August 28th, 2023

Statistical Analysis Plan:

The Impact of Surgeon Training on Patellofemoral Arthroplasty Revision Rates: A Target Trial Emulation Prospective Cohort Study

Trial registration

NCT04772625

Authors/Collaborators

Louise Eggers Rasmussen, Sabrina Mai Nielsen, Paul Blanche, Frederik Espersen, Tobias Freyberg Justesen, Lasse Enkebølle Rasmussen, Stine Hangaard, Robin Christensen, Anders Odgaard

Affiliations

LR: [1]Department of Orthopaedic Surgery, Rigshospitalet, Copenhagen University Hospital.

SMN: [1]Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen & [2]Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark.

PB: Section of Biostatistics, Department of Public Health, University of Copenhagen.

FE: Department of Orthopaedic Surgery, Hillerød Hospital.

TFJ: Department of Surgery, Zealand University Hospital Køge.

LER: Department of Orthopaedic Surgery, Vejle Hospital.

SH: Department of Radiology, Herlev Hospital.

RC: [1]Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen & [2]Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark.

AO: [1]Department of Orthopaedic Surgery, Rigshospitalet, Copenhagen University Hospital & [2]Department of Clinical Medicine, University of Copenhagen.

SAP FRAMEWORK AND CONTENTS

This statistical analysis plan (SAP) is framed by the recommendations described by Gamble et al. [Gamble 2017] and Hiemstra [Hiemstra 2019] that focuses on clinical trials.

1. ADMINISTRATIVE INFORMATION

1.1 Trial registration

Registered with ClinicalTrials.gov as “Failure Analysis of Patellofemoral Arthroplasty”, NCT04772625. This registration is an umbrella registration that covers more analyses from the cohort. The current study on the effect of surgeon training is the first study based on the cohort.

1.2 Funding

The Parker Institute, Bispebjerg Hospital and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL). Internal funds from the Department of Orthopaedics, Rigshospitalet, has been allocated to this project.

1.3 Versions

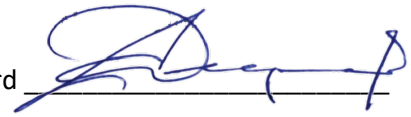
SAP v. 1.0	21.08.2023
Protocol 6.0	02.11.2022

1.4 Roles and responsibility

Louise Eggers Rasmussen	First author
Sabrina Mai Nielsen	Co-author, biostatistician
Paul Blanche	Co-author, statistician
Frederik Espersen	Co-author, data collection
Tobias Freyberg Justesen	Co-author, data collection, protocol draft
Lasse Enkebølle Rasmussen	Co-author, data collection, protocol draft
Stine Hangaard	Co-author, data collection
Robin Christensen	Co-author, senior statistician
Anders Odgaard	Senior author, project initiator, main investigator

1.5 Signatures

Robin Christensen  Aug 22, 2023

Anders Odgaard 

24 August 2023

2. INTRODUCTION

2.1 Background and rationale

Isolated patellofemoral osteoarthritis (PF-OA) is a degenerative disease of the knee that causes pain and limits patients' daily activities. When conservative treatment fails, arthroplasty may be indicated in severe cases, and the options are patellofemoral arthroplasty (PFA) and total knee arthroplasty (TKA). PF-OA may occur in as many as 10% of men and 25% of women older than 55 years of age [Davies 2002, McAlindon 1992]. Although PF-OA may be a frequent condition, the proportion of PFA is below 1% of all knee replacements in most national registers [Lewis 2020, AJRR 2021, NJR 2021, Odgaard 2021], suggesting that PF-OA is either under-diagnosed, under-treated or mostly treated with TKA.

Registries have consistently demonstrated that the early cumulative revision rate of PFA is more than three times that of TKA, prompting one multinational study to imply that PFA treatment should be abandoned [Lewis 2020]. In contrast to this, randomised clinical trials and cohort studies have concluded that PFA has outcomes that are at least on par with TKA [Clement 2019, Joseph 2020, Lin 2021, Odgaard 2022, Peng 2021]. One RCT has demonstrated knee-related quality of life differences in favour of PFA during the first two postoperative years [Odgaard 2018] and no real difference between TKA and PFA beyond two years up to six years [Odgaard 2022]. The RCT also found similar revision rates between the two implant types at six years [Odgaard 2022]. A review has furthermore suggested that complications from PFA recorded with registers differ from those observed in controlled studies [Bendixen 2019], suggesting either different practices among surgeons or validity problems when recording complications in registers. The core of the debate on whether to recommend PFA over TKA - or vice versa - is in essence whether "real world data", i.e., register data, is more reliable and should be given larger weight than data from cohorts and randomised studies when judging the benefit of specific implant types and their usage [Collins 2020, Groenwold 2021].

One explanation for the different observations of outcome from registers and experimental studies may be the level of surgeon training, and one may surmise that register observations reflect the level of surgeon training more than the inherent properties of the implants. Mastering surgical treatments requires knowledge of indications, surgical technique, management of complications etc. To the best of our knowledge, no implant register considers surgeons' level of expertise with a specific implant, which may thus represent a possible source of bias when analysing implant types in register studies.

2.2 Aim, Hypothesis and Objective(s)

In the current study, we aim to investigate whether surgeon training may bias the rate of complications, i.e., revisions and other reoperations, following PFA. We believe that conclusions may have wider implications to other implant types, when judging relative merits of implants from register data. Specifically, we will examine whether the cumulative revision rate of PFA-trained surgeons participating in the above-mentioned randomised clinical trial (trial surgeons) differs from that of surgeons not participating in the trial (non-trial surgeons) [Odgaard 2018, 2022]. In the RCT, all participating surgeons had undergone focused training, and no surgeon would independently perform surgery as part of the trial without prior supervision. We hypothesised that trial surgeons would have a 6-year cumulative revision rate different from that of non-trial surgeons, and we will approach this by using nationwide data combined with a manual review of hospital notes. The primary objective is to compare the effect of trial experienced

surgeons, relative to non-trial surgeons, on the cumulative revision rate from baseline to year 6, in patients having a Patellofemoral Arthroplasty. Secondary objectives are to compare the effect of trial experienced surgeons, relative to non-trial surgeons, on (*i*) cumulative reoperation rate (other than revision) and (*ii*) the cumulative mortality rate from baseline to year 6 in these patients.

3. STUDY METHODS

3.1 Study design

The present study is a retrospective population-based cohort study of consecutive patients operated with PFA in the 8-year accrual period from January 1, 2008, to December 31, 2015. The cohort was determined using a national arthroplasty register and administrative databases, and detailed information on treatment and outcome was obtained from the administrative databases in combination with extracting patient-specific data from individual patients' hospital notes.

The primary outcome is the 6-year cumulative revision rate (CRR6).

Secondary outcomes are the 6-year cumulative reoperation rate (other than revision) and the 6-year cumulative mortality rate.

The study will compare the outcomes of two groups of surgeons: 1) those surgeons, who were part of the randomised clinical trial comparing PFA and TKA, and 2) other surgeons. The surgeons who were part of the trial were specifically trained in indications, surgical technique, and management of complications. They had attended a seminar on PFA, and they were only allowed to operate independently in the RCT after having been supervised in the PFA procedure.

The study is registered with ClinicalTrials.gov (NCT04772625), and reporting will adhere to the STROBE statement [Vandenbroucke 2007]. The study was approved by the Danish Data Protection Agency (P-2020-914 of 19.10.2020), and permission to access individual patients' hospital notes was obtained from the National Board of Health (3-3013-2907/1 of 25.09.2018) and the Capitol Region (R-20078659 of 16.04.2021). According to Danish national ethical standards, individual patient consent was not necessary.

3.2 Data sources

The Danish Knee Arthroplasty Register (DKAR) [Pedersen 2012] has existed since 1998. Surgeons report the procedures directly to DKAR independent of administrative registers. The completeness of registration was 85.5% and 97.1% for primary arthroplasty procedures performed in 2008 and 2015, respectively. These estimates of completeness are routinely reported by DKAR and are based on independent registrations in the Danish National Patient Register (DNPR).

DNPR [Lynge 2011] records all contacts and treatment episodes in Danish public hospitals and private clinics, irrespective of patient citizenship and funding. The register has since 2000 formed the basis of public hospital reimbursement, and the registration from public hospitals is assumed to be complete since then [Lynge 2011]. Private hospitals and clinics have been obliged to record with the DNPR since 2003. The registration is done by the hospitals' administrative staff in a process independent of the DKAR registration. The overall completeness of the registration of patellofemoral arthroplasty is expected to be high, but it is unknown. Surgical procedures are coded by the NOMESCO Classification of Surgical Procedures [NOMESCO 2010]. No changes have been made to registration or coding practices during the accrual period or later.

The Central Personal Register (CPR) was established in 1968 and contains basic personal information (10-digit identifier, name, address, date of birth, citizenship, status, etc.) for all persons who have been or are currently residing in Denmark. The status is one of four: 1) alive and living in Denmark, 2) deceased, 3) missing, or 4) emigrated. Dates of status change are recorded.

Patient notes are stored at individual hospitals. Most notes were digitised but some hospitals had used paper formats in the early part of the accrual period. The research group of this study requested copies of the notes and radiographs from relevant hospitals, and a very large majority of hospitals sent the notes as requested. One private hospital had ceased to exist because of financial problems, and the notes had been destroyed before we requested the notes. A few hospitals failed to locate relevant notes or failed to react to our requests.

3.3 Population and data retrieval

The unit of observation in this study is a knee, and a patient may contribute with both knees. The index procedure is defined as the operation where the primary PFA was performed. We define PFA as a procedure where an implant consisting of two manufactured, man-made parts are inserted on the opposing articular faces of the patellofemoral joint (both patella and femoral trochlea) with the purpose of permanently replacing biological tissues.

All PFA operations performed during the accrual period were identified through DKAR and DNPR. Information about residence and vital status of patients was obtained from the CPR system. We used data from both DKAR and DNPR with the intention to obtain a cohort as complete as possible. CPR was used to determine death as a competing event in the competing risk analyses of cumulative incidence rates and to identify censoring events (emigration and disappearance). All Danish residents are identified by their unique 10-digit CPR number (equivalent to a social security number), and the CPR number is used to link the information in the accessed databases and hospital notes.

From the DKAR we retrieved the set of patients recorded with a primary procedure and the set of patients recorded with a revision from a PFA to any other arthroplasty. The information included the CPR number, the date of the procedure, and the hospital. This DKAR set was merged with patients recorded with a PFA procedure in the DNPR (NOMESCO procedure code KNGB13), also specifying the procedure date and the hospital. The DNPR was subsequently searched for all revisions (NOMESCO procedure codes KNGC*), other surgical knee procedures (NOMESCO procedure codes KNG* except KNGC*) or femoral amputations (NOMESCO procedure code KNFQ19) recorded for the identified patients with an operation date later than the date of the index procedure. The record retrieval process is shown as the upper four boxes in anticipated figure 1 (in section 6. Outline - anticipated tables and figures). Having identified the knees with a potential primary PFA from the registers, further information about individual knees' baseline condition, treatment, and a possible later revision procedure, was obtained by manually reviewing patient records and pre- and postoperative radiographs retrieved from the relevant hospitals. The goal was to confirm the primary PFA procedure, i.e., the index procedure, the operation date and the operated side. The review also recorded possible revisions. All censoring events and complications were recorded for the first six postoperative years. For patients recorded with a revision in a hospital different from the primary hospital, these notes were also requested and studied to verify the details of the revision procedure.

The review allowed us to collect details of the clinical history, physical findings, surgical procedure, and the postoperative course. We also accumulated relevant radiographs. In short, all information relevant to the operated knee was recorded. The information relating to the primary procedure was categorised in four: 1) demographics, 2) preoperative medical history, 3) preoperative physical examination, and 4) surgical data. This information is presented in table 1 and in supplementary table 1.

Some of the procedures were found to be wrongly coded, and a large majority of these came from a single hospital where TKAs with patella resurfacing customarily were recorded with codes for both TKA and PFA. Some knees were recorded twice because of side confusion, and in some cases hospital notes were unavailable. The final number of knees with validated primary PFA procedures available for analysis will be reported in the flow diagram (i.e., figure 1) and we also validated the details of revisions when these had occurred. The manual reviews were performed by four of the authors (LR, FE, TFJ and LER), and individual cases were consulted with AO in case of ambiguity. We did not collect the details of reoperations (other than revisions), and subsequent analyses of incidence rates of these were based on the database records only.

4. STATISTICAL PRINCIPLES

4.1 Statistical and clinical significance

All *p* values and 95% confidence intervals will be two-sided. We will pragmatically define a test result corresponding to $p \leq 0.05$ as being statistically significant (i.e., indicating that the null hypothesis should be rejected). *P* values will be stated exactly, apart from values less than 0.001, which will be expressed as $P < 0.001$. *P* values will be expressed with two significant figures and up to three decimal places.

Clinical insignificance: A 95% confidence interval excluding RRs greater than 1.25 units (or less than 0.80) between groups will be interpreted as indicating the absence of a clinically meaningful relative difference.

4.2 Adherence and protocol deviations

The present study is non-interventional and considerations of adherence to intervention are not applicable.

4.3 Analysis populations

The ‘analysis population’ for the primary analysis will include all knees with the main covariates mentioned in section 5.4 available.

4.4 Target trial approach

We will use the framework of Target Trial Emulation for causal inference from observational data [Hernan 2022] that involves emulating a hypothetical randomised controlled trial (RCT) that would have addressed the same research question we intend to answer. The idea is to emulate this proposed "target trial" that would have been conducted had the study population been randomised to the trial or non-trial surgeon group. The purpose of the framework is to improve from the anticipated limitations of causal inference from observational studies, such as selection bias and confounding, and thus attempt to estimate the causal effect of the treatment or exposure of interest.

We will emulate a randomised experiment (i.e., the target trial) that would answer the question whether trial and non-trial surgeons get the same results. This involves using the same analysis as for the corresponding target trial (i.e., crude analysis), except that there will be subsequent adjustment for baseline variables in an attempt to emulate random treatment assignment. This will partially enable us to estimate the causal effect of the treatment in a way that is less prone to bias and confounding.

4.5 Rubin causal model

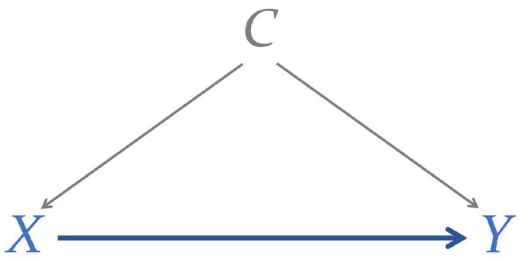
Strengthening causal inference from observational data requires that each knee's time zero is registered appropriately – i.e., the time when they are operated on, and thus assigned to a treatment strategy (trial surgeon vs non-trial surgeon). Using the potential outcomes framework, we will report causal effects of estimation approaches classified according to the no unmeasured confounding assumption; the assumption which states that the type of surgeon who operates is independent of the potential outcomes, given the covariates (i.e., variables collected at baseline).

The Rubin causal model is based on the idea of potential outcomes. To measure the causal effect of having X_{Exposed} (compared to X_{Control}) for a specific knee, we need to compare the outcome for the same knee in both alternative futures ($Y_i|_{\text{Exposed}}$. *vs* $Y_i|_{\text{Control}}$). Since it is impossible to see both potential outcomes at once, one of the potential outcomes is always missing. This dilemma is the fundamental problem of causal inference. Because of this, unit-level causal effects cannot be directly observed. However, this is the argument for why randomised experiments allow for the estimation of population-level causal effects [Rubin 2005]. A randomised experiment assigns people randomly to treatments: X_{Exposed} or X_{Control}. Because of this random assignment, the groups are (on average) equivalent, and the difference in outcome Y_i can be attributed to the group assignment since that was the only difference between the groups (X_{Exposed} and X_{Control}, respectively). An estimate of the Average Causal Effect (also referred to as the Average Treatment Effect) can then be obtained by computing the difference in means between the treated (X_{Exposed}) and control (X_{Control}) samples.

Randomised trials can be expensive, resource intensive, and under certain circumstances unethical. For these reasons, scientists often rely on observational data. The challenge with observational data is that treatments are not applied randomly, often leading to selection bias and confounding variables. Improved confounding variable balance between treatment (X_{Exposed}) and control (X_{Control}) groups can be achieved by, e.g., standardization (aka G-computation), matching observations from each group based on the propensity score, which in this case would be the probability that a patient received the experimental group (X_{Exposed}) given the observed covariates. Propensity score analysis (or similar) seeks to isolate the treatment as the only difference between our treatment and control groups. Thus, propensity score methods attempt to correct for the assignment mechanism by finding control units similar to treatment units ($Y_0|X_{\text{Exposed}} \approx Y_0|X_{\text{Control}}$).

To prespecify possible propensity scores we will use the following pragmatic definition of what makes a confounding variable (*C*), figure below [Christensen et al 2023]:

- The Covariate (C) is an ancestor (cause) of the outcome (Y)
- The Covariate (C) probably cause the exposure (X; e.g. group)
- The Covariate (C) is <u>not</u> a descendant (effect) of the exposure (X) or outcome (Y)



4.5 G-computation

G-computation (aka standardization or g-formula) is a powerful tool for estimating the causal effect of an exposure on an outcome in observational studies [Chatton 2020, Hernan and Robins 2020, Goetghebeur et al 2020]. The g-computation, as described in e.g., chapter 13 in Hernán and Robins (2020) or Goetghebeur et al (2020) relies on the estimation of the risk of outcome given treatment and covariates, via a regression model, the so-called "outcome model". Under correct model specification and the usual assumptions needed for causal inference (i.e., "no unmeasured confounding", "consistency", and "positivity") the method is unbiased to estimate an average treatment effect defined as either the difference or the ratio of the average counterfactual outcome between the two groups. To model our binary outcomes (e.g., revision surgery within 6 years) we will use a multiple logistic regression model, as commonly done. More details can be found in Section 5.6 below.

4.6 Propensity score weighting

Propensity score weighting is a statistical technique that is commonly used in observational studies to adjust for potential confounding variables and to improve the balance between the treatment and control groups [Thomas 2019, Haukoos 2015]. The propensity score is the conditional probability of receiving the treatment given a set of covariates. It is estimated using a logistic regression model, where the dependent variable is the treatment status (0 for control, 1 for treatment) and the independent variables are the covariates that may be potential confounders. The propensity score ranges from 0 to 1 and represents the individual's probability of being assigned to the treatment group given his or her covariate values. Propensity score weighting involves weighting the observations in the treatment and control groups based on their propensity scores.

We will use the inverse of the propensity score as the weight for each observation, to balance the distribution of covariates identified as potential confounders between the treatment and control groups. We will assess the balance of these covariates before and after propensity score weighting using standardized differences (StdDs). An StdD less than 0.1 will be considered a negligible imbalance [Haukoos 2015]. If there are still significant imbalances after weighting, we may consider including additional covariates or using other techniques.

5. ANALYSES

The primary outcome and secondary outcomes are all originally derived from time-to-event outcomes. In this context, logistic regression can be used to model the probability of an event occurring within a certain time frame, given a set of predictor variables.

Multiple logistic regression with g-computation adjusted for potential confounders will be used as the primary model, and a (univariate) logistic regression with propensity score weighting as sensitivity analysis.

Estimates from both unadjusted (i.e., crude) and adjusted analyses will be reported. Any observed difference from the unadjusted analyses - which simply compare the results of two populations - is likely to be caused by either a fundamental effect of surgeon affiliation (trial or non-trial) or by confounding factors of the patient selection, e.g., demography.

Initially, the cumulative revision, cumulative reoperation (other than revision), and cumulative mortality during the first 6 years will be visualized using Aalen-Johansen estimator. Note that we do not expect censored data within the 6 years follow-up. Hence, the Aalen-Johansen curves should simplify to the curves of the empirical cumulative distribution function.

5.1 Framework

Superiority testing, while a 95% confidence interval excluding RRs greater than 1.25 units (or less than 0.80) between groups will be interpreted as indicating the absence of a clinically meaningful relative difference [Christensen et al 2023].

5.2 Exposure and control

Exposure: Operation by a trial surgeon.
Control: Operation by a non-trial surgeon.

5.3 Power and sample size considerations

Not applicable. [Hernán 2022]

5.4 Potential confounders

Following a discussion among three of the trial surgeons about important confounders, the potential confounders listed below will be considered in the main analyses:

- Age (<50, 50-70, >70)
- Male (yes, no)
- Any history of either patella dislocation, knee trauma, or knee dysplasia (yes, no/not reported)
- Duration of symptoms (<5 years/not reported, ≥5 years)
- Any use of analgesics (yes, no/not reported)
- Previous surgery (yes, no/not reported)
- Year of surgery

The following potential confounders will be considered – in addition to the ones above - in the sensitivity analyses:

- BMI (kg/m^2)
- Tibiofemoral Kellgren-Lawrence (KL)grade (0-1, 2-4)
- Patellofemoral KL-grade (0-2, 3-4)

5.5 Cumulative incidence

Since the outcomes of interest are based on time-to-event outcomes, we will visualize the cumulative incidence function. To account for both censoring and competing risk, the Aalen-Johansen estimator will be used.

Time zero is well defined for all knees as the date of the primary procedure (the index procedure), and we then observe events at later time points. The observations of this study come from administrative registers and hospital notes. We assume that all occurrences of each outcome are unequivocally identified in the study.

If a patient emigrates or disappears, we will assume that their knee will have a risk of the outcomes similar to patients, who do not emigrate or disappear. Thus, these events will result in non-informative censoring. Death is a competing event, and neither revision nor reoperation will be observed after death. Consequently, we will consider death as a competing event to revision. Likewise, we will consider revision to be a competing event to other reoperations.

5.6 Multiple logistic regression with g-computation

The primary analysis will be based on multiple logistic regression with g-computation. The multiple logistic regression will include the occurrence of revision within 6 years as dependent variable, the treatment as fixed effect (Operation by a trial surgeon vs Operation by a non-trial surgeon), as well as potential confounders as listed above. Based on background clinical knowledge, it has been decided to pre-specify a logistic regression model without any interaction terms. We will use usual generalized estimating equation (GEE) (with independence working correlation) to fit the model, and we will use robust standard errors to compute confidence intervals and p-values. The robust standard error for the average treatment effect will be calculated by bootstrapping, by resampling the patients (not the individual knees) to properly account for the correlation between the knees of the same patients. Similar computation of robust standard errors will be performed for the other analyses (i.e., the crude “unadjusted” analysis and the propensity score analysis).

We do not anticipate censored observations as we expect that all knees are fully followed-up for at least 6 years. However, if the data contains censored observations, we will consider a direct extension of the method described above to right-censored data as described in Blanche et al [Blanche 2022] (for this main analysis and likewise for the other analyses, crude, and propensity score). As often recommended, we will compute confidence intervals for the log of the risk ratio (using a simple Wald-type confidence interval method) and then exponentiate it to compute the confidence interval of the risk ratio.

The secondary outcomes, 6-year cumulative reoperation rate (other than revision) and 6-year cumulative mortality rate, will be analysed using a similar approach.

5.7 Logistic regression with propensity score weighting

A (univariate) logistic regression model with propensity score weighting will be used as sensitivity analysis. The logistic regression will include the occurrence of revision within 6 years as dependent variable, the treatment as fixed effect (Operation by a trial surgeon vs Operation by a non-trial surgeon). Inverse propensity score weighting will be applied. A propensity score will be computed based on the potential

confounders listed above. Based on background clinical knowledge, it has been decided to pre-specify the logistic regression model (fitted to compute propensity scores) without any interaction terms.

The secondary outcomes, 6-year cumulative reoperation rate (other than revision) and 6-year cumulative mortality rate, will be analysed using a similar model.

5.8 Robustness

The robustness of the conclusions from the main analyses will be assessed using a larger set of potential de-confounding variables for the multiple logistic regression with g-computation and the logistic regression with propensity score weighting (listed above). Note that these analyses will be based on a smaller sample size. The analysis sets will consist of all knees for which we have full data for the larger set of confounders.

6. Outline - anticipated tables and figures

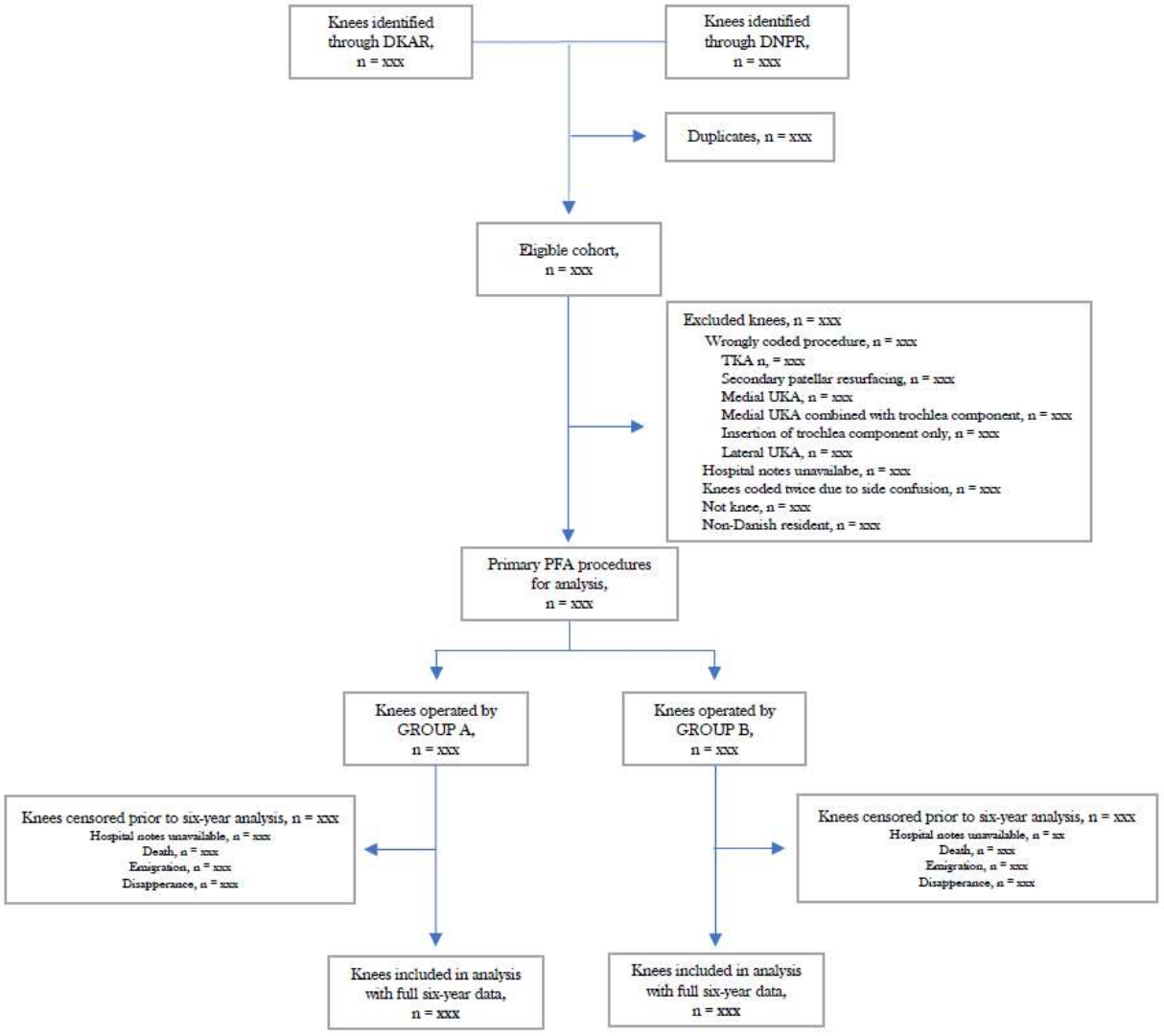


Figure 1: Illustrating knee flow from enrolment to full six-year outcome assessment.

Table 1: Baseline characteristics of all the included knees, stratified by exposure group.

Variable	PFA Trial Surgeon Group (XXX knees)		PFA Non-trial Surgeon Group (XXX knees)		Standardized Difference
	n		n		
Demographics					
Age (years)*					
Male, no. (%)*					
Height (cm)					
Weight (kg)					
BMI (kg/m^2)†					
Medical history					
Primary diagnosis:					
PF-OA (dysplasia or idiopathic), no. (%)					
PF-OA (post traumatic OA), no. (%)					
Traumatic lesion, no (%)					
Primary symptom:					
Pain, no. (%)					
Patellar instability, no. (%)					
History of patella dislocation*:					
Yes, no. (%)					
No / not reported, no. (%)					
History of knee trauma*:					
Yes, no. (%)					
No / not reported, no. (%)					
Knee dysplasia*:					
Yes, no. (%)					
No / not reported, no. (%)					
Duration of symptoms*:					
< 5 years, no. (%)					
≥ 5 years, no. (%)					
Not reported, no. (%)					
Use of analgesics* ‡:					
NSAID / paracetamol, no. (%)					
Opioids, no. (%)					
Tramadol / codeine, no. (%)					
Gabapentin, no. (%)					
Other, no. (%)					
Previous surgery*:					
Yes, no. (%)					
No / not reported, no. (%)					
Previous conservative treatment:					
Yes, no. (%)					
No / not reported, no. (%)					
Physical examination of the knee					
Right knee, no. (%)					
Range of motion:					
< 120 degrees, no. (%)					
≥ 120 degrees or to soft tissue, no. (%)					
Varus malalignment, no. (%)					
Valgus malalignment, no. (%)					
Effusion, no. (%)					
Positive patellar apprehension test, no. (%)					
Positive Clarke's test, no. (%)					

Retropatellar crepitus, no. (%)					
J-sign, no. (%)					
AP instability, no. (%)					
ML instability, no. (%)					
Hip pathology, no. (%)					
Radiology					
TF KL-grade†:					
0, no. (%)					
1, no. (%)					
2-4, no. (%)					
PF KL-grade†:					
0-2, no. (%)					
3-4, no. (%)					

Data are presented as mean (SD), unless otherwise stated.

*Baseline characteristics considered potential confounders.

† Baseline characteristics to be considered in the sensitivity analyses.

‡ The sum exceeds the total sample n and the proportions add up to more than 100% because patients can use more than one type of analgesic.

PF-OA, patellofemoral osteoarthritis; TF-OA, tibiofemoral osteoarthritis; AP instability, anteroposterior instability; ML instability, mediolateral instability.

Figure 2: Cumulative incidence of revision (panel A), reoperation other than revision (panel B), and death (panel C). Three curves in each panel showing the cumulative incidence of revision for the entire population (dotted line), the PFA trial surgeon group, and the PFA non-trial surgeon group.

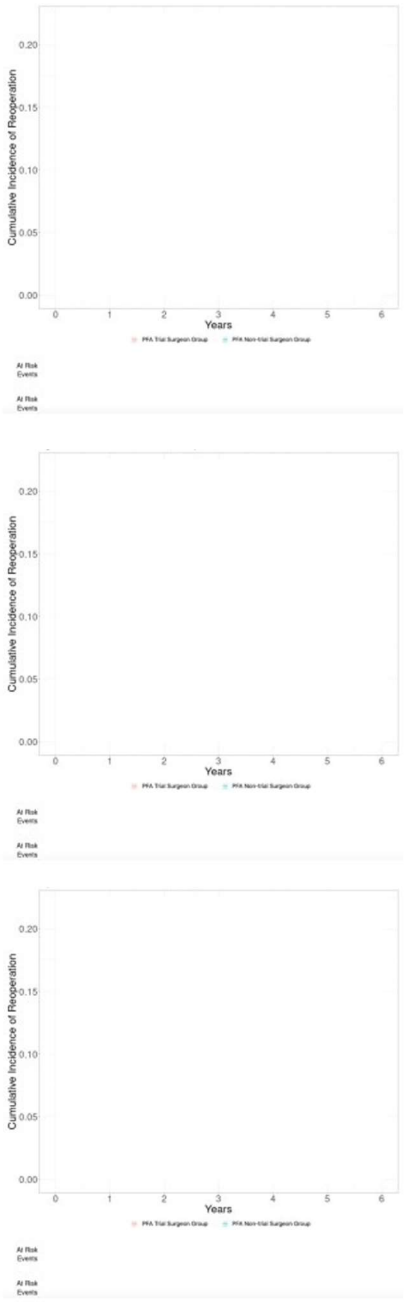


Table 2. Crude and adjusted risk ratios for the six-year cumulative reoperations, revisions, and mortality in the PFA Trial Surgeon Group vs. PFA Non-trial Surgeon Group.

Variable	PFA Trial Surgeon Group (XXX knees)	PFA Non-trial Surgeon Group (XXX knees)	Crude RR (95%CI)	Adjusted RR (95%CI)
Revision, no (%)				
Reoperation, no (%)				
Mortality, no (%)				

Risk ratios estimated from by the ratio of the observed risks in the two groups (crude RR) and multiple logistic regression with g-computation, adjusted for potential confounders (adjusted RR).

95%CI, 95% confidence interval; PFA, patellofemoral arthroplasty; RR, risk ratios.

Supplementary table 1. Surgical data of all the included knees, stratified by exposure group.

Variable	PFA Trial Surgeon Group (XXX knees)		PFA Non-trial Surgeon Group (XXX knees)		Standardized Difference
	n		n		
Year of PFA operation*					
Duration of surgical procedure (minutes)					
Intraoperative pathology					
PF-OA:					
Yes, no. (%)					
No / not reported, no. (%)					
TF-OA:					
Yes, no. (%)					
No / not reported, no. (%)					
Implant brand					
Avon, no. (%)					
Cartier, no. (%)					
Hemicap, no. (%)					
Journey, no. (%)					
LCS, no. (%)					
Nexgen, no. (%)					
Sigma, no. (%)					
Vanguard, no. (%)					
Wave, no. (%)					

Data are presented as mean (SD), unless otherwise stated.
* Baseline characteristic considered a potential confounder.
PF-OA, patellofemoral osteoarthritis; TF-OA, tibiofemoral osteoarthritis.

Supplementary table 2. Inverse propensity weighted risk ratios for the six-year cumulative reoperations, revisions, and mortality in the PFA Trial Surgeon Group vs. PFA Non-trial Surgeon Group.

Variable	PFA Trial Surgeon Group (XXX knees)	PFA Non-trial Surgeon Group (XXX knees)	IPW RR (95%CI)
Revision, no (%)			
Reoperation, no (%)			
Mortality, no (%)			

Risk ratios estimated from logistic regression with propensity score weighting (IPW RR).
95%CI, 95% confidence interval; IPW, Inverse propensity weighted; PFA, patellofemoral arthroplasty; RR, risk ratio.

Supplementary table 3. Further adjusted risk ratios based on multiple logistic regression for the six-year cumulative reoperations, revisions, and mortality in the PFA Trial Surgeon Group vs. PFA Non-trial Surgeon Group.

Variable	PFA Trial Surgeon Group (XXX knees)	PFA Non-trial Surgeon Group (XXX knees)	Further adjusted RR (95%CI)
Revision, no (%)			
Reoperation, no (%)			
Mortality, no (%)			

Risk ratios estimated from multiple logistic regression with g-computation, adjusted for additional potential confounders (further adjusted RR).
95%CI, 95% confidence interval; PFA, patellofemoral arthroplasty; RR, risk ratios.

Supplementary table 4. Inverse propensity weighted risk ratios for the six-year cumulative reoperations, revisions, and mortality in the PFA Trial Surgeon Group vs. PFA Non-trial Surgeon Group, based on additional potential confounders.

Variable	PFA Trial Surgeon Group (XXX knees)	PFA Non-trial Surgeon Group (XXX knees)	Further adjusted IPW RR (95%CI)
Revision, no (%)			
Reoperation, no (%)			
Mortality, no (%)			

Risk ratios estimated from logistic regression with propensity score weighting (IPW RR), based on additional potential confounders. 95%CI, 95% confidence interval; IPW, Inverse propensity weighted; PFA, patellofemoral arthroplasty; RR, risk ratio.

7. References

AJRR - The American Joint Replacement Registry. Annual Report 2021. Available at: https://www.aaos.org/registries/publications/ajrr-annual-report/

Bendixen NB, Eskelund PW, Odgaard A. Failure modes of patellofemoral arthroplasty - registries vs. clinical studies: a systematic review. Acta Orthop. 2019;90:473-478

Clement ND, Howard TA, Immelman RJ, et al. Patellofemoral arthroplasty versus total knee arthroplasty for patients with patellofemoral osteoarthritis. Bone Joint J. 2019;101:41-46

Collins R, Bowman L, Landray M, Peto R. The Magic of Randomization versus the Myth of Real-World Evidence. N Engl J Med. 2020 Feb 13;382(7):674-678. doi: 10.1056/NEJMsb1901642. PMID: 32053307.

Dunbar MJ, Robertsson O, Ryd L. What's all that noise? The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty. Acta Orthop Scand. 2004;75:119–126. doi: 10.1080/00016470412331294355

Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P, Berlin J, Senn S, Day S, Barbachano Y, Loder E. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017 Dec 19;318(23):2337-2343. doi: 10.1001/jama.2017.18556. PMID: 29260229.

Groenwold RHH. Trial Emulation and Real-World Evidence. JAMA Netw Open. 2021 Mar 1;4(3):e213845. doi: 10.1001/jamanetworkopen.2021.3845. PMID: 33783521.

Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758-64. doi: 10.1093/aje/kwv254. Epub 2016 Mar 18. PMID: 26994063; PMCID: PMC4832051.

Hiemstra B, Keus F, Wetterslev J, Gluud C, van der Horst ICC. DEBATE-statistical analysis plans for observational studies. BMC Med Res Methodol. 2019 Dec 9;19(1):233. doi: 10.1186/s12874-019-0879-5. PMID: 31818263; PMCID: PMC6902479.

Joseph MN, Achten J, Parsons NR, Costa ML; PAT Trial Collaborators. The PAT randomized clinical trial. Bone Joint J. 2020 Mar;102-B(3):310-318. doi: 10.1302/0301-620X.102B3.BJJ-2019-0723.R1. PMID: 32114806.

Lewis PL, Tudor F, Lorimer M, et al. Short-term revision risk of patellofemoral arthroplasty is high: an analysis from eight large arthroplasty registries. Clin Orthop Relat Res. 2020;478:1222-1231.

Lin W, Dai YK, Dong CL, Piao K, Hao K, Wang F. Joint awareness after patellofemoral arthroplasty evaluated with the forgotten joint score: a comparison study. Orthop Surg. 2021;13:833-839

Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011 Jul;39(7 Suppl):30-3. doi: 10.1177/1403494811401482. PMID: 21775347.

National Joint Registry for England, Wales and Northern Ireland. 18th annual report. 2021. Available at: https://reports.njrcentre.org.uk/Portals/0/PDFdownloads/NJR%2018th%20Annual%20Report%202021.pdf. Accessed December 30, 2021

Odgaard A, Lindberg-Larsen M, Schrøder H, et al. Annual report of the Danish Knee Arthroplasty Register. 2021. Available at: https://www.sundhed.dk/content/cms/99/4699_dkr_aarsrapport_2021.pdf. Accessed December 30, 2021

Odgaard A, Kappel A, Madsen F, Kristensen PW, Stephensen S, Attarzadeh AP. Patellofemoral Arthroplasty Results in Better Time-weighted Patient-reported Outcomes After 6 Years than TKA: A Randomized Controlled Trial. Clin Orthop Relat Res. 2022 Sep 1;480(9):1707-1718. doi: 10.1097/CORR.0000000000002178. Epub 2022 Mar 21. PMID: 35315804; PMCID: PMC9384928.

Pedersen AB, Mehnert F, Odgaard A, Schrøder HM. Existing data sources for clinical epidemiology: The Danish Knee Arthroplasty Register. Clin Epidemiol. 2012;4:125-35. doi: 10.2147/CLEP.S30050. Epub 2012 May 7. PMID: 22701092; PMCID: PMC3372970.

NOMESCO Classification of Surgical Procedures (NCSP), version 1.15
Available at: https://norden.diva-portal.org/smash/get/diva2:970547/FULLTEXT01.pdf

Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA. 2022 Dec 27;328(24):2446-2447. doi: 10.1001/jama.2022.21383. PMID: 36508210.

Christensen R, Ranstam J, Overgaard S, Wagner P. Guidelines for a structured manuscript: Statistical methods and reporting in biomedical research journals. *Acta Orthop*. 2023;94:243-249. Published 2023 May 10. doi:10.2340/17453674.2023.11656

Chatton A, Le Borgne F, Leyrat C, Gillaizeau F, Rousseau C, Barbin L, Laplaud D, Léger M, Giraudeau B, Foucher Y. G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. Sci Rep. 2020 Jun 8;10(1):9219. doi: 10.1038/s41598-020-65917-x. PMID: 32514028; PMCID: PMC7280276.

Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

Goetghebeur E, le Cessie S, De Stavola B, Moodie EE, Waernbaum I; “on behalf of” the topic group Causal Inference (TG7) of the STRATOS initiative. Formulating causal questions and principled statistical answers. Stat Med. 2020 Dec 30;39(30):4922-4948. doi: 10.1002/sim.8741. Epub 2020 Sep 23. PMID: 32964526; PMCID: PMC7756489.

Haukoos JS, Lewis RJ. The Propensity Score. JAMA. 2015 Oct 20;314(15):1637-8. doi: 10.1001/jama.2015.13480. PMID: 26501539; PMCID: PMC4866501.

Thomas L, Li F, Pencina M. Using Propensity Score Methods to Create Target Populations in Observational Clinical Research. JAMA. 2020 Feb 4;323(5):466-467. doi: 10.1001/jama.2019.21558. PMID: 31922529.

Hernán MA. Causal analyses of existing databases: no power calculations required. J Clin Epidemiol. 2022 Apr;144:203-205. doi: 10.1016/j.jclinepi.2021.08.028. Epub 2021 Aug 27. PMID: 34461211; PMCID: PMC8882204.

Blanche, Paul Frédéric, Anders Holt, and Thomas Scheike. "On logistic regression with right censored data, with or without competing risks, and its use for estimating treatment effects." *Lifetime Data Analysis* (2022): 1-42.

Peng G, Liu M, Guan Z, Hou Y, Liu Q, Sun X, Zhu X, Feng W, Zeng J, Zhong Z, Zeng Y. Patellofemoral arthroplasty versus total knee arthroplasty for isolated patellofemoral osteoarthritis: a systematic review and meta-analysis. J Orthop Surg Res. 2021 Apr 15;16(1):264. doi: 10.1186/s13018-021-02414-5. PMID: 33858458; PMCID: PMC8048312.

Donald B Rubin (2005) Causal Inference Using Potential Outcomes, Journal of the American Statistical Association, 100:469, 322-331, DOI: 10.1198/016214504000001880

Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology. 2007 Nov;18(6):805-35. doi: 10.1097/EDE.0b013e3181577511. PMID: 18049195.

APPENDIX 1: DATA HANDLING

A1.1 Data extraction

A1.1.1 DKAR

Data from DKAR was delivered in two files:

- primDKAR

  Contains all information recorded on the primary procedures

  - Procedure identification
    - Person identifier (CPR)
    - Hospital
    - Side
    - Procedure date
  - Background data
    - Height
    - Weight
    - Modified Charnley classification [Dunbar 2004]
    - Etiology
    - Previous procedures to knee
    - American Knee Society Score (function and knee indices)
  - Operation
    - Femur component manufacturer and make
    - Femur component fixation
    - Patella component type (all polyethylene or metal-backed polyethylene)
    - Patella component fixation
    - Bone cement manufacturer and make
    - Surgical procedure time
    - Type of operating theatre air change (laminar flow or conventional theatre)
    - Type of anaesthetics (local, regional or general)
    - Use of local infiltration anaesthetics (yes or no)
    - Intraarticular catheter for delivery of local anaesthetics
    - Antibiotic prophylaxis (none or antibiotic type)
    - Planned duration of AB prophylaxis
    - Thromboprophylaxis (drug name)
    - Use of tourniquet (yes or no). If yes, then also pressure and duration.
    - Complications (none or specify)
    - Use of navigation system (yes or no)
    - Surgical approach
    - Use of intraarticular drain (yes or no)
- revDKAR
  - Contains information about revisions. As above and
    - Previous revisions

        - Number of previous revisions
    - Current revision
        - Indication, also information about bacteriology if infection
        - Type of implant prior to current revision

A1.1.2 DNPR

The data from DNPR was delivered in the following files:

- t_sksopr_1.asc

  Contains information on primary procedure:

  - Unique contact identifier (id in table t_adm_1.asc)
  - Optional supplementary codes (side)
  - Date of the procedure
  - Hospital
  - Department
  - Procedure code (KNGB13 in all cases)
- t_adm_1

  Contains information about primary procedure contacts:

  - Unique contact identifier
  - Hospital
  - Department
  - Person identifier (CPR)
- t_sksopr_2.asc

  Contains information on secondary procedures (reoperations):

  - Unique contact identifier (id in table t_adm_2.asc)
  - Optional supplementary codes (side)
  - Date of the procedure
  - Hospital
  - Department
  - Procedure code (KNG* or KNFQ19)
- t_adm_2

  Contains information about secondary procedure contacts:

- sds_sghklasse

  Contains information about hospitals. This table is used to convert hospital codes into legible text.

- t_person.asc

  Contains information about the status of each person identifier. For the current data set, these codes were in use:

- 1: alive and residing in Denmark
- 80: emigrated
- 90: deceased

A1.1.3 Concordance check

For patients that existed in both primDKAR and primDNPR, agreement between records was checked for hospital, side, and date of procedure.

Many cases of disagreement were caused by hospital name inconsistency and could be automatically corrected. In cases, where hospital inconsistency could not be explained by name inconsistency, we retained both records in the data set. For all other cases of discordance, the records were retained in the data set.

A1.1.4 Hospital notes

For each record, we contacted the relevant hospital with the person identifier (CPR number) and requested hospital notes and radiographs relating to orthopaedic treatment of a knee and recorded from the beginning of the accrual period to 31st December 2020. All hospitals responded to our request with the exception of a single private hospital that had closed. Most hospitals delivered the requested information, but some observations were lost, as hospitals could not locate the notes (figure F2).

A1.1.1 Missing data set

A unit of observation is considered to be missing, if hospital notes for a given primary procedure cannot be retrieved in spite of being recorded in one of the source registers. Surgeon name and full verification of the procedure can only be obtained by reading the procedure note. The unit of observation is also considered to be missing in the hypothetical situation where hospital notes are available, but the minimal data set (defined below) cannot be determined.

A1.1.2 Minimal data set

The information is verified by access to hospital notes with a full description of the primary procedure. The minimal data set is defined as full and reliable when all of the following are verified:

- Patient ID (CPR number)
- Side
- Date of primary procedure
- Name of hospital where the primary procedure was performed
- Surgeon name
- Vital status per 01.01.2022

A1.1.3 Full data set

Hospital notes provide information for both the primary procedure and a possible subsequent revision or reoperation.

A1.1.4 Organisation of data for analyses

All units of observation will be organised into a data set according to a data template provided by the statistical analyst.