Statistical Analysis Plan	Version 1 12/10/2023
SAP revision history	None
SAP contributors with roles and responsibilities:	
Person writing SAP	Jacob Fyhring Mortensen
Senior statistician	Paul Blanche
Principal investigator	Anders Odgaard
Brief Title from Clinicaltrials.org:	Medial Unicondylar Knee Arthroplasty vs Total Knee Arthroplasty
Study title that matches the study protocol:	Randomized clinical trial of medial unicompartmentel versus total arthroplasty for medial tibio-femoral OA.
Version of protocol referenced:	English Version 3, March 4, 2017 Amended on the 08-01-2020 with filename: English Protocol-rent tillæg 2019-BBH-randers-silkeborg-horsens
ClinicalTrials.gov ID:	NCT03396640
Unique Protocol ID:	H-16037372

Author details:

Name	Roles & responsibility	Affiliation
Jacob Fyhring Mortensen	First author, project initiator and coordinator	1,2
Paul Blanche	Co-author, Statistician	3
Claes Sjørslev Blom	Co-author, data collection	4
Andreas Kappel	Co-author, data collection, protocol draft	5
Per Wagner Kristensen	Co-author, protocol draft	6
Martin Lindberg-Larsen	Co-author, data collection	7
Xiaodong Ma	Co-author, data collection	8
Frank Madsen	Co-author, data collection, protocol draft	9
Henrik Schrøder	Co-author, data collection, protocol draft	10
Snorre Stephensen	Co-author, data collection, protocol draft	1
Morten Vase	Co-author, data collection	11
Lasse Enkebølle Rasmussen	Co-author, data collection, protocol draft	6
Svend Erik Østgaard	Co-author, data collection, protocol draft	5
Anders Odgaard	Senior Author, project initiator, main investigator	12

Gentofte Hospital[1], Nykøbing Falster Sygehus[2], University of Copenhagen[3], Randers Sygehus[4], Aalborg Universitetshospital[5], Vejle Sygehus[6], Odense Univesitets Hospital[7], Bispebjerg Hospital[8], Århus Universitetshospital[9], Næstved Sygehus[10], Silkeborg Sygehus[11], Rigshospitalet[12].

Table of Contents

Introduction: Background, rationale, objectives and hypotheses, study type, scope of study

In the surgical treatment of isolated anteromedial unicompartmental osteoarthritis of the knee, it is possible to choose between well-documented treatmens such as a medial unicompartmental knee arthroplasty (mUKA), or a total knee arthroplasty (TKA). The demand for a blinded multicenter randomized clinical trial (RCT) with the comparison of mUKA and TKA has been increasing in recent years, to determine which treatment is better, as only few well-designed studies have been performed. Supporters of TKA suggest that this treatment gives more predictable results, whereas supporters of UKA suggest that it is unnecessary to remove functional articular cartilage in other non-affected compartments. If the mUKA wears or loosens, revision surgery will be relatively easy, whereas revision surgery after a TKA can be more demanding .

“mUKA vs TKA” is a double-blinded multicenter Randomized Clinical Trial. Ten hospitals throughout the five administrative regions of Denmark have participated in the study. It has been planned to include 350 patients prospectively. To limit bias, all participants except the theatre-staff were blinded .

Follow-up are with PROM-questionnaires and clinical controls up to 20 years seen in Tabel 1.

It was planned that the results will be assessed and published at different timepoints after inclusion of the last patient:

1) PROM-questionnaires (at 2, 6, 10, 20 years)

2) Clinical assessment of knee condition (at 2, 10, 20 years)

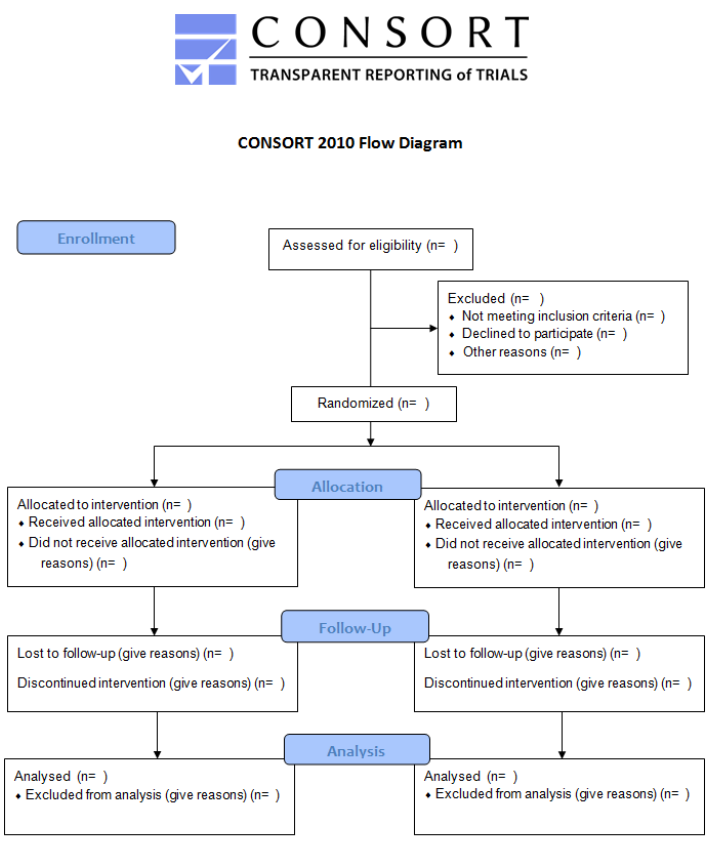
3) cost analysis. (at 5 years)

In this study we aimed for long-term evidence. Patients live longer, and outcomes should be assessed with a life-long perspective. Hence, we aimed for a total of 20 years follow-up. Another strength of our study is that the fate (revision, reoperation, death, emigration) is closely monitored in Denmark, thanks to extensive use of national registers and unique identifiers for all residents.

Presentation of withdrawal and follow-up data

A flowchart adhering to the consort-statement will be produced, as seen below in Figur 1.

Figur 1: Consort flow diagram, template



Baseline patient characteristics and how they will be descriptively summarized

Baseline data for age, sex, body mass index, previous procedures and clinical findings before randomization will be presented using descriptive statistics only, within each treatment group. No p-values or other inferential statistics will be used to compare the distribution of the baseline variables between two groups. This is in line with standard recommendations, e.g., those from CONSORT. Social factors such as accommodation, marital status, work status, and more will also be included in baseline summarization. PRO data and clinically observed data at baseline will also be summarized per group. This data will be reported as mean (standard deviation [SD]), median (interquartile range [IQR]), or count (percent) as appropriate. All of the above mentioned data will be presented in a table similar to Tabel 2.

Definitions of outcomes and sequence of measurement, specific measurements and units used, analysis method and expected presentation for each outcome effect.

Sequences of measurement of different outcomes can be seen in Figur 4.

All analyses will be based on an 'Intention-to-treat' basis and outcome results will be presented in a table similar to Tabel 3 (more details are presented below).

Statistical significance will be judged via either the presented 95% confidence intervals and matching p-values or adjusted p-values for multiple testing, as appropriate (see details below). For all analyses of area under a score curve (AUC, see below), we will divide the AUC by the observation time, to obtain an (easier) interpretation of the difference in AUCs as an average time-weighted improvement for the individual.

Primary outcome:

Area under the Oxford Knee Score curve within 2 years postoperatively

Scale range 0 (severe arthritis) - 48 (satisfactory joint function).

The primary clinical question of interest is: What is the treatment difference between UKA and TKA in the average area under the Oxford Knee Score curve within 2 years postoperatively, among patients alive 2 years after surgery?

The estimand is described by the following attributes:

- Treatment condition: the investigational intervention (UKA or TKA) regardless of post-surgery treatment and/or revision surgery.
- Population: all patients treated for anteromedial knee unicompartmental osteoarthritis alive 2 years after surgery.
- Endpoint: area under the Oxford Knee Score curve within 2 years postoperatively
- Remaining intercurrent events: the intercurrent events are addressed by the treatment condition of interest attribute. There are no remaining intercurrent events.
- Population-level summary: difference in mean between treatment conditions

Rationale for estimand: it aims at reflecting how patients are treated in clinical practice.

Primary outcome statistical Analysis: We will use a mixed-effects model for repeated measurements (MMRM), which is equivalent to a usual and simple ANCOVA model when there is no missing data. However, it is more suitable to handle missing data assumed missing at random . We will model the mean of the Oxford knee score at each visit given baseline covariates and treatment via this model. From the fit of the model, we will deduce the estimated mean difference in the area under the curve of the score as the difference of the area under the curve of the mean differences between the treatment groups. Similarly, we will compute a 95% confidence interval and matching p-value via usual Wald-type inference.

For this model, the repeated measurement outcome will be the Oxford score (at baseline and each visit) and covariates will consist of the treatment group, the follow-up time at which the score was measured, sex, hospital and age. For the hospital variable, we will merge the two locations "Frederikshavn" and "Farsø" into one group (e.g. labelled "Other in north Jutland") as very few surgeries took place at each of these two sites. In the unlikely case in which convergence issues would arise due to the site variable, we will further merge all sites with less than 10 surgeries (or 15, if needed) into one group.

We will use interaction terms between all follow-up visit times and treatment, to model a separate mean difference in Oxford score between the treatment group at each follow-up visit without restrictions. However, no difference at baseline will be modelled to reflect the randomization, as often recommended (2). Also, we will use interaction terms between all baseline covariates (I.e. sex, hospital and age) and follow-up time, to model different associations between the Oxford score at each time and the baseline covariates.

No other interaction terms will be used in the model. We will use a usual unstructured covariance matrix to model the variance-covariance matrix of the repeatedly measured outcome. This implies that we will model different variances in Oxford score at different follow-up visits and also different correlations between the scores measured at different follow-up visits, without restrictions, as it is also common (2, 3). Again, this model is equivalent to a usual and simple ANCOVA model when fitted with complete data. We will fit the model using restricted maximum likelihood (REML). Interestingly, this approach is often recommended and it has been shown to be robust against model-misspecification, when the sample size is relatively large, as it is in our study .

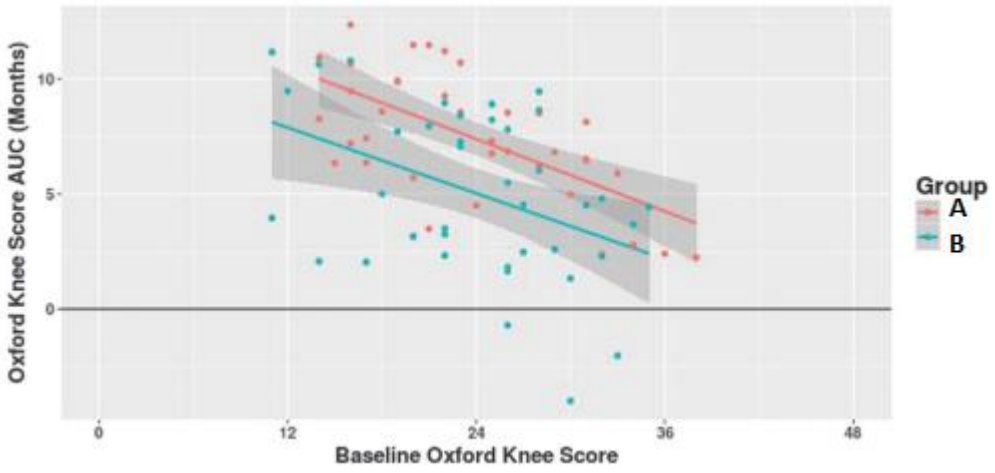
Analysis Sets: all randomized participants alive 2 years after surgery.

Software: we plan to use R and especially the package “nlme” and its function “gls”, to fit the MMRM model (Wang et al., 2019; Wang et al., 2021)).

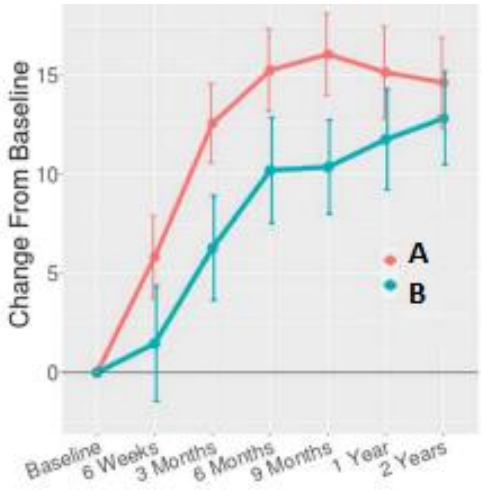
For making graphical representations of the main results obtained from the MMRM model described above, we will consider producing a similar figure to Figur 2 below. That is, a scatter plot displaying baseline Oxford score (x-axis) versus AUC (y-axis) for each patient, with one color per treatment group. In addition, we will consider adding two regression lines, which will represent the estimated mean AUC for any baseline score (within the range of the observed baseline values) and each treatment, when all other covariates entering the model are set to the observed mean value (rounded) for continuous covariates (age) or the level observed the most frequently (for categorical covariates). Confidence intervals will be displayed around the regression lines (computed from the same MMRM model fit).

In addition, to present the repeated measurements data on Oxford knee score collected up to 2 years, we will consider producing a figure similar to Figur 3 below. It will represent the curves of the mean change in score from baseline at each follow-up visit in each treatment group, for the same “average aged” patient as defined above. Confidence intervals around each curve at each follow-up visit will be added (computed from the same MMRM model fit).

Figur 2: AUC for primary outcome comparing interventions on improvement curve (generic example, not related to this study)



Figur 3: Paired differences relative to baseline for primary outcome (generic example, not related to this study)



Secondary outcomes:

We have twelve secondary outcomes (S1-S12, see Figure 4), which will be presented inTabel 3. The first is:

S1) the Oxford Knee Score at 2 years postoperatively.

The corresponding clinical question of interest is: What is the treatment difference between UKA and TKA in the average Oxford Knee score at 2 years postoperatively, among patients alive 2 years after surgery? This is similar to the research question about the primary outcome, except that here we are not interested in the AUC but in the single value at 2 years postoperatively. We will use a similar estimand as for the primary analysis (only the endpoint differs, now it is the score at 2 years, no longer the AUC) and the exact same analysis set and MMRM model will be used.

Ten of the eleven other secondary outcomes S2-S12 (I.e all but S5) consist of the area under the curve of the following scores, within 2 years postoperatively (similar AUC as for the primary outcome):

S2) Forgotten Joint Score, which also ranges from 0 to 100. High scores indicate a high degree of "forgetting" the artificial joint—i.e. a low degree of awareness.

S3) Flexion score of the Copenhagen Knee ROM scale. It measures how much a patient can bend his or her knee. Instead of the values 1, 2, 3, 4, 5 and 6 of the score, we will use the mid-points of the corresponding intervals for the angle. The intervals are 0-60, 60-75, 75-90, 90-105,105-120, 120-135 {Mørup-Petersen, 2018 #304}

S4) Range of movement, which ranges from 0 to 160. It is a measure of the passive knee range of movement in degrees, using a standard (30 cm) goniometer. It is calculated as the difference of two values recorded by the surgeon.

S6) Symptoms score of the Knee Osteoarthritis Outcome Score (KOOS). This is a sum-score ranging from 0 (worst) to 100 (best) calculated from the answer of the patient to seven questions.

S7) Bodily pain score of the Short Form (36) Health Survey (SF36). The higher the score the less bodily pain i.e., a score of zero is equivalent to maximum Bodily pain and a score of 100 is equivalent to no bodily pain.

S8) Physical functioning score of the SF36, which also ranges from 0 to 100, the higher the better.

S9) Pain score of the KOOS scale. It is also a sum-score ranging from 0 (worst) to 100 (best). It is calculated from the answer of the patient to nine questions. .

S10) Vitality score of the SF36, which also ranges from 0 to 100, the higher the more vitality and less energy/fatigue.

S11) Pain score of the EQ5D (subscale of EQ5D), which ranges from 1 to 5(1: no problems, 2:slight problems, 3:moderate problems, 4:severe problems and 5: extreme problems).

S12) UCLA activity scale, which ranges from 1 to 10. It is an assessment of the patient activity outcome evaluations of lower extremity joint reconstructions on a scale from 1-10, 10 being extremely high activity level within impact sports.

For each of these scores, the clinical question of interest is: What is the treatment difference between UKA and TKA in the average area under the score curve within 2 years postoperatively, among patients alive 2 years after surgery? (Similar to the research question about the primary outcome).

We will use similar estimands and analysis sets as those corresponding to the primary outcome are considered, for the same reasons.

For each of these scores, the corresponding statistical analysis will also be similar to that of the primary score (Oxford knee score). The same kind of MMRM will be fitted and used to compute the mean difference in AUCs, confidence intervals and p-values (and to make similar graphical representations). The same modelling strategy with regards to baseline covariates adjustment, inclusion of interaction terms etc will be used, for the same reasons. Only for the forgotten joint score (S2) will the modelling differ by a minor difference: the baseline value will not be used to fit the model (as it is not expected to be strongly predictive of the follow-up measurements, .

Additionally, the fifth secondary outcome is:

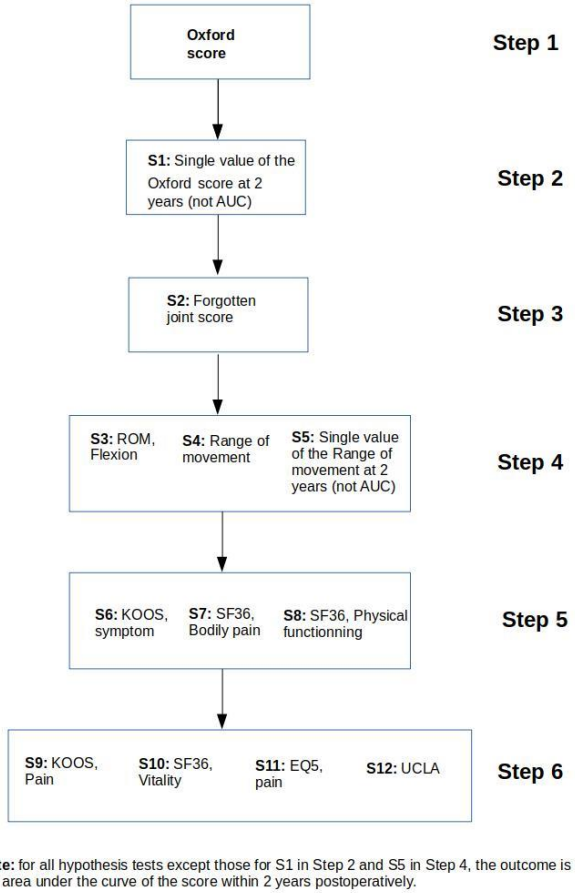
S5) the range of movement at 2 years postoperatively.

The corresponding clinical question of interest is: What is the treatment difference between UKA and TKA in the average range of movement score at 2 years postoperatively, among patients alive 2 years after surgery? This is similar to the research question about the secondary outcome S4, except that here we are not interested in the AUC but in the single value at 2 years postoperatively. We will use a similar estimand as for the secondary outcome S4 (only the endpoint differs, now it is the score at 2 years, no longer the AUC) and the exact same analysis set and MMRM model will be used.

Multiplicity adjustment in statistical hypotheses testing.

To obtain a strong control of the family-wise error rate (FWER) at 5%, we will use a Bonferroni parallel gatekeeping procedure, with equally weighted hypotheses, as described details in Section 4 of . We will test the null hypotheses sequentially, in six steps, as graphically illustrated on **Fejl! Henvisningskilde ikke fundet.** and detailed below.

Figur 4 Sequential testing to define the parallel gatekeeping procedure.



- Step 1: We will first compute a p-value for the primary outcome (Step 1). If it is not significant ($p>0.05$), we will consider the results of the remaining hypothesis tests as also not statistically significant ($p>0.05$). If it is significant ($p<0.05$), then we will continue to Step 2.

- Step 2: We will compute a p-value for the secondary outcome S1. If it is not significant (p>0.05), we will consider the results of the remaining hypothesis tests as also not statistically significant (p>0.05). If it is significant (p<0.05), then we will continue to Step 3.
- Step 3: We will compute a p-value for the secondary outcome S2. If it is not significant (p>0.05), we will consider the results of the remaining hypothesis tests as also not statistically significant (p>0.05). If it is significant (p<0.05), then we will continue to Step 4.
- Step 4: We will compute a p-value for each of the secondary outcomes S3, S4 and S5. Here we will compute p-values "Bonferroni adjusted" for multiple testing (as detailed below), because there are 3 hypothesis tests. If none of the three p-values are significant, we will consider the results of the remaining hypothesis tests as also not statistically significant (p>0.05). If at least one is significant, we will continue to Step 5.
- Step 5: We will compute a p-value for each of the secondary outcomes S6, S7 and S8. Here again, we will compute p-values "Bonferroni adjusted" for the multiple testing because there are3 hypothesis tests. To judge statistical significance at this step 5, we will keep using the 5% threshold only if all previously tested statistical hypotheses were statistically significant (p<0.05). If only k=1 among the 3 tested null hypotheses at step 4 were significant, we will use the significance threshold (1/3)*5%=1.67% instead. If k=2 were significant, we will use the threshold (2/3)*5%=3.33%. That is, with k denoting the number of previously rejected hypotheses at step 4 in any case, we will use the significance threshold (k/3)*5%. If none of the three p-values are significant, we will consider the results of the remaining hypothesis tests as also not statistically significant (p>0.05). If at least one is significant, we will continue to Step 6.
- Step 6: We will compute a p-value for each of the secondary outcomes S9, S10, S11 and S12. Here again, we will compute p-values "Bonferroni adjusted" for multiple testing because there are 4 hypothesis tests. In addition, here again we will use an appropriate significant threshold, lower than 5% if not all null hypotheses have been rejected at steps 4 or 5. Specifically, let k and l denote the number of null hypotheses that have been rejected at steps 4 and 5; the significance threshold at step 6 is now defined as (k/3)*(l/3)*5%.

Note that to compute p-values "Bonferroni adjusted" for multiple testing at steps 4, 5 and 6, we will multiply the unadjusted p-values by 3 at steps 4 and 5 and by 4 at step 6.

The above Bonferroni parallel gatekeeping procedure is described in details in Section 4 of .

Additional outcomes: complications and serious adverse events (SAE)

The main adverse events and complications will be summarized in a table such as Tabel 4. For the outcomes reoperaton, revision surgery and death, we will report counts and percentages, i.e. estimates of 2-year risks, together with exact binomial confidence intervals. Exact unconditional confidence intervals for the risk difference between treatment groups will also be reported, using Boschloo's test methodology (as implemented in the function "uncondExact2x2" of the R package "exact2x2" . No method to adjust for multiple testing will be used for these exploratory end points. Hence, reporting will be limited to point estimates of effects with 95% confidence intervals. For completeness, we will also present the counts and percentages of each type of reoperation and revision surgery observed in each treatment group.

Assessment of blinding

We will report descriptive statistics to assess the extent to which blinding during the first year seemed to have been as effective as anticipated. Specifically, for each treatment group, we will report:

- For how many patients was the randomization revealed within the first year? (Blinding was planned within the first year only)
- For each follow-up visit at which the information was collected (I.e. 2 weeks, 4 months, and 1 year after surgery), the proportion of patients who answered "yes" were also asked whether they knew which implant they had received to control for correctness.
- For each case in which the implant was revealed, the reason and timing of the premature reveal would be noted.

Calculations or transformations used to derive outcome

None.

Methods to check for distributional assumptions

The statistical analysis based on MMRM is known to be robust against model misspecification, asymptotically . As the sample size is relatively large (n=350), no model checking is pre-specified.

Method for handling missing data

The MMRM model was chosen specifically to handle missing data optimally. It will be fitted with the observed data, which will possibly be unbalanced if there are some missing data. This approach is often recommended .

Subgroup definition and analysis, if applicable

We plan to perform subgroup analyses in the following subgroups:

- patients aged above 70 and below 70

- patients with body mass index (BMI) above 30 and below 30

The analysis of the primary outcome and those of the secondary outcomes (S1,..,S12) for which a statistically significant difference was found, will be re-done within these subgroups. The same modeling and statistical methods will be used as for the analyses of the entire sample.

Definition of intervention adherence and how it will be presented

A patient's treatment has adhered to the protocol, *if* written consent was obtained *and* the patient was operated according to the randomization (as described above) *and* the patient did not earlier than two years after the operation decline permission to access national registers and hospital notes for outcomes.

Conversely, the trial treatment of a patient has not adhered to the protocol, *if* written consent was not obtained *or* the patient was not operated according to the allocation *or* the patient withdrew permission to access notes and registers within two years.

We will term the following occurrences as partial breach of adherence:

1. Operation type revealed to patient prior to 1-year follow-up.

2. Patient choses to withdraw from study earlier than two years after the operation, declining to provide PRO-data and to attend follow-up appointments, but accepting researchers' access to notes and registers.
3. Patient choses to withdraw from study later than two years after the operation without a valid reason, declining to provide PRO-data and to attend follow-up appointments, but accepting researchers' access to notes and registers.

Definition and summary of protocol deviations

Number of adherences, non-adherences and partial adherences within 2 years within each treatment group will be reported

Randomization details

Stratified, permuted block randomization with a 1:1 allocation ratio. Block sizes were be 4, 6, 8 and 10. Stratification was two-dimensional, where hospital is one dimension and patient sex the other. The randomization wase done as closely preceding surgery as possible at each hospital.

Sample size calculation

Based on data from the PFA vs. TKA study , the Oxford Knee Score (OKS) at precisely two years postoperatively has an SD of 8.4, while the timeweighted average (based on AUC) during the first two years has a smaller SD of 7.2. We found it reasonable to assume similar distributions of OKS in the current study. The minimally clinically important difference in the OKS is approximately 3 . By using the SD of 8.4 (the largest of the two) and setting the significance level (alpha) to 5% and the power (1-beta) at 90%, 165 patients were required in each group. We added a buffer of 10 patients in each group, resulting in a total of 350 patients. This sample size calculation should assure enough power to detect both a clinically meaningfull difference at 2 years (secondary outcome S1) and a clinically meaningfull difference in AUC during the first 2 years (primary outcome).

Superiority, equivalence, or noninferiority hypothesis testing framework.

Superiority study.

Timing of future, related analyses of this trial

Primary and secondary outcomes analysis and cost-benefit analysis at 5 years of follow-up:

A similar analysis to that of the 2 years follow-up data described in this document is planned after 5 years of follow-up. In addition, a cost-benefit analysis will be performed at 5 years of follow-up.

Primary and secondary outcomes and cost-benefit analysis at 10 years of follow-up:

A similar analysis to that of the 5 years follow-up is expected after 10 years of follow-up. Also, we anticipate similar analyses after 15 and 20 years of follow-up.

Details on summarizing safety data

Not applicable

Statistical packages used for analysis

The R software and especially the R package “nmle” and its “gls” function for the main analyses (as described above). We will also use the R package “exact2x2” to compare the proportions of SAE, as described above too.

Appendix

Tabel 1 W = weeks, M = months, Y = years, PROM = Patient Reported Outcome Measures, P/E = Physical examination, Comp = complications, # = Unblinding of patient treatment

	Enrolment	Allocation	PROM	P/E	X-ray
Eligibility screening	•				
Informed consent	•				
Baseline variables			•	•	•
Operation		•			
2w				•	•
1 m			•		
2 m			•		
3 m			•		
4 m				•	
6 m			•		
9 m			•		
1y		#	•	•	•
18 m			•		
2y			•	•	•
3y + 4y			•		
5y			•		
6y + 8y			•		
10y			•	•	•
12y + 14y + 16y + 18y			•		
20y			•	•	•

Tabel 2: Patient demographics

Characteristics	**Treatment A** **N=**	**Treatment B** **N=**	**Total** **N=**
Sex, n (%) -Males -Females			
Hospital, n (%) -A -B - etc.			
BMI, n (%) - <18.5 - 18.5 to <25 - 25 to <30 - 30 to <40 - ≥40			
Employment Status, n (%) -Employed -Not employed, retired -Not employed, other			
Marital status, n (%) -Unmarried -Married/partner -Divorced/widow -other -other			
Education, n (%) -None -Highschool -Bachelor -Academic -Other			
Residence, n (%) -Farm/ranch -House -owned Apartment -rented apartment -rented room -Elderly home -Other			
Work, n (%) -Independent -employed -Student -early pension -Retired -Politician -unemployed -Other			

Knee specific demographics			
Side, n (%) -Right -Left			
Ethiology -Idiopathic -Meniscectomy -Osteochondritis dissecans -Spontaneous osteonecrosis -tibial condyle fracture -Other trauma			
Length of symptoms (months)			
Previous knee Surgery, n (%) -No -Yes -Type			
Previous knee surgery on contralateral knee, n (%) -No -Yes -Type			
Surgery type, and year -Arthroscopy -ACL recon. -Ligament recon. (other) -Osteotomy, PTO -Osteotomy, DFO -Osteosynthesis -Open meniscectomy			
Non-surgical treatment before surgery, n (%) -None -Mild analgesics + NSAID -Opioids -Steroidinjections -Training/physical therapy -change in Knee loading (crutches,insoles, etc) -			
Knee problems as a teenager -No -Yes -Type			
Ipsilateral groin pain, n (%) -No -Yes			

-Type			
Ipsilateral Ischias, n (%) -No -Yes -Type			
Alignment, n (%) -Severe varus -Mild varus -Neutral -Mild Valgus -Severe Valgus			
Patella position, n (%) -Alta -Neutral -Infera (baja)			
Degree og knee effusion, n (%) -None -Fluctuation, no pressure -Fluctuation, pressure -Patella tap, no pressure -Patella tap, pressure -Patella tap not possible because of massive effusion.			
Circumference above patella, mean (SD)			
Soreness, n (%) -Medial joint line -Lateral joint line -Patellafacet, medially -Patellafacet, laterally			
Knee ROM, mean (SD) -Extension -Flexion			
AnteroPosterior stability, n (%) -Normal -Normal amplitude, unsure anterior stop -High amplitude, unsure stop -Posterior lag			
ML-stability, n (%) -Normal -Medial looseness -Lateral looseness			
Normal and painfree hip-ROM, n (%) -Yes -no pain, decreased ROM -Pain, normal ROM -Pain, decreased ROM			

Peripheral pulse, n (%) -Normal -weak/unsure -None			
Normal neurological inspection, n (%) -Yes -No			
Baseline PROM demographics			
OKS			
KOOS -dimensions			
SF36 -dimensions			
FJS			
EQ5D EQ5D-VAS			
UCLA			
CPH Knee ROM			

Tabel 3: Study outcomes by treatment group at 2 years.

	Total **N=**	**Treatment** **A** **N=**	**Treatment** **B** **N=**	**Mean** **Difference*** **(95% CI)**	**p-value**
	mean (SD)	**mean (SD)**	**mean (SD)**		
Primary Outcome OKS					
Secondary outcomes					

Tabel 4: Adverse events

Variable	**Treatment A** **N (%)**	**Treatment B** **N (%)**	**Total** **N (%)**	**Risk difference**
Reoperation -Type Revision -Type Death				

Tabel 5: Surgical details

Variable	**Treatment A** **N=**	**Treatment B** **N=**	**Total** **N=**
Side, n (%) Right Left			
Arthroplasty TKR mUKR			
Anesthetic Strategy, n (%) General Anesthetic Spinal Anesthetic Regional Block, n (%) -Yes -No			
Duration of surgery, n (minutes)			
Length of stay (days) Time of day surgery (0-24h)			
Blinding upheld (n, %) 2weeks 4month 1year			
Surgeon level -Consultant -Unit specialist			
Meniscus status -Medial -Lateral ACL status PCL status			
Knee pathology -None -Chondrocalcinosis -Osteocondrithis -Synovitis -Pigmented villonodulus			
Tourniquet -Yes -No			
Incision foil -Yes			

-No			
Supplemental procedures -synovectomy total - Partial -autotransplant of bone -Amotio atellae			
Total Knee arthroplasty type			
ROM at end of surgery -Flexion -Extension			

References

"Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products." 2023, https://www.fda.gov/media/148910/download.

Chow S, Shao J, Wang H. . "Sample Size Calculations in Clinical Research. ." *Chapman & Hall/CRC Biostatistics Series* 2nd Ed. (2008): page 89.

Dmitrienko, A., W. W. Offen, and P. H. Westfall. "Gatekeeping Strategies for Clinical Trials That Do Not Require All Primary Effects to Be Significant." [In eng]. *Stat Med* 22, no. 15 (Aug 15 2003): 2387-400. https://doi.org/10.1002/sim.1526.

Fay, Michael, and Sally Hunsberger. *Practical Valid Inferences for the Two-Sample Binomial Problem.* 2019.

Mallinckrodt, Craig H., Peter W. Lane, Dan Schnell, Yahong Peng, and James P. Mancuso. "Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials." *Drug information journal : DIJ / Drug Information Association* 42, no. 4 (2008/07/01 2008): 303-19. https://doi.org/10.1177/009286150804200402. https://doi.org/10.1177/009286150804200402.

Mortensen, J. F., L. E. Rasmussen, S. E. Ostgaard, A. Kappel, F. Madsen, H. M. Schroder, and A. Odgaard. "Randomized Clinical Trial of Medial Unicompartmentel Versus Total Knee Arthroplasty for Anteromedial Tibio-Femoral Osteoarthritis. The Study-Protocol." [In eng]. *BMC Musculoskelet Disord* 20, no. 1 (Mar 20 2019): 119. https://doi.org/10.1186/s12891-019-2508-1.

Odgaard, A., F. Madsen, P. W. Kristensen, A. Kappel, and J. Fabrin. "The Mark Coventry Award: Patellofemoral Arthroplasty Results in Better Range of Movement and Early Patient-Reported Outcomes Than Tka." [In eng]. *Clin Orthop Relat Res* 476, no. 1 (Jan 2018): 87-100. https://doi.org/10.1007/s11999.0000000000000017.

Wang, Bingkai, Elizabeth L. Ogburn, and Michael Rosenblum. "Analysis of Covariance in Randomized Trials: More Precision and Valid Confidence Intervals, without Model Assumptions." *Biometrics* 75, no. 4 (2019): 1391-400. https://doi.org/https://doi.org/10.1111/biom.13062. https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13062.

Wang, Bingkai, Ryoko Susukida, Ramin Mojtabai, Masoumeh Amin-Esmaeili, and Michael Rosenblum. "Model-Robust Inference for Clinical Trials That Improve Precision by Stratified Randomization and Covariate Adjustment." *Journal of the American Statistical Association* (2021): 1-12. https://doi.org/10.1080/01621459.2021.1981338. https://doi.org/10.1080/01621459.2021.1981338.


mUKA
VS.
TKA