis (see 1).
4. Concentric heel rise: For description see 5.
5. Eccentric-concentric heel rise: For the strength tests (4 and 5), a linear encoder was used. A spring-loaded string was connected to a sensor inside the linear encoder unit. When the string was pulled, the sensor gave a series of digital pulses proportional to the distance travelled. The resolution is approximately 1 pulse every 0.07 mm. By counting the number of pulses per time, the displacement as a function of time can be recorded and thus allow calculation of time, length, velocity, force, and power (force x velocity). In this experiment, the spring-loaded string of the linear encoder was attached to the heel of the participant’s shoe and thus the height (in centimeters) and time (in seconds) of the heel displacement of the heel-rise could be measured. The weight of the participant and the extra external weight were entered into the MuscleLab software and peak power in watts was calculated. The best trial (ie, with the highest power in watts) for each weight was used for data analysis.
6. Muscular endurance test: A standing heel-rise test. The total amount of work performed (in joules) and the maximum heel-rise height were used for data analysis.

For each of these endpoints, a similar measurement for the healthy Achilles is performed. The ratio of the test result in the treated foot versus the healthy foot is calculated and multiplied with 100. This measurement is referred to as the Limb symmetry index (LSI). These are also considered as secondary endpoints. Furthermore, the physical tests are performed both at the 6-month visit and at the 1-2 years visit, the results of which are both considered as secondary endpoints. Furthermore, the change in the test results, from the 6-month to the 1-2 years visit, are also considered secondary endpoints. Thus, there are 36 secondary endpoints relating to the physical tests.

5.1.4 Overview of Outcomes

Level	Outcome	Timeframe	Type
Primary	ATRS	1-2 years	Continuous
Secondary	SF-36	1-2 years	Continuous
	Re-rupture rate	12 months	Dichotomous
	Muscle-lab measurement system (see 5.1.3.4 for description)	6 and 1-2 years	Continuous

5.2 Analysis Methods

5.2.1 Primary Outcome

5.2.1.1 Primary Analysis

The primary outcome in this trial is the change in the ATRS score from baseline to the 1-2 years visit.

The primary analysis will be performed by applying a linear mixed effects model to model the longitudinal change from baseline ATRS measurements at the 3-month, 6-month and 1-2-year visits. A fixed interaction term between the follow-up visit and treatment group will be included in the model. The baseline ATRS measurement will be included as a fixed covariate in the model. Study centre, which was used as a stratification variable in the randomization, will also be included as a fixed covariate. An individual-specific random intercept term will be included in the model to account for the dependencies within individuals. An unstructured correlation structure will be assumed for the repeated measurements.

From the fitted model, the estimated marginal means will be computed for each time point and each treatment. Based on the estimated marginal means, an omnibus test, where the null hypothesis is that there is no difference between the three treatment groups, will be performed for the 1-2-year visit. This corresponds to step 1 in Section 2.4.1. If the test is statistically significant on the 5% level, pairwise testing between the three treatment groups will be performed for the 1-2-year visit (step 2 in Section 2.4.1). A 5% significance level will be used in each of these tests. As described in Sections 2.4.1-2.4.2, the hierarchical testing procedure maintains the overall type I error rate at 5%. The estimated marginal means with 95% confidence intervals of the change in ATRS from baseline will be presented for each of the treatment groups for the 1-2 years visit.

If the omnibus test at the 1-2 years visit does not show statistical significance, then no further hypothesis testing will be performed. However, the estimated marginal means and the accompanying 95% confidence intervals will still be presented, but should be considered as secondary analyses.

The analysis will be conducted in the FAS.

5.2.1.2 Summary Measures

The primary effect estimate will be the change in the ATRS from baseline to the 1-2 years visit. If the omnibus test of a treatment difference at the 1-2 years visit is significant, the estimated marginal mean change in ATRS from baseline will be displayed for the three treatment together with the p-values of the pairwise tests of the three treatment groups (See Sections 2.4.1, 2.4.2 and 5.2.1.1). If the omnibus test does not identify a treatment difference, then no pairwise testing will be performed, and no p-values will be presented. However, the 1-2 years estimated marginal mean change in ATRS from baseline will be reported for each treatment group, together with 95% confidence intervals, but these should be regarded as secondary analyses.

5.2.1.3 Assumption Checks and Alternative Analyses

Assumption checks of the fitted model will be done by visual inspection of the residuals. Marginal residuals will be plotted against the predicted marginal responses, and a check for trend will be performed. To check if the dependencies in the data has been sufficiently accounted for by the fitted

STATISTICAL ANALYSIS PLAN for the Achilles trial

model, inspection of scatter plots of conditional residuals will be done for all combinations of visits (i.e. visit 3-month vs. 6-month, 3-month vs 1-2 years and 6-month vs. 1-2 years). Normality of conditional residuals will be assessed by inspecting qq-plots for each visit.

If there are substantial deviation for normality, or there are indication of substantial residual dependence, then further random effects will be introduced in the mixed model.

5.2.1.4 Missing Data

At baseline and at the three follow-up visits (3-month, 6-month and 1-2-year), patients answer the ATRS questionnaire. Each questionnaire contains ten questions, and each question is answered on an 11-point Likert scale ranging from 0 to 10. The total score is calculated by summing the individual Likert items. For each question, the value is set according to the guidelines in the ATRS manual (8). Importantly, if at least one of the 10 questions lack an answer, the ATRS is regarded as missing for the patient for that particular visit.

Missing baseline ATRS data will be imputed using mean imputation (9). If baseline ATRS are skewed, median imputation will be used instead.

Missing values in follow-up ATRS measurements (at the 3-month, 6-month and 1-2-year visits) will be handled by the mixed modelling approach. However, individuals with no follow-up measurements will then be excluded from the analysis. If the proportion of patients that has no follow-up measurements is less than 5%, no action will be taken to include these in the analysis. If this proportion is larger than 5%, multiple imputation will be used to impute the 1-2 years ATRS for these individuals.

5.2.1.5 Sensitivity Analyses

The following sensitivity analyses will be performed:

1. If there are patients with missing baseline ATRS, a sensitivity analysis will be conducted in the patients that has baseline a measurement (i.e. excluding the patients with missing baseline measurement).
2. If the missing baseline ATRS measurements were imputed using median imputation (see Section 5.2.1.4), a sensitivity analysis will be conducted using mean imputation instead.
3. The ATRS questionnaire is known to sometimes be misinterpreted, and inverted, by patients (10, 11). If the study group suspects that this has happened for one or more patients at any time point, a sensitivity analysis will be conducted where these possibly inverted ATRS measurements are inverted back. I.e. if a patient e.g. has answered the value 2 on an item in the questionnaire, this is converted to $10-2=8$. This is only done for patients where the study group suspects that the scale has been inverted.
4. A sensitivity analysis will be conducted in the PPS, the per protocol set. That is, if any patients received a different treatment than randomized to, they will in this sensitivity analysis be placed in the treatment group corresponding to the treatment they actually received.
5. A sensitivity analysis will be conducted in patients that has complete follow-up ATRS measurements, the complete case set (CCS).

5.2.1.6 Subgroup Analyses

No subgroup analyses will be performed.

5.2.2 Dichotomous Secondary Outcome: Re-ruptures

5.2.2.1 Main Analysis

The probability of re-rupture will be compared pairwise between the three treatment groups.

A logistic regression model will be fitted to the data, where the treatment variable will be adjusted by study centre (the stratification factor used in the randomization). From the model, the estimated marginal means of the risk difference between the groups will be reported pairwise.

We expect that the number of re-ruptures will be low, which may lead to the normality approximation to the binomial distribution becoming questionable. If any of the expected numbers in one of the cells in the 3-by-2 contingency table, summarizing the number re-ruptures in the treatment groups, is less than 5, the following analysis will be performed in place of fitting the logistic regression model: The confidence intervals of the risk difference between the groups will pairwise be calculated using the Newcombe Hybrid Score[12]. Thus, the risk difference will not be adjusted for study centre.

The analysis will be conducted in the FAS.

5.2.2.2 Summary Measures

The 3-by-2 Contingency table (see Section 5.2.2.1) will be presented.

5.2.2.3 Assumption Checks

A check of the expected numbers in the 3-by-2 Contingency table (see Section 5.2.2.1) will be done. If any of these numbers are less than 5, the non-parametric Newcombe Hybrid Score will be used to calculate the risk differences between the groups pairwise

5.2.2.4 Missing Data

All patients are followed-up for re-ruptures until 12 months. The participating hospitals are instructed to report any re-rupture (total or partial) detected during the follow up period directly to the principal investigator. This will secure that the chance of missing the count of any re-rupture is regarded as insignificant.

5.2.2.5 Sensitivity Analyses

No sensitivity analyses will be performed.

5.2.2.6 Subgroup Analyses

No subgroup analyses will be performed.

STATISTICAL ANALYSIS PLAN for the Achilles trial

5.2.3 Continuous Secondary Outcome: SF-36v2

5.2.3.1 Main Analysis

The outcome is the change in the SF-36 score from baseline to the 6-month and the 1-2-year visit. The two summary scores, PCS and MCS, as well as the sub-score PF will be considered. Thus, there are three SF-36-based secondary outcome variables. Note that the SF-36 scores will be standardized to the 1998 US population, as provided by the ProCore software by Qualitymetric (<https://www.optum.com/business/solutions/life-sciences/answer-research/patient-insights/sf-health-surveys.html>).

The analyses will be conducted in the FAS.

5.2.3.2 Summary Measures

The 6-month and 1-2-year visit estimated marginal mean change in SF-36 score from baseline will be reported for each treatment group, together with 95% confidence intervals.

5.2.3.3 Assumption Checks

Similar as for the primary endpoint.

5.2.3.4 Missing Data

Partially answered SF-36 questionnaires will be treated according to the SF-36v2 user's manual [13]. Note that the manual allows for a certain degree of missingness in a given questionnaire. If the manual deemed a questionnaire as not sufficiently answered, then the score value for the individual at the visit in question will be treated as a missing value, in the same manner as completely unanswered questionnaires.

Missing baseline SF-36 score data will be imputed using mean imputation (9). Missing values in follow-up SF-36 scores (at the 6-month and the 1-2 years visits) will be handled by the mixed modelling approach. However, individuals with no follow-up measurements will then be excluded from the analysis. If the proportion of patients that has no follow-up measurements is less than 5%, no action will be taken to include these in the analysis. If this proportion is larger than 5%, multiple imputation will be used to impute SF-36 scores for these individuals.

5.2.3.5 Sensitivity Analyses

No sensitivity analyses will be performed.

5.2.3.6 Subgroup Analyses

No subgroup analyses will be performed.

5.2.4 Continuous Secondary Outcome: Muscle lab measurements

The six muscle lab measurement variables will be analysed in the same way.

5.2.4.1 Main Analysis

There are three outcomes for each physical test:

- The difference between the 6-month and the 1-2 years test performance.
- The 6-month test performance.

STATISTICAL ANALYSIS PLAN for the Achilles trial

- The 1-2-year test performance.

Linear regression models will be fitted to each of these outcomes, for each physical test. The treatment variable will be adjusted by study centre (the stratification variable used in the randomization). Note that for each of the 6 physical tests at each visit, a test is also performed in the healthy foot. The ratio between the performance in the treated foot and the healthy foot, multiplied with 100, (the LSI) is also considered an outcome variable for each physical test.

The analyses will be conducted in the FAS.

5.2.4.2 Summary Measures

The effect estimates will be the pairwise differences in the estimated marginal means (that is, mean difference adjusted for study centre) between the treatment groups, reported together with 95% confidence intervals.

5.2.4.3 Assumption Checks

Visual inspection of residual plots will be performed (qq-plots, histograms and residuals vs. fitted values). Substantial deviations from normality will be addressed by suitable transformations of the data.

5.2.4.4 Missing Data

We expect three types of missing data for these outcome variables:

1. Individuals is present in the gym, tries to perform the test, but do not manage to fulfil the minimum requirements to have the test approved. These values are in reality censored rather than missing.
2. Value may be missing due to e.g. non-working equipment in the gym where the test takes place, even if the patient were physically there.
3. Values may be missing because the patient did not show up for the physical tests.

The following handling of the three types of missingness will be performed:

1. Threshold imputation: These values will be imputed in the following way:
 - I. The 1% highest and 1% lowest observed values for the variable will be removed, and a normal distribution will be fitted to the 'trimmed' observations.
 - II. Values "klarte ikke" (did not manage) will be replaced by drawing values from the distribution in point I, truncated by 0 and the 10th percentile of the distribution.
2. These values are assumed to be missing at random, and multiple imputation by chained equation will be performed.
3. Multiple imputation using chained equations will be used to impute these values.

The multiple imputation in point 2 and 3 will be done simultaneously, after point 1 has been performed. The following variables will be used to impute the test scores at 6 months and 1-2 years: Sex, Age, Weight at baseline, weight at the time of the test, study centre, results in the other 5 physical tests, results from all 6 physical test in the healthy foot, and ATRS scores at 6 and 1-2 years.

STATISTICAL ANALYSIS PLAN for the Achilles trial

Note that the LSI variables, which is the ratio between the injured and the healthy foot multiplied by 100, will not be imputed as described above. For the threshold imputation, the value in the injured and healthy foot will be imputed separately, and the LSI value is obtained from these. In the multiple imputation procedure, the LSI variables for the different tests will be imputed by passive imputation.

5.2.4.5 Sensitivity Analyses

Performing the analysis described in Section 5.2.4.1 on the population of patients that showed up for the physical test. I. e. after the threshold imputation described in Section 5.2.4.4 has been performed.

5.2.4.6 Subgroup Analyses

No subgroup analyses will be performed.

5.2.5 Time to event secondary outcomes

Not applicable.

5.2.6 Additional Analyses

Not applicable.

6 Safety Analyses

Complications and adverse events were continuously registered. For instance, thromboembolic events, wound healing problems, infections and nerve-damage.

6.1 Adverse Events

Any complications and adverse events were continuously documented and any new rupture on the injured side (re-rupture) or the healthy side were immediately reported to the study coordinator. Re-ruptures are considered as a secondary outcome (see Section 5.1.3.2). Adverse events other than re-ruptures will be tabulated.

6.2 Clinical Laboratory Parameters

Not applicable.

6.3 Vital Signs

Not applicable.

7 Statistical Software

All statistical analyses will be done using R version 3.6.3 (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>).

8 References

8.1 Literature References

1. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. 2017;318(23):2337-2343. doi:10.1001/jama.2017.18556.
2. Nilsson-Helander K, Thomee R, Silbernagel KG, Thomee P, Faxen E, Eriksson BI et al (2007) The Achilles tendon Total Rupture Score (ATRS): development and validation. Am J Sports Med 35:421-426
3. Myhrvold SB, Sandnes Ø, Hoelsbrekken SE (2018) Validity and reliability of the Norwegian translation of the Achilles tendon Total Rupture Score. Knee Surg Sports Traumatol Arthrosc 26:2045-2050
4. Tests of Data Quality, Scaling Assumptions, and Reliability of the SF-36 in Eleven Countries: Results From the IQOLA Project, International Quality of Life Assessment. J Clin Epidemiol. 1998 Nov;51(11):1149-58. doi: 10.1016/s0895-4356(98)00106-1
5. Measurement properties and normative data for the Norwegian SF-36: results from a general population survey. Health and Quality of Life Outcomes 15, Article number: 51 (2017)
6. Evaluation of Lower Leg Function in Patients With Achilles Tendinopathy. Knee Surg Sports Traumatol Arthrosc. 2006 Nov;14(11):1207-17.doi: 10.1007/s00167-006-0150-6
7. Major Functional Deficits Persist 2 Years After Acute Achilles Tendon Rupture. Kne Surg Sports Traumatol Arthrosc. 2011 Aug;19(8):1385-93. doi: 10.1007/s00167-011-1511-3
8. Performance of the Achilles Tendon Total Rupture Score Over Time in a Large National Database: Development of an Instruction Manual for Accurate Use. Am J Sports Med. 2020 May;48(6):1423-1429. doi: 10.1177/0363546520912222.
9. Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials?. *Statistical methods in medical research*, 27(9), 2610-2626).
10. Cross cultural adaptation of the Achilles tendon total rupture score with reliability, validity and responsiveness evaluation. Knee Surg Sports Traumatol Arthrosc 2013;21(6):1356-1360.
11. Reliability and validation of the Dutch Achilles tendon Total Rupture Score. Knee Surgery, Sports Traumatology, Arthroscopy. July 2016
12. Fagerland, Morten, Stian Lydersen, and Petter Laake. *Statistical analysis of contingency tables*. CRC press, 2017.
13. User's manual for the SF36v2 Health Survey. ME Maruish, M Maruish, M Kosinski, JB Bjorner, B Gandek, DM Turner-Bowker, JE Ware. 2011. www.scienceopen.com.

8.2 Reference to Data Handling Plan

Not applicable.