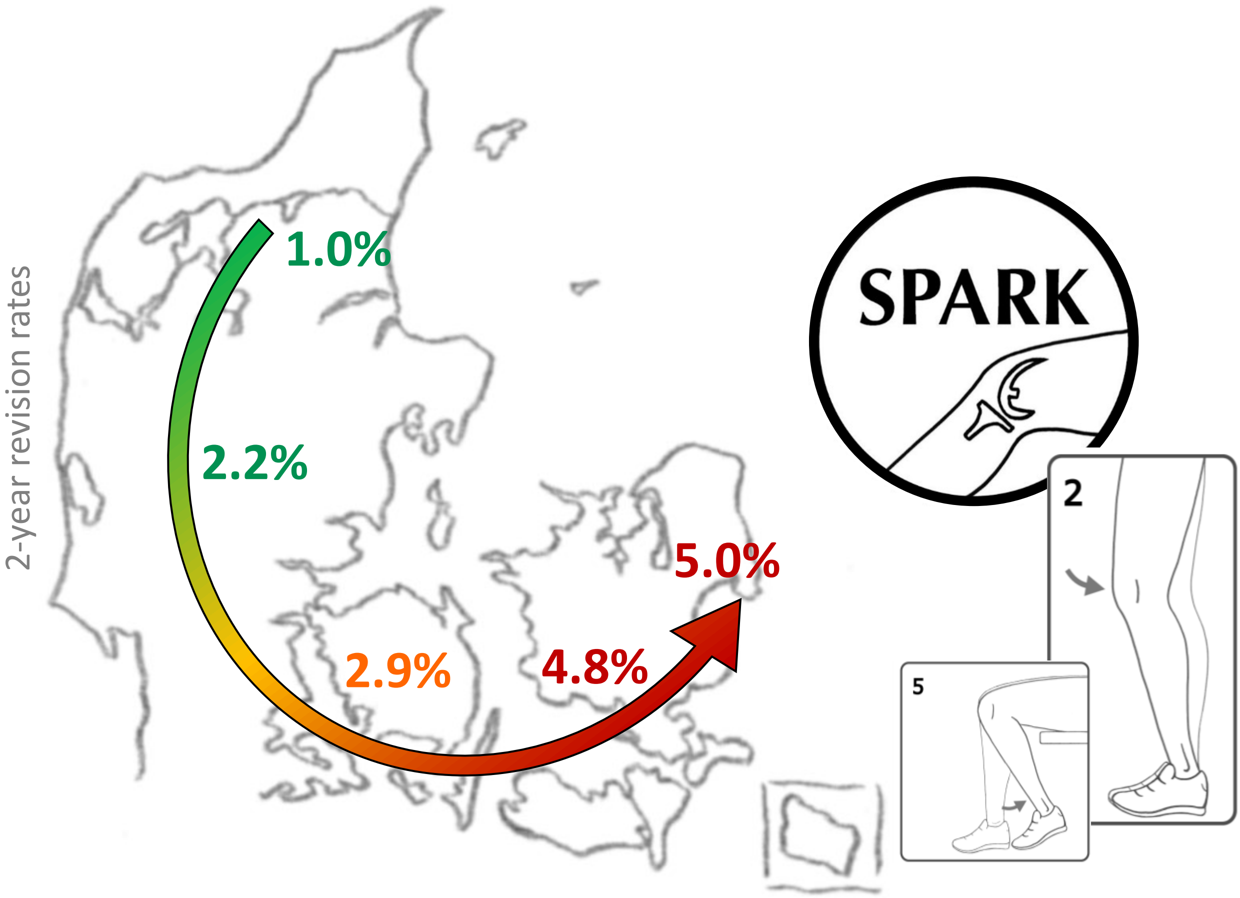
UNIVERSITY OF COPENHAGEN
THE GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES

PhD Thesis

Anne Mørup-Petersen



Evaluation of Danish knee replacement surgery: Patient-reported outcomes versus register data

The SPARK Study

Variation in patient Satisfaction, Patient-reported outcome measures, radiographic signs of Arthritis and Revision rates in Knee arthroplasty patients in three Danish regions

This thesis has been submitted to the Graduate School of Health and Medical Sciences
University of Copenhagen, 14 February 2020

PhD Thesis

Evaluation of Danish knee replacement surgery:

Patient-reported outcomes versus register data

The SPARK study

Variation in patient **S**atisfaction, **P**atient-reported

outcome measures, radiographic signs of **A**rthritis, and **R**evision rates

in **K**nee arthroplasty patients in three Danish regions

Anne Mørup-Petersen

Author

Anne Mørup-Petersen, MD
University of Copenhagen
Department of Orthopaedic Surgery, Herlev and Gentofte Hospital
Kildegaardsvej 28, 2900 Hellerup, Denmark
Mobile phone: (+45) 26 13 15 10
Email: anne.moerup-petersen.02@regionh.dk/ morup.anne@gmail.com

Academic supervisors

Primary supervisor: Prof. Anders Odgaard, Consultant Orthopaedic Surgeon, FRCS, DMSc
University of Copenhagen/ Department of Orthopaedic Surgery
Herlev and Gentofte Hospital/ Rigshospitalet, Copenhagen, Denmark

Primary co-supervisor: Prof. Michael Rindom Krogsgaard, Consultant Orthopaedic Surgeon, DMSc, PhD
University of Copenhagen/ Department of Orthopaedic Surgery
Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

Co-supervisor: Ass.prof. Mogens Laursen, Consultant Orthopaedic Surgeon, PhD
Aalborg University/ Department of Orthopaedic Surgery
Aalborg University Hospital, Aalborg & Farsø, Denmark

Assessment committee

Chairman

Prof. Jes Bruun Lauritzen, Consultant Orthopaedic Surgeon, DMSc
University of Copenhagen/ Department of Orthopaedic Surgery
Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

Opponents

Prof. David Beard, GDPhys, MCSP, MA, MSc, DPhil, FRCS (hon)
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
University of Oxford, United Kingdom

Prof. Søren Kold, Consultant Orthopaedic Surgeon, DMSc
Aalborg University, Department of Orthopaedic Surgery
Aalborg University Hospital, Aalborg, Denmark

Funding

The PhD project was primarily funded by the Health Research Fund of the Capital Region of Denmark, the 2015 grant. Contributions have been made from The Department of Orthopaedic Surgery, Herlev and Gentofte Hospital and The Fund of The Kjærsgaard Family, Sunds.

Printed in Copenhagen, 2020, in fonts Calibri and Calibri Light.

The Graduate School of Health and Medical Sciences, University of Copenhagen, February 2020

List of papers

Study no.	Short name	Paper
Study I	SPARK Pre	**Large variation in revision rates after primary knee arthroplasty: A matter of patient selection in primary surgery? Baseline data from 1452 patients in the prospective multicenter cohort study, SPARK.** Mørup-Petersen A, Laursen M, Madsen F, Krogsgaard MR, Mongelard KBG, Roemer L, Winther-Jensen M, Odgaard A. *Submitted.*
Study II	SPARK Post	**Revision rates differ widely, yet patient-reported outcomes after primary knee replacement are the same: 1-year results from the prospective multicenter cohort study, SPARK.** Mørup-Petersen A, Laursen M, Madsen F, Krogsgaard MR, Winther-Jensen M, Odgaard A. *Submitted.*
Study III	UCLA	**Translation, cultural adaptation and measurement properties of the Danish version of the UCLA Activity Scale for use in hip and knee replacement patients.** Mørup-Petersen A, Skou ST, Holm CE, Holm PM, Varnum C, Krogsgaard MR, Laursen M, Odgaard A. *Manuscript draft.*
Study IV	ROM	**Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale.** Mørup-Petersen A, Holm PM, Holm CE, Klausen TW, Skou ST, Krogsgaard MR, Laursen M, Odgaard A. *The Journal of Arthroplasty, 2018. 33(9): p. 2875-2883.e3.* ***Corrigendum added**:* *The Journal of Arthroplasty, 2019. 34(8): p. 1860-1861.*

Preface

So many people deserve to be thanked as I complete this work. The project would not have been possible without generous donations from The Health Research Fund of The Capital Region of Denmark. I thank Steen Mejdahl and Claus Munk Jensen, my present and my former Consultant Orthopaedic Surgeon, Head of Department of Orthopaedic Surgery, Herlev and Gentofte Hospital, for allowing me the opportunity to perform this study in your department. I have been met with great enthusiasm by nurses, physiotherapists and secretaries, and also by consultant orthopaedic surgeons, whom I particularly thank for sharing knowledge and research ideas with me. Special thanks to secretaries Teresa Mastrakoulis, Aviaaja Højer, Camilla Stokkebro and Belinda Ierst for valuable help, encouragement and good laughs. In the years leading up to this work, I was lucky to be inspired and informally mentored by Camilla Ryge who advised me to seize this opportunity, and by late Charlotte Asperud, whose work and personality I deeply admired.

I have had the joy of working on three different clinical research projects as part of this thesis. For fruitful cooperation and good advice, I thank Karen Dyreborg, Pætur Mikal Holm, Christina Enciso Holm, Søren T. Skou, Claus Varnum, Charlotte Fredborg, Tobias Wirenfeldt-Klausen, Lars Jansen, Jonathan Comins, Karl Bang Christensen, Carsten Bogh Juhl, Michelle Möger Andersen, Finn Steffens at AIS Sprog, Gert Galster at Sund-IT and Henrik Schrøder and staff at the Department of Orthopaedic Surgery, Næstved Hospital. Special thanks to physiotherapist Hanne Hornshøj for performing measurements in the ROM study, to Anna Lilja Secher and Anna Pors Nielsen for valuable critical reviews, and to Susanne Boel and Anette Enemark Larsen, whom I have enjoyed following ever since we worked together during the preparatory studies.

The SPARK study was highly dependent on surgeons and local enthusiasts in Aarhus and Farsø/Aalborg, without whose efforts no patients would have entered or stayed in the study. For this help, I thank Ulla Hornum, Gitte Broholm, Patricia Svare, Tina Isaksen, Gitte Raae Høybye and colleagues, and all surgeons who invited patients to participate: Andreas Kappel, Jesper Villefrance, Poul Torben Nielsen, Poul Hedevang Christensen, Ole Simonsen, Claes Blom, Claus Fink Jepsen, Søren Rytter and medical students at Gentofte Hospital. Special thanks to Frank Madsen who went out of his way for the SPARK study and taught me a lot about knee replacement surgery during my short stay in Aarhus. Warm thanks to Philip Helmer, Anette Liljensøe and IT-supporters at Procordo Software Aps for offering data collection services free of charge and for support during the project.

Knee radiographs were meticulously sought out by Birthe Schøning and Heidi Ladefoged Poulsen and sorted by Naima Elsayed and Camilla Grube. Radiograph analyses were coordinated by Anders Odgaard and performed by radiologists, Kristian Mongelard and Lone Rømer, and by surgeons, Svend Erik Østgaard, Mogens Laursen, Andreas Kappel, Lasse Enkebølle Rasmussen, Snorre Stephensen, Thomas Bjerno, Lars Peter Møller, Thomas Lind, Anders Odgaard, Henrik Schrøder, Frank Madsen, Claus Fink Jepsen and Søren Rytter. Thank you for your many hours of hard work.
I am grateful to the many patients who offered their time and efforts, and without whose overwhelming interest and active participation we would not have succeeded with our studies.

During my enrollment as a PhD student at University of Copenhagen, I have attended 15 different courses, all inspiring and well prepared. I thank the statisticians at Department of Biostatistics, University of Copenhagen for well-planned courses and subsequent phone consultations. As frustrating as it has been to become familiar with statistical analysis through two complex software systems (SAS and R), mostly on my own, I would not have been without this lesson. Every battle with the machine (a few excluded) helped me in the overall understanding of analyses and rationales. I consider it a gift to have brought mathematical thinking back into my life and to bring a sense of statistics and scientific rationale into my future clinical and scientific work.

I thank Matilde Winther-Jensen at Center for Clinical Research and Prevention, Frederiksberg Hospital, for kindly and wittingly sharing your wide epidemiological and statistical knowledge with me and for helping me gain an overview of my data. I learned so much from working with you, and I treasure the weeks I spent in your professional and warm work environment.

Office work has been so different from the clinical work I knew. Without good company in the PhD office, this journey would have been so lonely. I thank my wonderful colleagues for helping me stay sane and make me look forward to going to work every day. Special thanks to Anne Brun Hesselvig for the little extra insights and mad sense of humor.

To my supervisor, Anders Odgaard, I am truly grateful to have been given this opportunity to learn to do research and work in depth with a subject of my interest. While offering me your kind, insightful supervision and brilliant ideas, you have trustfully let me handle this project on my own which was good for me. Warm thanks to my co-supervisors, Michael Rindom Krogsgaard and Mogens Laursen for genuine participation and excellent feedback on my work. I hope to be able to work with you all again in the future.

I thank my dear family, in-laws and friends for tremendous support; my friends for still being there, my mother for teaching me to face challenges with enthusiasm, and my late father and grandfather for teaching me to be curios and study hard -how I would have loved to share this with you both. I thank my brother and his wife for reminding me of the (other) joys of life, my husband for being so smart, nice and constantly encouraging, and my beloved children for bearing with me
-I can’t wait to spend more time with you.

Anne Mørup-Petersen
Copenhagen, February 2020

Dictionary

Definitions

Term	Definition
Goniometer	Device used to measure joint position (angle between two bones) (DK: vinkelmåler)
Implant/ prosthesis	An artificial construction to permanently replace a joint surface
Procedure	Operation
Revision	Any subsequent exchange, removal, addition or modification of implants
Revision rate	Proportion of patients who had their implant revised within a certain time after primary surgery

Abbreviations

Abbreviation	English	Danish
OA	Osteoarthritis	Artrose/osteoartrose/slidgigt
KA	Knee arthroplasty	Knæalloplastik: knæprotese af enhver type
pKA	Primary knee arthroplasty	Førstegangs-knæprotese
TKA	Total knee arthroplasty	Helprotese, erstatter alle knæets tre ledkamre
UKA	Unicompartmental knee arthroplasty	Samlebetegnelse for de tre typer delprotese
MUKA	Medial unicompartmental knee arthroplasty	Delprotese på knæets inderside (medial)
LUKA	Lateral unicompartmental knee arthroplasty	Delprotese på knæets yderside (lateralt)
PFA	Patellofemoral (knee) arthroplasty	Delprotese mellem knæskallen og lårbenet
ROM	Range of motion	Bevægeudslag (bøje-/strækkeevne)
PROM	Patient-reported outcome measure	Spørgeskema til at måle personens subjektive tilstand

Table of contents

English Summary

Results after primary knee arthroplasty (pKA) surgery are traditionally evaluated by comparing revision rates, e.g. among hospitals, implant types or surgeons. In the last decade, revision rates have persistently differed among Danish regions and high-volume hospitals, e.g. between 1 and 5% per 2 years in 2015. The SPARK study was initiated a) to determine if these variations were a sign of true variations in quality of pKA surgery and b) to identify possible differences in patient selection that might offer an explanation to revision rate differences.

A prospective observational triple-center cohort study followed pKA patients in two low-revision-rate hospitals (Aarhus and Farsoe/Aalborg) and one high-revision-rate hospital (Gentofte/ Copenhagen) from baseline to 1-year postoperatively (2016-19). Patients completed 5 sets of patient-reported outcome measures (PROMs) including the primary outcome, Oxford Knee Score (OKS). All available standing postero-anterior knee radiographs were classified according to the Ahlbäck and Kellgren-Lawrence classifications of osteoarthritis (OA) and also by a new method where 13 KA surgeons (through 17,767 head-to-head radiograph comparisons) ranked all 1051 available knee radiographs from no. 1, “the knee expected to cause the most symptoms” to no. 1051, “the least symptoms”.

Baseline data were available for 1452 patients (56% of operated patients, response rate 89%) and revealed some differences in patient selection among the hospitals, though most were paradox to known revision risk factors. For example, patients were 1.3-2.0 years older in the high-revision-rate hospital, and in one low-revision-rate hospital (Aalborg), more patients were male (56 vs. 43-45%) and BMI was 1.5-1.7 points higher. The second low-revision-rate hospital (Aarhus) operated fewer patients with mild degrees of OA but used unicompartmental implants far more frequently (49% compared to 14-22%). Regional pKA incidences were up to 28% higher in the high-revision-rate region. At 1-year follow-up, 90% of participants responded and 97% responded at least once postoperatively. Patient satisfaction, willingness to repeat surgery and OKS were the same across the three hospitals at 1-year, and the same was the case for most other PROM results.

In conclusion, this study of pKA patients in three Danish hospitals known to present very different revision rates found some differences in patient selection but as most variations were contradictory to known revision risk factors, these differences could not explain the observed hospital variations in revision rates. PROM results after surgery were the same in the three hospitals, and therefore, on a hospital level, high revision rates were not associated to poor quality of surgery, or vice versa.

This thesis also describes a) the translation, cultural adaptation and test of measurement properties of UCLA Activity Scale for use in Danish hip and knee OA patients, and b) the development and validation of Copenhagen Knee ROM Scale; a short illustration-based questionnaire for patients to report their ability to passively flex and extend the knee.

Dansk Resumé

Hvert år får 8.500 danskere med slidgigt (artrose) i knæet indsat et kunstigt knæ (en primær knæalloplastik). Udskiftningen af knæets ødelagte ledflader med en protese af metal og plastik fører hos de fleste patienter til smertefrihed, øget bevægelighed og forbedret livskvalitet, men en vis andel ender med at blive genopereret (revideret), dvs. få udskiftet (dele af) protesen. Det kan f.eks. skyldes åbenlyse komplikationer (f.eks. infektion), eller at patienten fortsat har smerter. Dansk Knæalloplastikregister modtager indberetning om 97% af de førstegangsoperationer og 93% af de revisioner, der foretages på offentlige og private sygehuse i Danmark. Siden registrets grundlæggelse i 1997 har man observeret klare forskelle i revisionsrater (andel af reviderede patienter pr. tid efter operation) mellem danske regioner. Der er nogen variation mellem hospitalerne indenfor hver region, men den overordnede tendens er klar: Jo tættere man kommer på København, des flere patienter gennemgår en revision. F.eks. blev 5% af de patienter, der var førstegangsopereret i Region Hovedstaden i 2012, revideret indenfor to år, mens det kun gjaldt 2% af patienterne i Region Midtjylland og 1% i Region Nordjylland. Disse forskelle er så store, at det ikke alene kan skyldes tilfældig variation, og det er nærliggende at fortolke tallene således, at kvaliteten af behandlingen er bedre, jo længere væk fra København operationen bliver foretaget.

Registerdata gør det muligt at overvåge udviklingen indenfor behandlingsområdet ”kunstige knæ” over tid, og med revisionsrater kan man tilsyneladende følge behandlingskvaliteten og identificere dårligt fungerende proteser eller uhensigtsmæssige arbejdsgange sikkert og hurtigt. Til at sammenligne patienternes operationsresultater er revisionsrater dog ikke nødvendigvis et tilstrækkeligt eller retvisende mål. Når en operation er foretaget for at afhjælpe smerter og give bedre knæfunktion, kan det undre, at operationens kvalitet måles på risikoen for revision – en risiko, der ikke alene bestemmes af operationens resultat, men også er påvirket af faktorer som sundhedsvæsenets kapacitet, patientens motivation for fornyet operation og kirurgens vurdering af at kunne bedre et problem. At vurdere, hvorvidt en revision skal tilbydes, er langtfra enkelt, og mange patienter med ringe resultat er slet ikke er egnede til fornyet operation. Af disse årsager er revisionsrater ikke nødvendigvis et relevant kvalitetsmål for patienterne, og man risikerer at gå glip af nuanceret information om resultaterne hos de ca. 90%, der aldrig bliver revideret.

For at foretage en fyldestgørende undersøgelse af operationsresultater efter indsættelse af kunstige knæ på tværs af danske regioner tog vi i 2015 initiativ til ”SPARK-studiet” (The SPARK study: Variation in patient Satisfaction, Patient-reported outcome measures, radiographic signs of Arthritis, and Revision rates in Knee arthroplasty patients in three Danish regions). Spørgeskemaer om den subjektive tilstand, udfyldt af patienterne (patient-reported outcome measures = PROM), var det primære effektmål, der skulle belyse hvorvidt kvaliteten af førstegangs-operationer varierede mellem regionerne, sådan som registeroplysningerne kunne indikere.

Studiet omfattede 1452 patienter opereret for knæartrose på tre store ortopædkirurgiske afdelinger i Århus (Region Midtjylland), Gentofte (Region Hovedstaden) og Farsø (Aalborg, Region

Nordjylland). Patienterne blev fulgt med omfattende (overvejende elektroniske) spørgeskemasæt fem gange i forløbet, fra de blev opskrevet til at få indsat et kunstigt knæ til ét år efter protesen var indsat. Det skete for at kunne sammenligne både patienternes udgangspunkt og deres resultater af operationen på tværs af de tre afdelinger. Spørgeskemaet forud for operationen blev besvaret af 89% af de tilmeldte patienter (svarende til 56% af alle opererede på de tre afdelinger), og af disse svarede 97% på ét eller flere skemaer efter operationen (90% ved 1 år). Derudover blev patienterne sammenlignet på tværs af hospitalerne ud fra sværhedsgraden af knæartrose på stående røntgenoptagelser taget før operationen; i alt 1051 patienters billeder blev rangeret fra nr. 1 (sværeste) til nr. 1051 (mildeste grad af artrose) af 13 kirurger fra forskellige landsdele. Desuden klassificerede to røntgenlæger graden af artrose på de samme røntgenbilleder efter to traditionelle rangskalaer.

Data forud for operation viste, at patienterne i Gentofte var ca. 2 år ældre end patienterne på de øvrige hospitaler, og i Farsø var de i højere grad overvægtige (gennemsnitligt ca. 1,5 BMI-point højere). Alle tre metoder til klassifikation af artrosens sværhedsgrad ud fra røntgenbilleder viste, at færre patienter blev opereret på baggrund af mild artrose i Århus end på de to øvrige hospitaler. Ifølge 2017-data fra Landspatientregistret fik 28% flere hovedstadspatienter i aldersgruppen 60-79 år et kunstigt knæ i forhold til Region Midtjylland (og 13% flere end i Region Nordjylland), og samme tendens var gældende i de foregående år (justeret for indbyggertal). Tilsammen kan disse oplysninger indikere, at der er en lavere tærskel for at tilbyde førstegangsoperation med knæprotese i Region Hovedstaden, særligt i sammenligning med Region Midtjylland. Det understreges i den forbindelse, at SPARK-undersøgelsen alene omfatter de patienter, der er blevet tilbudt (og har ønsket) operation. Patienternes begrundelser for at ønske operation (f.eks. smerter, arbejde, motion, problemer med at klare sig selv, osv.) varierede ikke på tværs af hospitalerne.

Såvel før som efter operationen angav patienterne samme symptomgrad, aktivitetsniveau og bevægelighed i knæet på tværs af de tre hospitaler, dog med enkelte udsving. Patienterne i Århus, hvor delproteser benyttes hyppigere, opnåede marginalt bedre strækkeevne. Farsø-patienterne oplevede en større forbedring i generel helbredstilstand, og på Gentofte kom de sig generelt hurtigere efter operation; en forskel der dog var udjævnet ved 3-måneders opfølgningen.

I alt blev 28 patienter (1,9%) revideret indenfor det første år, og de udgik af 1-års-opgørelsen af patientrapporterede data. I Århus blev færre patienter revideret end på de øvrige hospitaler (0,6% mod 2,0-2,4%). Det var dog ikke en statistisk sikker forskel, bl.a. pga. det lave antal patienter. Ved 1-års kontrollen havde 19% af patienterne i Århus en ringe fremgang i knæscore (defineret som mindre end 8 points forbedring i Oxford Knee Score) mod hhv. 13 og 14% på de to øvrige hospitaler - en forskel, der heller ikke var statistisk sikker. Vi fandt heller ingen statistisk sikker forskel mellem hospitalerne, hvis man forestillede sig, at de 28 reviderede patienter (som ikke indgik i 1-års opfølgningen) alle hørte til i gruppen med dårlig fremgang, eller når man sammenlignede de senest afgivne svar fra alle de 1414 patienter (97%), der havde svaret på mindst ét af spørgeskemaerne efter operationen.

Patienternes overordnede tilfredshed varierede ikke mellem hospitalerne. Det samme gjaldt andelen af patienter, der efter ét år ville gentage operationen, hvis de fik mulighed for at vælge om: 92% svarede enten ”ja, helt afgjort” eller ”ja, formentlig”. Patienternes egen umiddelbare vurdering af deres knægener på en skala fra 0-100 adskilte sig heller ikke mellem hospitalerne, eller mellem patienter i storbyer, i mindre byer og på landet. Når patienterne blev inddelt i grupper efter sværhedsgrad af knæartrose på røntgen, var der ingen overordnede forskelle mellem hospitalernes resultater målt på Oxford Knee Score eller viljen til at ”gentage” operationen. Dog havde Århus-patienterne ringere resultater end de øvrige i gruppen af i alt 64 patienter med sværest artrose, vurderet på én af de røntgen-skalaerne (Kellgren-Lawrence).

SPARK-undersøgelsen har vist, at patient-rapporterede resultater efter indsættelse af kunstigt knæ er lige gode på tværs af tre store hospitaler beliggende i hver af de tre danske regioner, der udviser de største indbyrdes forskelle i revisionsrater. Der er således ikke holdepunkter for, at patienter opereret på et hospital med høj revisionsrate får behandling af ringere kvalitet end patienter behandlet på ét med lav revisionsrate. Disse fund står i kontrast til, hvad mange års indsamlede registerdata umiddelbart kunne tyde på, og understreger således, at man skal være varsom med at drage konklusioner om operationskvalitet alene på baggrund af revisionsrater.

Undersøgelsens resultater tyder på, at man på tværs af hospitaler og landsdele har forskellig tilgang til viderebehandling i de tilfælde, hvor der er et utilfredsstillende resultat efter indsættelse af kunstigt knæ. Det kan gælde for knækirurger og andet sundhedspersonale, men også patienter, hvis forventninger kan variere afhængigt af f.eks. geografi og information givet forud for operationen. En sådan tendens kan være yderligere forstærket af regionernes forskellige tilbud om efterkontrol efter operation; i de år, hvor SPARK-undersøgelsen foregik, blev patienter på Gentofte Hospital tilbudt 3-måneders kontrol hos kirurgen, mens lægerne på de øvrige hospitaler kun så de patienter, der havde problematiske forløb. Desuden er det tænkeligt, at revisioner avler revisioner, dvs. at kirurger, der er vant til at foretage mange revisioner, er mere tilbøjelige til at tilbyde revision, idet en overskridelse af revisionstærsklen i højere grad bliver almindeligt.

Fremtidige studier bør således fokusere på indikationer for revision. Her kan patient-rapporterede data og systematisk gennemgang af beslutningsprocesser forud for revisioner hjælpe med at kortlægge den gavn, som patienter kan forvente af en revision, afhængigt af tidspunkt, indikation og undersøgelsesfund.

Udover SPARK-studiet inkluderer denne PhD afhandling to forberedende understudier. Det ene, ”Studie III: UCLA” bestod i oversættelse, kulturel tilpasning og afprøvning (validering) af UCLA Aktivitetsskala, hvor knæ- og hofteprotesepatienter selv rangerer deres daglige fysiske aktivitetsniveau på en skala fra 1 til 10 ved hjælp af eksempler på velkendte aktiviteter. I lighed med udenlandske studier har vi blandt 264 patienter vist, at 66% af knæpatienterne og 79% af hoftepatienterne øger deres selvrapporterede aktivitetsniveau i løbet af det første år efter indsættelse af en ledprotese, og typisk med 1-2 trin på skalaen. Det andet understudie, ”Studie IV: ROM” bestod i udvikling af en simpel, tegningsbaseret skala, der gør patienterne i stand til at indberette deres bøje- og strækkeevne i knæet. Det kan være nyttigt i situationer, hvor en traditionel vinkelopmåling ikke kan lade sige gøre, f.eks. ved større forskningsprojekter og ved direkte patientindberetning til registre. Studiet viste, at patienterne forstod metoden og angav deres bevægelighed med større præcision og sikkerhed, end man tidligere har opnået med mere komplicerede spørgeskemaer. I SPARK-studiet kunne skalaen bl.a. demonstrere forskelle i bevægelighed blandt grupper af patienter, der tilbydes forskellige protesetyper. I takt med, at besparelser i sundhedsvæsenet har ført til afskaffelse af ambulant lægelig kontrol efter proteseoperation, er skemaet sidenhen taget i brug på flere danske afdelinger som screening for bevægeproblemer efter operation.

Samlet har arbejderne i denne PhD afhandling vist, at kvaliteten af den operative behandling af knæartrose ikke bør vurderes uden inddragelse af patient-rapporterede resultater. Understudierne har samtidig demonstreret, at inddragelse af patienter i udviklingen af en PROM er kompliceret, men altafgørende for målemetodens anvendelighed. Forfatteren går ind for en systematisk indsamling af relevante PROM-data til brug for fremtidig forskning og kvalitetssikring, forudsat at både indsamling og præsentation af data foregår på en måde, der opleves som vedkommende og gavnlig for såvel den enkelte patient som for de læger, der skal vurdere resultaterne.

1. Introduction

Knee arthroplasty (KA) surgery is an increasingly common and most often successful operation. Every year, approximately 8,500 Danish knee osteoarthritis (OA) patients of mean age 69 years receive this treatment with the purpose to relieve pain and improve knee function and quality of life [95, 112]. To most patients, replacing the destroyed knee cartilage with an artificial knee joint means the end of long-standing knee pain and decreasing mobility [76]; after a few days in the hospital and subsequent recovery, no further treatment is needed.

For some patients, however, the operation is unsuccessful. Previous studies have reported that up to 1 in 5 patients experience long-term pain and are dissatisfied with the result of surgery [5, 20, 86, 90, 91]. A minority of patients have to undergo one or more reoperations (revisions) for reasons including loosening of the implant, infection, pain and instability [76]. This happens to about 7% of patients operated at age 65 years, but in the growing population of knee replacement patients younger than 60 years, the lifetime risk of revision is suggested to be as high as 20% in women and 35% in men [27, 67, 76].

To keep track of knee arthroplasty (KA) operations, all primary knee arthroplasties (pKA) and revisions are principally to be registered in the Danish Knee Arthroplasty Register (DKAR) along with surgical details reported directly from surgeons [74, 81]; current completeness is 97% in pKAs and 93% in KA revisions [113]) Every year, aggregated results are gathered in a report and discussed among surgeons at the annual meeting in the Danish Hip and Knee Arthroplasty Society [114]. As the main surgical outcomes, revision rates provide information of the proportion of patients who have undergone revision during a certain time period after pKA. All five Danish regions keep revision rates below the national standards of 3, 5 and 8% per 1, 2 and 5 years, respectively [112], but single knee arthroplasty centres within regions occasionally fluctuate above these levels. Register data effectively provide an overview of trends in knee replacement surgery and offer fast and detailed results, which are essential for surveillance of e.g. new implants or techniques [8].

Revision rates do have limitations as indicators of surgical quality, though [93]. Patients undergo pKA with a hope of becoming free of pain and improve their knee function and quality of life, but with revision rates, we measure the treatment results by merely counting the number of patients who are revised [34]. Revisions are done for a variety of reasons and apart from cases with deep infection, the decision on whether or not to reoperate and, if so, which methods and implants to use, is rarely straightforward. Previous publications have demonstrated how patients are offered revisions at very different symptom stages depending on primary implant type; patients with medial unicompartmental knee arthroplasties (MUKA) were five times as likely to be revised compared to total knee arthroplasty (TKA) patients with the same Oxford Knee Score (OKS) [34], partly because revision of MUKA's are perceived as less complicated than TKA's and because persistent pain in MUKA patients may be attributed to arthritis of the non-operated joint compartments. For some patients with persistent pain after primary operation, there is no obvious surgical possibility to relieve their suffering. Yet, when using revision rates as the only endpoint, such patients appear as successful cases because their prosthesis is still in place. Thus, ideally, rather than a dichotomous question of revision or not, pKA results should be measured as a spectrum of various degrees of symptom relief [76]. Such information can be provided by valid patient-reported outcome measures (PROMs), i.e. questionnaires that translate symptoms and observations reported directly from patients into scores on a standardized scale.

Rationale

These were the initial concerns, when it was observed that revision rates continuingly were differing among Danish regions [115]. In 2015, revision rates varied significantly, from 1.0% at 2 years in North Denmark Region to 5% in the Capital Region. There was no research to confirm whether regional revision rate variations were 1) a sign of truly varying quality of pKA surgery across Danish regions, 2) a consequence of surgeons offering pKA to patients with different knee OA severities and varying symptom states, or 3) a matter of patients and surgeons in different regions reacting differently to suboptimal results after surgery.

To seek answers to these questions, we initiated the SPARK-study: "Variation in patient Satisfaction, Patient-reported outcome measures, radiographic signs of Arthritis, and Revision rates in Knee arthroplasty patients in three Danish regions". Studies concerning the first two questions form the basis of this thesis (Studies "I: SPARK Pre" and "II: SPARK Post") while data on the third question, i.e. concerning revision patients included in the SPARK project, are yet to be analysed and reported in future publications.

The thesis also includes two preparatory studies. "Study III: UCLA" concerns the translation, cultural adaptation and validation of the UCLA Activity Scale, a questionnaire assessing patient-reported levels of physical activity. "Study IV: ROM" concerns the development of a questionnaire that allows patients to report their own ability to flex (bend) and extend (straighten) the knee; an important surgical outcome measure, which normally requires hospital visits for goniometer measurements performed by a trained health professional.

2. Background

Study I & II: SPARK Pre & Post

The Danish Knee Arthroplasty Register (DKAR)

Following Sweden, Finland and Norway, Denmark was the fourth country in the world to implement a systematic registration of knee replacement procedures when The Danish Knee Arthroplasty Register (DKAR) was founded in 1997 [81]. Data from DKAR is linked to The National Patient Register through patients' unique CPR (civil registration) number, which is used whenever Danish citizens are in contact with healthcare providers, public as well as private [74]. Reporting to DKAR became mandatory in 2006 and completeness reached 90% in 2009 [74]. Most recently, 97% of all

KA procedures were registered in 2017: 97% of the 8,584 pKA procedures and 93% of the 1,037 revision procedures [113]. There was an increasing number of pKA procedures until 2010 where incidence stabilized until a sudden 14% rise was noted in 2018 [113]. Revision incidence reached a maximum in 2012 and was decreasing until a slight rise was noted in 2018.

Organization of Danish knee arthroplasty surgery

Far the majority (94%) of Danish primary knee replacements and almost all revisions (99%) are currently performed in public hospitals, organized in five regions (fig. 1) [112]. In year 1997 to 2000, KA surgery was performed in 39 public and 3 private Danish hospitals. In 2009, numbers had increased to 43 and 22 hospitals, respectively [115]. Since then, a growing focus on specialization led to closing of half of Danish hospitals, and reimbursements were reduced in the private sector. Thus in 2017, only 23 public and 12 private hospitals were responsible for all Danish KA surgery [112].

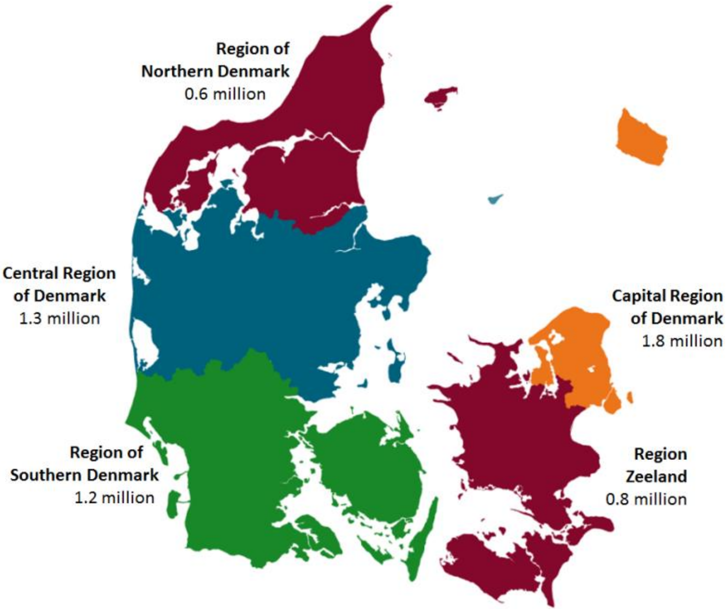


Figure 1. *Number of inhabitants in the five Danish regions (based on an illustration provided by North Denmark Region).*

Almost all public hospitals, that perform pKA surgery, do revisions as well, and only patients with major bone loss or skin defects must be referred to a highly specialized department, of which there is one in each region [116]. In contrast to USA, where only 70% of revisions are performed in the same hospital that inserted the primary knee arthroplasty [23], Danish patients do not often attend a second hospital for further treatment or a second opinion. In case they do, any revision still contributes to the revision rate of the pKA hospital. This thesis concerns pKAs performed in public hospitals only.

Regional variance in revision rates

For more than a decade, DKAR has reported persistent differences in revision rates among the five Danish Regions (fig. 2). With few exceptions, the variance among regions exceed the variance within regions, and in rough figures, the general trend has been a two-fold difference between western and eastern parts of the country [112, 115, 117]. When this study was initiated, the 2-year revision rates differed widely between regions, with the Capital Region presenting the higher rates (fig. 3).

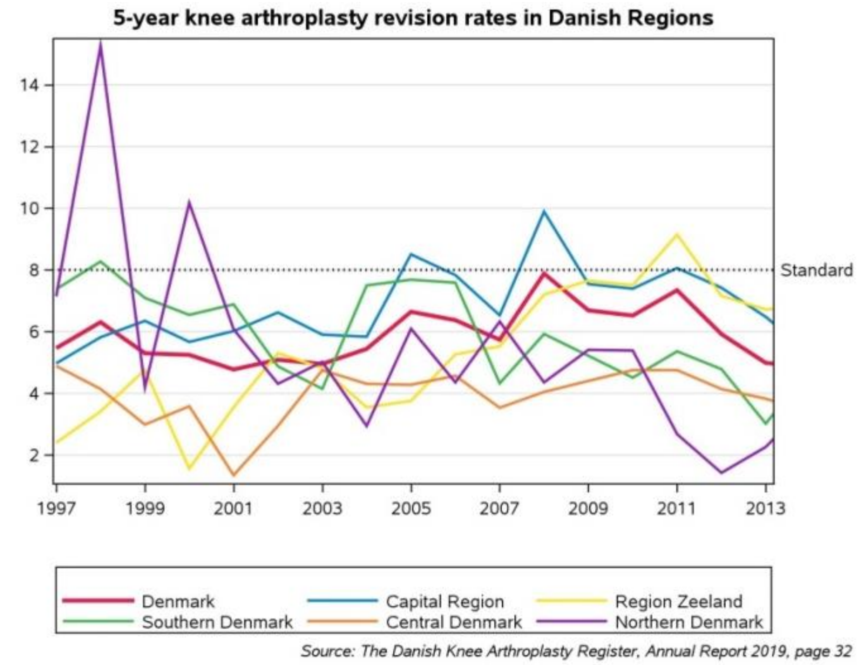


Figure 2. *Revision rates per region at 5 years for patients who had primary knee replacements in each region, displayed per year of primary knee replacement. (Illustration by the Danish Knee Arthroplasty Register)*

A similar geographical trend is noted in Danish hip surgery [118], indicating that a general difference in patient culture or surgeons' approach can contribute to these differences, rather than specific technical aspects of knee surgery. Among Danes, it is often said that people in the western and rural parts of Denmark have a higher pain threshold than the urban people living near Copenhagen. This has, however, to the best of our knowledge, never been proven. An Australian study of 1955 primary hip and knee replacement patients demonstrated how patients from rural presented at an earlier age and with more advanced OA, but since Danish geographical distances are without comparisons to Australian, there is no saying that this would be the case in Denmark [19]. Statistics Denmark have, however, reported how employees have more days on sick leave the closer they live to Copenhagen [58]. The annual Danish National Survey of Patient Experiences ("LUP") consistently reports how patient satisfaction is higher in hospitals in the northern and western parts of the country compared to the eastern parts [119]. In 2009, the highest level of overall patient satisfaction among patients in all medical fields was reported in Farsoe Hospital, Aalborg (North Denmark) and the lowest in Gentofte Hospital (Copenhagen, Capital Region). It was suggested that the regional differences in general patient satisfaction might be a matter of differences in expectations rather than differences in treatment quality [52]. An interregional correction index was even suggested, but never implemented.

Knowledge gap in Danish knee arthroplasty surgery

In the field of Danish KA surgery, there was no existing research to indicate whether differences in revision rates were a reflection of cultural aspects or a sign of true differences in results across regions, or both. The interpretation of research results and register data across Danish KA centres was blurred by these constant and unexplained differences in revision rates. Thus, it was of great interest to compare outcomes after pKA surgery across hospitals. Naturally, if some hospitals offered better treatment than others, surgeons should disseminate their knowledge to the benefit of patients throughout the country.

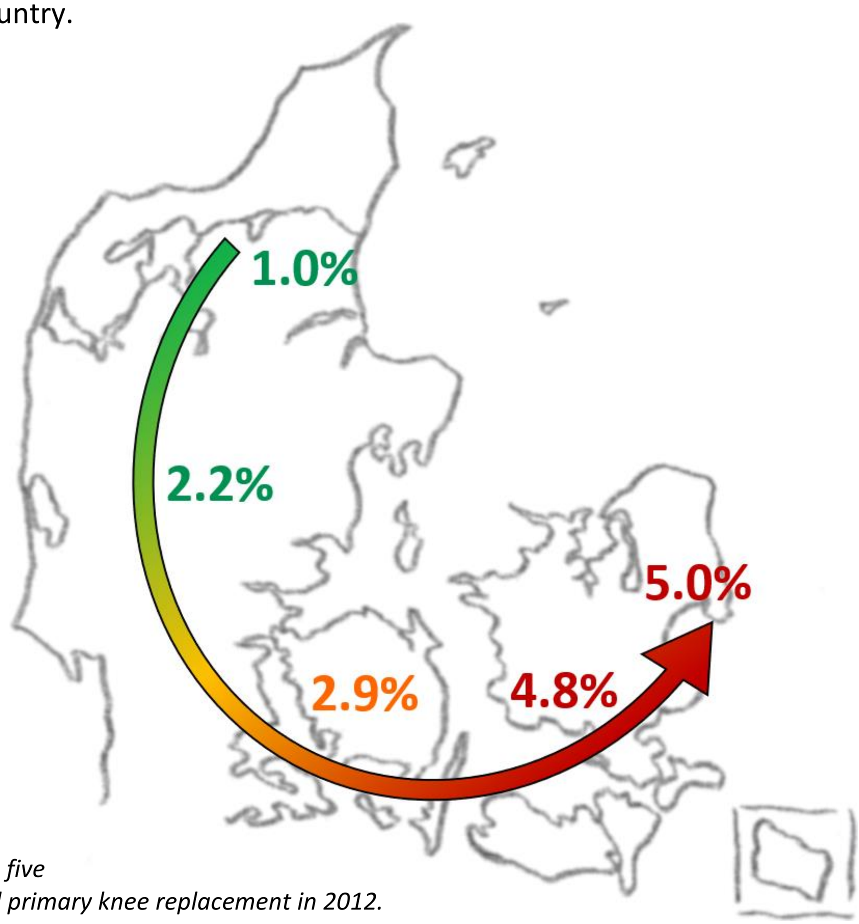


Figure 3. *2-year revision rates in the five Danish regions for patients who had primary knee replacement in 2012.*

Radiographic severity of knee OA

In the Danish National Clinical Guidelines on treatment of knee OA published in 2012, it was advised that KA should be considered when conservative treatment modalities (e.g. weight loss, analgesics and physiotherapy) had failed in patients with "more than minimal" radiographic OA [96]. Nonetheless, there were no registers or studies to confirm that patients were offered KA surgery on the same grounds across regions or hospitals. Inequalities would not be too surprising since the optimal timing of KA remains to be established [9, 17, 28, 73, 76, 94] and divergent opinions on pKA indications have previously been reported among individual Danish knee surgeons [102]. Several studies have shown that PROM results and patient satisfaction are higher in patients with severe radiographic OA prior to surgery compared to patients with low degrees of knee OA [18, 50, 73, 84] whereas other studies have found surgical results to be independent of OA grading

[59, 101]. Patients operated at a younger age are at higher risk of subsequent revision, though the causes are more complicated than merely a longer lifespan ahead. On the other hand, there can be solid arguments not to wait too long to offer KA to patients with decreasing knee function, as the function may not recover again after surgery [28, 53].

It was plausible that the presence of any regional differences in threshold for primary KA might contribute to the explanation of the revision rate variations. Thus, as part of the investigation of the clinical reality behind the regional revision rate variation, we found it necessary to explore whether the radiological threshold for pKA varied among hospitals [10].

Initiation of the SPARK project

In 2015, funds from The Health Research Fund of the Capital Region of Denmark made it possible to initiate a PhD project to explore the clinical reality behind the regional revision rate variations. The project name, "SPARK" was an abbreviation of "Variation in patient **S**atisfaction, **P**atient-reported outcome measures, radiographic signs of **A**rthritis, and **R**evision rates in **K**nee arthroplasty patients in three Danish regions".

Study III: UCLA

The pain and functional impairment caused by knee OA affects patients' ability to stay physically active. After pKA surgery, there is most often a potential to increase the level of physical activity [87, 92], though for some patients, the physical habits and interests have been gradually modified with the progression of the disease and are not easily changed even if the knee condition allows [72]. Physical activity has been recognized as important to KA patients [96], so to perform a proper comparison of patient-reported outcomes across regions, there was a need for a measure of physical activity in the SPARK study both before and after surgery.

However, measuring physical activity is a demanding activity in itself. Accelerometers are sometimes described as the gold standard for this purpose, yet, for both logistic and economic reasons, this was not an option in the SPARK study [39]. Secondly, keeping in mind the nature of knee OA symptoms, the concept of physical activity is not unambiguous; one person may consider one hour of walking more physically demanding than carrying heavy luggage to the third floor, whereas another may feel the opposite. Not only do these activities pose different burdens to the knee, but their difficulty also depends on other joint conditions, comorbidities (e.g. heart, lung and weight problems) and personal preferences. Having realized this challenge, we decided to aim for a measure to compare patients' physical activity level across regions, that allowed patients to decide which activities were most important to their evaluation.

A PubMed literature search was made to identify a relevant, patient-friendly, brief and valid PROM to assess patient-reported physical activity in knee replacement patients. The University of California Activity Scale (UCLA) (fig. 4) was chosen for several reasons [2, 111]. Firstly, in a systematic review by Terwee and colleagues in the COSMIN group (COnsensus-based Standards for the selection of health status Measurement Instruments) [120], UCLA was recommended as one of two most useful scores for use in populations of hip and knee replacement patients [99], mainly due to better construct validity and higher completion rates when compared to similar measures [69, 99]. Secondly, based on the reported scores in international publications, the difficulty of UCLA activity levels appeared suitable for knee (and hip) replacement patients both pre- and postoperatively [87, 92]. This was opposed to e.g. the Tegner score, in which activity examples were chosen with younger patients in mind (anterior cruciate ligament injuries), leading to a floor effect in hip and knee OA patients, i.e. too many patients were in the lowest end of the scale, meaning that a decrease in physical activity could not be detected [69].

The development process behind UCLA was not published, so the original intention of the score is unknown. When use of UCLA was first described [2, 111], surgeons sought to correlate physical activity to polyethylene wear, which was the most common reason for revision and thus a major concern in the first decades of joint replacement surgery [54]. Thus, activities that particularly stressed lower extremity joints (e.g. running and jumping) may have been the main objects of interest. With the later improvements in polyethylene durability [76] and an increasing focus on the health benefits of physical activity in all ages, the focus has shifted towards using UCLA to describe the positive effects of joint replacement on physical activity [99]. Physical activity can also be considered a lifestyle parameter when patient groups are compared, and improvements in this parameter can become part of health-economic consideration when different treatment options are prioritized by health authorities. Thus, today, the specific joint stress may play a minor role in both patients' and health staff's interpretation of the concept of physical activity.

	Activity Level
1	Wholly inactive: dependent on others; cannot leave residence
2	Mostly inactive: very restricted to minimum activities of daily living
3	Sometimes participates in mild activities such as walking, limited housework, and limited shopping
4	Regularly participates in mild activities
5	Sometimes participates in moderate activities such as swimming and can do unlimited housework or shopping
6	Regularly participates in moderate activities
7	Regularly participates in active events such as bicycling
8	Regularly participates in very active events such as bowling or golf
9	Sometimes participates in impact sports such as jogging, tennis, skiing, acrobatics, ballet, heavy labor, or backpacking
10	Regularly participates in impact sports

Figure 4. *UCLA Activity Scale, American version published by Amstutz in 1984. In the 1998 version, level 10 was presented first and a short patient instruction was added.*

With the simplicity and brevity of UCLA come some obvious drawbacks. Levels are neither mutually exclusive (the same patient can fit in to more than one level), nor are they exhaustive to all patients (some patients may not find a level that suits them). UCLA is an attempt to rate physical activity through a mix of several aspects of the construct "physical activity": 1) type of activity (various sports, hobbies and activities of daily living), 2) intensity (high or low impact), 3) duration time ("how long?") and 4) frequency ("how often?"). Naturally, this leaves room for personal interpretation just as a patients' perception of the levels may vary with age, sex, cultural environment and personal expectations. Recognizing these challenges, we did not expect a high degree of discrimination among patients, but the advantages of (expected) patient-friendliness and brevity made UCLA the best compromise for the current purpose.

Danish version of the UCLA Activity Scale

Several Danish hip and knee arthroplasty studies reported UCLA results, but no publication described a Danish UCLA version or the translation and cultural adaptation process behind. Through emails and phone calls with authors, it became clear that at least two Danish versions were in circulation, but translations were undocumented. The version used in the GLA:D (Good Life with Osteoarthritis in Denmark) Annual Report 2014 revealed a bimodal distribution of scores in hip and knee OA patients who were referred to physiotherapy (fig. 5) [89]. With the second Danish translation, no score distribution was available. Based on these findings, in order to be able to use UCLA results in the SPARK study, we found it necessary to produce a new Danish translation, to cross-culturally validate it and to test the measurement properties in a study sample outside the SPARK study.

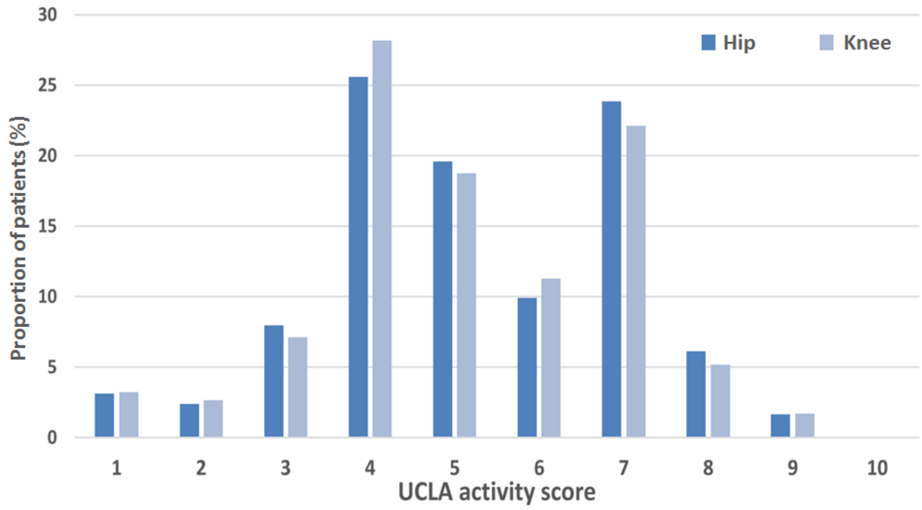


Figure 5. *Distribution of scores in hip and knee OA patients assessed with a previous Danish translation of UCLA Activity Scale (GLA:D Annual Report 2014)*

Study IV: ROM

Besides inducing pain, knee osteoarthritis affects knee function and reduces range of motion (ROM) [110]. Both flexion and extension are compromised as the destruction of cartilage is followed by osteophytes, inflammation, swelling, stiffness and increasing bone deformity; all factors that inhibit full motion. Knee ROM is one of the main objective outcomes noted at follow-up visits after knee replacement and it is important to patients [60]. Hence, it was highly desirable to include ROM measures in the SPARK study. However, as the study was based on questionnaires, having patients come in for professional goniometer measurements five times during the study course would add largely to study expenses. Also, the increased burden laid on patients would undoubtedly decrease the number of patients completing the study.

We searched the existing literature for patient-reported ROM tools as a substitution for goniometer measurements [107]. Two results were found, but both were not satisfactory for our purpose despite fair validation processes. The main problem was face validity in both cases, i.e. challenges with the initial impression of the scores; one drawing-based method by Borgbjerg et al. required patients to examine their ROM while sitting down on a flat surface such as the floor; a procedure which is difficult or impossible to many knee OA patients [7]. With a second method, a photo-based scale by Gioe et al. [32], we found it difficult to separate ROM angles from each other, as only half of the thigh and calf were visible in pictures. Both scores were accompanied by thorough instructions which, we found, would take too much of patients' time and effort, just as any smartphone goniometer application would have [44, 64]. Thus, inspired by these two scales and their individual advantages and shortcomings, we set out to create a new scale for patients to report their knee range of motion and, subsequently, to test its measurement properties among KA patients.

3. Aims

The overall aim of the SPARK study was to explore whether regional differences in knee arthroplasty (KA) revision rates were a sign of true variations in surgical results after primary KA or a sign of differences in patient selection prior to primary KA across Danish regions and hospitals. The existing literature did not suggest differences in these parameters, and therefore, all analyses were based on the null-hypothesis that there was no difference among hospitals.

Study I: SPARK Pre

With SPARK study baseline data, we explored whether Danish high-volume hospitals in three separate regions with very different revision rates differed in patient selection for primary KA surgery.

Study II: SPARK Post

The analysis of postoperative outcomes in the SPARK study served to determine if patient-reported results after primary KA differed among patients in the three hospitals.

Study III: UCLA

Translation and cultural adaptation of UCLA Activity Scale for use in Danish hip and knee arthroplasty patients was performed to provide a patient-reported measure of physical activity to aid comparisons among patient groups, e.g. in the SPARK study.

Study IV: ROM

This study aimed to produce a patient-friendly and reliable tool for patients to self-report their ability to flex and extend the knee, intended for use in registers and large trials (e.g. SPARK) where goniometer measurements by health professionals are not a feasible option.

3. Patients & Methods

Study I & II: SPARK Pre & Post

An observational cohort study was chosen as the proper design to draw a realistic picture of the clinical reality in current Danish KA surgery [24, 78]. Patients were included from 1 September 2016 and 15-17 months ahead in three Danish regions: the two with the lowest revision rates (North and Central Denmark) and the one with the highest (Capital Region) (fig. 2 & 3). Hospital revision rates express the proportion of patients who were revised (in any hospital) during a certain time period after pKA surgery in the hospital at question. In each of the three regions, the largest KA centre at the time (2015) agreed to join the study [115] (fig. 6). In the three hospitals, 2-year revision rates for patients who had a pKA in 2012 were 1.6, 1.2 and 6.1%, respectively, which were all in line with each region's revision rate (2.2, 1.0 and 5.0%, respectively) (fig. 3). Thus, we expected to include comparable numbers of patients from high- and low-revision-rate hospitals. The hospital differences in revision burden (fig. 6), i.e. revisions' proportion of all KA procedures, were not only caused by the revision rate differences in question but were also a consequence of patients with pKA from other hospitals being referred to these centres for revision surgery.

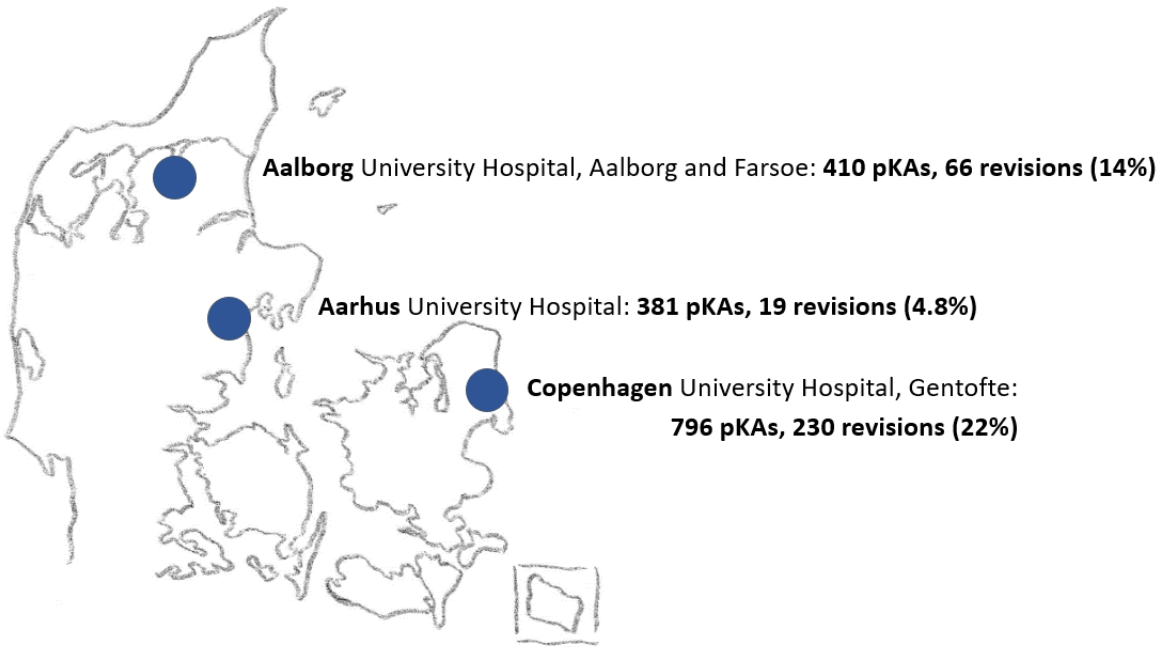


Figure 6. *Total number of primary knee arthroplasty operations (pKA), revisions and revision burden (in %) in the three hospitals who hosted the SPARK study (2014).*

Due to the observational design, no alterations were made in hospitals' daily practice, only, surgeons (medical students in Copenhagen) were to inform and include patients when planning pKA surgery. Patients were given oral and written information (Appendix 1) about the study and signed the informed consent form on site. Surgeons had been informed of the study both verbally and in writing, and the project leader (the author) was readily available in case of questions or doubts, either by phone, per email or in person (Copenhagen only).

Patients met the inclusion criteria if they were scheduled for pKA of any type: total KA (TKA), medial UKA (MUKA), lateral KA (LUKA) or patellofemoral knee arthroplasty (PFA) (partial surface replacements were not performed). Due to practical matters, only patients able to receive PROMs via email and answer electronically were allowed inclusion. This was chosen after weighing risk of bias against feasibility [124]. The elderly in Denmark generally have a high IT-literacy compared to the rest of Europe; in 2017, 2 out of 3 persons above 65 years used the internet every day [124]. This probably reduces (but does not eliminate) inclusion bias using an IT-based solution. Nonetheless, in the final 6 months of the inclusion period, also patients without an email address were invited to participate on paper questionnaires. This was done to obtain some information about this otherwise lost group of patients.

Questionnaires were sent as unique links to patients' email address at baseline (preoperatively), at 6 weeks and at 3, 6 and 12 months postoperatively (fig. 7) followed by reminders in case of no reply. At the 1-year follow-up, if necessary, reminders were followed by paper questionnaires as well. Later, 2-year questionnaires were sent out and 5-year follow-up is scheduled, but this thesis concerns results in the first postoperative year only.

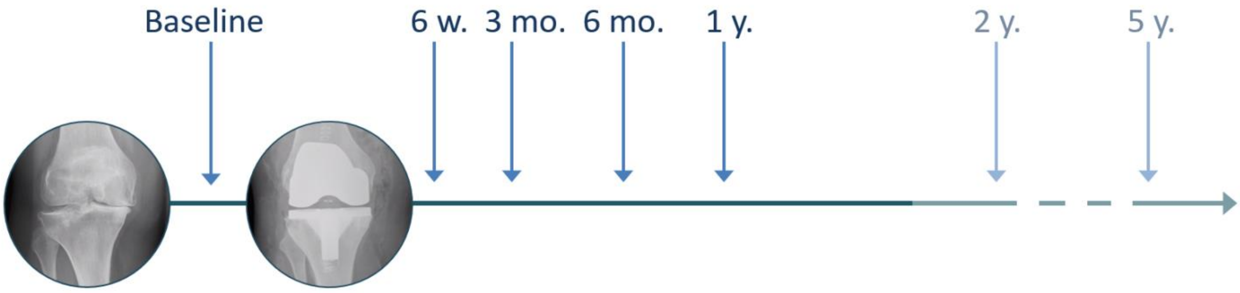


Figure 7. *Timeline of the SPARK-study PROM sets. 2-year results are not included in this thesis and 5-year data are yet to be collected.*

Patient-reported outcome measures

PROMs were chosen as the main outcome of the SPARK study to gain a subjective measure of knee pain and function directly from patients and avoid the bias that would inevitably be introduced by interference of the surgeon or any other staff in reporting of results [93, 104]. We selected PROMs based on several factors, of which proper development and validation among osteoarthritis and KA patients were the main concerns [22, 29, 30, 61, 83, 106]. Also, previous (Danish) use of questionnaires was considered, as was time consumption for the patients. We found it very important to keep questionnaires both short and highly relevant to patients as they were expected to complete several PROM sets in a study with no intervention or benefit to patients and without meeting the primary investigator. After reviewing the literature, time consumption alone did not become the deciding factor; questionnaires were chosen by relevance and measurement properties, and as it turned out, the appointed ones were all relatively brief. Only in the choice of physical activity scale was brevity a dominating factor in the choice of scale.

After reviewing the literature, the SPARK study was preceded by a pilot study to decide on the final PROM set. A proposed set of relevant questionnaires and additional single questions (e.g. regarding smoking, use of analgesics, willingness to repeat surgery, etc.) were all tested among 30 knee OA patients before and 3 months after pKA, and among 34 pKA patients contacted at 5 years postoperatively. The study was conducted in cooperation with occupational therapist Anette Enemark Larsen as part of the validation of the Danish version of the interview-based Canadian Occupational Performance Measure (COPM) [25]. Information from patient interviews helped in the mapping of patients' symptoms and functional challenges, and, based on patient feedback and PROM answers, single questions were re-worded and final PROMs were selected for the SPARK-study.

The resulting PROM set included both knee-specific and general health-related (generic) PROMs (table 1). The main outcome was Oxford Knee Score (OKS), a knee-specific questionnaire consisting of 12 questions with 5 answer options in Likert-boxes. The score was developed for patients awaiting knee replacement by use of patient interviews in 1998 [3, 15, 16, 22, 38, 65, 108]. OKS has been translated into various languages and has proven valid, responsive, reliable and acceptable to patients in multiple studies including Rasch validation [12, 38]. The Danish version was translated by recommended methods and evaluated among 22 patients with knee dysfunction, knee OA or KA in 2009 but has not been published [70]. OKS can also be split in two separate scores (pain and function) [37], but in this study, only the overall score is used.

The amount of ceiling effect with Oxford Knee Score was debated, so to be sure to detect differences among the patients with the best outcomes, we included The Forgotten Joint Score (FJS)[4, 100] which was developed with participation of patients to meet the rising expectations among modern KA patients.

We would have appreciated goniometer measurements of passive range of motion (ROM) at every time point in all patients. As this was not a feasible option, we used the Copenhagen Knee ROM Scale (CKRS) to provide us with a substitute of this information. Patients reported their ability to flex and extend the knee on the scale ranging from 0-6 and 0-5, respectively, with high numbers representing normal ROM (or hyperextension). As the extension part of CKRS was changed after the SPARK study had started, extension measures at baseline were only recorded in the last 699 patients.

The generic health questionnaire, EQ-5D-5L covers five dimensions of self-reported general health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression as well as the EQ visual analogue scale (EQ VAS) where patients rate their own general health on a scale from 0-100 (100 best). This PROM was chosen among other validated generic PROMs (e.g. SF-36, SF-12 or WHO) as it offers cost-benefit comparison with other health care services through calculation of QALYs (quality-adjusted life years) and it has no floor or ceiling effects in TKA patients [45, 121]. Moreover, it's appealing systematic layout and brevity were particularly beneficial in this study. The 5L version was chosen over the 3L version due to higher sensitivity [40, 45].

Additional stand-alone questions were made as simple as possible, yet still with exclusive and exhaustive answer options. For example, data concerning type of analgesics would have been valuable, but mentioning the names or groups of possible drugs made the question and answer options too complicated. Therefore, only the frequency of analgesics use was recorded with five answer options ranging from "more than once daily" to "rarely or never". At the six months' follow-up, patients were asked, "Did you attend physical therapy for rehabilitation after hospital discharge?" (yes/no). Further specifications of duration, number of visits and source of payment were recorded but not referred in this thesis. In every set of questionnaires, the most open questions were asked first to prevent patients from being primed by the more specific questions that followed [47]. For example, every PROM set was started with a self-made "global knee anchor" question, "How is your knee at the moment?", answered on a visual analogue scale (VAS) ranging from "My knee is not functioning at all or it is very painful" in one end to "My knee is painless and normally functioning" in the other, providing scores from 0-100 (100 best).

Table 1. Timing of PROMs and main questions in the SPARK study

	Baseline	6 w.	3 mo.	6 mo.	1 y.
PROMs					
✓ Oxford Knee Score	+	+	+	+	+
✓ Copenhagen Knee ROM Scale	+	+	+	+	+
✓ EQ-5D (5L and VAS)	+	+	+	+	+
UCLA Activity Scale	+		+	+	+
Forgotten Joint Score			+	+	+
Single questions					
✓ Global knee anchor *"How is your knee at the moment?", answered on VAS 0-100*	+	+	+	+	+
✓ Use of painkillers *"How often do you take painkillers because of your knee?"* *5 answer options from "more than once daily" to "rarely or never"*	+	+	+	+	+
Motivations for surgery *"What made you choose to undergo surgery?* *Pick up to 5 of 13 answer options (e.g. pain, mobility, work, hobbies)*	+				
Overall patient satisfaction *"How satisfied are you with the overall experience of the operation and its result?" 5 answer options (one neutral)*		+	+	+	+
Willingness to repeat surgery *"Suppose you could turn back time: now that you know the result, would you still choose to have a knee replacement?"* *5 answer options (one neutral)*					+

✓ *denotes questions or PROMs that were included in all PROM sets.*

Radiographic classification of knee OA

As an objective measure of severity of knee OA, we collected preoperative weight bearing semi-flexed (15-30°) standing postero-anterior (PA) radiographs. In Aarhus, radiographs were taken with full body weight put on one leg at a time, whereas Aalborg and Copenhagen practiced weight bearing on both legs at the same time. Patients who were to have a LUKA or PFA inserted and patients with predominantly lateral OA were excluded from this part of the study. Two separate radiologists, blinded to patient and hospital information, viewed all available pictures and classified the degree of knee OA by two established classifications; both the Alhbäck (ranging from 0-5, where 0 represents absence of OA) and the Kellgren-Lawrence (K-L) (0-4) classifications were assessed (table 2). In cases with disagreement, the radiologists met and reached consensus [SPARK Pre].

Table 2. Radiographic classifications of knee osteoarthritis

Classification	Kellgren-Lawrence (K-L)	Ahlbäck
0	No OA	(No OA)
1	Doubtful JSN, possible OP lipping	JSN < 3 mm
2	Possible JSN (AP weight-bearing), definite OP	Joint space obliteration
3	Definite JSN, multiple OP, sclerosis, possible bony deformity	Minor bone attrition (0-5 mm)
4	Marked JSN, large OP, severe sclerosis, definite bony deformity	Moderate bone attrition (5-10 mm)
5	-	Severe bone attrition (>10 mm)

OA = osteoarthritis, JSN = Joint space narrowing, OP = osteophyte(s) [71] .

Furthermore, a third classification was made using a new method where 13 individual experienced KA surgeons from all over Denmark were presented to multiple PA-radiographs in random pairs from the SPARK database. For each pair, the surgeon was to select the radiograph of the knee that was thought to be most symptomatic. Through use of the mathematical Bradley-Terry model, the process led to a complete ranking of all available radiographs, as documented in an article outside the present thesis [62].

Implant choice

From the Danish Knee Arthroplasty Register, it was well-known that use of unicompartmental implants was far more common in Aarhus than in Copenhagen and Aalborg. Implant type (TKA, MUKA, LUKA or PFA) was registered for each SPARK patient based on surgeon's postoperative registration -or patient chart data in case of mismatch between planned and registered procedure. Specific subtype or product name was not registered, as the bias that this could introduce was considered inseparable from information about which hospital the operation was performed in [SPARK Pre].

Incidence of primary knee arthroplasty

As a supplementary approach to the question of possible patient selection differences, incidence of pKA was retrieved from the National Patient Register for procedures performed in year 2017 for a) all inhabitants above 40 years of age and b) for the subgroup of patients aged 60-79 years. Reports were available on regional level only.

Study III: UCLA

The UCLA study work process included three main steps (fig. 8). In step 1, translation was conducted by use of a modified dual panel method [26, 56, 97] by a professional translator, a physiotherapist and an orthopaedic senior house officer. The new Danish version was then refined and culturally adapted in cooperation with 22 lay persons and 55 hip and knee OA patients.

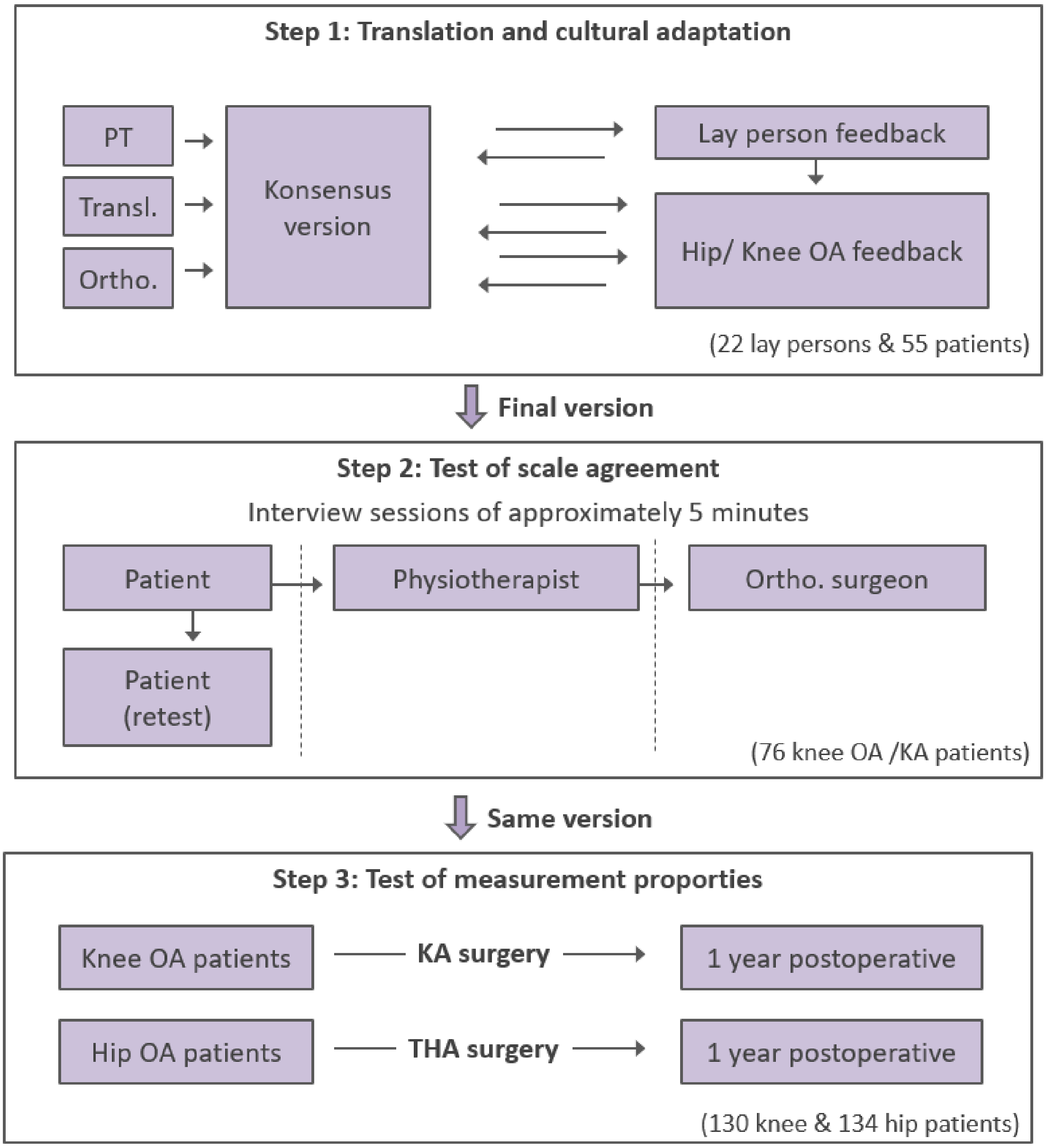


Figure 8. *UCLA translation and validation process steps. Abbreviations: PT = physiotherapist, Transl. = professional translator, Ortho. (surgeon) = orthopaedic surgeon, THA = Total hip arthroplasty. Dotted lines denote blinding between test sections.*

In Step 2, the degree of correlation with external assessment of patients' physical activity level was studied; a physiotherapist and either of two orthopaedics (one senior house officer, one specialist registrar) interviewed 76 knee OA and KA patients and, blinded to patients' UCLA answers, estimated their UCLA level. Also, test-retest reliability was studied for patients' own estimates 7-10 days after the initial assessment.

Step 3 was a test of UCLA measurement properties (construct validity and responsiveness) in a sample of 134 total hip arthroplasty (THA) and 130 KA patients, who completed UCLA, Oxford Hip/Knee Score (OHS/OKS), the generic EQ-5D questionnaire both pre- and 1-year postoperatively. Overall patient satisfaction was recorded at 1-year. For further methodological details, please consult "Study III: UCLA".

Study IV: ROM

The development and validation of a questionnaire for patients to estimate knee range of motion (ROM) was conducted in two main steps (fig. 9). To ensure validity and patient-friendliness, patients were involved in the entire process, which was thoroughly accounted for in "Study IV: ROM". The final version of the Copenhagen Knee ROM Scale (CKRS) was tested against blinded goniometer measurements of passive motion in 100 knee OA/ KA patients [57, 68]. For retest purposes, patients completed the questionnaire again 7-10 days after the initial assessment.

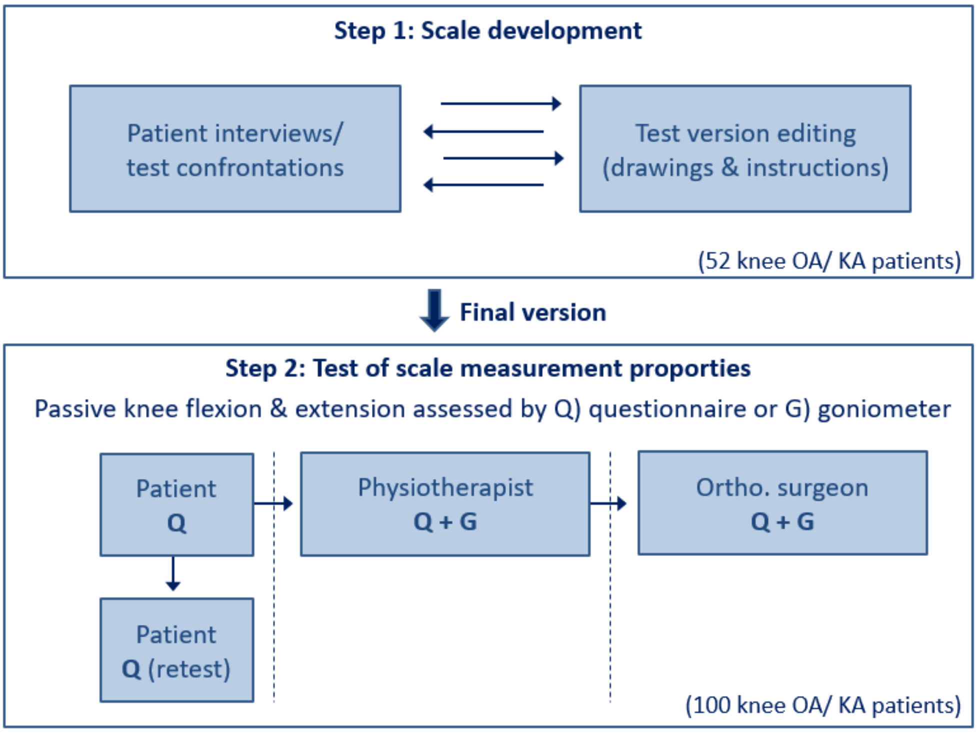


Figure 9. *Overview of the development and validation process of Copenhagen Knee ROM Scale (Study IV: ROM). Dotted lines denote blinding between test sections.*

Statistical analyses

In general, PROM scores and radiographic classifications were treated as non-parametric measures, but exceptions were made in cases where original developers have recommended use of parametric methods (e.g. OKS, FJS, EQ-VAS and EQ-5D-index) or when assumptions were met (e.g. global knee anchor). In some cases (e.g. UCLA Activity Scale), parametric methods were used in statistical computations but in order to present results as detailed as possible, means and standard deviations (SD) were provided in addition to medians and ranges. For sample size estimates and specific analyses, please consult each manuscript.

Level of significance was set to alpha 0.05 (two-sided) and 95% confidence intervals (CI) were provided when relevant. Means and standard deviations (SD) were reported as mean ± 1 SD. In comparison of hospitals, Aarhus was set as the reference hospital, as this was between the two others both geographically and with respects to revision rate levels and degree of urbanization. Only the overall p-value from hypothesis test comparing all three hospitals was reported. In "Study II: SPARK Post", OKS results were compared as both absolute and change scores, and we compared the proportions of patients reaching Minimal Important Change (MIC) of 8 OKS points (defined as the gain in OKS at 1 year that is considered an important change to the average patient) [3, 42, 98]. As 1-year results were missing from patients who had undergone revision surgery during the 1-year observation period, analyses were repeated with imputed scores, where all revision patients were hypothetically added to the group of patients who had not reached MIC. This was done to prevent high-revision-rate hospitals from being favored in comparisons. Use of imputation is clearly stated in the text. Moreover, to avoid wasting previous postoperative answers from 1-year non-responders, comparisons were repeated for "last available postoperative OKS", i.e. 1-year, 6- or 3-months, or, if needed 6-weeks scores, thereby including some of the revision patients as well.

Data collection and Case Report Forms were handled by Procordo ApS, Copenhagen. The author computed all statistical analyses and graphs, in SPARK and UCLA studies by use of R Statistical Programming (RStudio Inc., version 1.0.153) [79] and in the ROM study by use of SAS Statistical Software (SAS University Edition, version 3.6, Cary, NC) [122].

Ethics and funding

All studies were approved by The National Committee of Health Research Ethics and the Danish Data Protection Agency. The PhD study was financed by The Health Research Fund of the Capital Region of Denmark (2015 Grant). Office facilities were provided by the Department of Orthopaedic Surgery, Herlev and Gentofte Hospital, and travel grants to cover dissemination of postoperative SPARK results were donated by the Fund of the Kjærsgaard Family, Sunds.

5. Results

Study I: SPARK Pre

Inclusion

Preoperative SPARK questionnaires were sent to 1624 patients awaiting KA (1704 patients had originally accepted participation), and 1452 (89% of these) answered preoperatively (fig. 10). 53 patients participated with first one knee and subsequently the other knee, accounting for 7.3% of answers (106 knees) [82]. Radiographs were available from 1051 patients after subtraction of 224 patients with LUKA, PFA or predominantly lateral OA (177 pictures were unavailable due to logistic matters not related to patient factors). The 13 KA surgeons’ pairwise comparisons of radiographs summed up to a total of 17,767.

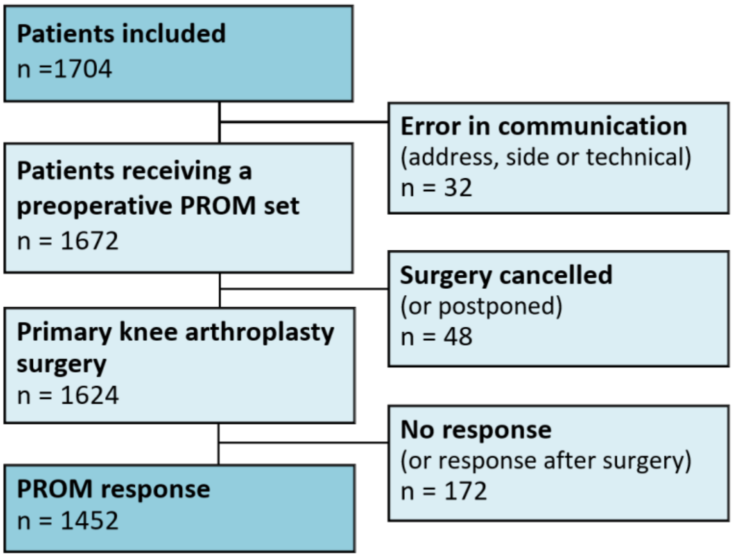


Figure 10. *SPARK (Pre) inclusion flowchart*

Patient representativeness

Based on data from the Danish Knee Arthroplasty Register (2017), which holds information of 97% of Danish pKA procedures, the preoperative PROM answers represented approximately 56% of all patients undergoing pKA in the three hospitals during the study period; 62% in Aarhus/Copenhagen and 37% in Aalborg (table 3) [SPARK Pre]. SPARK-participants were 1.1 year younger than non-participants (p=0.020) and more participants were male (42 vs. 38%, p=0.016). On a hospital level, however, the differences in sex distribution and age between participants and non-participants were only significant in Aarhus. Distribution of implant types were the same in participants and non-participants in all hospitals (p>0.230). The 41 patients that participated by letter were 8.1 years older than others (CI: 6.3 - 9.8, p<0.001) and insignificantly more were women (71 vs. 54%, p=0.052).

Table 3. Inclusion analysis based on complete surgical activity in 2017

		Complete primary KA population 2017	SPARK participation		
			Yes	No	p
Patients (n (%))	Total	1924 (100)	1083 (56.3)	841 (43.7)	-
	Aarhus	391 (100)	243 (62.1)	148 (37.9)	-
	Aalborg	429 (100)	161 (37.5)	268 (62.5)	-
	Copenhagen	1104 (100)	679 (61.5)	425 (38.5)	-
Age (mean ± SD)	Total	68.19 ± 9.8	67.7 ± 9.2	68.8 ± 10.5	**0.020**
	Aarhus	67.10 ± 10.6	66.1 ± 9.9	68.7 ± 11.5	**0.019**
	Aalborg	67.62 ± 9.8	66.7 ± 8.8	68.2 ± 10.3	0.141
	Copenhagen	68.80 ± 9.4	68.5 ± 9.0	69.2 ± 10.2	0.256
Male sex (n (%))	Total	779 (40.5)	459 (42.4)	320 (38.1)	**0.016**
	Aarhus	153 (39.1)	105 (43.2)	48 (32.4)	**0.043**
	Aalborg	202 (47.8)	83 (51.6)	119 (44.4)	0.070
	Copenhagen	424 (38.4)	271 (39.9)	153 (36.0)	0.125

P-values <0.05 *indicate a skewness in distribution of participants and non-participants.*

Hospital variation in patient selection

The SPARK baseline patient characteristics and PROM data revealed some differences in patient selection among the three hospitals (table 4). As many variables show considerable variation between males and females, and sex distribution varied among hospitals, the baseline data were stratified by sex as well (table 4). In Copenhagen, SPARK patients were 1.4 and 2.0 years older than those in Aalborg and Aarhus, respectively. In Aalborg, the proportion of male patients was higher (56 vs. 43-45%), as was BMI (1.5 - 1.7 kg/cm^2), also when adjusted for age and sex. By contrast, patients in the three hospitals were similar with respects to self-reported general health (EQ parameters), physical activity level (UCLA) and smoking and alcohol habits.

The primary outcome, OKS, was no different across hospitals before surgery (23.3 ± 7, p=0.884) (table 4, fig. 11), also when adjustments were made for age, sex and BMI. There was also no hospital difference in the other knee-specific variables, global knee anchor (0-100), frequency of use of analgesics against knee pain or proportion of patients who felt that the knee in question was their main physical disability. Patient-reported flexion and extension (table 4) were equal across hospitals, but when viewed dichotomously, more patients in Aalborg reported an extension deficit (p=0.007) (CKRS extension no. 0-3, providing a 78% sensitivity and 70% specificity of identifying extension deficits above 10°). Patients reported the same motivations for undergoing surgery in all hospitals (p>0.127) (table 5), but motivations differed with implant type and sex [Study I: SPARK Pre; table 3].

Table 4. Preoperative demographics and patient-reported outcomes

		Hospital				Sex		
Hospital/group	**Total sample**	**Aarhus**	**Aalborg**	**Copenhagen**	**p**	**Female**	**Male**	**p**
Revision rate level		Low	Low	High				
Patients (%)	1452 (100)	321 (22)	202 (14)	929 (64)		793 (55)	659 (45)	
Demographics								
Age	68.0 ± 9.3	66.6 ± 9.7	67.3 ± 9.1	68.6 ± 9.1	**0.002**	67.7 ± 9.7	68.3 ± 8.7	0.214
Male sex (%)	659 (45)	145 (45)	114 (56)	400 (43)	**0.002**	0 (0)	659 (100)	
Health & lifestyle								
Weight (kg)	86 ± 17	85 ± 17	90 ± 15	85 ± 17	**0.002**	81 ± 16	92 ± 15	**<0.001**
BMI (kg/m^2)	28.9 ± 5.0	28.5 ± 4.6	30.2 ± 5.1	28.7 ± 5.1	**<0.001**	29.2 ± 5.7	28.5 ± 4.1	**0.009**
BMI group (%)					**<0.001**			0.601
Normal (< 25)	329 (23)	77 (24)	26 (13)	226 (24)		196 (25)	133 (20)	
Overweight (25-29.9)	589 (41)	140 (44)	78 (39)	371 (40)		284 (36)	305 (46)	
Obese (≥ 30)	529 (37)	102 (32)	98 (49)	329 (36)		309 (39)	220 (3)	
Alcohol (> 2 units per day) (%)	164 (11)	36 (11)	15 (7)	113 (12)	0.154	43 (5)	121 (18)	**<0.001**
Daily smoking (%)	159 (11)	41 (13)	21 (10)	97 (11)	0.499	87 (11)	72 (11)	1.000
Urbanization (%)					**<0.001**			0.773
Countryside	78 (5)	18 (6)	33 (16)	27 (3)		43 (5)	35 (5)	
Small town or village	354 (24)	75 (23)	111 (55)	168 (18)		190 (24)	164 (25)	
City or suburb	1019 (70)	228 (71)	58 (29)	733 (79)		559 (71)	460 (70)	
Participation by letter (%)	41 (2.8)	5 (1.6)	10 (5.0)	26 (2.8)	0.074	29 (3.7)	12 (1.8)	0.052
EQ-VAS	61 ± 22	62 ± 21	58 ± 24	62 ± 22	0.091	59 ± 22	65 ± 21	**<0.001**
EQ-5D-5L Index	0.59 ± 0.15	0.59 ± 0.15	0.61 ± 0.12	0.59 ± 0.15	0.144	0.58 ± 0.15	0.60 ± 0.14	**0.028**
UCLA Activity Scale	4.7 [4] ± 1.9	4.8 [4] ± 1.9	4.8 [4] ± 1.9	4.7 [4] ± 1.8	0.551	4.5 [4] ± 1.7	5.1 [5] ± 2.0	**<0.001**
Knee-specific PROMs								
Oxford Knee Score (OKS)	23.3 ± 6.7	23.5 ± 7.0	23.2 ± 6.5	23.3 ± 6.7	0.884	22.0 ± 6.4	24.8 ± 6.8	**<0.001**
Median OKS	24	24	24	23		22	25	
Global knee anchor	28 ± 18	27 ± 17	30 ± 18	29 ± 18	0.193	28 ± 18	29 ± 18	0.132
Copenhagen Knee ROM Scale[1]								
Flexion	4.9 [5] ± 1.2	4.8 [5] ± 1.2	4.8 [5] ± 1.1	4.9 [5] ± 1.2	0.236	4.9 [5] ± 1.2	4.9 [5] ± 1.2	0.645
Deficit (CKRS 0-4) (%)	416 (29)	97 (30)	58(29)	261 (28)	0.774	223 (28)	193 (29)	0.678
Extension[2]	3.5 [4] ± 1.0	3.4 [4] ± 1.0	3.4 [3] ± 0.9	3.5 [4] ± 0.9	0.188	3.5 [4] ± 1.0	3.5 [4] ± 0.9	0.401
Deficit (CKRS 0-3) (%)[2]	340 (49)	63 (45)	72 (62)	205 (46)	**0.007**	200 (49)	140 (49)	1.000
Knee is main disability (%)	1261 (87)	289 (90)	176 (87)	796 (86)	0.146	680 (86)	581 (88)	0.327
Analgesics due to knee pain (%)					0.094			**<0.001**
More than once daily	667 (46)	145 (45)	83 (41)	439 (47)		407 (51)	260 (40)	
Once daily	187 (13)	34 (11)	32 (16)	121 (13)		95 (12)	92 (14)	
More than once weekly	218 (15)	42 (13)	27 (13)	149 (16)		129 (16)	89 (14)	
More than once monthly	142 (10)	39 (12)	19 (9)	84 (9)		70 (9)	72 (11)	
Rarely or never	237 (16)	61 (19)	41 (20)	135 (15)		91 (11)	146 (22)	

Abbreviations: BMI = Body Mass Index (BMI group "underweight" (<18.5 kg/m^2) comprised only two patients, who were thus included in the "normal" group). UCLA: UCLA Activity Scale (1-10, 10 most active). Global knee anchor: Patients' overall knee assessment, recorded on VAS (0-100, 100 best). [1]) Copenhagen Knee ROM Scale: Flexion 0-6 (6 is max), Extension 0-5 (5 is max). [2]) n = 699.

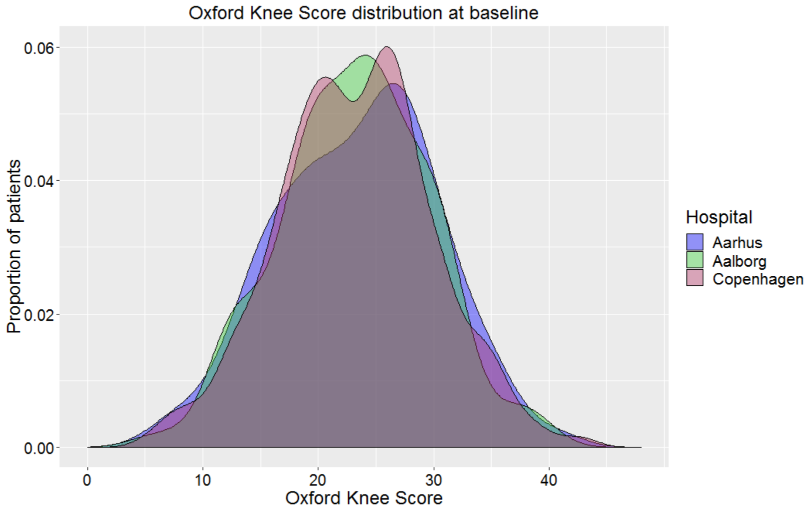


Figure 11. *Oxford Knee Score distribution at baseline in 1452 patients in the three hospitals (Kernel density plot).*

Table 5. Patients' motivations for surgery		
SPARK sample (n=1452)	**n**	**%**
Pain	1174	82
Mobility (walking, stairclimbing, bicycling)	784	55
Sports, exercise & physical activity	580	41
Knee motion and stability	521	37
The surgeons' advice	516	36
Hobbies (leisure time, travelling)	474	33
Mood and energy	471	33
Tired of taking medication	404	28
Duties (housework, gardening, helping others)	440	31
Independency and selfcare	390	27
Work	242	17
Being with family and friends	176	12
Marital (incl. sexual) life	66	4.6
Missing answer	16	1.1

Answers to the question, "Which factors or problems made you choose surgery? Pick up to five motivations". Options are listed by frequency.

As expected, implant choice varied among hospitals. In Copenhagen and Aalborg, 22 and 14% of patients, respectively, had a unicompartmental implant (MUKA, LUKA or PFA) inserted compared to 49% in Aarhus (p<0.001) (fig. 12, table 6). Hospital difference was noted in radiographic knee OA severity as well; Aarhus generally had fewer patients with mild OA, e.g. 2% with K-L grade 0-1 as opposed to 7% in Copenhagen and 9% in Aalborg (p=0.013) (fig. 13a, table 6). Corresponding figures for Ahlbäck classification 0-1 were 25, 34 and 38% (p=0.015) (fig. 13b). Moreover, with surgeons' ranking of knee OA severity, Aarhus patients were overrepresented in the advanced end of the spectrum (fig. 13c) (p<0.001). Urbanization was not associated to radiographic classifications, but with all three methods, males had more advanced radiographic OA than their female counterparts. Between MUKA and TKA patients, most of the examined variables differed significantly; however, in Aarhus, the two groups were similar with respects to age, sex and BMI [Study I: SPARK Pre; table 5].

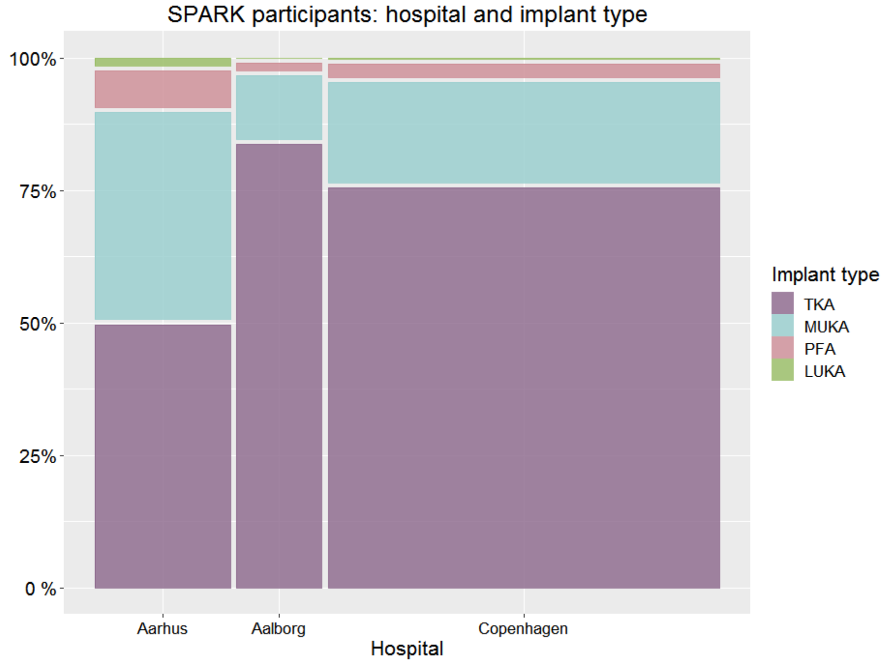


Figure 12. *Hospital and implant type in the 1452 SPARK participants.*

Table 6. Implant choice and radiographic severity of knee OA

		Hospital				Sex		
Hospital/group	**Total sample**	**Aarhus**	**Aalborg**	**Copenhagen**	**p**	**Female**	**Male**	**p**
Revision rate level		Low	Low	High				
Patients (%)	1452 (100)	321 (22)	202 (14)	929 (64)		793 (55)	659 (45)	
Implant type (%) (n=1452)					**<0.001**			**0.001**
TKA	1059 (73)	164 (51)	174 (86)	721 (78)		590 (74)	469 (71)	
MUKA	336 (23)	129 (40)	25 (12)	182 (20)		160 (20)	176 (27)	
PFA	50 (3.4)	23 (7.2)	3 (1.5)	24 (2.6)		38 (4.8)	12 (1.8)	
LUKA	7 (0.5)	5 (1.6)	0 (0.0)	2 (0.2)		5 (0.6)	2 (0.3)	
Radiographic severity of knee OA[3] (n=1051)								
K-L classification (%)					**0.016**			**0.011**
0	7 (0.7)	0 (0.0)	2 (1.2)	5 (0.7)		5 (0.9)	2 (0.4)	
1	57 (5.4)	4 (1.9)	13 (7.6)	40 (5.9)		35 (6.5)	22 (4.3)	
2	136 (13)	24 (12)	23 (14)	89 (13)		76 (14)	60 (12)	
3	787 (75)	160 (78)	123 (72)	504 (75)		395 (73)	392 (76)	
4	64 (6.1)	18 (8.7)	10 (5.8)	36 (5.3)		27 (5.0)	37 (7.2)	
K-L classification ≥ 2 (%)	987 (94)	202 (98)	156 (91)	629 (93)	**0.013**	498 (93)	489 (95)	0.082
K-L classification ≥ 3 (%)	851 (81)	178 (86)	133 (78)	540 (80)	0.067	422 (78)	429 (84)	**0.039**
Ahlbäck score (%)					0.104			**0.010**
0	56 (5)	7 (3)	9 (5)	40 (6)		37 (7)	19 (4)	
1	289 (28)	44 (21)	56 (33)	189 (28)		158 (29)	131 (26)	
2	401 (38)	94 (46)	61 (36)	246 (37)		198 (37)	203 (40)	
3	291 (28)	57 (28)	43 (25)	191 (28)		140 (26)	151 (29)	
4	12 (1.1)	3 (1.5)	2 (1.2)	7 (1.0)		4 (0.7)	8 (1.6)	
5	2 (0.2)	1 (0.5)	0 (0.0)	1 (0.1)		1 (0.2)	1 (0.2)	
Ahlbäck score ≥ 2 (%)	704 (67)	154 (75)	106 (62)	444 (66)	**0.015**	342 (64)	362 (71)	**0.019**
Ahlbäck score ≥ 3 (%)	305 (29)	61 (30)	45 (26)	199 (30)	0.696	145 (27)	160 (31)	0.148
Surgeons' ranking								
Mean	540	380	598	561	**<0.001**	575	503	**0.002**
Range [25-75%]	[270-808]	[188-718]	[315-864]	[293-824]		[318-845]	[238-778]	

Abbreviations: K-L = Kellgren-Lawrence. Surgeons' ranking = 13 surgeons' ranking of radiographic knee OA severity, total range 1- 1051 (1 is most severe).

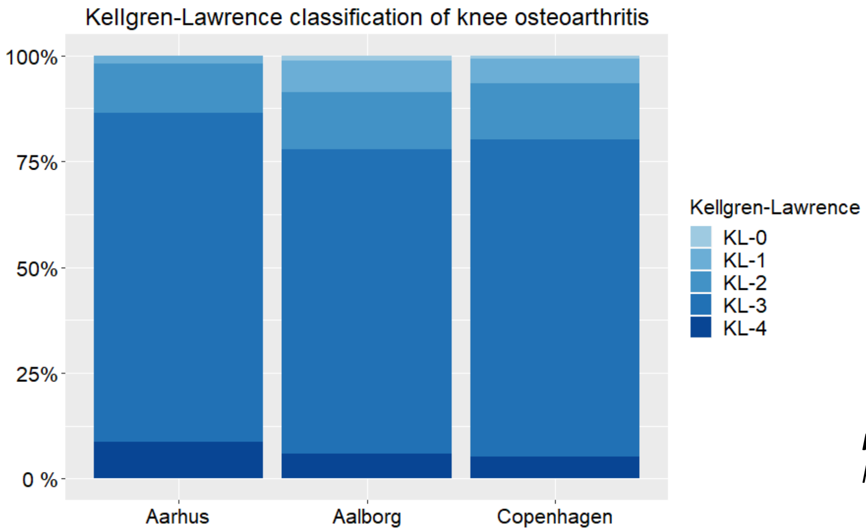


Figure 13a. *K-L classification of radiographic knee OA in the three hospitals.*

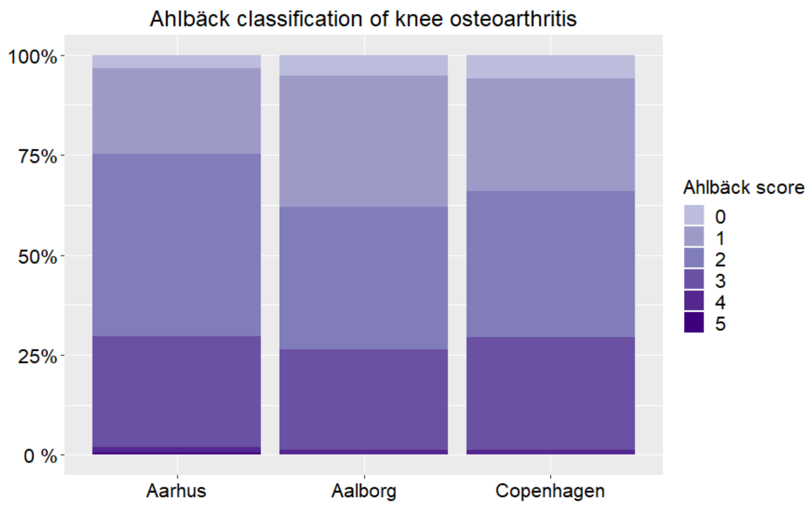


Figure 13b. *Ahlbäck classification of radiographic knee OA in the three hospitals.*

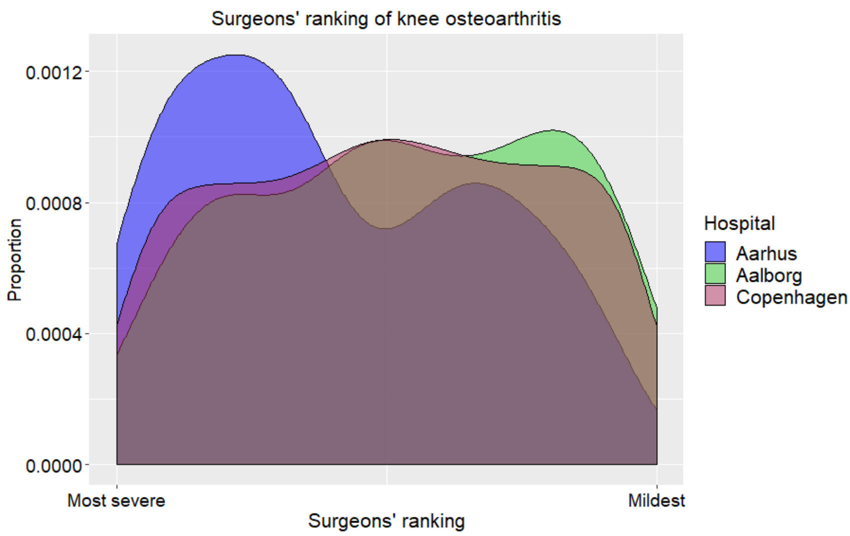


Figure 13c. *Surgeons' ranking of radiographic knee OA severity in the three hospitals (Kernel density plot).*

On a regional level, the incidence of pKA in 2017 was up to 28% higher in the Capital Region compared to the two other regions in this study (table 7) [SPARK Pre].

Table 7. Regional incidence of primary knee arthroplasty per region in year 2017

	Central Denmark	**North Denmark**	**Capital Region**	**p**
Regional revision rate	Low	Low	High	
SPARK example (hospital)	Aarhus	Aalborg	Copenhagen	
Incidence per 100.000 inhabitants				
All patients aged > 40 y.	235	276	285	**<0.001**
Subgroup: ages 60-79 y.	416	463	534	**<0.001**

Data retrieved from the National Patient Register (May 2019). Differences were stable in the preceding years.

Study II: SPARK Post

Follow-up completeness

During the first postoperative year, three patients had asked to leave the study, seven had died, and 28 patients who had been revised were excluded on the day of revision but were followed until then. Of the 1452 patients who participated preoperatively, 1307 (90%) answered the 1-year follow-up questionnaire. 1414 patients (97%) answered at least one postoperative questionnaire [Study II: SPARK Post; table 2].

1-year revisions in the SPARK sample

Revision surgery was performed in 28 patients (1.9%) during the first postoperative year with no significant difference in incidence among hospitals: 2 patients were revised in Aarhus (0.6%), 4 in Aalborg (2.0%) and 22 in Copenhagen (2.4%) (p=0.141) [Study II: SPARK Post; table 3]. Infections caused 13 revisions (1 in Aarhus (0.3%), 1 in Aalborg (0.5%) and 11 in Copenhagen (1.1%), p=0.169). Other causes than deep infection were responsible in 15 cases (no hospital difference, p=0.386). SPARK revision patients were no different from others with respects to age, sex or implant type, but they had a lower BMI (27 vs. 29 kg/m^2, p=0.008).

Hospital variation in PROM results at 1 year postoperatively

The primary outcome, OKS reached 39.0 ± 7.4 at 1-year follow-up with no difference among hospitals (p=0.096) (table 8, fig. 14a). Adjustment for confounders 1) age and sex, or 2) age, sex, baseline OKS, EQ-VAS, BMI and K-L (or Ahlbäck) classification did not change this conclusion. Change in OKS from pre- to 1-year postoperative was, however, lower in Aarhus (1.3-1.6 points compared to Copenhagen and Aalborg, overall p=0.038) (fig. 14b), still when adjusted for age and sex but not when adjusted for all the aforementioned confounders (i.e. item 2 above).

Table 8. Patient-reported outcomes at 1-year follow-up

		Total sample	Hospital			p
			Aarhus (Low rev. rate)	**Aalborg** (Low rev. rate)	**Copenhagen** (High rev. rate)	
Patients at baseline (n)		1452	321	202	929	
Responders at 1 year follow-up (n (%))		1307 (90)	288 (90)	187 (93)	832 (90)	0.424
Oxford Knee Score (OKS)						
	1 y. (n=1306)	39.0 [41] ± 7.4	38.1 [40] ± 8.3	39.1 [41] ± 7.2	39.2 [41] ± 7.2	0.092
	Last available postop. (n=1414)	38.3 [40] ± 8.0	37.5 [40] ± 8.7	38.7 [40] ± 7.5	38.5 [40] ± 7.8	0.119
	Change (n=1307)	15.4 ± 8.1	14.3 ± 8.7	15.9 ± 7.8	15.7 ± 8.0	**0.038**
OKS change < MIC (8 points) (total no. of patients in analysis (%))						
	1 y. (n = 1307)	195 (15)	56 (19)	25 (13)	114 (14)	0.051
	1 y. imputed* (n = 1335)	223 (17)	58 (20)	29 (15)	136 (16)	0.231
	Last available postop. (n = 1414)	237 (17)	66 (21)	31 (16)	140 (16)	0.074
Overall assessments						
	Willing to repeat surgery (%)					0.124
	"Yes, certainly"	1005 (77)	211 (73)	150 (80)	644 (77)	
	"Yes, probably"	200 (15)	46 (16)	26 (14)	128 (15)	
	"I don't know"	52 (4)	14 (4.9)	6 (3)	32 (3.9)	
	"No, probably not"	32 (2.5)	12 (4.2)	3 (1.6)	17 (2.0)	
	"No, absolutely not"	17 (1.3)	5 (1.7)	2 (1.1)	10 (1.2)	
	"Satisfied" or "very satisfied" (%)	1125 (86.2)	238 (82.6)	161 (86.6)	726 (87.4)	0.624[a]
Global knee anchor	1 y. (0-100)	80 ± 21	78 ± 24	81 ± 21	80 ± 19	0.082
	Change	51 ± 26	50 ± 29	51 ± 26	51 ± 25	0.769
Forgotten Joint Score, 1y		59.8 ± 27	59.1 ± 29	59.7 ± 25	60.1 ± 26	0.862
Patient-reported knee range of motion (CKRS units [b])						
Flexion	1y.	5.35 [6] ± 0.76	5.41 [6] ± 0.76	5.30 [5] ± 0.76	5.34 [5] ± 0.77	0.324
	Deficit (CKRS 0-4) (n (%))	165 (13)	32 (11)	21 (11)	112 (13)	0.483
	Change	0.48 [0] ± 1.2	0.57 [0] ± 1.2	0.55 [0] ± 1.2	0.43 [0] ± 1.1	0.160
Extension	1 y.	4.14 [4] ± 0.67	4.24 [4] ± 0.65	4.10 [4] ± 0.61	4.12 [4] ± 0.68	**0.016**
	Deficit (CKRS 0-3) (n (%))	161 (12)	29 (10)	24 (13)	108 (13)	0.420
	Change [c]	0.67 [1] ± 1.0	0.73 [1] ± 1.0	0.72 [1] ± 0.9	0.64 [1] ± 1.0	0.595
UCLA Activity Scale	1 y.	6.0 [6] ± 1.9	5.8 [6] ± 1.9	6.0 [6] ± 1.8	6.0 [6] ± 1.9	0.499
	Change	1.2 [1] ± 1.9	1.0 [1] ± 1.9	1.3 [1] ± 1.9	1.3 [1] ± 1.9	0.064
EQ-VAS	1 y.	79 ± 18	78 ± 20	82 ± 15	79 ± 18	0.079
	Change	17.4 ± 23	16.1 ± 24	24.3 ± 24	16.3 ± 22	**<0.001**
EQ-5D-5L Index	1 y.	0.81 ± 0.15	0.80 ± 0.17	0.83 ± 0.14	0.82 ± 0.14	**0.040**
	Change	0.22 ± 0.17	0.20 ± 0.18	0.23 ± 0.15	0.22 ± 0.17	**0.049**
Daily use of analgesics against knee pain (n (%))		166 (13)	41 (14)	22 (12)	103 (12)	0.364[a]
Supervised physiotherapy in rehabilitation (n (%)) [d]		702 (73)	115 (51)	92 (70)	495 (81)	**<0.001**

*When no unit is noted, means ± SD [and medians] are provided. *) "1y. imputed": Here, all 28 revised patients are assumed to be in the group with OKS change < MIC (8 points). [a]) Patient satisfaction is dichotomized for presentation, but P-value refers to tests of all five ordinal answer options. [b]) CKRS: With Copenhagen Knee ROM Scale. [c]) n = 699 due to delay of scale development. [d]) Only the last 966 patients were asked about physiotherapy.*

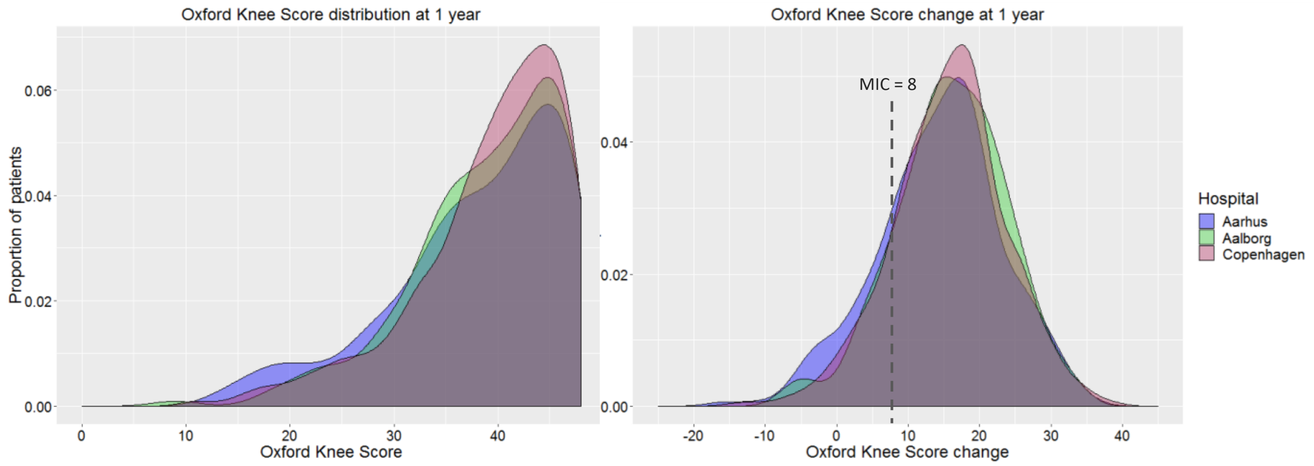


Figure 14a+b. *Oxford Knee Score (OKS) per hospital at 1 year: a) absolute score, b) change score (Kernel density plots). "MIC = 8" denotes the Minimal Important Change (MIC) in OKS considered to represent the minimum change perceived as important to the average patient.*

The proportion of patients who had not improved at least 8 OKS points (Minimal Important Change, MIC) after 1 year, was 19% in Aarhus, 13% in Aalborg and 14% in Copenhagen and (p=0.051) (fig. 14b). As revision patients were not represented in this calculation and were not evenly distributed among hospitals, the analysis was repeated with imputation by hypothetically adding all 28 revision patients to the group with OKS change scores below 8 points. In that scenario, proportions of patients not reaching MIC would be 20, 15 and 16%, respectively (p=0.231). Comparisons were also made using the change score of the last available OKS of the 1414 patients (97.4%) who had answered at any time postoperatively (including 17 revision patients); here, the according proportions were 21, 16 and 16% (p=0.074).

Patient satisfaction and willingness to repeat surgery at 1 year were both independent of hospital (p=0.624/0.124) (table 8). 86% of patients declared themselves either "very satisfied" (53%) or "satisfied" (33%) with the "overall experience and the result of surgery", while 8.8% were "neither satisfied or dissatisfied", 3.2% were "dissatisfied" and 1.7% were "very dissatisfied". When patients were asked, "Suppose you could turn back time: now that you know the result, would you still choose to have a knee replacement?", 92% of patients were positive; 77% replied "Yes, certainly" and 15% "Yes, probably".

In the remaining PROMs, some hospital differences were noted. In Aalborg, patients improved far more on EQ parameters (general health), e.g. 24 points on EQ-VAS compared to 16 points in the other hospitals (p<0.001), which was not altered with adjustments for age, sex and BMI. Physiotherapy during rehabilitation varied widely with Copenhagen presenting the largest utilization rates of 81%, followed by 70% in Aalborg and 51% in Aarhus (p<0.001). A minor difference was noted in patient-reported extension at 1-year favouring Aarhus (0.12-0.14 CKRS points higher, probably corresponding to 0.4-0.8° [Study IV: ROM]).

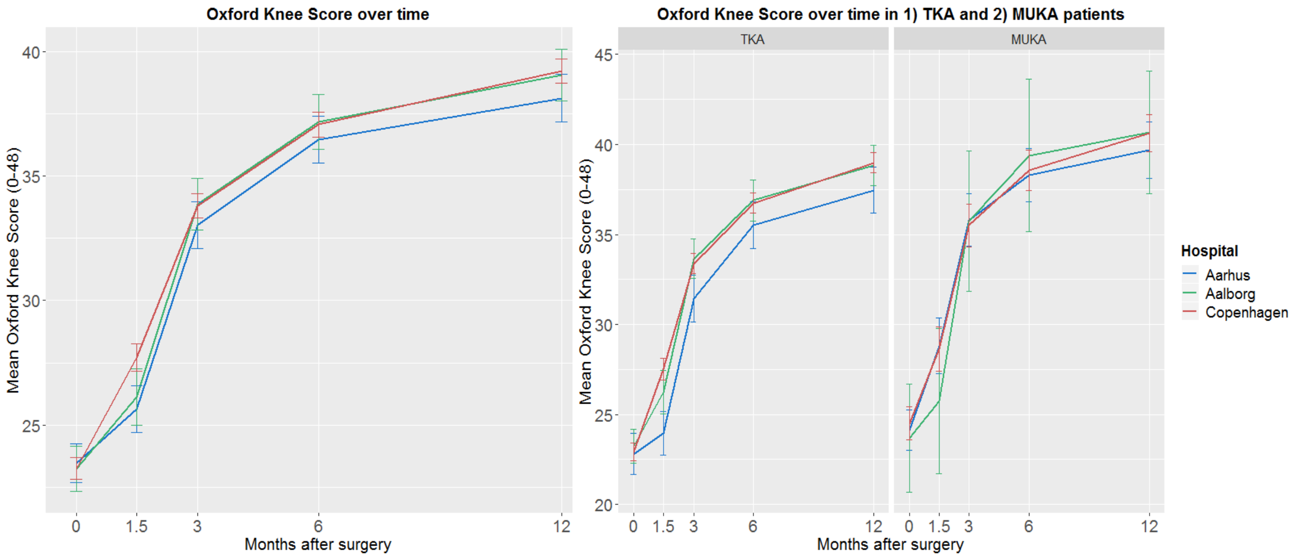


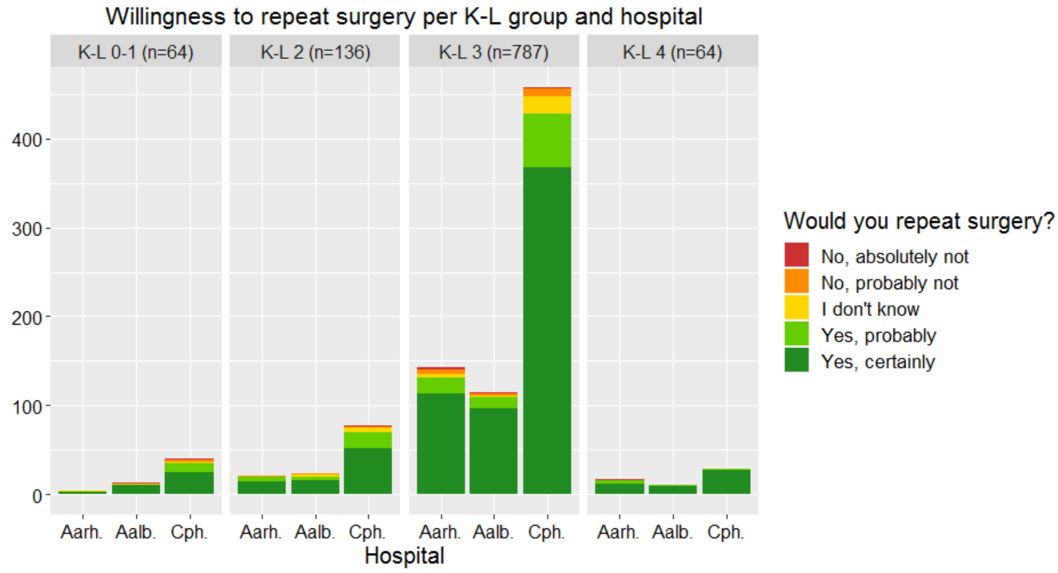
Figure 15a+b. *Oxford Knee Score in the first postoperative year in a) all patients, and b) TKA and MUKA patients alone. Whiskers denote mean ± 2 x standard error of the mean.*

PROM scores over time in the first postoperative year

At six weeks, OKS was 1.6-2.1 points higher in Copenhagen (p=0.001), but this difference was leveled out from 3 months and forward (fig. 15a). When OKS results over time were stratified by implant type (MUKA and TKA only), the hospital differences were further nuanced (fig. 15b). The observed OKS difference at 6 weeks was not accompanied by similar fluctuations in other PROMs or in patient-reported ROM.

Hospital variation in 1-year results within patient groups

With patients grouped by K-L (or Ahlbäck) classification, we compared willingness to repeat surgery, 1-year OKS and patients' last available postoperative OKS among hospitals (fig. 16a+b). No differences in these results were observed within patient groups (p>0.087), except in the 62 patients with K-L grade 4, where the 17 Aarhus patients had approximately 4-6 OKS points worse results (p≤0.045) (the 2 MUKA patients in the Aarhus group had OKS scores of 37 and 40).



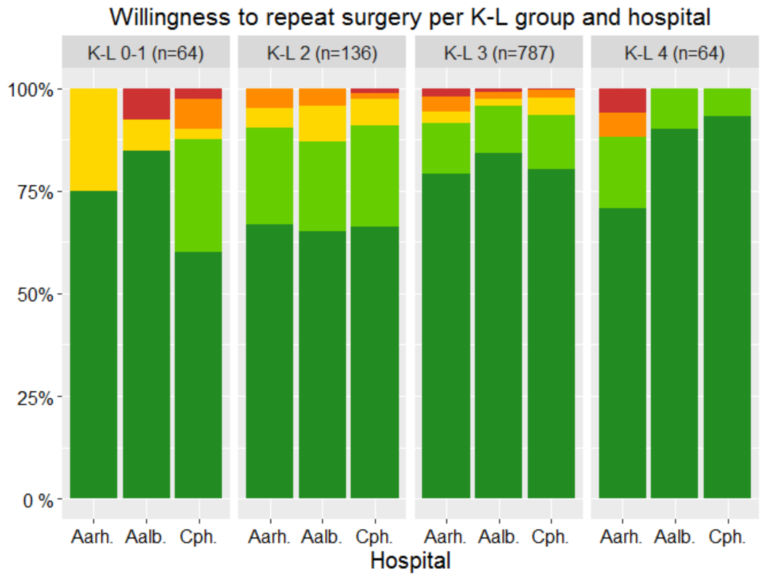


Figure 16a+b. *Willingness to repeat surgery 1 year postoperatively grouped by hospital and preoperative Kellgren-Lawrence classification of knee OA, displayed as a) counts and b) proportion of patients in each hospital.*

When patients were grouped by baseline OKS (0-20/21-30/31-48), there were no hospital differences in the same outcomes ($p \geq 0.224$) (fig. 17a+b). Nor did willingness to repeat surgery vary across the three hospitals in patients who had achieved the same OKS result (10-point intervals) at 1-year follow-up ($p \geq 0.157$) (fig. 18a+b).

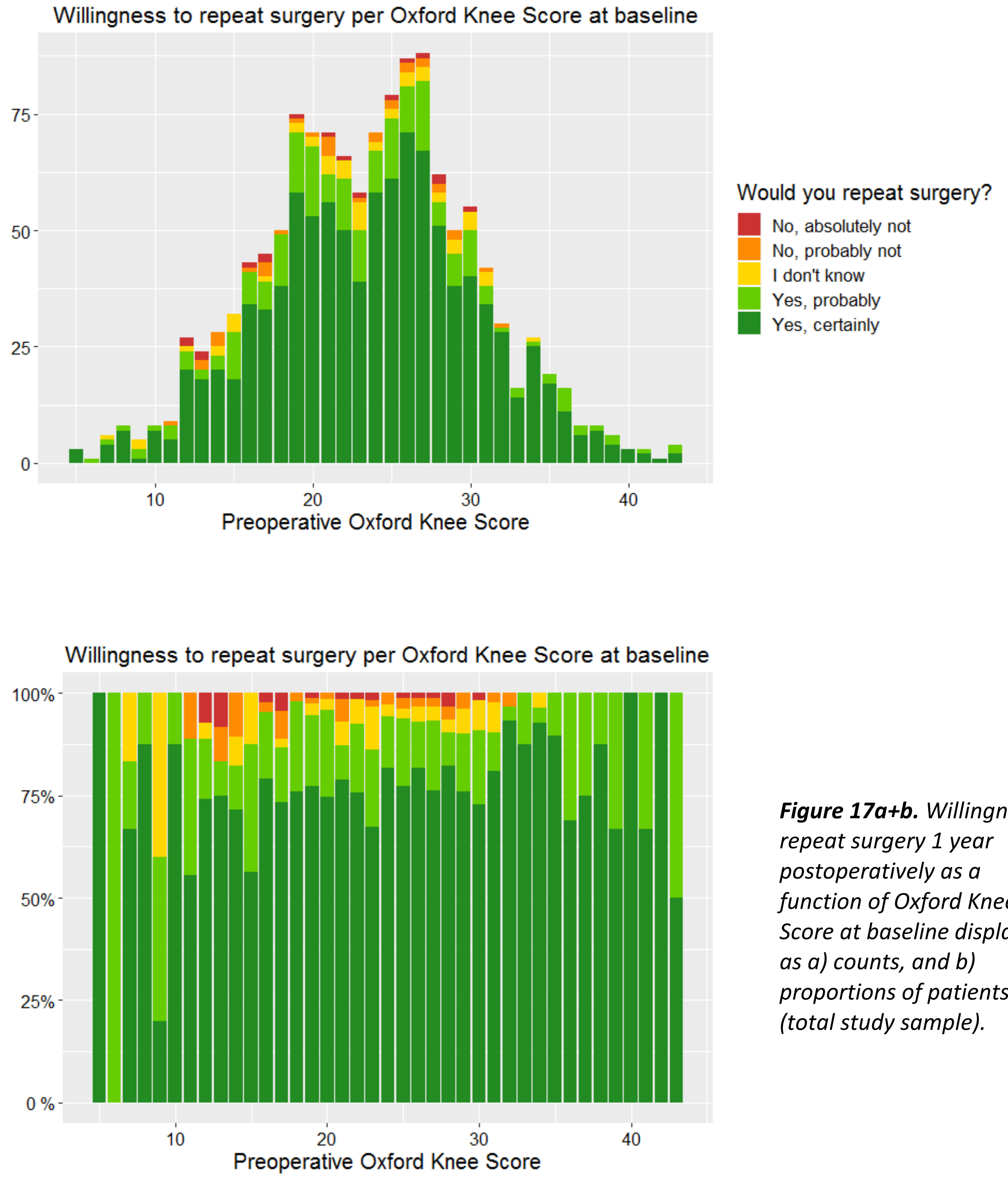


Figure 17a+b. *Willingness to repeat surgery 1 year postoperatively as a function of Oxford Knee Score at baseline displayed as a) counts, and b) proportions of patients (total study sample).*

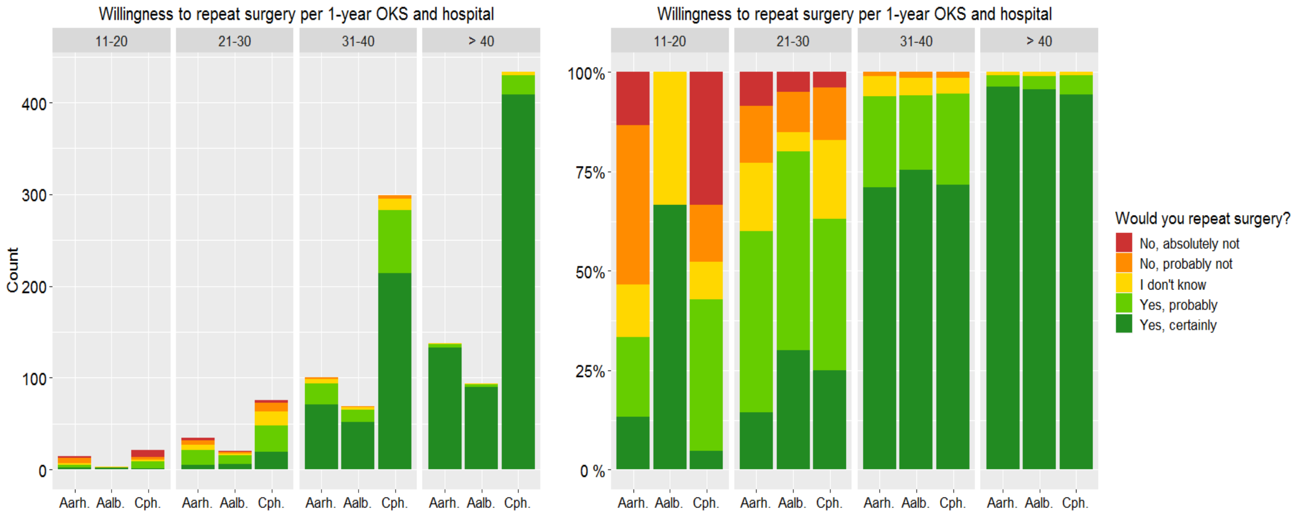


Figure 18a+b. *Willingness to repeat surgery at 1 year grouped by 1-year OKS and hospital, displayed as a) counts, and b) proportions of patients in each hospital (one patient with OKS < 10 is not shown).*

Study I and II: Additional findings

Self-evaluation

To address the hypothesis that patients in the western parts and more of rural parts of Denmark underestimated their symptoms compared to their eastern and more urban counterparts, we studied the relation between patients' overall assessment of their knee condition (global knee anchor, 0-100) and their (more specific) OKS score across all five PROM-sets from baseline to 1-year follow-up. "Self-evaluation", i.e. global knee anchor divided by OKS, did not differ among hospitals (p=0.092) (fig. 19a). When self-evaluation was compared among patient-reported degrees of urbanization, patients in the middle category, "Small town or village", had slightly higher self-evaluation than others (1.96 ± 0.8 compared to 1.90 ± 0.7 in the "Countryside" and 1.89 ± 0.7 in "City or suburb", p=0.047) (fig. 19b).

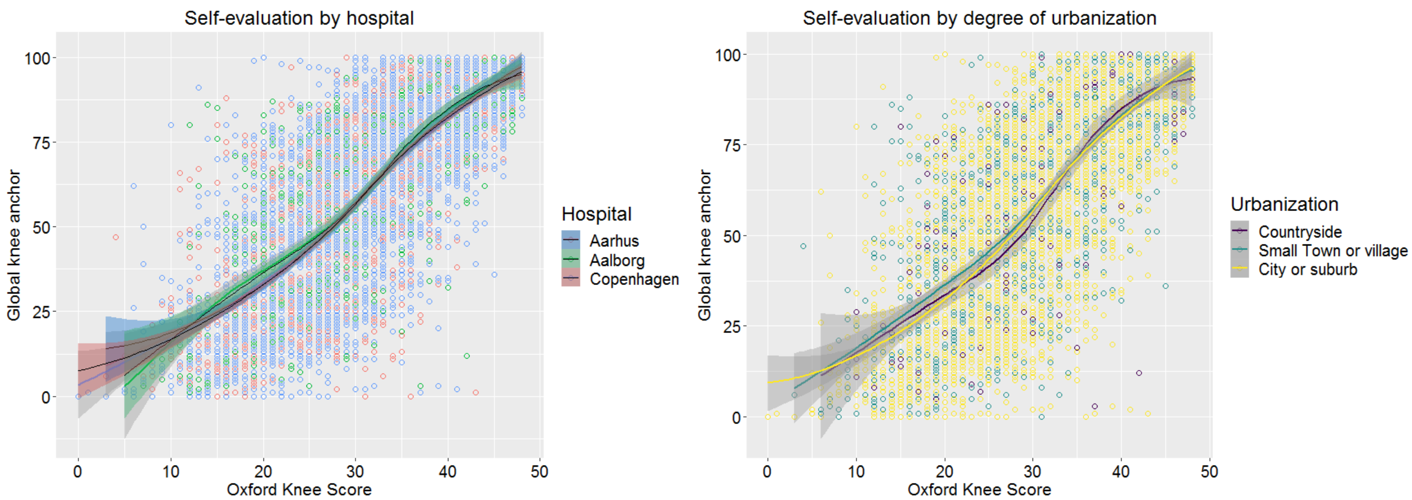


Figure 19a+b. *Relation between Oxford Knee Score and patients' perception of the knee condition ("global knee anchor") in the complete study period, stratified by a) hospital and b) degree of patient-reported urbanization.*

Study III: UCLA

Translation and cultural adaptation

A short patient instruction text was added to the original American 1984-version, and wording and layout were changed for the scale to appropriately reflect the activities of today's Danish KA patients (Appendix E) [Study III: UCLA]. Activity examples needed major changes and some activities (e.g. bicycling) were moved to other levels of the scale as a consequence of cultural differences. After the initial translation process, the new Danish version of UCLA needed 8 rounds of editing among lay persons and hip and knee patients in order to be sufficiently self-explanatory and acceptable to patients.

Correlation with external assessment of physical activity level

In the interview studies of 76 patients (mean age 66 y.), 11 patients (15%) misunderstood the questionnaire and marked more than one answer option. There was a systematic underestimation of 0.2-1.6 UCLA points by the physiotherapist and orthopaedic surgeons compared to patients' own estimates, despite a moderate to strong correlation between estimates (fig. 20) [Study III: UCLA; table 2].

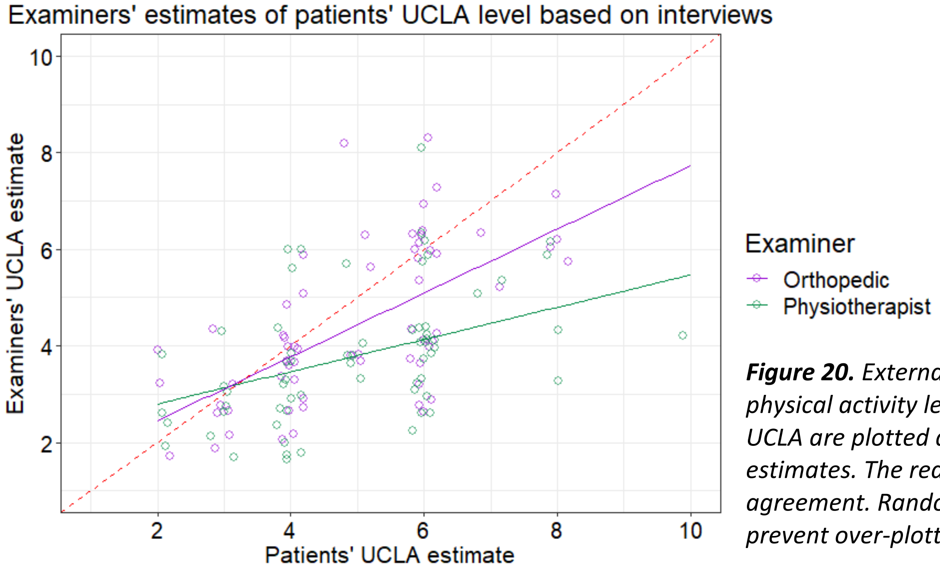


Figure 20. *External assessment of patients' physical activity level. Examiners' estimates of UCLA are plotted against patients' own estimates. The red dotted line indicates perfect agreement. Random variance (jitter) is added to prevent over-plotting.*

Test-retest reliability

Out of 38 valid test-retests, 21 patients reported not to have changed their subjective physical activity level at mean 8 days after the initial test round and were thus eligible for retest analysis. 13 of these 21 patients had marked the same level as in the first round.

Construct validity and responsiveness

Among 134 THA and 130 KA patients, who had completed both pre- and 1-year postoperative PROM sets, most patients (79% of hip and 66% of knee patients) had an increase in UCLA level during the first year after surgery (fig. 21). Only results from knee patients are graphically presented; please consult the "Study III: UCLA" manuscript for full overview of results. Mean improvement in UCLA level was 1.7 ± 2.3 (median 1) in knee patients and 2.3 ± 2.0 (2) in hip patients, with large Effect Sizes of 0.96 and 1.2, respectively (mean UCLA change divided by SD of baseline UCLA) [Study III: UCLA; table 2]. The proportion of knee patients reporting physical activity levels in the upper half of the scale (≥ level 6), went from 19% at baseline to 63% at 1-year (fig. 22) [Study III: UCLA; table 3].

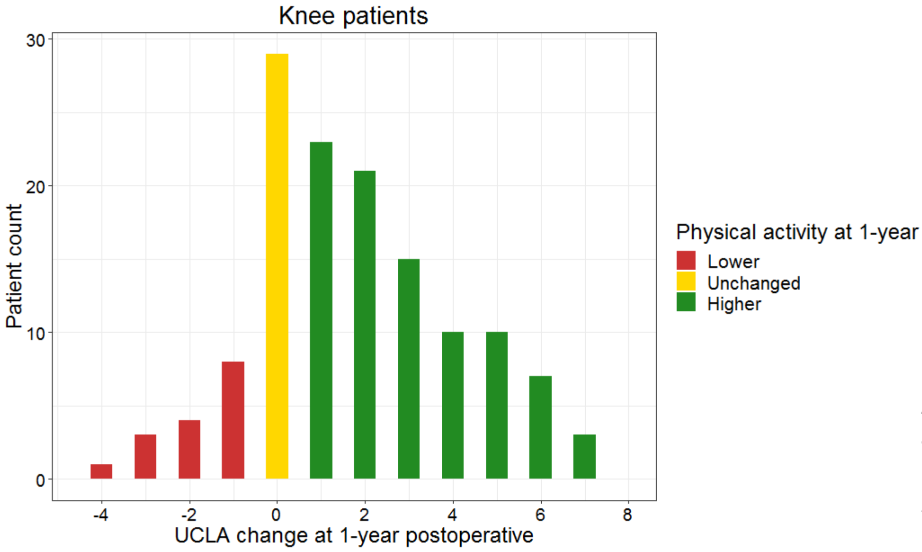


Figure 21. *Distribution of UCLA change scores from baseline to 1-year follow-up in knee replacement patients.*

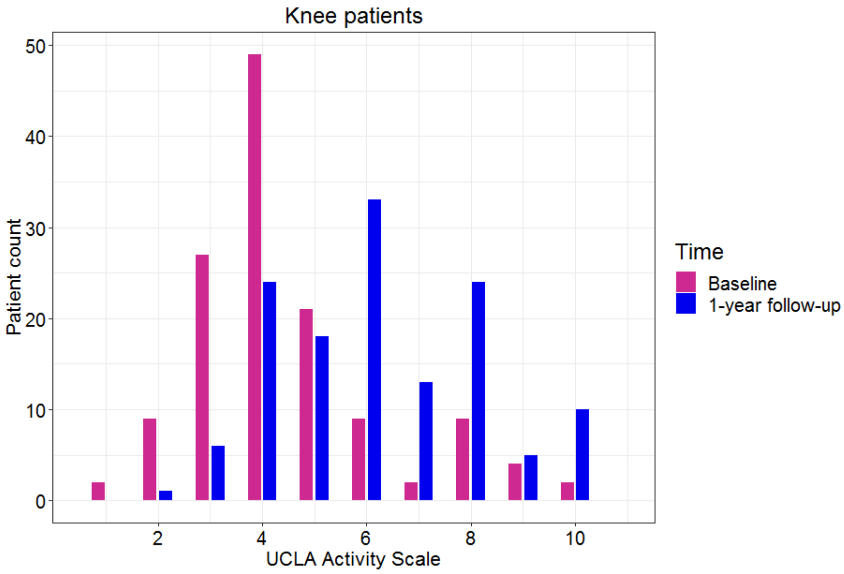


Figure 22. *Distribution of pre- and 1-year postoperative UCLA scores in knee replacement patients.*

We found no floor or ceiling effects in either patient group. There was a moderate correlation between UCLA change scores and change in Oxford Knee Score, and UCLA change scores proved a fair correlation with patient satisfaction and change in EQ-VAS. Knee patients who had reached the Minimal Important Change (MIC) of 8 OKS points at 1-year follow-up had a larger UCLA improvement (mean +2.1 points) than those who had not (mean -0.2 points, p<0.001). In hip patients, the difference between the two groups was not statistically significant (2.4 vs. 1.5 points, p=0.3).

Study IV: ROM

The development steps led to Copenhagen Knee ROM Scale (CKRS), a two-item scale with 6 illustrations (drawings) of knee flexion and 5 of knee extension (Appendix C: Danish original version & D: English translation). All illustrations of knee motion were made with 15° intervals between knee positions. The subsequent validation studies included 100 knee patients (41 with OA, 59 with KA) at mean age 71 years. In flexion, 94(%) of patients marked either the correct or the adjacent answer and no one were more than two pictures away from the right answer. In extension, 99(%) answered either correctly or one apart, and 1(%) was two away. The main results are summarized (table 9). For detailed results, please consult Study IV: ROM [63, 66].

Table 9. Summarized measurement properties of Copenhagen Knee ROM Scale

Mean differences ± SD (LoA)	**Flexion**	**Extension**
Patient vs. goniometer	-0.7 ± 12.3° (24°)	1.1 ± 11.6° (23°)
Correlation (Spearman's / Pearson's)	0.80 / 0.79	0.57 / 0.63
Professionals' CKRS estimate vs. goniometer mean	1.6 ± 6.7° (13°)	1.1 ± 4.6° (9°)
Physiotherapist vs. surgeon (both goniometer)	0.8 ± 4.2° (8°)	1.1 ± 3.0° (6°)
Thresholds (examples)	**Flexion 100°**	**Extension 10°**
Sensitivity / sensitivity (%)	95 / 81	78 / 70
Positive / negative predictive value (%)	58 / 98	36 / 93
Patient retest[1]	**Flexion**	**Extension**
Perfect agreement	45 (83%)	36 (68%)
Weighted Kappa (range -1 to 1)	0.84	0.66

LoA: Limits of agreement. For definitions of sensitivity etc., please consult [Study IV: ROM, table 3]. [1]) n = 54.

6. Discussion

Study I & II: SPARK Pre & Post

Key findings

The SPARK study found some hospital variation in patient selection for primary knee arthroplasty (pKA) surgery among the three hospitals. Some were in line with, but most were paradox to known revision risks on a hospital level (table 10) [43, 76]. It seems unlikely that variation in radiographic knee OA severity and regional incidence of pKA of such (modest) size could be responsible for significant amounts of the persistent hospital differences in revision rates. Nonetheless, high revision rates seem to follow high regional pKA incidence. Both might reflect an underlying geographical difference in expectations or in risk aversion in patients or surgeons (or both). This, we cannot say from the present data.

Table 10. Overview of main findings in the SPARK study				
Hospital	**Aarhus**	**Aalborg**	**Copenhagen**	**Revision risks**
Revision rate level (past decade)	Low	Low	High	
Patient selection (SPARK Pre)				
Age			1.7 y. older	opposite
Sex		More males		opposite
BMI		1.6 kg/m^2 higher		opposite
Radiographic classification of knee OA	Advanced			in line
Use of unicompartmental implants	High	Low	Intermediate	opposite
Oxford Knee Score (preoperative)	=	=	=	
General health (EQ-5D and -VAS)	=	=	=	
Patient-reported knee flexion/extension	=	=	=	
UCLA Activity Scale	=	=	=	
Regional incidence of pKA	Low	Intermediate	High	(in line)
Postoperative PROM results 1 y. (SPARK Post)				
Oxford Knee Score	=	=	=	
Patients with OKS change > MIC (8 points)	(Fewer)			(opposite)
Willingness to repeat surgery	=	=	=	
Overall patient satisfaction	=	=	=	
General health (EQ-5D and -VAS)	=	Better	=	
Forgotten Joint Score	=	=	=	
UCLA Activity Scale	=	=	=	
Patient-reported knee flexion	=	=	=	
Patient-reported knee extension	(Better)			(in line)

"=" indicate results that are similar in the three hospitals. Known revision risks were in line with or opposite to the current findings. Results in parenthesis were doubtful or only present before adjustments for confounders.

1-year postoperative results were, apart from minor exceptions, similar across the three hospitals that were known to perform very differently when surgical results were evaluated by revision rates alone. Patients with the same severity of preoperative knee OA were equally satisfied after surgery in all hospitals. From the present results, we are unable to evaluate the surgical quality in the 28 specific revision cases but is seems unlikely that an important general difference in surgical quality among the three hospitals would exist without leading to detectable PROM differences in a population of this magnitude. Thus, based on the SPARK results, there is no reason to believe that high revision rates reflect lower quality of pKA surgery in general [34] Naturally, it would not be sound to make inferences to populations with higher levels of revision rates (i.e. beyond approx. 5% per 2 years).

Thus, the SPARK results point to possible variations in revision thresholds rather than varying results of surgery as a plausible explanation of the persistent revision rate differences among hospitals and regions. It may be that a common lower threshold for surgery exists in the high-revision-rate region, which is reflected by both the higher pKA incidence and the higher revision rate. Future research should focus on revision thresholds, timing and patients' benefit from revision surgery depending of indication and preoperative findings to optimize treatment of patients with suboptimal results [6, 49]. For example, we included approximately 80 revision patients along with pKA patients in the SPARK study. We plan to analyze their PROM results and revision causes in the near future to find out if patients benefit equally from revision surgery in the three hospitals. Hypothetically, equal benefit of revisions across hospitals despite variations in revision rates would suggest that very low revision rates are not necessarily to the best interest of patients.

An exception may be cases with deep infection; in the SPARK sample, also revisions due to deep infection were (insignificantly) more common in Copenhagen. Infection cases, where the decision to revise is only sparsely influenced by patient and surgeon preferences, could be compared across hospitals and regions in a separate study.

Strengths and limitations

We chose an observational cohort study with very few exclusion criteria as the appropriate study design to compare routine pKA treatment in the three hospitals [24]. This implies the presence of confounders that could not meaningfully be separated from the hospital variable in analyses of results, e.g. implant selection, implant manufacturer, antibiotic prophylaxis, physiotherapeutic assistance in rehabilitation, etc. The observational nature of the study alone prevents us from drawing conclusions about causality between observed differences and surgical outcomes across hospitals and patients. Thus, the associations observed in this study merely serve to nuance the picture of (Danish) present day pKA surgery and generate hypotheses for later studies.

The aim was to include practically all pKA patients who had an email address. Yet, baseline results were obtained from only approximately 56% of patients (62% in Aarhus/Copenhagen and 37% in Aalborg) which introduces a risk of inclusion (selection) bias. Attempts were made to collect information about the patients, who were *not* included in the study, but these reports were inadequately filled out and, thus, the collection was discontinued. As a consequence our speculations of the reasons for this incomplete inclusion rely on informal feedback from the including surgeons and medical students alone: According to these reports, almost all invited patients gladly accepted participation if they were only invited, but some patients (allegedly 1 in 5 to 1 in 10) had to withdraw from participation because of lack of email address (disregarding the 41 patients who *were* allowed participation by letter late in the inclusion period). The decision to use an IT-based solution was based on both feasibility matters and a trust in the elderly Danish patients to be familiar with IT; it has previously been described that knee OA patients prefer electronic questionnaires over paper [35]. 64% of Danish citizens above age 65 years used the internet every day in 2017 (21% did not use it at all) [124]. Based on the observation that letter participants were 8 years older than email participants and more likely to be female (71 vs. 54%), the IT-based setup may have been responsible for some of the observed skewness in inclusion: SPARK patients were 1.1 year younger and males were overrepresented (42 vs. 38%).

Patient participation was essential to this study. Among the patients who joined the study, we had very high response rates. 89% answered the first questionnaire, and 97% of those gave at least one answer postoperatively. At the 1-year follow-up, 90% responded (92% of patients available). Much effort was put into PROM selection, patient information, layout and wording of the questionnaires, and response rates indicate that this was time well spent. Patients have contacted me for technical support and explanatory comments; the 245 unique email correspondences in my mailbox (and a 56 cm pile of paper questionnaires) emphasize that even electronic PROM measurements are not free of the need for human intervention for a research project to run. Many patients might have dropped out if no-one were there to help with technical or understanding issues, and some may have felt more obliged to answer knowing that "real people" were there to receive the answers.

Comorbidity is known to influence Oxford Knee Score results [21, 108]. In all three hospitals, comparable proportions of patients (87-90%) reported that the knee planned for surgery was their main mobility restriction. Other or more objective measures of comorbidity, e.g. ASA score (American Society of Anaesthesiologists) or Charlson Comorbidity Index might have nuanced the comparison of hospitals just as data concerning socioeconomic conditions might have [21, 36, 91].

Radiographic classifications

Radiographic classifications pointed to more advanced stages of knee OA in Aarhus despite the finding that patients were older in Copenhagen and more overweight in Aalborg. Differences were mostly prominent in the lower end of the scales, as fewer Aarhus patients presented with mild degrees of OA (K-L or Ahlbäck grade 0-1). The radiographic variance may have been exaggerated by technical differences in weight bearing during the recording of radiographs: Aarhus patients were instructed to carry weight on one leg only while Aalborg and Copenhagen patient carried their weight on both knees at the same time. The author is not aware of studies estimating the influence of complete vs. partial weight bearing on radiographic classifications in PA radiographs of semi-flexed knees. However, in a previous study of standing AP radiographs of 100 *extended* knees, 32 (%) were assigned a K-L classification one level higher when weight was carried on one leg instead of two [75]. In the Ahlbäck classification, joint space narrowing plays a larger role than in the K-L classification (table 2) [1, 48, 71]. Thus, we cannot rule out that methodological differences may have caused (some of) the classification differences observed in our study. However, the large hospital differences in surgeons' rankings, that were under the influence of many other morphological characteristics beside joint space narrowing (e.g. tibiofemoral axis, bone cysts, chondrocalcinosis), support that radiographic knee OA was indeed more advanced in Aarhus.

Incidence of pKA

The relative regional incidence of pKA in the Capital Region was 21-28% higher than in Central Denmark, and 3-15% higher than in the North Denmark Region. We have no knowledge of variations within each region, thus, numbers do not necessarily reflect the local hospital practice. When mentioning the Danish regional variation in pKA utilization rates (factor-1.3) it must be remembered that these numbers were lower than national variations reported from other countries, e.g. a factor-1.6 in Finland, factor-1.8 in Germany and factor-27 in Spain [33, 55, 85]. Determining whether variation in incidence is a problem to Danish OA patients would require comparable data on symptoms, radiology and subsequent outcomes in the patients who were treated by non-surgical means as well.

PROM results

The wide selection of PROMs used in our comparison of patients across hospitals provides high sensitivity to group differences and a security against missing a difference which is in fact present (type II error). This could happen due to ceiling effects in e.g. OKS; in a former study comparing patients with different TKA designs, there was no OKS difference between groups but significant differences in ROM measures and only measuring the ability to perform high impact activities revealed group differences [39]. In our study, Forgotten Joint Score and patient-reported knee motion (CKRS) results were the same across hospitals, with the exception, though, that Aarhus patients had better 1-year knee extension (assuming this did not represent patients with problematic hyperextension [77]). As the CKRS was under development in the beginning of the study, we only have baseline extension data from half of patients (699), which underlines the significance of the result. The difference appears to be small (estimated 0.4-0.8°) but we are not aware of which amount of patient-reported extension represents a clinically relevant difference on either individual or group level. Extension differences among hospitals disappeared when 1-year extension was adjusted for confounders including preoperative extension.

Remarkably, no patients with very high (or low) preoperative OKS scores were unwilling to repeat surgery (fig. 17a+b). This supports the findings of Judge et al., where patient satisfaction was not correlated to preoperative OKS, and it indicates that patients presenting with knee OA symptoms should not be denied pKA surgery based on high OKS scores alone [39, 46]. Further, this underlines that OKS was designed to evaluate KA treatment, not to diagnose knee OA [39]. The association between radiographic signs of arthritis and patient satisfaction were slightly clearer, but the high rate of patient satisfaction even in mild K-L groups (fig. 16a+b) underline, how radiographs alone cannot predict results. A combination of the three (symptoms, radiology and willingness to repeat) might reveal that patients were overall well selected for surgery. However, prediction of results was beyond the scope of this project.

PROM perspectives

Patients are increasingly involved in the decision making process and in the evaluation of treatment results in all medical fields [13, 41]. During recent years, the advantages of using PROMs to collect this valuable information have become visible to most health care providers, but so has the challenges of selecting, collecting and analyzing PROM data. Several Danish PROM initiatives have been started to [80, 103, 105, 123], and hopefully within a few years, we will be able to routinely collect relevant PROMs in the Danish Knee Arthroplasty Register to the benefit of patients, surgeons and researchers [8, 109, 113].

Study III: UCLA

Translating and culturally adapting UCLA Activity Scale (UCLA) for use in Danish hip and knee patients was challenging, mostly because the underlying construct and the intention behind the original measurement tool was unclear [2, 111]. The Danish version of UCLA has inherited this weakness, and the mix of activity constructs (type, frequency, intensity and duration) leave room for large variation in perception of levels among individuals. This was underlined by the finding of systematic errors when clinicians tried to estimate patients' UCLA answers based on interviews [Study III: UCLA]. The imperfect normal distribution of results with this version is more acceptable than the previous bimodal distribution provided by another Danish translation, and it might be further improved with small changes in wording. Publications of other international versions have not provided histograms or score distribution tables, thus, uneven distributions may well be a universal problem with this scale [31, 87, 92]. In our studies of construct validity evaluated by correlation with OKS, EQ-5D and patient satisfaction, the Danish version had measurement properties comparable to international versions (for details, please consult the discussion section of Study III: UCLA).

With today's knowledge of PROMs, ideally, a scale of this kind should have been developed with use of patient involvement to ensure content validity. This is a factor that cannot be compensated for by other measurement properties [104] and the lack of proper documentation of development would be a sound argument not to use the scale. More valid information of physical activity (restrictions) might be obtained by use of e.g. the OKS-APQ (Oxford Knee Score – Activity and Participation Questionnaire), an 8-item questionnaire which was developed in cooperation with patients and guided by Rasch analyses [14]. The main advantage of UCLA appears, instead, to be good responsiveness to change. It is plausible that results from relevant patient populations may be useful to future patients as part of preoperative patient information. However, as we had no anchor question concerning physical activity alone, MIC could not be provided. Also, it must be remembered, that we have not studied the validity of UCLA against other measures of activity, e.g. accelerometers, performance-based measures or other more comprehensive activity questionnaires, and therefore, we cannot rule out that the reported improvements after joint replacement are not only subjective, i.e. a matter of ability, rather than representing a true increase in performed physical activity [39, 88].

Study IV: ROM

Copenhagen Knee ROM Scale (CRKS) provided measures of knee flexion and extension with far better correlation to goniometer measurements than what had been reported by other self-reporting ROM instruments [32, 51]. In all fairness, some of the correlation could be mathematically contributed to the wider range of angle measurements which came from actively seeking out patients with poor ROM to be able to test the whole range of the scale. However, we succeeded in obtaining a patient-reported measure of passive ROM with no overall difference in means of either flexion or extension. Extension measures were, however, somewhat problematic as described in detail in the study [Study IV: ROM]. As expected, patient-reported ROM estimates on CKRS had larger measurement errors (i.e. wider limits of agreement) than both goniometer measurements and professionals' estimates on CKRS. However, in comparison with similar instruments, CKRS provided the same or slightly lower measurement errors and a comparable sensitivity and specificity in detecting poor ROM results despite its larger increments in knee angles between illustrations (15 vs. 5-10° intervals) [7, 11, 32, 51]. Overall, CKRS is the best documented scale to date and at present the most precise tool to record patient-reported knee motion, e.g. for registers and large trials [107].

In reporting of patients' CRKS results (e.g. in the SPARK study), we used the illustration numbers (0-6 and 0-5, respectively) to clearly indicate that flexion and extension emerged from patients' estimates; reporting degrees might lead readers to think that goniometer measurements had been made. Also, any future CKRS validations, e.g. in other patient populations or languages, would most likely provide slightly different relations to goniometer measurements, thus, with use of CKRS units, patients' results remain directly comparable across versions. For detailed discussion of results, please see Study IV: ROM.

7. Conclusion

The SPARK study has demonstrated that patient-reported results after knee arthroplasty surgery were the same in three high-volume hospitals known to perform very differently when results were measured by revision rates only. The study results outline that revision rates, however useful they are at detecting implant or surgeon outliers, do not provide the full picture of surgical quality. In this study, the uniformity in PROM results across hospitals indicate that persistent revision rate differences among hospitals were not a reflection of hospital variations in overall surgical quality but more likely a sign of varying thresholds for revision surgery. Thus, had surgical treatment strategies been changed on the basis of register data alone, it could have led to lower quality of treatment. To improve future results of knee arthroplasty surgery and help patients with inferior outcomes, future studies should focus on indication, timing and patient benefit of revision surgery.

Altogether, the works in this thesis demonstrate both the complexity and the necessity of using patient-reported outcomes in the evaluation of results after knee replacement surgery. The author advocates for a systematic collection of relevant PROMs from all knee arthroplasty (and osteoarthritis) patients as long as collection and presentation of results are perceived as relevant and useful to the individual patients.

8. Future perspectives

The SPARK data set is, to the best of our knowledge, the largest of its kind in Denmark, including radiographic classifications and several different PROMs prospectively collected from preoperatively to (so far) 1 year after knee arthroplasty surgery. Together with the newly collected 2-year data and the scheduled 5-year data, this cohort may serve as a valuable reference set for later comparisons. Much of the information could be combined to produce updated information for patients as well as for health professionals, e.g. general practitioners who are the first to advice patients about knee OA treatment. For example, patients are currently most often informed that 1 in 5 patients are dissatisfied after knee replacement surgery; based on SPARK data (willingness to repeat surgery and overall patient satisfaction) this proportion is closer to 1 in 10.

Based on OKS, patients with high degrees of pain and low preoperative knee function (OKS < 20 points), can expect improvements already at 6 weeks, whereas others on average find themselves improved at 3 months (fig. 23).

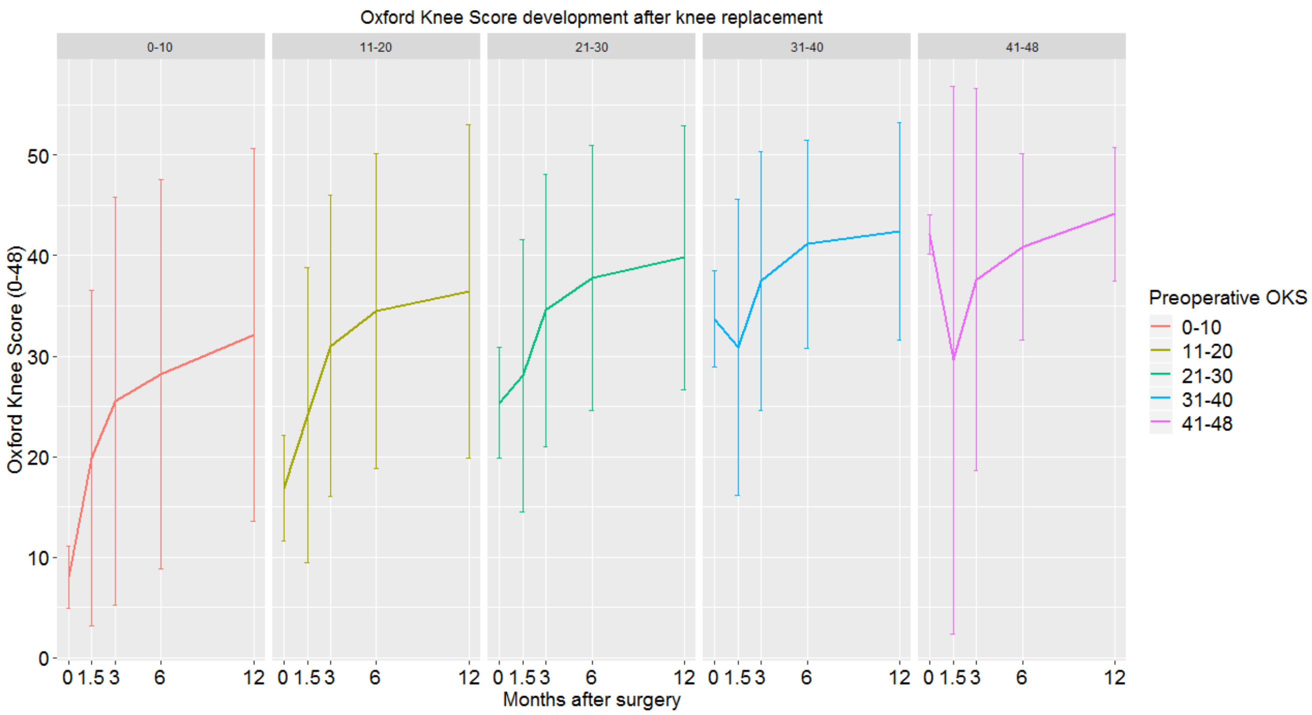


Figure 23. *OKS during the first year after surgery stratified by baseline OKS. Whiskers denote mean ± 2 SD.*

We plan to look more into range of motion (ROM) data. With Copenhagen Knee ROM Scale, we have an opportunity to map estimates of knee motion free of bias from the professional observer and from more patients than it is otherwise possible within the normal clinical setting. For example, the scale has been able to demonstrate pre- and postoperative differences among groups of patients with different implant types (fig. 24a+b). Follow-up data stratified by ROM at baseline should be interesting to both patients and might adjust preoperative expectations.

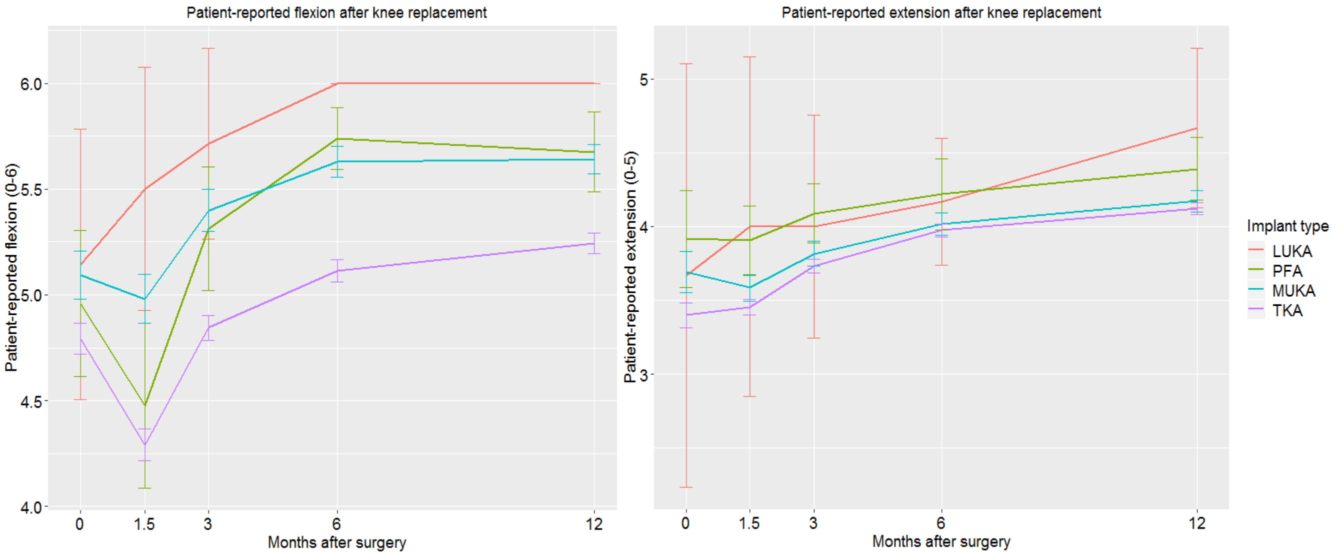


Figure 24a+b. *Patient-reported a) flexion and b) extension (Copenhagen Knee ROM Scale units) over time grouped by implant type. Whiskers denote mean ± 2 standard error of the mean.*

9. References

1. Ahlbäck S, Rydberg J. Röntgenologisk klassifikation och undersökningsteknik vid gonartros. *Lakartidningen*. 1980;77:2091–2093, 2096.

2. Amstutz HC, Thomas BJ, Jinnah R, Kim W, Grogan T, Yale C. Treatment of primary osteoarthritis of the hip. A comparison of total joint and surface replacement arthroplasty. *J Bone Jt. Surg Am*. 1984;66:228–241.

3. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol*. 2015;68:73–79.

4. Behrend H, Giesinger K, Giesinger JM, Kuster MS. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplast.* 2012;27:430–436.e1.

5. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open*. 2012;2:e000435.

6. Bieger R, Kappe T, Fraitzl CR, Reichel H. The aetiology of total knee arthroplasty failure influences the improvement in knee function. *Arch Orthop Trauma Surg*. 2013;133:237–241.

7. Borgbjerg J, Madsen F, Odgaard A. Patient Self-Assessed Passive Range of Motion of the Knee Cannot Replace Health Professional Measurements. *J Knee Surg*. 2017.

8. Carr A. Will registries slow down or accelerate innovation? *EFORT Open Rev*. 2019;4:416–422.

9. Chang CB, Yoo JH, Koh IJ, Kang YG, Seong SC, Kim TK. Key factors in determining surgical timing of total knee arthroplasty in osteoarthritic patients: age, radiographic severity, and symptomatic severity. *J Orthop Traumatol*. 2010;11:21–27.

10. Claessens AA, Schouten JS, van den Ouweland FA, Valkenburg HA. Do clinical findings associate with radiographic osteoarthritis of the knee? *Ann Rheum Dis*. 1990;49:771–774.

11. Collins JE, Rome BN, Daigle ME, Lerner V, Katz JN, Losina E. A comparison of patient-reported and measured range of motion in a cohort of total knee arthroplasty patients. *J Arthroplast.* 2014;29:1378–1382.e1.

12. Conaghan PG, Emerton M, Tennant A. Internal construct validity of the Oxford Knee Scale: evidence from Rasch measurement. *Arthritis Rheum*. 2007;57:1363–1367.

13. Coulter A. Measuring what matters to patients. *Bmj*. 2017;356:j816.

14. Dawson J, Beard DJ, McKibbin H, Harris K, Jenkinson C, Price AJ. Development of a patient-reported outcome measure of activity and participation (the OKS-APQ) to supplement the Oxford knee score. *Bone Jt. J*. 2014;96–b:332–338.

15. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Jt. Surg Br*. 1998;80:63–69.

16. Dawson J, Fitzpatrick R, Murray D, Carr A. A response to issues raised in a recent paper concerning the Oxford knee score. *Knee*. 2006;13:66–68.

17. Dieppe P, Lim K, Lohmander S. Who should have knee joint replacement surgery for osteoarthritis? *Int. J. Rheum. Dis.* 2011;14:175–180.

18. Dowsey MM, Nikpour M, Dieppe P, Choong PFM. Associations between pre-operative radiographic changes and outcomes after total knee joint replacement for osteoarthritis. *Osteoarthr. Cartil.* 2012;20:1095–1102.

19. Dowsey MM, Petterwood J, Lisik JP, Gunn J, Choong PFM. Prospective analysis of rural-urban differences in demographic patterns and outcomes following total joint replacement. *Aust. J. Rural Health*. 2014;22:241–248.

20. Dunbar MJ. Subjective outcomes after knee arthroplasty: Thesis. 2001.

21. 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.

22. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. *J Bone Jt. Surg Br*. 2001;83:339–344.

23. Dy CJ, Bozic KJ, Padgett DE, Pan TJ, Marx RG, Lyman S. Is changing hospitals for revision total joint arthroplasty associated with more complications? *Clin Orthop Relat Res*. 2014;472:2006–2015.

24. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4:e296.

25. Enemark-Larsen A. Exploring client-centred practice in Danish occupational therapy and the influence of the Canadian Occupational Performance Measure. 2019.

26. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J. Clin. Epidemiol.* 2015;68:435–441.

27. Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet*. 2019;393:655–663.

28. Fortin PR, Penrod JR, Clarke AE, St-Pierre Y, Joseph L, Belisle P, Liang MH, Ferland D, Phillips CB, Mahomed N, Tanzer M, Sledge C, Fossel AH, Katz JN. Timing of total joint replacement affects clinical outcomes among patients with osteoarthritis of the hip or knee. *Arthritis Rheum*. 2002;46:3327–3330.

29. Gagnier JJ, Mullins M, Huang H, Marinac-Dabic D, Ghambaryan A, Eloff B, Mirza F, Bayona M. A Systematic Review of Measurement Properties of Patient-Reported Outcome Measures Used in Patients Undergoing Total Knee Arthroplasty. *J Arthroplast.* 2017;32:1688–1697 e7.

30. Garratt AM, Brealey S, Gillespie WJ. Patient-assessed health instruments for the knee: a

structured review. *Rheumatol.* 2004;43:1414–1423.

31. Ghomrawi HMK, Lee Y, Herrero C, Joseph A, Padgett D, Westrich G, Parks M, Lyman S. A Crosswalk Between UCLA and Lower Extremity Activity Scales. *Clin. Orthop. Relat. Res.* 2017;475:542–548.

32. Gioe TJ, Pomeroy D, Suthers K, Singh JA. Can patients help with long-term total knee arthroplasty surveillance? Comparison of the American Knee Society Score self-report and surgeon assessment. *Rheumatol.* 2009;48:160–164.

33. Gómez-Barrena E, Padilla-Eguiluz NG, García-Rey E, Cordero-Ampuero J, García-Cimbrelo E. Factors influencing regional variability in the rate of total knee arthroplasty. *Knee*. 2014;21:236–241.

34. Goodfellow JW, O'Connor JJ, Murray DW. A critique of revision rate as an outcome measure: re-interpretation of knee joint registry data. *J Bone Jt. Surg Br*. 2010;92:1628–1631.

35. Gudbergsen H, Bartels EM, Krusager P, Waehrens EE, Christensen R, Danneskiold-Samsoe B, Bliddal H. Test-retest of computerized health status questionnaires frequently used in the monitoring of knee osteoarthritis: a randomized crossover trial. *BMC Musculoskelet Disord*. 2011;12:190.

36. Harcourt WG, White SH, Jones P. Specificity of the Oxford knee status questionnaire. The effect of disease of the hip or lumbar spine on patients' perception of knee disability. *J Bone Jt. Surg Br*. 2001;83:345–347.

37. Harris K, Dawson J, Doll H, Field RE, Murray DW, Fitzpatrick R, Jenkinson C, Price AJ, Beard DJ. Can pain and function be distinguished in the Oxford Knee Score in a meaningful way? An exploratory and confirmatory factor analysis. *Qual. Life Res.* 2013;22:2561–2568.

38. Harris K, Dawson J, Gibbons E, Lim CR, Beard DJ, Fitzpatrick R, Price AJ. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016;7:101–108.

39. Hossain FS, Konan S, Patel S, Rodriguez-Merchan EC, Haddad FS. The assessment of outcome after total knee arthroplasty: are we there yet? *Bone Jt. J*. 2015;97–b:3–9.

40. van Hout B et al. JMF. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health 2012 Jul-Aug;15(5):708-15.

41. Højgaard B, Kjellberg J, Det Nationale Institut for Kommuners og Regioners Analyse og Forskning (KORA). *Fem megatrends der udfordrer fremtidens sundhedsvæsen*. 2017.

42. Ingelsrud LH, Roos EM, Terluin B, Gromov K, Husted H, Troelsen A. Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. *Acta Orthop*. 2018;89:541–547.

43. Jasper LL, Jones CA, Mollins J, Pohar SL, Beaupre LA. Risk factors for revision of total knee arthroplasty: a scoping review. *BMC Musculoskelet Disord*. 2016;17:182.

44. Jenny JY, Bureggah A, Diesinger Y. Measurement of the knee flexion angle with smartphone applications: Which technology is better? *Knee Surg Sport. Traumatol Arthrosc*. 2016;24:2874–2877.

45. Jin X, Al Sayah F, Ohinmaa A, Marshall DA, Smith C, Johnson JA. The EQ-5D-5L Is Superior to the -3L Version in Measuring Health-related Quality of Life in Patients Awaiting THA or TKA. *Clin Orthop Relat Res*. 2019;477:1632–1644.

46. Judge A, Arden NK, Price A, Glyn-Jones S, Beard D, Carr AJ, Dawson J, Fitzpatrick R, Field RE. Assessing patients for joint replacement: can pre-operative Oxford hip and knee scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection? *J Bone Jt. Surg Br*. 2011;93:1660–1664.

47. Kahneman D. *Thinking fast and slow*. Farrar, Straus and Giroux; 2011.

48. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. *Ann. Rheum. Dis.* 1957;16:494–502.

49. van Kempen RW, Schimmel JJ, van Hellemondt GG, Vandenneucker H, Wymenga AB. Reason for revision TKA predicts clinical outcome: prospective evaluation of 150 consecutive patients with 2-years followup. *Clin Orthop Relat Res*. 2013;471:2296–2302.

50. Keurentjes JC, Fiocco M, So-Osman C, Onstenk R, Koopman-Van Gemert AW, Poll RG, Kroon HM, Vliet Vlieland TP, Nelissen RG. Patients with severe radiographic osteoarthritis have a better prognosis in physical functioning after hip and knee replacement: a cohort-study. *PLoS One*. 2013;8:e59500.

51. Khanna GMD, Singh JAMDMPH, Pomeroy DLMD, Gioe TJMD. Comparison of Patient-Reported and Clinician-Assessed Outcomes Following Total Knee Arthroplasty. *J. Bone Jt. Surg. - Am. Vol.* 2011;93:e117.

52. Kristensen N. Er tilfredse borgere lig med kvalitet? Vive/KORA, AKF nyt Dec 2011.

53. Lavernia C, D'Apuzzo M, Rossi MD, Lee D. Is postoperative function after hip or knee arthroplasty influenced by preoperative functional levels? *J Arthroplast.* 2009;24:1033–1043.

54. Lavernia CJ, Sierra RJ, Hungerford DS, Krackow K. Activity level and wear in total knee arthroplasty: a study of autopsy retrieved specimens. *J Arthroplast.* 2001;16:446–453.

55. Makela KT, Peltola M, Sund R, Malmivaara A, Hakkinen U, Remes V. Regional and hospital variance in performance of total hip and knee replacements: a national population-based study. *Ann Med*. 2011;43 Suppl 1:S31-8.

56. McKenna SP, Doward LC. The translation and cultural adaptation of patient-reported outcome measures. *Value Heal.* 2005;8:89–91.

57. McLaughlin P. Testing agreement between a new method and the gold standard—How do we test? *J. Biomech.* 2013;46:2757–2760.

58. MedHelp. Sygefravær i Danmark -en analyse af sygefraværet på det danske arbejdsmarked,

2011.

59. Meding JB, Ritter MA, Faris PM, Keating EM, Harris W. Does the preoperative radiographic degree of osteoarthritis correlate to results in primary total knee arthroplasty? *J Arthroplast.* 2001;16:13–16.

60. Miner AL, Lingard EA, Wright EA, Sledge CB, Katz JN. Knee range of motion after total knee arthroplasty: how important is this as an outcome measure? *J Arthroplast.* 2003;18:286–294.

61. Mokkink LB, Prinsen CAC, Patrick D, Alonso J, Bouter L, de Vet HCW, Terwee CB. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). User manual, version 1.0. 2018.

62. Mongelard K, Morup-Petersen A, Roemer L, Kristensen KB, Odgaard A. Heuristic ranking delivers more detail than than ordinal grading of knee osteoarthritis radiographs. *Submiss. Process*.

63. Morup-Petersen A, Holm PM, Holm CE, Klausen TW, Skou ST, Krogsgaard MR, Laursen MB, Odgaard A. Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale. *J Arthroplast.* 2018;33:2875–2883.e3.

64. Mourcou Q, Fleury A, Diot B, Franco C, Vuillerme N. Mobile Phone-Based Joint Angle Measurement for Functional Assessment and Rehabilitation of Proprioception. *Biomed Res Int*. 2015;2015:328142.

65. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. *J Bone Jt. Surg Br*. 2007;89:1010–1014.

66. Mørup-Petersen A, Holm PM, Holm CE, Klausen TW, Skou ST, Krogsgaard MR, Laursen MB, Odgaard A. Corrigendum to “Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale” [The Journal of Arthroplasty 33 (2018) 2875-2883]. *J. Arthroplasty*. 2019.

67. Niemelainen MJ, Makela KT, Robertsson O, A WD, Furnes O, Fenstad AM, Pedersen AB, Schroder HM, Huhtala H, Eskelinen A. Different incidences of knee arthroplasty in the Nordic countries. *Acta Orthop*. 2017;88:173–178.

68. Norkin C. *Measurement of Joint Motion, 5e, A Guide to Goniometry*. 5. ed. (Norkin C, White DJ, eds.). Philadelphia: F.A. Davis; 2016.

69. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? *Clin Orthop Relat Res*. 2009;467:958–965.

70. Odgaard A, Paulsen A. Translation and Cross-Cultural Adaptation of the Danish Version of Oxford Knee Score (OKS). Abstract 2009.

71. Open-edit radiology resource. www.radiopedia.org.

72. Paxton RJ, Melanson EL, Stevens-Lapsley JE, Christiansen CL. Physical activity after total knee arthroplasty: A critical review. *World J. Orthop.* 2015;6:614–622.

73. Peck CN, Childs J, McLauchlan GJ. Inferior outcomes of total knee replacement in early radiological stages of osteoarthritis. *Knee*. 2014;21:1229–1232.

74. Pedersen AB, Mehnert F, Odgaard A, Schroder HM. Existing data sources for clinical epidemiology: The Danish Knee Arthroplasty Register. *Clin Epidemiol*. 2012;4:125–135.

75. Pinsornsak P, Naratrikun K, Kanitnate S, Sangkomkamhang T. The one-leg standing radiograph. *Bone Joint Res.* 2016;5:436–441.

76. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, Carr A, Beard D. Knee replacement. *Lancet*. 2018;392:1672–1682.

77. Pua YH, Ong PH, Chong HC, Yeo W, Tan CI, Lo NN. Associations of self-report physical function with knee strength and knee range-of-motion in total knee arthroplasty possible nonlinear and threshold effects. *J Arthroplast.* 2013;28:1521–1527.

78. Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. *Acta Orthop.* 2019;90:297.

79. RCoreTeam. RStudio version 1.1.463. https://www.R-project.org/. R: A Language and Environment for Statistical Computing.

80. Regeringen, Danske Regioner. *Aftale om regionernes økonomi for 2016*. 2015.

81. Robertsson O, Bizjajeva S, Fenstad AM, Furnes O, Lidgren L, Mehnert F, Odgaard A, Pedersen AB, Havelin LI. Knee arthroplasty in Denmark, Norway and Sweden. A pilot study from the Nordic Arthroplasty Register Association. *Acta Orthop*. 2010;81:82–89.

82. Robertsson O, Ranstam J. No bias of ignored bilaterality when analysing the revision risk of knee prostheses: Analysis of a population based sample of 44,590 patients with 55,298 knee prostheses from the national Swedish Knee Arthroplasty Register. *BMC Musculoskelet. Disord.* 2003;4:1.

83. Rolfson O, Bohm E, Franklin P, Lyman S, Denissen G, Dawson J, Dunn J, Eresian Chenok K, Dunbar M, Overgaard S, Garellick G, Lubbeke A, Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty R. Patient-reported outcome measures in arthroplasty registries Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for selection, administration, and analysis. *Acta Orthop*. 2016;87 Suppl 1:9–23.

84. Schnurr C, Jarrous M, Gudden I, Eysel P, Konig DP. Pre-operative arthritis severity as a predictor for total knee arthroplasty patients' satisfaction. *Int Orthop*. 2013;37:1257–1261.

85. Schäfer T, Pritzkuleit R, Jeszenszky C, Malzahn J, Maier W, Günther KP, Niethard F. Trends and geographical variation of primary hip and knee joint replacement in Germany. *Osteoarthr. Cartil.* 2013;21:279–288.

86. Scott CE, Oliver WM, MacDonald D, Wade FA, Moran M, Breusch SJ. Predicting dissatisfaction following total knee arthroplasty in patients under 55 years of age. *Bone Jt. J*. 2016;98–b:1625–1634.

87. Scott CEH, Turnbull GS, MacDonald D, Breusch SJ. Activity levels and return to work following total knee arthroplasty in patients under 65 years of age. *Bone Jt. J*. 2017;99–b:1037–1046.

88. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sport. Med*. 2003;37:197–206; discussion 206.

89. Skou ST, Roos E. GLA:D (Good Life with osteoArthritis in Denmark) - Annual Report. 2014:28.

90. Skou ST, Roos EM, Laursen MB, Rathleff MS, Arendt-Nielsen L, Rasmussen S, Simonsen O. Response to Letter to Editor: "2-year outcome from two parallel randomized controlled trials. Reporting considerations." *Osteoarthr. Cartil.* 2018.

91. Skrejborg P, Petersen KK, Kold S, Kappel A, Pedersen C, Ostgaard SE, Simonsen O, Arendt-Nielsen L. Presurgical Comorbidities as Risk Factors For Chronic Postsurgical Pain Following Total Knee Replacement. *Clin J Pain*. 2019;35:577–582.

92. SooHoo NF, Li Z, Chenok KE, Bozic KJ. Responsiveness of patient reported outcome measures in total joint arthroplasty patients. *J Arthroplast.* 2015;30:176–191.

93. Stammers JG, Kuo A, Hart AJ, Smeeth L, Skinner JA. Registry Data-Valuable Lessons But Beware the Confounders. *J Arthroplast.* 2017;32:S63-s67.

94. Stone OD, Duckworth AD, Curran DP, Ballantyne JA, Brenkel IJ. Severe arthritis predicts greater improvements in function following total knee arthroplasty. *Knee Surg Sport. Traumatol Arthrosc*. 2017;25:2573–2579.

95. Sundhedsdatastyrelsen. The Danish National Health Register. (Landspatientregistret) https://www.esundhed.dk/Registre/Landspatientsregisteret/Operationer. 2019.

96. Sundhedsstyrelsen. *Knæartrose – nationale kliniske retningslinjer og faglige visitationsretningslinjer (Danish National Clinical Guidelines on treatment of knee osteoarthritis)*.

97. Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting quality of life instruments. *Value Heal.* 2004;7 Suppl 1:S27-30.

98. Terluin B, Eekhout I, Terwee CB, De Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *J. Clin. Epidemiol.* 2015;68:1388–1396.

99. Terwee CB, Bouwmeester W, van Elsland SL, de Vet HC, Dekker J. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthr. Cartil.* 2011;19:620–633.

100. Thomsen MG, Latifi R, Kallemose T, Barfod KW, Husted H, Troelsen A. Good validity and reliability of the forgotten joint score in evaluating the outcome of total knee arthroplasty. *Acta Orthop*. 2016;87:280–285.

101. Tilbury C, Holtslag MJ, Tordoir RL, Leichtenberg CS, Verdegaal SH, Kroon HM, Fiocco M, Nelissen RG, Vliet Vlieland TP. Outcome of total hip arthroplasty, but not of total knee arthroplasty, is related to the preoperative radiographic severity of osteoarthritis. A prospective cohort study of

573 patients. *Acta Orthop*. 2016;87:67–71.

102. Troelsen A, Schroder H, Husted H. Opinions among Danish knee surgeons about indications to perform total knee replacement showed considerable variation. *Dan Med J*. 2012;59:A4490.

103. Den Tværregionale Task Force, Danske Regioner. Styring efter værdi for patienten. 2015.

104. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011.

105. Videnscenter for Brugerinddragelse. *Program PRO. Anvendelse af PRO-data i kvalitetsudviklingen af det danske sundhedsvæsen*. 2016.

106. Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. *J Knee Surg*. 2010;23:137–151.

107. Weick J, Bawa HS. The potential utility of patient-reported range of motion after total knee arthroplasty. *Ann Transl Med*. 2015;3:193.

108. Whitehouse SL, Blom AW, Taylor AH, Pattison GT, Bannister GC. The Oxford Knee Score; problems and pitfalls. *Knee*. 2005;12:287–291.

109. Wilson I, Bohm E, Lubbeke A, Lyman S, Overgaard S, Rolfson O, A WD, Wilkinson M, Dunbar M. Orthopaedic registries with patient-reported outcome measures. *EFORT Open Rev*. 2019;4:357–367.

110. Wylde V, Lenguerrand E, Brunton L, Dieppe P, Gooberman-Hill R, Mann C, Blom AW. Does measuring the range of motion of the hip and knee add to the assessment of disability in people undergoing joint replacement? *Orthop Traumatol Surg Res*. 2014;100:183–186.

111. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. *J Arthroplast.* 1998;13:890–895.

112. The Danish Knee Arthoplasty Register. Annual Report 2018.

113. The Danish Knee Arthoplasty Register. Annual Report 2019.

114. Danish Hip and Knee Arthroplasty Society: http://dshk.ortopaedi.dk/.

115. The Danish Knee Arthoplasty Register. Annual Report 2015.

116. Sundhedsstyrelsens specialevejledning for Ortopædisk kirurgi (Organization of Orthopaedic Surgery in Denmark). June 2019.

117. The Danish Knee Arthoplasty Register. Annual Report 2010.

118. The Danish Hip Arthoplasty Register. Annual Report 2018.

119. LUP Somatik: Landsdækkende Undersøgelse af Patientoplevelser 2017. Kompetencecenter for Patientoplevelser. www.patientoplevelser.dk/lup.

120. COSMIN: COnsensus-based Standards for the selection of health status Measurement

Instruments. Homepage: www.cosmin.nl.

121. https://euroqol.org/wp-content/uploads/2019/09/EQ-5D-5L-English-User-Guide_version-3.0-Sept-2019-secured.pdf.

122. SAS Statistical Software, University Edition, version 3.6, SAS Institute Inc., Cary, NC.

123. Kulturrevolution på vej på de danske sygehuse. *Ugebrevet Mandag Morgen*. 2015.

124. Statistics Denmark: Use of IT in the Danish population (IT-anvendelse i befolkningen). 2017:8–9,37.

Study I: SPARK Pre

Large variation in revision rates after primary knee arthroplasty: A matter of patient selection? Baseline data from 1452 patients in the prospective multicenter cohort study, SPARK

Anne Mørup-Petersen[1], Mogens Laursen[2], Frank Madsen[3], Michael R. Krogsgaard[4], Kristian B.G. Mongelard[5], Lone Rømer[6], Matilde Winther-Jensen[7], Anders Odgaard[1]

[1)] Department of Orthopedic Surgery, Herlev and Gentofte Hospital, Copenhagen, Copenhagen University
[2)] Department of Orthopedic Surgery, Aalborg University Hospital, Aalborg and Farsoe
[3)] Department of Orthopedic Surgery, Aarhus University Hospital
[4)] Section for Sports Traumatology M51, Bispebjerg and Frederiksberg Hospital, Copenhagen, Copenhagen University
[5)] Department of Radiology, Herlev and Gentofte Hospital, Copenhagen, Copenhagen University
[6)] Department of Radiology, Aarhus University Hospital
[7)] Center for Clinical Research and Prevention, Department of Clinical Epidemiology, Bispebjerg and Frederiksberg Hospital, Copenhagen, Copenhagen University

Corresponding author: Anne Mørup-Petersen, email: morup.anne@gmail.com

Abstract

Background/Rationale

Revision rates after primary knee arthroplasty (KA) differ widely among countries, regions and hospitals, e.g. between 1 and 5% at 2 years among the five Danish regions. We explore whether this reflects true variations in surgical treatment.

Questions / Purposes

This study asks, "Do hospitals with different revision rates differ in selection of patients for KA?", as measured by 1) demography and lifestyle, 2) patient-reported outcome measures (PROMs) and motivations for surgery, 3) radiographic classification of knee osteoarthritis and 4) KA incidence and implant choice. A separate article will report postoperative results.

Methods

The prospective cohort study included patients scheduled for primary KA of any kind in three high-volume hospitals (2016-18). Among outcomes were Oxford Knee Score (OKS), Copenhagen Knee ROM Scale (knee motion) and blinded radiographic osteoarthritis classifications. 1452 patients provided baseline data (89% of patients contacted, 56% of all operated).

Results

1) Patients were 1.3-2.0 y. older (p=0.002) in Copenhagen (high-revision), and BMI was 1.5-1.7 kg/m^2 higher (p<0.001) in Aalborg (low-revision). 2) PROMs, ROM and motivations for surgery were the same across hospitals (OKS 23.3 ± 7) but varied with implant type. 3) In Aarhus (low-revision), osteoarthritis was more severe (Kellgren-Lawrence/Ahlbäck ≥ 2 in 98/75% vs. 92/64%, p<0.015). KA incidence was 28% higher in the high-revision region (p<0.001).

Conclusion

Preoperative PROMs were similar across hospitals with very different revision rates. Radiographic classifications and KA incidence indicate higher thresholds for primary surgery in one low-revision hospital, but the remaining variations were paradox to well-known revision risk factors.

Introduction

The quality of knee arthroplasty (KA) surgery is traditionally evaluated by comparing revision rates. Large variation is seen among nations as expressed in results from arthroplasty registers. The Swedish register report a 10-year cumulative revision rate for TKA (indication osteoarthritis) of approximately 3.6% for operations performed after 2006, and comparable figures for Denmark and England/Wales are 6.3% and 3.1%, respectively [3, 47, 55]. This variation implies differences in quality of service. All registers also show considerable variation at national levels. An example is the Danish Knee Arthroplasty Register [32, 45], where revision rate differences have persistently been observed among the five Danish administrative regions. The highest revision rates are seen in the capital (5.0% at 2 years in 2015), and rates fall with increasing distance from the capital (2.2% in Central Denmark and 1.0% in North Denmark). This variation has been present for more than a decade (Appendix A)[48, 49], and a similar variation is noted for hips [50]. Paradoxically, some of the hospitals with the lowest revision rates after KA are those with the most frequent use of unicompartmental implants, that are otherwise up to five times more likely to be revised than total knee arthroplasties [2, 14, 31, 51].

It is tempting to guess at explanations for regional difference, e.g. that surgeons in high-revision regions or hospitals perform poorer surgery. Another option could be that revision rate differences might reflect regional variance in surgeons' thresholds for offering revision surgery or underlying variations in patients' characteristics or disease states at the time of primary surgery. No previous study has to our knowledge compared KA patients across national regions or hospitals with respects to demography, knee symptoms and severity of radiographic knee osteoarthritis (OA) prior to surgery; all factors which have been shown to be associated with subsequent postoperative patient satisfaction and risk of revision [12, 20, 24]. Thus, in 2015, we initiated the observational cross-regional triple-center study, SPARK, to investigate "Variation in patient Satisfaction, Patient-reported outcome measures, radiographic signs of Arthritis, and Revision rates in Knee arthroplasty patients in three Danish regions". The purpose of studying preoperative conditions was to reach a better understanding of subsequent regional variation in revision rates [8].

Preoperative data from the SPARK study aims to answer the question: "Does patient selection for primary knee arthroplasty differ among three high-volume hospitals with consistently differing revision rates?", as measured by 1) demography and lifestyle, 2) patient-reported outcome measures (PROMs) and motivations for surgery, 3) radiographic classification of knee OA and 4) primary KA incidence and implant choice. All analyses were based on the null-hypothesis that patient selection was the same across the three hospitals. As the study was explorative in origin, further data-driven analyses were allowed when appropriate [36]. Postoperative results will be reported in a separate publication.

Methods

A prospective, observational cohort study was initiated to investigate patients listed for knee arthroplasty at three large hospitals using PROMs and radiographic analyses. The reporting of the study follows the STROBE guidelines for observational cohort studies [11].

Patient selection

We selected the largest knee arthroplasty center at the time in each of the three regions with the most extreme revision rates. All three hospitals had revision rates comparable to the particular region as a whole [48]. Patients who were examined and subsequently scheduled for any type of primary KA were eligible for inclusion from 1 September 2016 to 30 November (Aarhus University Hospital in Central Denmark and Aalborg University Hospital Farsoe in North Denmark) or to 31 December 2017 (Herlev-Gentofte, Copenhagen in Capital Region). Exclusion criteria were severe developmental knee deformities, malignant knee tumor, hemophilia, dementia or language barriers that could not be overcome by help from relatives. More importantly, patients unable to receive the questionnaire by email were not included in the study during most of the study period. During the last 6 months of the inclusion period, we also included patients without an email in order to unveil any differences in variables in this group of patients.

Inclusion process

In Aalborg and Aarhus, surgeons invited patients to participate in the study when surgery was scheduled. In Copenhagen, patients were included by one of the 20-25 employed medical students who performed the routine pre-assessment a few weeks before primary KA. Before giving their written consent, patients were given one page of study information. Two days after inclusion, patients received an email with a unique link to the preoperative PROM set (or a letter with the same content). If necessary, up to two email reminders were sent three days apart. To prevent patients with bilateral knee OA from being confused as to which knee to think of, the email clearly stated that the knee planned for surgery was the object of study. Patients scheduled for operation in both knees were allowed to participate twice if the operations were performed on separate occasions. The six patients scheduled for simultaneous bilateral surgery were asked to choose in advance which knee to participate with [30]. Since PROM was the cornerstone in this study, patients who failed to fill in the questionnaire prior to surgery were excluded from data analyses.

Questionnaires were sent to 1704 patients (figure 1), 52 of those by letter. In 32 cases, the email address or laterality was wrong, or a technical error occurred. In 48 cases, surgery was cancelled or postponed beyond the study period. Of the remaining 1624 patients, who all received a questionnaire, 1452 patients (89.4%) answered before surgery (fig. 1). In this group, 53 patients participated with one knee first and the contralateral knee later, hence 1399 individuals were represented in the study. Patients spent 12:30 minutes on average filling out the preoperative questionnaire, at mean 29 days (median 18 days) before surgery.

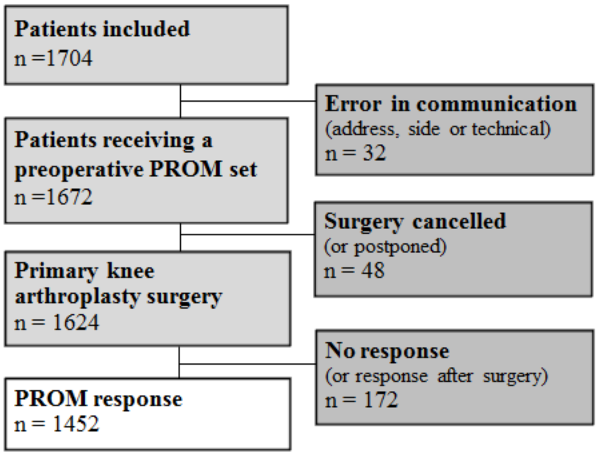


Figure 1. *Inclusion flowchart.*

Inclusion rate

We did not collect information about non-participants. Instead, post-hoc quantification of inclusion rates and comparisons of participants vs. non-participants were carried out. Since time from inclusion to surgery varied, the analysis had to be based upon registered surgical activity in a relevant time period [41]. Among 1924 patients operated in 2017, 1083 patients (56.3 %) provided PROM data for this study, 62% in Aarhus/Copenhagen and 38% in Aalborg (table 1). Non-responders were equally distributed among the three hospitals (Aarhus 7.0, Aalborg 8.2 and Copenhagen 10.0%, $p = 0.210$). In 2017 overall, patients operated in Copenhagen were older than patients in the other two hospitals ($p=0.006$). In Aalborg, more patients were male (48 vs. 38-39 %, $p=0.005$) despite only limited interregional variation in sex distribution in the background population (between 47 and 49% males in ages 60-79 years) [56]. These differences were reflected in the SPARK samples of Aalborg and Copenhagen. In Aarhus and in the SPARK sample as a whole, males and younger patients were more likely to participate than their female and older counterparts. Further analyses (not shown) revealed that the distribution of KA implant types (medial/lateral unicompartmental (MUKA/LUKA), patellofemoral (PFA) and total (TKA)) was no different between participants and non-participants within each hospital ($p \geq 0.230$).

Table 1. Inclusion analysis based on complete surgical activity in 2017					
		Complete primary KA population 2017	SPARK participation		
			Yes	No	p
Patients (n (%))	Total	1924 (100)	1083 (56.3)	841 (43.7)	-
	Aarhus	391 (100)	243 (62.1)	148 (37.9)	-
	Aalborg	429 (100)	161 (37.5)	268 (62.5)	-
	Copenhagen	1104 (100)	679 (61.5)	425 (38.5)	-
Age (mean ± SD)	Total	68.19 ± 9.8	67.7 ± 9.2	68.8 ± 10.5	**0.020**
	Aarhus	67.10 ± 10.6	66.1 ± 9.9	68.7 ± 11.5	**0.019**
	Aalborg	67.62 ± 9.8	66.7 ± 8.8	68.2 ± 10.3	0.141
	Copenhagen	68.80 ± 9.4	68.5 ± 9.0	69.2 ± 10.2	0.256
Male sex (n (%))	Total	779 (40.5)	459 (42.4)	320 (38.1)	**0.016**
	Aarhus	153 (39.1)	105 (43.2)	48 (32.4)	**0.043**
	Aalborg	202 (47.8)	83 (51.6)	119 (44.4)	0.070
	Copenhagen	424 (38.4)	271 (39.9)	153 (36.0)	0.125

P-values <0.05 indicate a skewness in distribution of participants and non-participants.

PROM selection

Several generic and knee specific PROMs were collected. Oxford Knee Score (OKS) was the primary outcome [6, 10]. It was developed in 1998 for knee OA patients having a total knee replacement. Items were developed through patient interviews, and the score has been thoroughly validated, e.g. through Rasch analysis [5, 7, 17, 19, 44]. The Danish translation was made in 2009 using recommended methods [16] (not published). UCLA Activity Scale allowed patients to report their level of physical activity on a 10-level scale [42, 46], and Copenhagen Knee ROM Scale assessed patient-reported range of knee motion where flexion was assessed on a scale from 0-6 (6 max) and extension from 0-5 (5 max) [29, 32]. Based on validation studies, "flexion deficit" was set to identify 95% of patients with passive flexion below 100° (sensitivity) and exclude 81% of those with flexion above that limit (specificity). Similarly, "extension deficit" would identify 78% of patients with passive extension worse than 10° and exclude 70% of those with extension better that 10°. The knee-specific questions were preceded by the generic EQ-5D-5L and EQ-VAS measuring self-reported general health [18, 22, 40, 45, 52]. Also, a global knee anchor question assessed patients' overall perception of the knee condition, worded, "How is your knee at the moment?". Answers to the anchor question were given on a visual analogue scale (VAS) marked in one end, "My knee is not functioning at all or it is very painful" and in the other end, "My knee is painless and normally functioning" returning scores from 0-100 (100 best). Patients also indicated their motivation for undergoing surgery; 13 common reasons had been listed based on explorative interviews of 35 knee OA patients (unpublished). From this list, patients were asked to choose up to five, or add one as free text.

Height and weight was reported by patients along with lifestyle data concerning urbanization (”city/suburb", "small town/village" or "countryside"), daily smoking ("yes"/"no") and alcohol consumption (more or less than 2 standard drinks (12 g alcohol) per day). We asked whether the knee planned for surgery was the patients' main physical disability, with answer options, "yes", "no" or "difficult to say". The use of analgesics of any kind was assessed by the question: "How often do you take painkillers because of your knee?" answered in a 5-point Likert scale ranging from "more than once daily" to "rarely or never" (full wording in table 2).

Radiographic severity of OA

Patients' severity of tibiofemoral OA was assessed in preoperative weight bearing postero-anterior knee radiographs taken perpendicular to the tibial plateau with the knee in 15-30° flexion [4]. In Aalborg and Copenhagen, both knees were in the same picture with weight put on both legs, whereas in Aarhus, one leg was filmed at a time carrying almost total body weight. Only radiographs of patients with PROM answers were analyzed. Patients who were listed for PFA (50 of 1452 patients = 3.4 %) or LUKA (7 of 1452 patients = 0.5%) and patients with radiographs showing predominantly lateral OA (167 of the remaining 1395 patients = 12%) were excluded from radiographic analyses since these isolated OA types are usually better assessed in a tangential (skyline, Merchant) or flexed (Rosenberg) view, respectively. In 177 cases, radiographs were unavailable due to logistic matters that were independent of patient-related factors. The 1051 (86%) available radiographs (Aarhus 206, Aalborg 171 and Copenhagen 674) were viewed in random order by two separate, blinded radiologists with special interest and experience in musculoskeletal radiology. First, the Ahlbäck score [1], and secondly, in a round of new order, the Kellgren-Lawrence classification (K-L) [23] was noted for each patient. The interrater agreement reached a weighted Kappa value of 0.59 for both K-L and Ahlbäck. In 29% (K-L) and 41% (Ahlbäck) of cases, there was disagreement, and reassessment by both radiologists together was required to reach consensus. Finally, radiographs were evaluated free of classifications by 13 experienced knee arthroplasty surgeons from five Danish regions. Each surgeon was presented to the radiographs in multiple random pairs and was told each time to choose the radiograph of the knee expected to cause the more severe knee symptoms. Surgeons were told not to consider any formal grading system and instead use their personal experience and heuristics. The result of this heuristics-based assessment (17,767 comparisons) was a complete ranking of all available radiographs from 1 to 1051 (no. 1 most severe) [28].

Incidence of primary KA

Public hospitals perform 95% of all primary KA operations in Denmark [51]. The incidence of primary KA on a regional level (both private and public hospitals) was retrieved from the National Patient Register by NOMESCO classification KNGB and subgroups [53].

Statistical Analysis, Study Size

Sample size was based on clinical relevance and feasibility rather than a formal calculation. We expected approximately 1800 operations to be carried out in the study period, representing 15% of all primary KA's in Denmark. If 75% of patients were included and 80% answered the PROM set, 1080 answers would be ready for analysis. Any regional differences in examined parameters that could not be detected in a representative material of such size were considered clinically irrelevant to the overall study question.

Descriptive statistics were presented as total data and stratified by hospital and sex, and later by arthroplasty type (TKA vs. MUKA) in hospitals grouped by frequency of MUKA use. Significance of difference was tested depending on the type and structure of data: Chi-square test for dichotomized variables, unpaired t-test or one-way analysis of variance (ANOVA) for parametric variables and Mann-Whitney U or Kruskal-Wallis test for nonparametric (ordinal) data. In analogy with previous studies, OKS and EQ-5D data were treated as a numeric variable [30] as was knee flexion and extension [29, 32]. K-L, Ahlbäck, surgeons' ranking and UCLA scores were treated as ordinal variables. A separate article describes the heuristics-based assessment of radiographs[28]. In order to adjust for confounders, multiple regression analyses were carried out.

In regression models, Aarhus was set as the reference hospital, as it is placed between the two other hospitals both geographically and with respects to urbanization and revision rates. All tests were unpaired (as if every knee belonged to a unique participant); the 53 patients who participated twice accounted for only 7.3% of answers [38]. The significance level was set at 0.05 (two-sided) and 95% confidence intervals (CI) were reported when relevant. Data collection and Case Report Forms (CRF) etc. were handled by Procordo Aps, Copenhagen. Analyses were carried out in R (RStudio)[37].

Ethics and funding
The National Committee of Health Research Ethics provided ethical approval (Protocol no. 16038343, 2 September 2016) and The Danish Data Protection Agency (Jr. no. 2012-58-0004, HGH-2016-087, I-Suite no. 04819) approved data management. EQ-5D-5L permission was provided by EuroQol (ID 28583). The Health Research Fund of the Capital Region of Denmark funded this research (granted July 2015).

Results

Demography and lifestyle

Participants from Copenhagen were 1.4-2.0 years older than those in Aalborg and Aarhus, and male sex was more frequent in Aalborg (56 vs. 43-45%) (table 2). In the Aalborg group, males were 3.4 years older than females (95% CI, 0.8-6.0, p=0.007), whereas there was no difference within the other hospitals. BMI was lower in the elderly (-0.13 kg/m^2/year, 95% CI, -0.11-0.16, p<0.001) and higher in females (+0.69 kg/m^2, 95% CI, 0.18-1.2, p=0.007) as well as in Aalborg patients (+1.5-1.7 kg/m^2, p<0.001); the latter still when adjusted for age and sex (+1.4-1.9 kg/m^2). There were no hospital differences in smoking, alcohol consumption, self-reported general health (EQ-5D-5L) or physical activity level (UCLA). Males reported higher levels of all these parameters except smoking. When adjustments were made for age, sex and BMI, no variances in these lifestyle and general health parameters were explained by hospital, i.e. hospitals were comparable with respects to smoking, alcohol consumption, EQ-5D-5L and physical activity. The 41 patients participating by letter were 8.1 years older than others (95% CI, 6.3-9.8, p<0.001) and 29 (71%) of them were female (as opposed to 54% in the email group, p=0.052).

Table 2. Preoperative data from 1452 responding patients.

		Hospital				Sex		
	Total sample	**Aarhus** (Low rev.rate)	**Aalborg** (Low rev.rate)	**Copenhagen** (High rev.rate)	**p**	**Female**	**Male**	**p**
Patients (%)	1452 (100)	321 (22)	202 (14)	929 (64)		793 (55)	659 (45)	
Demographics & implants								
Age	68.0 ± 9.3	66.6 ± 9.7	67.3 ± 9.1	68.6 ± 9.1	**0.002**	67.7 ± 9.7	68.3 ± 8.7	0.214
Male sex (%)	659 (45)	145 (45)	114 (56)	400 (43)	**0.002**	0 (0)	659 (100)	
Implant type (%)					**<0.001**			**0.001**
TKA	1059 (73)	164 (51)	174 (86)	721 (78)		590 (74)	469 (71)	
MUKA	336 (23)	129 (40)	25 (12)	182 (20)		160 (20)	176 (27)	
PFA	50 (3.4)	23 (7.2)	3 (1.5)	24 (2.6)		38 (4.8)	12 (1.8)	
LUKA	7 (0.5)	5 (1.6)	0 (0.0)	2 (0.2)		5 (0.6)	2 (0.3)	
Health & lifestyle*								
Weight (kg)	86 ± 17	85 ± 17	90 ± 15	85 ± 17	**0.002**	81 ± 16	92 ± 15	**<0.001**
BMI (kg/m^2)	28.9 ± 5.0	28.5 ± 4.6	30.2 ± 5.1	28.7 ± 5.1	**<0.001**	29.2 ± 5.7	28.5 ± 4.1	**0.009**
BMI group (%)					**<0.001**			0.601
Normal (< 25)	329 (23)	77 (24)	26 (13)	226 (24)		196 (25)	133 (20)	
Overweight (25-29.9)	589 (41)	140 (44)	78 (39)	371 (40)		284 (36)	305 (46)	
Obese (≥ 30)	529 (37)	102 (32)	98 (49)	329 (36)		309 (39)	220 (3)	
Alcohol (> 2 units per day) (%)	164 (11)	36 (11)	15 (7)	113 (12)	0.154	43 (5)	121 (18)	**<0.001**
Daily smoking (%)	159 (11)	41 (13)	21 (10)	97 (11)	0.499	87 (11)	72 (11)	1.000
Urbanization (%)					**<0.001**			0.773
Countryside	78 (5)	18 (6)	33 (16)	27 (3)		43 (5)	35 (5)	
Small town or village	354 (24)	75 (23)	111 (55)	168 (18)		190 (24)	164 (25)	
City or suburb	1019 (70)	228 (71)	58 (29)	733 (79)		559 (71)	460 (70)	
Participation by letter (%)	41 (2.8)	5 (1.6)	10 (5.0)	26 (2.8)	0.074	29 (3.7)	12 (1.8)	0.052
EQ-VAS	61 ± 22	62 ± 21	58 ± 24	62 ± 22	0.091	59 ± 22	65 ± 21	**<0.001**
EQ-5D-5L Index	0.59 ± 0.15	0.59 ± 0.15	0.61 ± 0.12	0.59 ± 0.15	0.144	0.58 ± 0.15	0.60 ± 0.14	**0.028**
UCLA	4.7 [4] ± 1.9	4.8 [4] ± 1.9	4.8 [4] ± 1.9	4.7 [4] ± 1.8	0.551	4.5 [4] ± 1.7	5.1 [5] ± 2.0	**<0.001**
Knee-specific PROMs*								
OKS	23.3 [24] ± 6.7	23.5 [24] ± 7.0	23.2 [24] ± 6.5	23.3 [23] ± 6.7	0.884	22.0 [22] ± 6.4	24.8 [25] ± 6.8	**<0.001**
Global knee anchor	28 ± 18	27 ± 17	30 ± 18	29 ± 18	0.193	28 ± 18	29 ± 18	0.132
Range of motion (Copenhagen Knee ROM Scale) [1]								
Flexion	4.9 [5] ± 1.2	4.8 [5] ± 1.2	4.8 [5] ± 1.1	4.9 [5] ± 1.2	0.236	4.9 [5] ± 1.2	4.9 [5] ± 1.2	0.645
Deficit (CKRS 0-4) (%)	416 (29)	97 (30)	58(29)	261 (28)	0.774	223 (28)	193 (29)	0.678
Extension[2]	3.5 [4] ± 1.0	3.4 [4] ± 1.0	3.4 [3] ± 0.9	3.5 [4] ± 0.9	0.188	3.5 [4] ± 1.0	3.5 [4] ± 0.9	0.401
Deficit (CKRS 0-3) (%)[2]	340 (49)	63 (45)	72 (62)	205 (46)	**0.007**	200 (49)	140 (49)	1.000
"My knee is my main disability" (%)	1261 (87)	289 (90)	176 (87)	796 (86)	0.146	680 (86)	581 (88)	0.327
Analgesics due to knee pain (%)					0.094			**<0.001**
More than once daily	667 (46)	145 (45)	83 (41)	439 (47)		407 (51)	260 (40)	
Once daily	187 (13)	34 (11)	32 (16)	121 (13)		95 (12)	92 (14)	
More than once weekly	218 (15)	42 (13)	27 (13)	149 (16)		129 (16)	89 (14)	
More than once monthly	142 (10)	39 (12)	19 (9)	84 (9)		70 (9)	72 (11)	
Rarely or never	237 (16)	61 (19)	41 (20)	135 (15)		91 (11)	146 (22)	
Degree of radiographic OA[3]								
K-L classification (%)					**0.016**			**0.011**
0	7 (0.7)	0 (0.0)	2 (1.2)	5 (0.7)		5 (0.9)	2 (0.4)	
1	57 (5.4)	4 (1.9)	13 (7.6)	40 (5.9)		35 (6.5)	22 (4.3)	
2	136 (13)	24 (12)	23 (14)	89 (13)		76 (14)	60 (12)	
3	787 (75)	160 (78)	123 (72)	504 (75)		395 (73)	392 (76)	
4	64 (6.1)	18 (8.7)	10 (5.8)	36 (5.3)		27 (5.0)	37 (7.2)	
K-L classification ≥ 2 (%)	987 (94)	202 (98)	156 (91)	629 (93)	**0.013**	498 (93)	489 (95)	0.082
K-L classification ≥ 3 (%)	851 (81)	178 (86)	133 (78)	540 (80)	0.067	422 (78)	429 (84)	**0.039**
Ahlbäck score (%)					0.104			**0.010**
0	56 (5)	7 (3)	9 (5)	40 (6)		37 (7)	19 (4)	
1	289 (28)	44 (21)	56 (33)	189 (28)		158 (29)	131 (26)	
2	401 (38)	94 (46)	61 (36)	246 (37)		198 (37)	203 (40)	
3	291 (28)	57 (28)	43 (25)	191 (28)		140 (26)	151 (29)	
4	12 (1.1)	3 (1.5)	2 (1.2)	7 (1.0)		4 (0.7)	8 (1.6)	
5	2 (0.2)	1 (0.5)	0 (0.0)	1 (0.1)		1 (0.2)	1 (0.2)	
Ahlbäck score ≥ 2 (%)	704 (67)	154 (75)	106 (62)	444 (66)	**0.015**	342 (64)	362 (71)	**0.019**
Ahlbäck score ≥ 3 (%)	305 (29)	61 (30)	45 (26)	199 (30)	0.696	145 (27)	160 (31)	0.148
Surgeons' ranking (mean [25-75%]) [4])	540 [270-808]	380 [188-718]	598 [315-864]	561 [293-824]	**<0.001**	575 [318-845]	503 [238-778]	**0.002**

** Patient-reported data. Abbreviations: OKS: Oxford Knee Score (version 0-48, 48 best). BMI: Body Mass Index (BMI group "underweight" (<18.5 kg/m^2) comprised only two patients, who were thus included in the "normal" group). UCLA: UCLA Activity Scale (1-10, 10 most active). K-L: Kellgren-Lawrence. Global knee anchor: Patients' overall knee assessment, recorded on VAS (0-100, 100 best). [1]) Copenhagen Knee ROM Scale: Flexion 0-6 (6 is max), Extension 0-5 (5 is max), see text for details. [2]) n = 699. [3]) n=1051. [4]) Surgeons' ranking: radiographic knee OA severity, total range 1- 1051(1 most severe). When nothing else is stated, means [and medians] ± SD are reported.*

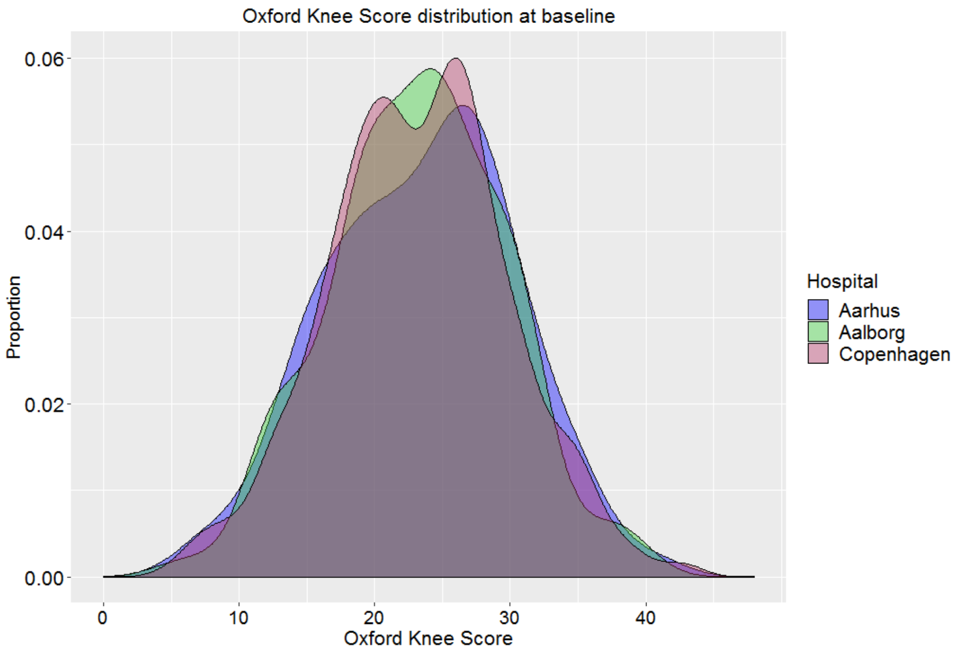


Figure 2. *Distribution of preoperative Oxford Knee Score per hospital (Kernel density plot).*

PROMs and motivations for surgery

The primary PROM, OKS did not differ among patients in the three hospitals at baseline (mean 23.3, p=0.884), nor when adjusted for age, sex and BMI (table 2, figure 2). The same was true for flexion, analgesics use and the global knee anchor. Extension deficits were more common in Aalborg (62 vs. 45-46%, p=0.007). Males scored 2.8 OKS points higher than females (95% CI, 2.1-3.5, p<0.001) and they reported less frequent use of analgesics but did not rate their knee better on the global knee anchor. OKS scores were significantly lower in obese patients (-2.6; 95% CI, -1.9-(-3.3), p<0.001) and in smokers (-1.5; 95% CI, -0.4-(-2.7), p=0.009). No hospital differences were observed in patients' motivations for surgery (p≥0.127), yet large variances were observed between implant groups and sexes (table 3).

Table 3. Patients' motivations for surgery

	Total	TKA	MUKA	p	Females	Males	p
n (%)	1452	1059 (73)	336 (23)		793 (55)	659 (45)	
Pain	1174 (82)	843 (81)	288 (88)	**0.009**	649 (84)	525 (81)	0.146
Mobility (walking, stairclimbing, bicycling)	784 (55)	592 (57)	159 (48)	**0.008**	450 (58)	334 (51)	**0.012**
Sports, exercise & physical activity	580 (41)	414 (40)	140 (43)	0.412	298 (39)	282 (43)	0.070
Knee motion and stability	521 (37)	392 (38)	107 (33)	0.103	290 (38)	231 (36)	0.485
The surgeons' advice	516 (36)	376 (36)	117 (36)	0.897	276 (36)	240 (37)	0.661
Hobbies (leisure time, travelling)	474 (33)	347 (33)	113 (34)	0.794	236 (31)	238 (37)	**0.017**
Mood and energy	471 (33)	337 (32)	115 (35)	0.429	273 (35)	198 (31)	0.062
Tired of taking medication	404 (28)	281 (27)	106 (32)	0.079	246 (32)	158 (24)	**0.002**
Duties (housework, gardening, helping others)	440 (31)	330 (32)	98 (30)	0.552	221 (29)	219 (34)	**0.042**
Independency and selfcare	390 (27)	310 (30)	72 (22)	**0.006**	239 (31)	151 (23)	**0.002**
Work	242 (17)	156 (15)	75 (23)	**0.001**	117 (15)	125 (19)	**0.047**
Being with family and friends	176 (12)	141 (14)	32 (10)	0.084	111 (14)	65 (10)	**0.016**
Marital (incl. sexual) life	66 (4.6)	46 (4.4)	17 (5.2)	0.681	27 (3.5)	39 (6.0)	**0.034**
Missing answer	16 (1.1)	11 (1.0)	4 (1.2)	-	13 (1.6)	3 (0.5)	-

Answers to the question,"Which factors or problems made you choose surgery? Pick up to five motivations". Options are listed by overall frequency.

Radiographic classification of OA

When measured by K-L classification and surgeons' ranking, knee OA severity was unevenly distributed among hospitals (table 2). Mild degrees of radiographic knee OA (K-L/Ahlbäck < 2) were less common in Aarhus patients (p=0.013/0.015). There was no association between radiographic classifications and urbanization, except in the group of 62 patients reporting to live in the countryside: here, specifically, low Ahlbäck scores (0-1) were less common (21%) than in patients in "small town/village" (38%) or "city or suburb" (32%) (p=0.040). With all three radiographic evaluation methods, males had more advanced OA than females. On a hospital level, however, this was only true in Copenhagen (p=0.006-0.031), partly true in Aarhus (p=0.009-0.094) and not true in Aalborg (p=0.884-0.935).

Primary KA incidence and implant choice

In the Capital Region, the incidence of primary KA in ages 60-79 years exceeds that of Central Denmark Region by 28% and that of North Denmark Region by 15% (table 4).

Table 4. Regional incidence of primary knee arthroplasty per region in year 2017

	Central Denmark	North Denmark	Capital Region	p
Regional revision rate	Low	Low	High	
SPARK example (hospital)	Aarhus	Aalborg (Farsø)	Copenhagen (Gentofte)	
Incidence per 100.000 inhabitants				
All patients aged > 40 y.	235	276	285	<0.001
Subgroup: ages 60-79 y.	416	463	534	<0.001

In the SPARK sample, surgery was carried out by 22 surgeons: four in Aarhus, six in Aalborg and twelve in Copenhagen. Apart from five surgeons in training programs, who were responsible for less than six operations each (and evenly distributed among hospitals), all surgeons were exclusively occupied with joint replacement surgery, and with few exceptions, the staffs had been stable in the preceding years. Overall, MUKA patients were 1.7 years younger (95% CI, 0.6-2.8) than TKA patients, more likely to be male (52 vs. 44%, p=0.011), had a lower BMI (-1.1 kg/m^2; 95% CI, -0.5-(-1.7), p<0.001) and scored 1.4 points higher (95% CI, 0.6-2.2, p<0.001) in OKS and 3.9 points higher (95% CI, 1.3-6.4, p=0.003) in general health (EQ-VAS). However, subgroup analyses demonstrated that in Aarhus, where 40% of patients were offered MUKA, patients in the two implant groups did not differ in age, sex or BMI (table 5). Instead, the group differences were more pronounced in self-reported health (EQ-VAS), global knee anchor and patient-reported knee range of motion, e.g. 0.5 points higher flexion in MUKA patients corresponding to approximately 5-10 degrees [29].

Table 5. Characteristics of TKA vs. MUKA patients in hospitals grouped by frequency of MUKA use.

	Aarhus			Aalborg/Copenhagen		
Sample	**TKA**	**MUKA**	**p**	**TKA**	**MUKA**	**P**
n (% of TKA+MUKA group)	164 (56)	129 (44)		895 (81)	207 (19)	
Demographics						
Age	67.0 ± 9.4	67.4 ± 8.5	0.720	68.9 ± 8.9	66.7 ± 9.0	**0.001**
Male sex (%)	76 (46)	59 (46)	1.000	393 (44)	117 (57)	**0.001**
BMI (kg/m^2)	28.8 ± 4.7	28.2 ± 4.6	0.261	29.2 ± 5.3	28.0 ± 4.4	**0.002**
UCLA	4.6 [4] ± 1.9	5.0 [5] ± 1.9	**0.042**	4.6 [4] ± 1.8	5.1 [5] ± 1.9	**0.001**
Patient-reported outcomes						
OKS	22.8 ± 7.5	24.1 ± 6.4	0.112	23.0 ± 6.6	24.4 ± 6.5	**0.005**
Global knee anchor	24 ± 17	29 ± 17	**0.010**	28 ± 17	31 ± 18	0.062
EQ-VAS	59 ± 22	65 ± 20	**0.016**	61 ± 22	64 ± 21	**0.046**
EQ-5D-5L Index	0.58 ± 0.16	0.61 ± 0.14	0.065	0.58 ± 0.15	0.61 ± 0.12	**0.008**
Flexion	4.6 [5] ± 1.2	5.1 [5] ± 1.1	**<0.001**	4.8 [5] ± 1.2	5.1 [5] ± 1.0	**0.008**
Deficit	63 (38)	24 (19)	**<0.001**	268 (30)	47 (23)	**0.045**
Extension	3.1 [3] ± 1.1	3.7 [4] ± 0.9	**0.009**	3.4 [3] ± 0.9	3.7 [4] ± 0.9	**0.007**
Deficit	105 (64)	60 (47)	**0.004**	529 (59)	84 (41)	**<0.001**
Radiographic assessments of knee OA						
K-L grade ≥ 2 (%)	99 (98)	103 (98)	1.000	618 (94)	167 (90)	**0.087**
K-L grade ≥ 3 (%)	81 (80)	97 (92)	**0.019**	527 (80)	146 (79)	0.735
Ahlbäck score ≥ 2 (%)	74 (74)	80 (76)	0.841	444 (68)	106 (57)	**0.010**
Ahlbäck score ≥ 3 (%)	41 (41)	20 (19)	**0.001**	220 (33)	24 (13)	**<0.001**
Surgeons' ranking (mean [IQR]))	315 [118; 654]	486 [279; 463]	**0.003**	518 [253; 809]	710 [505; 887]	**<0.001**

Test results refer to comparisons within the hospital group. For abbreviations, see table 2.

Discussion

The SPARK study was initiated to explore the clinical reality behind large differences in revision rates after primary knee arthroplasty (KA) among Danish regions. In this part of the study, patient selection in three high-volume hospitals with either high (Copenhagen) or low (Aarhus and Aalborg) revision rates was compared through preoperative PROMs, patient demographics, lifestyle, motivations for surgery, severity of radiographic OA, and implant choice at primary KA in 1452 patients. Primary KA incidence was retrieved on a regional level. We found that preoperative PROM scores, primarily OKS, were the same in all hospitals, with the exception of patient-reported knee extension deficit of which the clinical relevance is uncertain. Incidence of primary KA was up to 28% higher in the high-revision region and radiographic OA was more advanced in one low-revision hospital (Aarhus) though age was 1.4-2.0 years higher in Copenhagen and BMI was 1.5-1.7 points higher in Aalborg [26]. In Aarhus, where MUKA was offered to 40% of all patients, the choice between TKA and MUKA did not seem to be influenced by age, sex or BMI. It would appear from the summarized findings (table 6), that no generally accepted important differences in preoperative patient characteristics seem to explain consistent differences in revision rates among the three centres studied. A very high proportion of patients from a low-revision hospital were treated with unicompartmental implants, which is contrary to expectation. The mean age of patients in the high-revision hospital is higher than in the low-revision hospitals, which is contrary to expectation. The mean BMI and the proportion of male patients in one low-revision hospital are higher than in the high-revision hospital, which is also contrary to expectation.

Table 6. Summary of main findings.

Predictor	High- vs. low-revision hospitals
PROM	No difference
Incidence of knee arthroplasty	High incidence in high-revision hospital area
Radiographic OA	More severe OA in one of two low-revision hospitals
Age	Higher in high-revision hospital
BMI	Higher in one of two low-revision hospitals
Sex	More males in one of two low-revision hospitals
Implant type	More unicompartmental implants in one low-revision hospital

The dataset serves as a description of a recent knee arthroplasty cohort distributed at three hospitals, where one has consistently had a very high revision rate, and two have consistently had low revision rates. None of the hospitals have had major staff changes, and none of the surgeons are aware of changed practice, so we consider the description of baseline patient characteristics to be a fair representation of hospital routines. We consider it an important strength that these data concern patients who were treated in routine clinical settings. Simultaneous responses to several PROMs from nine out of ten participants in three geographical regions along with radiographic classifications of knee OA offers a unique reference set for later comparisons [8]. When interpreting the above results, it should be kept in mind that when several parameters are explored, some significant differences will be found without necessarily representing a reproducible or, for that matter, clinically important difference. Due to the observational nature of the study, conclusions about causality should not be made. Not all potential candidates for the study were included, and this could create bias. To make inclusion feasible, we did not require information about patients who were not invited or who declined participation and why so. It can be argued that electronical collection of PROMs might threaten the patient representativeness. Danish citizens are among the most IT-literate in Europe (two out of three Danish citizens above 65 years of age use the internet every day) [54] and previous studies have shown that knee OA patients prefer electronical questionnaires over paper [15]. The sample patients were close to the target population with respects to demography. Nonetheless, when participants without email were allowed in the study (final six months), they were eight years older than the other participants. Thus, some of the eldest and perhaps the least resourceful patients may have been left out as a consequence of the study design. In Aalborg, particularly, the relatively poor inclusion rate may have affected the generalizability of results. Obtaining information about comorbidity or socioeconomic factors might have revealed some inclusion bias or hospital differences [44]. As a proxy of socioeconomic factors, 10% of men and 8% of women in age group 65-74 years reported daily smoking in this sample; a proportion which was lower than the 17, respectively 14%, reported in the National Health Profile 2018 [21]; yet smoking is also known to be associated to lower risk of OA [25].

The urban-rural variations in radiographic classifications were very sparse in this study. This may partly be explained by the relatively small geographical distances in Denmark: almost all Danes live less than 1.5 hours' drive from a KA centre [9, 39], but also the small number of patients (n=62, 4.3%) living in the countryside in this sample must be kept in mind. As knee radiographs in Aalborg and Copenhagen are taken with weight bearing on both legs, the joint space width may have been exaggerated when compared to Aarhus. This may explain some of the hospital variance in radiographic knee OA severity, however, both K-L and Ahlbäck classifications rely on other morphological features than joint space narrowing, e.g. osteophytes and bone attrition[1, 23]. Nonetheless, the finding that there were only few patients with mild degrees of OA in Aarhus Hospital, which is situated in a region with a KA incidence that was 18-22% lower than in the Capital Region (Copenhagen), suggests that not all patients operated in Copenhagen would have been offered surgery if they had been living in Aarhus (or Aalborg) instead. Utilization of primary KA is known to vary across economies and countries, e.g. with a factor-10 among the Organization for Economic Co-operation and Development (OECD) countries alone [34, 35]. Apart from economy, tradition and accessibility to orthopaedic surgeons, also the general uncertainty of the optimal timing of primary KA plays a role, probably also among Scandinavian countries where a factor-2 difference is observed (Finland highest, Norway lowest) [33]. Even within countries, large variations are observed: the Finnish regions vary by a factor-1.6 in KA utilization[27], the 16 German federal states differ with a factor-1.8 [39] and the 16 Spanish regions with a factor-27 [13]. In this light, the Danish variation in incidence (factor-1.3) is small. If there is a higher threshold for primary KA surgery in Central (and North) Denmark Region, it is not necessarily explained by the acts of knee surgeons alone [43]; any possible regional differences in approach to knee OA treatment among patients, physicians, physiotherapists, and other caretakers would influence the number of patients admitted to hospital for surgical evaluation to begin with. Thus, the ideal comparison of regional patient selection for KA should also include studies on knee OA patients who were treated outside hospitals and by other means than surgery.

Conclusion

In this prospective cohort study of 1452 knee arthroplasty patients, baseline PROMs (primarily Oxford Knee Score) and patients’ motivations for surgery did not vary among high-volume hospitals with very different revision rates but patient demographics, BMI and radiographic knee OA severity did; one low-revision hospital operated fewer patients with mild radiographic knee osteoarthritis. Compared to the high-revision hospital, the two low-revision hospitals were situated in areas with lower incidence of primary knee arthroplasty. Ideally, future studies on surgical thresholds should also include the knee osteoarthritis patients who were not offered surgical treatment. In conclusion, differences in patient selection for primary knee arthroplasty seem to exist among Danish high-volume hospitals, but based on this study, none of the baseline differences are likely to explain the large and persistent differences in revision rates; most of the observed differences are paradox to well-known revision risk factors of revision. Follow-up studies on the SPARK population will clarify whether revision rate differences are reflected in PROM results after surgery and to which degree results are associated to surgical timing.

Acknowledgements

We thank the following for their contributions: Birthe Schøning, Heidi Ladefoged Poulsen, Camilla Grube, Naima Elsayed, Lars Jansen, Ulla Hornum, Gitte Broholm, Patricia Svare, Tina Isaksen, Gitte Raae Høybye, Anette Enemark-Larsen, Svend Erik Østgaard, Lasse Enkebølle Rasmussen, Thomas Bjerno, Lars Peter Møller, Snorre Stephensen, Thomas Lind, Henrik Schrøder, Søren Rytter, Claus

Fink Jepsen, Andreas Kappel, Procordo ApS, medical students at Gentofte Hospital, The Health Research Fund of the Capital Region of Denmark, and all participating patients.

References

1. Ahlbäck S, Rydberg J. Röntgenologisk klassifikation och undersökningsteknik vid gonartros. *Lakartidningen*. 1980;77:2091–2093, 2096.
2. Baker PN, Petheram T, Avery PJ, Gregg PJ, Deehan DJ. Revision for unexplained pain following unicompartmental and total knee replacement. *J Bone Jt. Surg Am*. 2012;94:e126.
3. Brittain R, Young E, Mccormack V, Swanson M. 16th Annual Report. 2019.
4. Buckland-Wright C. Which radiographic techniques should we use for research and clinical practice? *Best Pr. Res Clin Rheumatol*. 2006;20:39–55.
5. Conaghan PG, Emerton M, Tennant A. Internal construct validity of the Oxford Knee Scale: evidence from Rasch measurement. *Arthritis Rheum*. 2007;57:1363–1367.
6. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Jt. Surg Br*. 1998;80:63–69.
7. Dawson J, Fitzpatrick R, Murray D, Carr A. A response to issues raised in a recent paper concerning the Oxford knee score. *Knee*. 2006;13:66–68.
8. Dieppe P, Lim K, Lohmander S. Who should have knee joint replacement surgery for osteoarthritis? *Int. J. Rheum. Dis.* 2011;14:175–180.
9. Dowsey MM, Petterwood J, Lisik JP, Gunn J, Choong PFM. Prospective analysis of rural-urban differences in demographic patterns and outcomes following total joint replacement. *Aust. J. Rural Health*. 2014;22:241–248.
10. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. *J Bone Jt. Surg Br*. 2001;83:339–344.
11. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4:e296.
12. Fortin PR, Penrod JR, Clarke AE, St-Pierre Y, Joseph L, Belisle P, Liang MH, Ferland D, Phillips CB, Mahomed N, Tanzer M, Sledge C, Fossel AH, Katz JN. Timing of total joint replacement affects clinical outcomes among patients with osteoarthritis of the hip or knee. *Arthritis Rheum*. 2002;46:3327–3330.
13. Gómez-Barrena E, Padilla-Eguiluz NG, García-Rey E, Cordero-Ampuero J, García-Cimbrelo E. Factors influencing regional variability in the rate of total knee arthroplasty. *Knee*. 2014;21:236–241.
14. Goodfellow JW, O’Connor JJ, Murray DW. A critique of revision rate as an outcome measure: re-interpretation of knee joint registry data. *J Bone Jt. Surg Br*. 2010;92:1628–1631.
15. Gudbergsen H, Bartels EM, Krusager P, Waehrens EE, Christensen R, Danneskiold-Samsoe B, Bliddal H. Test-retest of computerized health status questionnaires frequently used in the monitoring of knee osteoarthritis: a randomized crossover trial. *BMC Musculoskelet Disord*. 2011;12:190.
16. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46:1417–1432.
17. Harris K, Dawson J, Doll H, Field RE, Murray DW, Fitzpatrick R, Jenkinson C, Price AJ, Beard DJ. Can pain and function be distinguished in the Oxford Knee Score in a meaningful way? An exploratory and confirmatory factor analysis. *Qual. Life Res.* 2013;22:2561–2568.
18. van Hout B et al. JMF. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health 2012 Jul-Aug;15(5):708-15.
19. Ingelsrud LH, Roos EM, Terluin B, Gromov K, Husted H, Troelsen A. Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. *Acta Orthop*. 2018;89:541–547.

20. Jasper LL, Jones CA, Mollins J, Pohar SL, Beaupre LA. Risk factors for revision of total knee arthroplasty: a scoping review. *BMC Musculoskelet Disord*. 2016;17:182.
21. Jensen HAR, Davidsen M, Ekholm O, Chirstensen AI. *Danskernes Sundhed 2017*. 2018.
22. Jin X, Al Sayah F, Ohinmaa A, Marshall DA, Smith C, Johnson JA. The EQ-5D-5L Is Superior to the -3L Version in Measuring Health-related Quality of Life in Patients Awaiting THA or TKA. *Clin Orthop Relat Res*. 2019;477:1632–1644.
23. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. *Ann. Rheum. Dis.* 1957;16:494–502.
24. Keurentjes JC, Fiocco M, So-Osman C, Onstenk R, Koopman-Van Gemert AW, Poll RG, Kroon HM, Vliet Vlieland TP, Nelissen RG. Patients with severe radiographic osteoarthritis have a better prognosis in physical functioning after hip and knee replacement: a cohort-study. *PLoS One*. 2013;8:e59500.
25. Kong L, Wang L, Meng F, Cao J, Shen Y. Association between smoking and risk of knee osteoarthritis: a systematic review and meta-analysis. *Osteoarthr. Cartil.* 2017;25:809–816.
26. Laxafoss E, Jacobsen S, Gosvig KK, Sonne-Holm S. Case definitions of knee osteoarthritis in 4,151 unselected subjects: relevance for epidemiological studies. *Skeletal Radiol.* 2010;39:859–866.
27. Makela KT, Peltola M, Sund R, Malmivaara A, Hakkinen U, Remes V. Regional and hospital variance in performance of total hip and knee replacements: a national population-based study. *Ann Med*. 2011;43 Suppl 1:S31-8.
28. Mongelard K, Morup-Petersen A, Roemer L, Kristensen KB, Odgaard A. Heuristic ranking delivers more detail than than ordinal grading of knee osteoarthritis radiographs. *Submiss. Process*.
29. Morup-Petersen A, Holm PM, Holm CE, Klausen TW, Skou ST, Krogsgaard MR, Laursen MB, Odgaard A. Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale. *J Arthroplast.* 2018;33:2875–2883.e3.
30. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. *J Bone Jt. Surg Br*. 2007;89:1010–1014.
31. Murray DW, Liddle AD, Dodd CA, Pandit H. Unicompartmental knee arthroplasty: is the glass half full or half empty? *Bone Jt. J*. 2015;97–b:3–8.
32. Mørup-Petersen A, Holm PM, Holm CE, Klausen TW, Skou ST, Krogsgaard MR, Laursen MB, Odgaard A. Corrigendum to “Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale” [The Journal of Arthroplasty 33 (2018) 2875-2883]. *J. Arthroplasty*. 2019.
33. Niemelainen MJ, Makela KT, Robertsson O, A WD, Furnes O, Fenstad AM, Pedersen AB, Schroder HM, Huhtala H, Eskelinen A. Different incidences of knee arthroplasty in the Nordic countries. *Acta Orthop*. 2017;88:173–178.
34. Pabinger C, Lothaller H, Geissler A. Utilization rates of knee-arthroplasty in OECD countries. *Osteoarthr. Cartil.* 2015;23:1664–1673.
35. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, Carr A, Beard D. Knee replacement. *Lancet*. 2018;392:1672–1682.
36. Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. *Acta Orthop.* 2019;90:297.
37. RCoreTeam. RStudio version 1.1.463. https://www.R-project.org/. R: A Language and Environment for Statistical Computing.
38. Robertsson O, Ranstam J. No bias of ignored bilaterality when analysing the revision risk of knee prostheses: Analysis of a population based sample of 44,590 patients with 55,298 knee prostheses from the national Swedish Knee Arthroplasty Register. *BMC Musculoskelet. Disord.* 2003;4:1.
39. Schäfer T, Pritzkuleit R, Jeszenszky C, Malzahn J, Maier W, Günther KP, Niethard F. Trends and geographical variation of primary hip and knee joint replacement in Germany. *Osteoarthr. Cartil.* 2013;21:279–288.

40. Sorensen J, Davidsen M, Gudex C, Pedersen KM, Bronnum-Hansen H. Danish EQ-5D population norms. *Scand J Public Heal.* 2009;37:467–474.
41. Sundhedsdatastyrelsen. The Danish National Health Register. (Landspatientregistret) https://www.esundhed.dk/Registre/Landspatientsregisteret/Operationer. 2019.
42. Terwee CB, Bouwmeester W, van Elsland SL, de Vet HC, Dekker J. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthr. Cartil.* 2011;19:620–633.
43. Troelsen A, Schroder H, Husted H. Opinions among Danish knee surgeons about indications to perform total knee replacement showed considerable variation. *Dan Med J*. 2012;59:A4490.
44. Whitehouse SL, Blom AW, Taylor AH, Pattison GT, Bannister GC. The Oxford Knee Score; problems and pitfalls. *Knee*. 2005;12:287–291.
45. Wilson I, Bohm E, Lubbeke A, Lyman S, Overgaard S, Rolfson O, A WD, Wilkinson M, Dunbar M. Orthopaedic registries with patient-reported outcome measures. *EFORT Open Rev*. 2019;4:357–367.
46. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. *J Arthroplast.* 1998;13:890–895.
47. The Danish Knee Arthoplasty Register. Annual Report 2019.
48. The Danish Knee Arthoplasty Register. Annual Report 2015.
49. The Danish Knee Arthoplasty Register. Annual Report 2010.
50. The Danish Hip Arthoplasty Register. Annual Report 2018.
51. The Danish Knee Arthoplasty Register. Annual Report 2018.
52. https://euroqol.org/wp-content/uploads/2019/09/EQ-5D-5L-English-User-Guide_version-3.0-Sept-2019-secured.pdf.
53. *NOMESCO Classification of Surgical Procedures, version 1.15*. Nordic Centre for Classifications in Health Care; 2010.
54. Statistics Denmark: Use of IT in the Danish population (IT-anvendelse i befolkningen). 2017:8–9,37.
55. *Swedish Knee Arthroplasty Register Annual Report 2018*. 2018.
56. Danish National Health Register (Statistikbanken). 2018.

Study II: SPARK Post

Revision rates differ widely, yet patient-reported outcomes after primary knee replacement are the same: 1-year results from the prospective multicenter cohort study, SPARK

Anne Mørup-Petersen[1], Mogens Laursen[2], Frank Madsen[3], Michael R. Krogsgaard[4], Matilde Winther-Jensen[5], Anders Odgaard[1]

[1)] Department of Orthopedic Surgery, Copenhagen University Hospital, Herlev and Gentofte
[2)] Department of Orthopedic Surgery, Aalborg University Hospital, Aalborg and Farsoe
[3)] Department of Orthopedic Surgery, Aarhus University Hospital
[4)] Section for Sports Traumatology M51, Bispebjerg and Frederiksberg Hospital, Copenhagen
[5)] Center for Clinical Research and Prevention, Department of Clinical Epidemiology, Bispebjerg and Frederiksberg Hospital, Copenhagen

Corresponding author: Anne Mørup-Petersen, email: morup.anne@gmail.com

Abstract

Background/Rationale

Knee arthroplasty (KA) revision rates differ widely among regions and high-volume hospitals. We explore whether revision rates are associated to surgical results as measured by patient-reported outcome measures (PROMs).

Questions / Purposes

1) Do 1-year PROMs after primary KA vary among three Danish hospitals known to differ in revision rates? 2) Do changes in PROM scores during the first postoperative year differ among hospitals? 3) Can patients with the same degree of knee osteoarthritis (OA) expect the same outcome across hospitals?

Methods

This prospective cohort study followed 1452 primary KA patients in three high-volume hospitals from pre- to 1-year postoperatively (2016-18) with Oxford Knee Score (OKS), Forgotten Joint Score, EQ-5D-5L, Copenhagen Knee ROM (range of motion) Scale, UCLA Activity Scale, analgesics use and willingness to repeat surgery. Response rate was 90%.

Results

1) 1-year OKS was no different among hospitals (39.0 ±7.4, p=0.096), nor when adjusted for age, sex, BMI, OKS and OA grading. In one low-revision hospital, 19% of patients had an OKS change <8 points compared to 13-14% in the other two hospitals (overall p=0.051). 2) PROMs changed similarly over time from 3 months. 3) Except for Kellgren-Lawrence grade 4, patients with the same OA grading or baseline OKS had similar results across hospitals (OKS/ willingness to repeat (92% positive), p>0.087).

Conclusions

Patient-reported results after primary knee arthroplasty were the same across hospitals with different revision rate levels. Thus, revision rate variations may be caused by different revision thresholds rather than variations in quality of primary surgery.

Introduction

When regions or hospitals differ in revision rates after primary knee arthroplasty (KA), it leads to assumptions of differences in the quality of surgery. This is the case for variation both among and within countries. As an example, some regions of Denmark have consistently had much higher revision rates than others. In short, revision rates decrease as the distance from Copenhagen increases (e.g. from 5 to 1% at 2 years in 2015) [19, 31]. We initiated a prospective cohort study to investigate whether differences in revision rates could be attributed to differences in baseline patient characteristics or to differences in the quality of surgery, defined as patients' subjective improvement following operation. Our first publication from the SPARK study ("Variation in patient Satisfaction, Patient-reported outcome measures, radiographic signs of Arthritis, and Revision rates in Knee arthroplasty patients in three Danish regions") reporting baseline data on 1452 patients revealed how patient demography, incidence of knee replacement, implant choice and radiographic classification of knee osteoarthritis varied among high-volume hospitals in three regions, whereas preoperative patient-reported outcome measures (PROMs) did not [19]. The study was unable to identify baseline differences that could reasonably explain the persistent differences in revision rates. The current study presents the one-year follow-up PROM results from the same patient cohort.

The general focus of this article is to explore whether the observed variations in revision rates correspond to differences in the quality of primary KA surgery. We approach the quality question by using patient-reported outcomes, and the specific research questions are as follows: 1) Do 1-year PROM results (primarily Oxford Knee Score) after primary KA vary among hospitals known to differ in revision rates? 2) Do changes in PROM scores during the first postoperative year differ among hospitals? 3) Can patients with the same degree of knee osteoarthritis (OA) expect the same outcome across hospitals? All research questions were explored from the null-hypothesis perspective that there was no difference among hospitals.

Methods

Study Design and Setting

From September 2016 and 15-16 months forward, patients scheduled for primary knee arthroplasty of any type were invited to participate in this prospective cohort study in three high-volume centers across Denmark. The three hospitals had revision rates that were representative of their particular region. Central and North Denmark Region (represented by Aarhus and Aalborg Hospitals, respectively), had relatively low revision rates (e.g. 2.2 and 1.0% per 2 y., 2015) whereas the Capital Region (Copenhagen) had a high revision rate (5.0%). Patients were followed with PROM-sets from preoperative to one year postoperative.

Participants

All patients scheduled for any type of primary knee replacement were eligible if they were accessible by email. In the last six months of the inclusion period, also patients without an email address were included and received questionnaires by letter. Only the 1452 patients who provided PROM-data prior to surgery were enrolled in this study (89% of those initially enrolled and contacted, 56% of all operated). The inclusion process, exclusion criteria, inclusion rate and representativeness of the sample are all documented in the aforementioned publication[19]. The reporting of the study followed STROBE guidelines[9, 30].

Data Sources

Questionnaires were sent out by emails with unique links before surgery and 6 weeks and 3, 6 and 12 months after surgery. As 53 patients had an operation in both knees during the study period and a substantial number had bilateral knee trouble, we sought to avoid confusion by clearly addressing "right" or "left knee" along with the current follow-up time in each email. In case of no reply, two reminders were sent 2-4 days apart. Additionally, at the 1-year follow-up, patients who failed to answer electronically were sent a printed questionnaire along with a pre-paid envelope to their home address. Information regarding surgical procedures was continuously collected from the routine registrations made by surgeons immediately after surgery. Any missing or erroneous surgical data were meticulously sought and corrected through patient charts and, whenever possible, follow-up questionnaire sequences were resumed on time. Patients who went through revision surgery (removal, exchange or addition of any implant material) were excluded on the day of revision. In these cases, a new sequence of questionnaires was started (not reported in the present study). Minor surgery or incidents, e.g. wound debridement or manipulation under anesthesia did not prompt exclusion. All revisions were attributed to the primary hospital, no matter which hospital (public or private) performed the revision.

Radiographic classification of knee osteoarthritis

Radiographic severity of knee OA in postero-anterior weight bearing radiographs was graded in 1051 of patients (only TKA and MUKA, when available) according to Kellgren-Lawrence (K-L) classification (0-4, where 4 is most severe) and Ahlbäck score (0-5, 5 most severe) [1, 15]. Moreover, 13 experienced knee OA surgeons ranked the radiographs based on clinical experience and free of traditional classification systems from the mildest (no. 1051) to the most severe case (no. 1) [17]. For details, we refer to our previous publication [19].

Outcome Measures

Both absolute and change scores of the primary outcome, Oxford Knee Score (OKS, 0-48 version) were recorded [2, 5, 8, 12] and the proportion of patients reaching the Minimal Important Change (MIC) of 8 points for "important improvement" was compared across hospitals [14]. Copenhagen Knee ROM Scale (CKRS) assessed patient-reported range of motion (ROM) with flexion ranging from 0-6 (6 max) and extension from 0-5 (5 max) [18, 21]. CKRS also estimated the proportion of patients with flexion deficits <100° (sensitivity 95%/ sensitivity 81%) or extension deficits >10° (78/70%). OKS and CKRS were preceded by a global knee anchor question, "How is your knee at the moment?" with answers on a Visual Analogue Scale (VAS) from 0-100 (100 best) [19], the generic EQ-5D-5L, EQ-VAS[13], and a question concerning frequency of use of any type of analgesics against knee pain with answer options; "More than once daily", "Once daily", "More than once weekly", "More than once monthly" or "Rarely or never".

Beside these questions, that were included in every pre- and postoperative PROM set, further questions or PROMs were added at varying time points. The Forgotten Joint Score (FJS) [3, 29] and UCLA Activity Scale (UCLA) were used in all sets from 3 months postoperatively, the latter also preoperatively. At 6 months, patients were asked whether they had received physiotherapeutic assistance in rehabilitation after hospital discharge. At 3, 6 and 12 months, patients were asked about overall satisfaction, “How satisfied are you with the overall experience of the operation and its result?” and given five answer options (Likert boxes, one neutral). These answers were expected to be influenced by experiences related to hospital service, kindness of caretakers, etc. [4], so, to achieve a more specific satisfaction measure, we asked for “willingness to repeat” at the 1-year follow-up: “Suppose you could turn back time: now that you know the result, would you still choose to have a knee replacement?” Five answer options were given: “Yes, certainly” or “yes, probably”, “I don’t’ know”, “No, probably not” and “No, absolutely not”. In each round, patients had an opportunity to write free text or to contact the first author.

Implant types, aftercare, follow-up routine and bias

Due to the observational nature of the study, we did not interfere with normal hospital routines regarding e.g. analgesics, aftercare or choice of implants. Inevitably, this has introduced some bias as these confounding factors were inseparable from the hospital variable. Each hospital used a different selection of cemented, uncemented and hybrid implants that had been on the market for at least ten years and had proven good survivorship in registers [32]. The predominant implant systems were NexGen^{TM} (Zimmer Biomet), PFC^{TM} Sigma (DePuy Synthes), Triathlon^{TM} (Stryker), Oxford^{TM} Mobile Bearing and ZUK^{TM} (Zimmer Biomet) and Avon^{TM} (Stryker).

In all three hospitals, tranexamic acid, glucocorticoids and prophylactic antibiotics (dicloxacillin in Copenhagen, cefuroxime in Aarhus and Aalborg) was administered preoperatively. Periarticular infiltration of local anesthetics was given intraoperatively. The postoperative oral analgesics of choice were paracetamol, non-steroid anti-inflammatory drugs (NSAID) and opioid for up to four weeks postoperatively. In 2017, the mean length of stay for Total KA (TKA) patients in Aarhus, Aalborg and Copenhagen was 2.4, 1.4 and 2.2 nights, respectively. For medial unicompartmental KA (MUKA) patients, numbers were 0.6, 1.3 and 0.7 nights, respectively [31].

Preparatory training with physiotherapists (crutch walking, stairclimbing etc.) was part of the multidisciplinary patient seminar taking place approximately two weeks before surgery in all three centers. Additionally, patients in Copenhagen and Aalborg were trained by a physiotherapist during the hospital stay. After discharge, Copenhagen patients were routinely offered free of charge supervised physiotherapy (typically more than ten sessions). By contrast, in Aarhus and Aalborg, a physiotherapist screened patients 2-6 weeks after discharge to identify those with poor progress and offered them physiotherapy if needed. If after 6-8 weeks, results were still not satisfactory, patients were referred to the surgeon for a follow-up visit. All Copenhagen patients met the surgeon for a routine clinical follow-up examination after three months, while in Aalborg and Aarhus, only those with abnormal findings on a 1-year postoperative knee radiograph (TKA patients only) were seen by a surgeon.

Demographics, Description of Study Population

The main characteristics of participating patients were outlined in table 1. For in-depth characteristics and sample representativeness, we refer to the publication of baseline data [19].

Table 1. Baseline characteristics of participants

		Value	%
Patients		1452	100
Age (years) (mean [median] ± SD)		68.0 [69] ± 9.3	
Male sex (n)		659	45
Body Mass Index (kg/m2) (mean [median] ± SD)		28.9 [28.3] ± 5.0	
Hospital		1452	100
Copenhagen (Gentofte, Capital Region)		929	64
Aarhus (Central Denmark Region)		321	22
Aalborg (Farsoe, North Denmark Region)		202	14
Radiographic severity of knee osteoarthritis (n)		1051	(100)
Kellgren-Lawrence classification	≥ 2	987	94
	≥ 3	851	81
Ahlbäck classification	≥ 2	704	67
	≥ 3	305	29

For complete baseline results per hospital, we refer to the previous publication [19]. *TKA: Total knee arthroplasty, MUKA/LUKA: medial/lateral unicompartmental knee arthroplasty, PFA: Patellofemoral knee arthroplasty.*

Response rates/Lost to follow-up

During the first postoperative year, three patients decided to leave the study, seven died and nine were not contacted for follow-up due to errors, e.g. wrong laterality or change of email address (table 2). 1414 patients (97.4%) responded at least once postoperatively. At 1 year, 1307 patients (90%) replied, and non-responders were equally distributed among hospitals (p=0.424).

Table 2. Questionnaire response rates

	Baseline	6 weeks	3 months	6 months	1 year	Any postop.
Responding patients (n)	1452[a]	1147[b]	1237	1241	1307	1414
Available patients (n) (revised/dead)	1452 [a] (0/0)	1296[b] (9/1)	1435 (15/2)	1433 (17/2)	1417 (28/7)	1443 (9/1)
Response rate						
per available patients (%)	(100) [a]	89	86	87	92	98
per 1452 baseline responders (%)	100 [a]	79	89	86	90	97
Days from surgery (mean [median] ± SD)	-29 [-18] ± 32	39 [38] ± 7	87 [84] ± 14	179 [176] ± 14	368 [359] ± 27	-

[a]) For response rates at inclusion, we refer to publication of baseline data [19]. *[b]) The 6-week questionnaire was delayed and thus not sent to the first 146 included patients. Non-responders at 1-year were further contacted by letter.*

Revision was conducted in 28 patients during the first postoperative year (1.9%); 2 (0.6%) in Aarhus, 4 (2.0%) in Aalborg and 22 (2.4%) in Copenhagen (p=0.141) (table 3). These patients were not the object of study and will be accounted for in a later publication when 2-year follow-up results are available, yet some details regarding time of revision, revision indication and latest postoperative OKS score are provided (table 3). Revision for other causes than deep infection was done in 15 cases: 1 in Aarhus, 3 in Aalborg and 11 in Copenhagen (p=0.386).

.

Table 3. Characteristics of patients who were revised during the first postoperative year

Patient group	**n** (%)	**Male sex** (n (%))	**Age** years (mean ± SD)	**BMI** kg/m^2 (mean ± SD)	**Implant type** (TKA/MUKA/other)
No revision	1424 (98.1)	642 (45)	68.0 ± 9	28.9 ± 5.0	1039/328/57
Revision	28 (1.9)	17 (61)	66.4 ± 10	26.9 ± 3.6	20/8/0
P-value	-	0.125	0.411	**0.008**[a]	0.723

Revision time		Aarhus	Aalborg	Copenhagen	Total sample
0 - 6 w.	n	**1**	**1**	**7**	**9**
	Indication	A	A	A, A, A, A, A, B, B	7A, 2B
	Last OKS before revision	-	-	-	-
6 w. - 3 mo.	n	**0**	**0**	**6**	**6**
	Indication			A, A, A, A, B, C	4A, 1B, 1C
	Last OKS before revision			A: 34,39,*NA*,*NA*. B: 25. C: 28.	Mean: (28)
3 - 6 mo.	n	**1**	**1**	**0**	**2**
	Indication	C	C		2C
	Last OKS before revision	10	20		Mean: 15
6 – 12 mo.	n	**0**	**2**	**9**	**11**
	Indication		C, C	A, A, B, C, C, C, C, C, C	2A, 1B, 8C
	Last OKS before revision		11,35	A: 45,18. B: 26. C: 16,28,29,31,32,34.	Mean: 28
Total	Revisions during year 1	**2**	**4**	**22**	**28**
	Indications	1A, 1C	1A, 3C	11A, 4B, 7C	13A, 4B, 11C
	Revision rate in sample (%)	0.6	2.0	2.4	1.9
	95% CI (%)	0.7 - 2.2	0.5 - 5.0	1.5 - 3.6	1.3 - 2.8

Indication: Revision due to A=deep infection, B=fracture or liner dislocation, C=other cause. "Last OKS before revision"= Patient's last postoperative Oxford Knee Score before revision. NA= Missing (not available). [a]*) BMI in revision group: 95% CI, -0.6-(-3.4).*

Statistical Analysis

The primary outcome was 1-year OKS, subsequently OKS change. As our null-hypothesis was that there was no difference among hospitals, all analyses compared all three hospitals and mainly overall p-values were reported. All p-values were two-sided with alpha level 0.05. Standard deviations were presented as (± SD). Aarhus was as chosen as the reference hospital, because it was in-between the two other hospitals with respects to geography, urbanization and revision rates. As reported in validation studies, "flexion deficit" identified 95% of patients with passive flexion below 100° (sensitivity) and 81% of patients with flexion above 100° (specificity) [18, 21], and likewise, "extension deficit" identified 78% of patients with passive extension above 10° and 70% of patients with extension better that 10°.

Tabular data were analyzed by Chi-square test (with Monte-Carlo correction for expected cell counts < 5), and Clopper-Pearson confidence intervals were provided when relevant. Non-parametric methods (Kruskal-Wallis or Wilcoxon /Mann-Whitney U test) were used for ordinal measures UCLA Activity Scale, patient satisfaction, willingness to repeat, use of analgesics and radiographic OA severity classifications (Ahlbäck and K-L), whereas the outcomes OKS, FJS, EQ-5D, CKRS and global knee anchor were compared using parametric statistical methods (analysis of variance (ANOVA) or t-test) [20]. For ease of reading, some ordinal outcomes were presented as dichotomous variables, yet only when detailed ordinal analysis provided no valuable additional information, and only P-values based on full ordinal comparisons were reported. Multiple general linear regression analysis was made for 1-year OKS and OKS change. Here, analyses were made for both Ahlbäck and K-L scales, and since the radiographic classification method did not change the overall result, CI's were reported based on one score only (Ahlbäck). To study patients with poor results, dichotomization of change scores was made at 8 OKS points, which has been identified as the Minimal Important Change (MIC) in Danish KA patients, i.e. the minimal change score considered to be an “important improvement” by the notional average patient, based on the predictive modelling approach [2, 14, 20, 28]. As 1-year change scores were unavailable in revision patients, attempts were made to compensate hospitals with few revisions: the analysis was repeated with a) imputed poor results in revision patients (use of imputation is clearly marked in the text) and with b) all “last available” postoperative OKS change scores.

All observations were treated as independent data, though a minority of answers (7.3%) came from patients included twice (e.g. first left and subsequently right knee) [25]. All analyses were carried out in R (RStudio) [24]. Data collection and Case Report Forms (CRF) were handled by Procordo Software Aps, Copenhagen.

Ethics and Funding

The study was funded by The Health Research Fund of the Capital Region of Denmark. The National Committee of Health Research Ethics approved the ethical aspects (Protocol no. 16038343, 2 September 2016) and data storage was approved by the Danish Data Protection Agency (Jr. no. 2012-58-0004, HGH-2016-087, I-Suite no. 04819). All patients gave written consent for authors to access hospital charts. Permissions to use restricted questionnaires were obtained through each license provider.

Results

Do 1-year PROM results after primary KA vary among hospitals known to differ in revision rates?

Patient satisfaction and willingness to repeat surgery were both independent of hospital (table 4). OKS at 1-year follow-up did not differ among patients in the three hospitals (39.0 ± 7.4, p=0.096) (figure 1A). This remained the case when adjusted for age and sex, and also when further adjusted for baseline OKS and EQ-VAS (baseline) and variables that differed among hospitals preoperatively: BMI and radiographic classification (Ahlbäck or K-L). Change in 1-year OKS was lower in Aarhus (+1.6 in Aalborg, 95% CI, 0.07-3.1, p=0.041, or +1.3 in Copenhagen, 95% CI, 0.24-2.4, p=0.017, respectively) (figure 1B). This conclusion was partly changed with adjustment for age, sex and baseline OKS (+1.0 in Aalborg, 95% CI, -0.27-2.3, p=0.121, or +1.1 in Copenhagen, 95% CI, 0.18-2.0, p=0.022), and when BMI, EQ-VAS (baseline) and radiographic classification was further added, there were no differences among hospitals (Aalborg 95% CI, -0.4-2.5/ Copenhagen 95% CI, -0.2-2.0, p>0.101). In Aarhus, 19 % of patients did not reach the minimal important change (MIC) of 8 OKS points at 1 year, as did not 13/14% of Aalborg/Copenhagen patients (p=0.051). To fairly take the uneven distribution of revised (excluded) patients into account, all 28 revision patients were assigned an imputed (hypothetic) change score < 8 points (MIC), and the analysis was repeated: proportions of patients not reaching MIC were now 20% vs. 15-16% in the same hospitals (p=0.231). Alternatively, when comparing “last available” postoperative OKS change score in 1414 patients (97.4% of participants including non-responders at 1 year and 17 revision patients), there was no difference among hospitals either (21, 16 and 16 %, p=0.074). Some hospital variance was noted in knee extension, in physiotherapeutic assistance during rehabilitation, and, as previously reported, in implant choice [19], and Aalborg patients gained more in general health (EQ-VAS, p<0.001).

Table 4. Patient-reported outcomes at 1-year follow-up.

			Hospital			
		Total sample	**Aarhus** (Low rev. rate)	**Aalborg** (Low rev. rate)	**Copenhagen** (High rev. rate)	**p**
Patients at baseline (n)		1452	321	202	929	
Implant type (n (%))						**<0.001**
	TKA	1059 (73)	164 (51)	174 (86)	721 (78)	
	MUKA	336 (23)	129 (40)	25 (12)	182 (20)	
	PFA	50 (3.4)	23 (7.2)	3 (1.5)	24 (2.6)	
	LUKA	7 (0.5)	5 (1.6)	0 (0.0)	2 (0.2)	
Oxford Knee Score (OKS)						
	1 y. (n=1306)	39.0 [41] ± 7.4	38.1 [40] ± 8.3	39.1 [41] ± 7.2	39.2 [41] ± 7.2	0.092
	Last available postop. (n=1414)	38.3 [40] ± 8.0	37.5 [40] ± 8.7	38.7 [40] ± 7.5	38.5 [40] ± 7.8	0.119
	Change (n=1307)	15.4 ± 8.1	14.3 ± 8.7	15.9 ± 7.8	15.7 ± 8.0	**0.038**
OKS change < MIC (8 points) (total no. of patients in analysis (%))						
	1 y. (n = 1307)	195 (15)	56 (19)	25 (13)	114 (14)	0.051
	1 y. imputed* (n = 1335)	223 (17)	58 (20)	29 (15)	136 (16)	0.231
	Last available postop. (n = 1414)	237 (17)	66 (21)	31 (16)	140 (16)	0.074
Overall assessments						
	Willing to repeat surgery (%)					0.124
	"Yes, certainly"	1005 (77)	211 (73)	150 (80)	644 (77)	
	"Yes, probably"	200 (15)	46 (16)	26 (14)	128 (15)	
	"I don't know"	52 (4)	14 (4.9)	6 (3)	32 (3.9)	
	"No, probably not"	32 (2.5)	12 (4.2)	3 (1.6)	17 (2.0)	
	"No, absolutely not"	17 (1.3)	5 (1.7)	2 (1.1)	10 (1.2)	
	"Satisfied" or "very satisfied" (%)	1125 (86.2)	238 (82.6)	161 (86.6)	726 (87.4)	0.624[a]
Global knee anchor	1 y. (0-100)	80 ± 21	78 ± 24	81 ± 21	80 ± 19	0.082
	Change	51 ± 26	50 ± 29	51 ± 26	51 ± 25	0.769
Forgotten Joint Score, 1y		59.8 ± 27	59.1 ± 29	59.7 ± 25	60.1 ± 26	0.862
Patient-reported knee range of motion (CKRS units [b])						
Flexion	1y.	5.35 [6] ± 0.76	5.41 [6] ± 0.76	5.30 [5] ± 0.76	5.34 [5] ± 0.77	0.324
	Deficit (CKRS 0-4) (n (%))	165 (13)	32 (11)	21 (11)	112 (13)	0.483
	Change	0.48 [0] ± 1.2	0.57 [0] ± 1.2	0.55 [0] ± 1.2	0.43 [0] ± 1.1	0.160
Extension	1 y.	4.14 [4] ± 0.67	4.24 [4] ± 0.65	4.10 [4] ± 0.61	4.12 [4] ± 0.68	**0.016**
	Deficit (CKRS 0-3) (n (%))	161 (12)	29 (10)	24 (13)	108 (13)	0.420
	Change [c]	0.67 [1] ± 1.0	0.73 [1] ± 1.0	0.72 [1] ± 0.9	0.64 [1] ± 1.0	0.595
UCLA	1 y.	6.0 [6] ± 1.9	5.8 [6] ± 1.9	6.0 [6] ± 1.8	6.0 [6] ± 1.9	0.499
	Change	1.2 [1] ± 1.9	1.0 [1] ± 1.9	1.3 [1] ± 1.9	1.3 [1] ± 1.9	0.064
EQ-VAS	1 y.	79 ± 18	78 ± 20	82 ± 15	79 ± 18	0.079
	Change	17.4 ± 23	16.1 ± 24	24.3 ± 24	16.3 ± 22	**<0.001**
EQ-5D-5L Index	1 y.	0.81 ± 0.15	0.80 ± 0.17	0.83 ± 0.14	0.82 ± 0.14	**0.040**
	Change	0.22 ± 0.17	0.20 ± 0.18	0.23 ± 0.15	0.22 ± 0.17	**0.049**
Daily use of analgesics against knee pain (n (%))		166 (13)	41 (14)	22 (12)	103 (12)	0.364[a]
Supervised physiotherapy in rehabilitation (n (%)) [d]		702 (73)	115 (51)	92 (70)	495 (81)	**<0.001**

*When no unit is noted, means ± SD [and medians] are provided. *) "1y. imputed": Here, all 28 revised patients are assumed to be in the group with OKS change < MIC (8 points). [a]) Patient satisfaction is dichotomized for presentation, but P-value refers to tests of all five ordinal answer options. [b]) CKRS: With Copenhagen Knee ROM Scale, patients rate flexion from 0 (unable) to 6 (full flexion ability), and extension from 0 (unable) to 5 (full extension or slight hyperextension). [c]) Only the last 699 patients included in this analysis due to delay of scale development. [d]) Only the last 966 patients were asked about physiotherapy.*

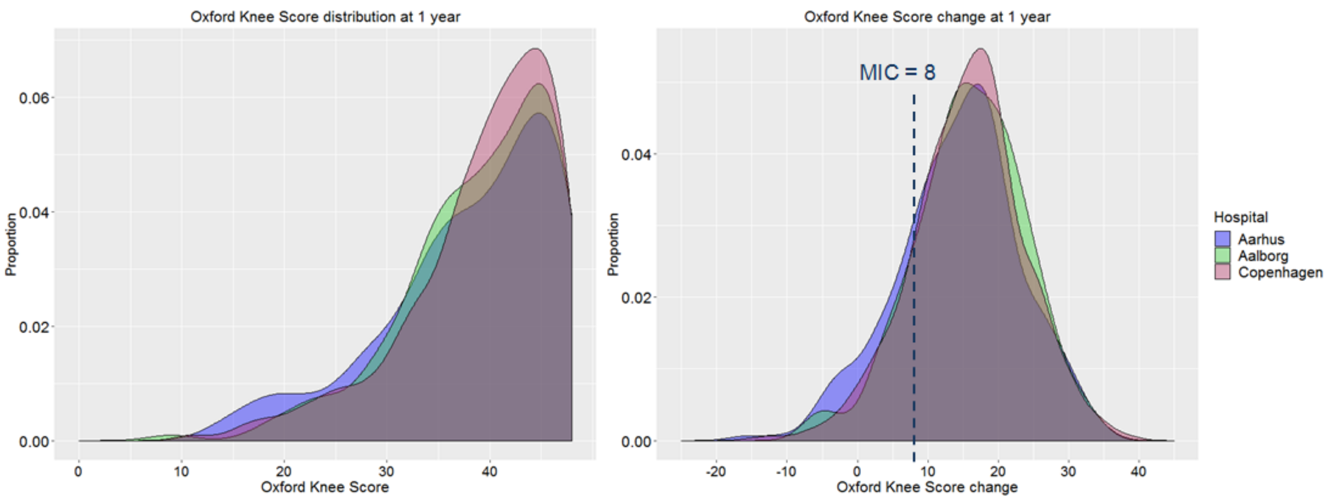


Figure 1A+B. *Oxford Knee Score at 1 year: A) absolute score, and B) change score with marking of minimal important change (MIC) = 8 points.*

Do changes in PROM scores during the first postoperative year differ among hospitals?

When OKS was studied over time, patients in Copenhagen had better OKS six weeks after surgery; 27.7 ± 7.3 compared to Aarhus (25.6 ± 8.1) and Aalborg (26.1 ± 7.3, p=0.001, unadjusted) (figure 2A), and fewer Copenhagen patients had change scores below MIC (66 vs. 72/76%) at six weeks. No similar pattern was noted in the other PROMs at six weeks (or later), and the hospital difference was nuanced when MUKA and TKA patients were studied separately (figure 2B). From three months, no hospital differences were observed in absolute OKS (table 5 displays the total sample). Through the entire study, OKS differed between TKA and MUKA patients in the overall sample, e.g. 1-y. OKS was 38.7 vs. 40.3, respectively (95% CI, 0.6-2.5, p=0.002).

Table 5. Development of main PROMs over time after surgery in the total sample.

	Baseline (preop.)	6 weeks	3 months	6 months	1 year
Oxford Knee Score (OKS)	23.3 ± 6.7	27.0 ± 7.6	33.6 ± 7.5	37.0 ± 7.4	39.0 ± 7.4
OKS change score	-	3.6 ± 8.1	10.2 ± 8.1	13.5 ± 8.0	15.4 ± 8.1
OKS change score < MIC (8 points) (n (%))	-	788 (69)	462 (36)	262 (21)	195 (15)
Global knee anchor (0-100)	28 ± 18	60 ± 21	71 ± 22	76 ± 21	70 ± 21
Range of motion (Copenhagen Knee ROM Scale (CKRS) units)					
Flexion	4.9 ± 1.2	4.5 ± 1.1	5.0 ± 0.9	5.3 ± 0.8	5.4 ± 0.8
Deficit (CKRS 0-4) (n (%))	416 (29)	525 (46)	317 (25)	188 (15)	165 (13)
Extension	3.5 ± 0.9	3.5 ± 0.7	3.9 ± 0.8	4.0 ± 0.7	4.1 ± 0.7
Deficit (CKRS 0-3) (n (%)) [a]	340 (49) [a]	336 (42) [a]	246 (27) [a]	202 (17)	161 (12)
Forgotten Joint Score	-	-	43 ± 25	53 ± 26	60 ± 27
UCLA Activity Scale	4.7 [4] ± 1.9	-	5.4 [5] ± 1.7	5.8 [6] ± 1.8	6.0 [6] ± 1.9
Daily use of analgesics against knee pain (n (%))	854 (59)	870 (76)	498 (39)	274 (22)	166 (13)
EQ-5D VAS	61 ± 22	71 ± 18	76 ± 17	78 ± 18	79 ± 18
EQ-5D-5L Index	0.59 ± 0.15	0.70 ± 0.13	0.76 ± 0.12	0.79 ± 0.14	0.81 ± 0.15

When no unit is noted, means ± SD and [medians] are provided. [a]) Total n is increasing during the study due to concomitant scale development.

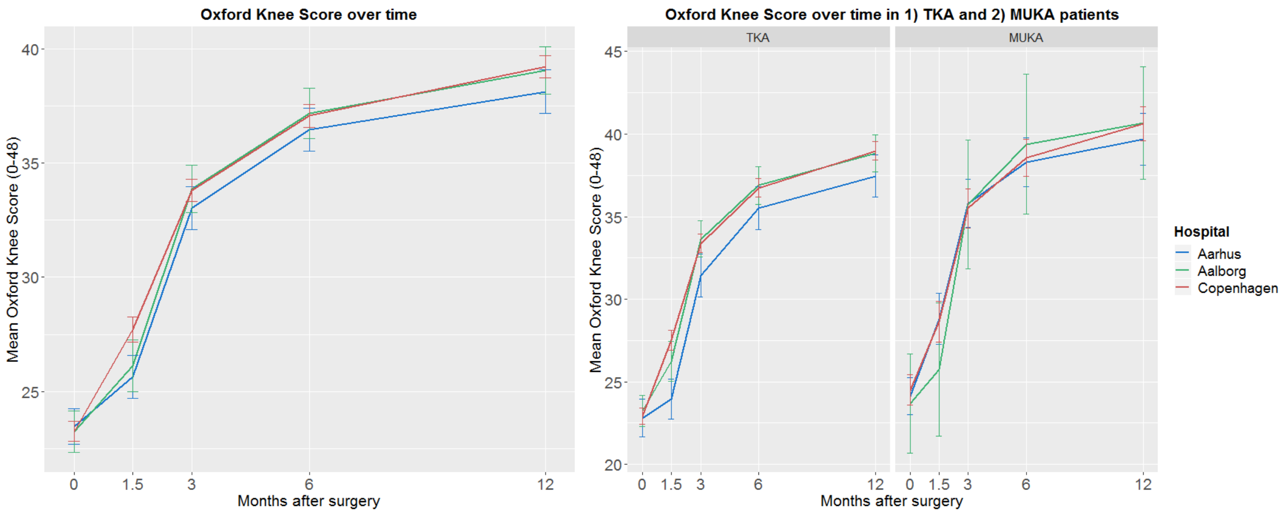


Figure 2A+B. *Oxford Knee Score during the first postoperative year in A) all patients, and in B) TKA and MUKA patients separately. Whiskers denote mean ± 2 x std. error of the mean.*

Can patients with the same degree of knee OA expect the same outcome across hospitals?

When patients were grouped by preoperative Ahlbäck or K-L classification, neither willingness to repeat surgery, 1-year OKS or last postoperative OKS differed among hospitals (p=0.087-0.980) (figure 3). An exception, though, was the K-L 4 group: here, the 17 Aarhus patients had 3.7-5.6 points lower 1-year OKS (95% CI, 0.04-10.5, p=0.045) and 3.5-5.6 points lower last postoperative OKS (95% CI, 0.03-10.3, p=0.040) It should be noted that in this group, the only two Aarhus patients with MUKA scored 37 and 40 points. When all patients were grouped by baseline OKS (0-20, 21-30 or 30-48) the same three outcomes did not differ among hospitals (P=0.224-0.482) (total sample displayed in figure 4). In Aarhus, where unicompartmental implants were twice as frequently used, patients had better 1-year extension, but extension and flexion were not associated to hospital when adjustments for baseline motion were made. However, in the overall sample, after adjustments for baseline motion, MUKA was associated to a larger increase in flexion (+0.34 CKRS points corresponding to approximately 5 degrees, p<0.001) [18] when compared to TKA at 1 year, but not to better extension (p=0.311) (figure 5). Finally, willingness to repeat surgery was independent of hospital for patients in the same 1-year OKS group (10 points' intervals) (p=0.157-0.821).

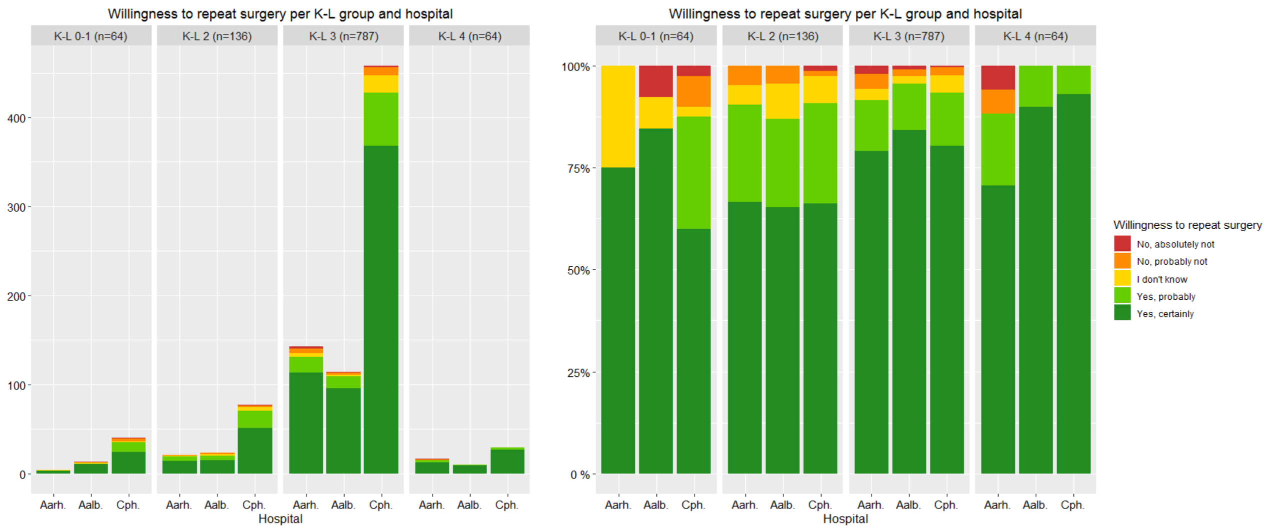


Figure 3A+B. *Willingness to repeat surgery at 1 year postoperatively grouped by Kellgren-Lawrence classification of preoperative knee OA and hospital, displayed as A) counts and B) proportion of patients.*

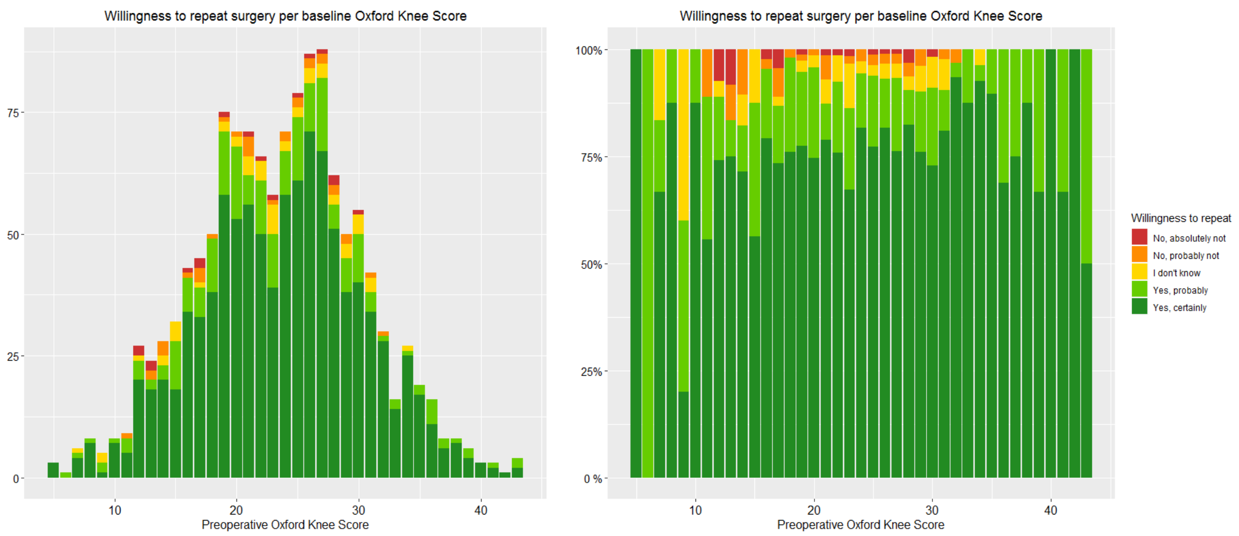


Figure 4A+B. *Willingness to repeat surgery at 1 year postoperatively as a function of Oxford Knee Score at baseline displayed as A) counts, and B) proportions of patients (total sample).*

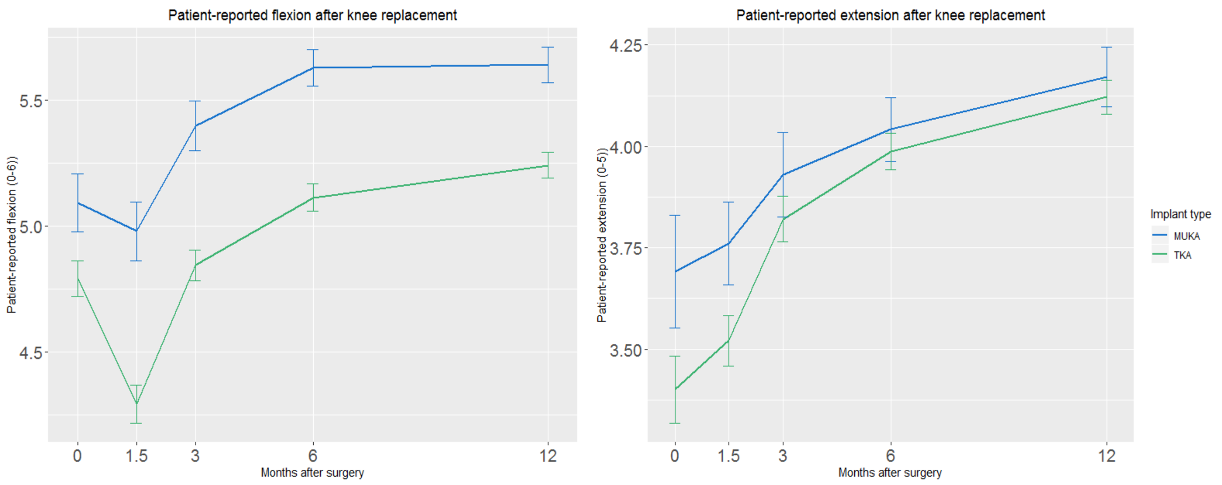


Figure 5A+B. *Patient-reported A) flexion and B) extension in the total sample, grouped by implant type (MUKA or TKA only), assessed with Copenhagen Knee ROM Scale. Whiskers denote mean ± 2 x std. error of the mean. Based on validation studies, flexion "4" corresponds to mean 101°, "5" to 121°, and 6 to 131°. In extension, "3" refers to mean 7°, "4" to 5°, and "5" to 1°.*

Discussion

National knee arthroplasty registers offer revision rates as simple and easily obtainable outcomes of knee replacement surgery [10, 23]. They are fast and efficient means to detect poor performance of implants, hospitals or even surgeons. However, with pain relief being the primary goal of knee arthroplasty surgery, revision rates are irrelevant measures of treatment quality in the large majority of patients, who are never revised [10, 26]. Also, surgical quality is not a yes-no question, but rather a wide spectrum ranging from a satisfied patient with a perfectly functioning prosthesis to an ill, infected patient in definite need of revision surgery. Outcomes measures should reflect this reality in order to measure and ultimately improve the quality of surgery. In this study, three high-volume knee arthroplasty centers with very different revision rates were found to have the same results after primary knee replacement when measured with well-established PROMs, patient satisfaction and willingness to repeat surgery. This contradicts the conclusion that could be drawn from implant survival data alone where high revision rates generally represent poorer surgical outcomes.

We found an observational study to be the most favorable design to answer our research questions [6]. With this design, however, conclusions about casual relations, e.g. between implant choice and outcomes, cannot be drawn. Another limitation was that the study was conducted in three hospitals, not three whole regions, thus, results do not necessarily reflect the circumstances on regional level. Despite our intention to invite practically all primary KA patients, the participation rate was restricted to approximately 56% of the patients operated in that period of time (62% in Aarhus/ Copenhagen, and 37% in Aalborg) [19]. Though sample patients were close to being representative with respects to age, sex and implant types, we were unaware of the socioeconomic distribution among hospitals, and comorbidity information was derived from the EQ-5D alone [7, 11]. We consider it an important strength that 89% of patients responded prior to surgery and we had a very low loss to follow-up; 90% of participants responded at 1 year, and 97.4% responded at least once postoperatively. Information about subsequent revision surgery was considered complete. Patients who were revised participated only until exclusion. Thus 1-year results were unknown in these 28 patients, but efforts were made to compensate for this lack of information when research questions were answered.

The well-known hospital differences in revision rates have not been confirmed in this sample of patients. Nor was there any detectable hospital difference in the proportion of patients revised for other causes than deep infection. This may in part be due to sample size but confirming revision rate differences was never the purpose of the study: surgical thresholds and revision patterns may have been affected by the initiation of a study with these matters in focus; 2- and 5-year follow-up of the SPARK cohort will clarify if this is the case. However, as surgeon staffs have been largely stable, we find it unlikely that surgical quality in the three hospitals have changed between past years and the study period.

1-year PROM results and willingness to repeat surgery were similar across the three hospitals with minor exceptions that could largely be explained by known confounding baseline characteristics. The six months mean OKS of 37.0 was in line with results from other countries, e.g. the 37.7 OKS reported from the National Joint Registry for England and Wales in 2015 [16]. One low-revision hospital, Aarhus, which operated fewer patients with mild radiographic OA and performed only two revisions (0.6%) in this sample, had 19% patients with 1-year OKS change score below the MIC of 8 points for “important improvement” [14] compared to 13-14% in the other hospitals. This was not a significant difference (overall p=0.051) and we are unable to say whether the patients with poor progress would have benefited from revision surgery.

Overall, the patients improved similarly across hospitals over time in the first year after surgery. Copenhagen patients recovered faster (higher OKS) in the first 6 weeks, which might be associated to the more frequent use of physiotherapy in rehabilitation. However, no differences were observed in flexion or extension, and when results were stratified by implant type, a different pattern was seen among hospitals. Also, the hospital difference in rehabilitation efforts did not seem to be associated to 1-year OKS. Thus, we withdraw from further speculations about the cause of the observed hospital difference in OKS at six weeks.

Finally, we found that patients with the same level of knee OA disease (as measured by either baseline OKS or radiographic OA classification) achieved the same outcome (1-year OKS) and had the same degree of willingness to repeat surgery across the three hospitals. This confirms an overall uniformity in the quality of surgery across hospitals. In the whole sample, there were several differences in outcomes between MUKA and TKA patients, but although Aarhus Hospital used unicompartmental implants twice as often as the other two hospitals, these implant-related differences were not directly visible in the overall comparison of hospitals. An exception, though, was a tendency of better range of motion in Aarhus, yet, this was only sporadically significant, and it must be kept in mind that the clinical relevance of patient-reported extension differences is yet to be clarified.

Conclusion

The SPARK study has shown that patient-reported outcomes one year after primary knee arthroplasty are the same in three high-volume centers that have differed in revision rates for a decade. Our results suggest that variance in revision rates may reflect regional variations in thresholds for revision rather than differences in the quality of primary surgery. Thus, any attempt to rank the quality of treatment in knee replacement centers solely resting upon implant survival data would have led to false conclusions. The same concerns should be raised when comparing revision rates among nations. Further studies should determine whether patients are offered revision surgery on the same clinical grounds across regions, and moreover to which extent patients benefit from revision surgery that is not motivated by deep infection. Such studies followed by thorough discussions about revision indications and techniques might serve patients with the poorest results as much as the ongoing attempts to refine primary knee replacement surgery.

Acknowledgements

We thank the following for their contributions to the study: Kristian Mongelard, Lone Rømer, Birthe Schøning, Heidi Ladefoged Poulsen, Camilla Grube, Naima Elsayed, Ulla Hornum, Gitte Broholm, Patricia Svare, Tina Isaksen, Gitte Raae Høybye, Anette Enemark-Larsen, Svend Erik Østgaard, Lasse Enkebølle Rasmussen, Thomas Bjerno, Lars Peter Møller, Snorre Stephensen, Thomas Lind, Henrik Schrøder, Søren Rytter, Claus Fink Jepsen, Andreas Kappel, Procordo ApS, medical students at Gentofte Hospital, The Health Research Fund of the Capital Region of Denmark, and all patients involved.

References

1. Ahlbäck S, Rydberg J. Röntgenologisk klassifikation och undersökningsteknik vid gonartros. *Lakartidningen*. 1980;77:2091–2093, 2096.
2. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol*. 2015;68:73–79.
3. Behrend H, Giesinger K, Giesinger JM, Kuster MS. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplast.* 2012;27:430–436.e1.
4. Clement ND, Macdonald D, Burnett R, Simpson A, Howie CR. A patient's perception of their hospital stay influences the functional outcome and satisfaction of total knee arthroplasty. *Arch Orthop Trauma Surg*. 2017;137:693–700.
5. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Jt. Surg Br*. 1998;80:63–69.
6. Dieppe P, Lim K, Lohmander S. Who should have knee joint replacement surgery for osteoarthritis? *Int. J. Rheum. Dis.* 2011;14:175–180.
7. 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.
8. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. *J Bone Jt. Surg Br*. 2001;83:339–344.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4:e296.
10. Goodfellow JW, O'Connor JJ, Murray DW. A critique of revision rate as an outcome measure: re-interpretation of knee joint registry data. *J Bone Jt. Surg Br*. 2010;92:1628–1631.
11. Harcourt WG, White SH, Jones P. Specificity of the Oxford knee status questionnaire. The effect of disease of the hip or lumbar spine on patients' perception of knee disability. *J Bone Jt. Surg Br*.

2001;83:345–347.
12. Harris K, Dawson J, Gibbons E, Lim CR, Beard DJ, Fitzpatrick R, Price AJ. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016;7:101–108.
13. van Hout B et al. JMF. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health 2012 Jul-Aug;15(5):708-15.
14. Ingelsrud LH, Roos EM, Terluin B, Gromov K, Husted H, Troelsen A. Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. *Acta Orthop*. 2018;89:541–547.
15. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. *Ann. Rheum. Dis.* 1957;16:494–502.
16. Liddle AD, Pandit H, Judge A, Murray DW. Patient-reported outcomes after total and unicompartmental knee arthroplasty. *Bone Joint J*. 2015;97–B:793–801.
17. Mongelard K, Morup-Petersen A, Roemer L, Kristensen KB, Odgaard A. Heuristic ranking delivers more detail than than ordinal grading of knee osteoarthritis radiographs. *Submiss. Process*.
18. Morup-Petersen A, Holm PM, Holm CE, Klausen TW, Skou ST, Krogsgaard MR, Laursen MB, Odgaard A. Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale. *J Arthroplast.* 2018;33:2875–2883.e3.
19. Morup-Petersen A, Laursen M, Madsen F, Krogsgaard MR, Winther-Jensen M, Mongelard KBG, Roemer L, Odgaard A. Large variation in revision rates after primary knee arthroplasty: A matter of patient selection? Baseline data from 1452 patients in the prospective multicenter cohort study, SPARK. *Submitted.*
20. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. *J Bone Jt. Surg Br*. 2007;89:1010–1014.
21. Mørup-Petersen A, Holm PM, Holm CE, Klausen TW, Skou ST, Krogsgaard MR, Laursen MB, Odgaard A. Corrigendum to "Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale" [The Journal of Arthroplasty 33 (2018) 2875-2883]. *J. Arthroplasty*. 2019.
22. Pedersen AB, Mehnert F, Odgaard A, Schroder HM. Existing data sources for clinical epidemiology: The Danish Knee Arthroplasty Register. *Clin Epidemiol*. 2012;4:125–135.
23. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, Carr A, Beard D. Knee replacement. *Lancet*. 2018;392:1672–1682.
24. RCoreTeam. RStudio version 1.1.463. https://www.R-project.org/. R: A Language and Environment for Statistical Computing.
25. Robertsson O, Ranstam J. No bias of ignored bilaterality when analysing the revision risk of knee prostheses: Analysis of a population based sample of 44,590 patients with 55,298 knee prostheses from the national Swedish Knee Arthroplasty Register. *BMC Musculoskelet. Disord.* 2003;4:1.
26. Robertsson O, Ranstam J, Lidgren L. Variation in outcome and ranking of hospitals: an analysis from the Swedish knee arthroplasty register. *Acta Orthop*. 2006;77:487–493.
27. Sundhedsdatastyrelsen. The Danish National Health Register. (Landspatientregistret) https://www.esundhed.dk/Registre/Landspatientsregisteret/Operationer. 2019.
28. Terluin B, Eekhout I, Terwee CB, De Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *J. Clin. Epidemiol.* 2015;68:1388–1396.
29. Thomsen MG, Latifi R, Kallemose T, Barfod KW, Husted H, Troelsen A. Good validity and reliability of the forgotten joint score in evaluating the outcome of total knee arthroplasty. *Acta Orthop*. 2016;87:280–285.
30. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M. Strengthening the Reporting of Observational Studies in Epidemiology

(STROBE): explanation and elaboration. *PLoS Med*. 2007;4:e297.
31. The Danish Knee Arthoplasty Register. Annual Report 2018.
32. *Swedish Knee Arthroplasty Register Annual Report 2018*. 2018.

Study III: UCLA

Translation, cultural adaptation and measurement properties of the Danish version of the UCLA Activity Scale for use in hip and knee replacement patients

Anne Mørup-Petersen[1], Søren T. Skou[2], Christina E. Holm[3], Pætur M. Holm[2], Claus Varnum[4], Michael R. Krogsgaard[5], Mogens Laursen[6], Anders Odgaard[1].

1) Department of Orthopedic Surgery, University of Copenhagen, Herlev and Gentofte Hospital, Copenhagen
2a) Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, 2b) Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Region Zealand
3) Department of Orthopedic Surgery, Copenhagen University Hospital, Rigshospitalet
4a) Department of Orthopedic Surgery, Vejle Hospital, University of Southern Denmark
4b) Department of Regional Health Research, Faculty of Health Science, University of Southern Denmark
5) Section for Sports Traumatology, Bispebjerg and Frederiksberg Hospital, Copenhagen, University of Copenhagen
6) Department of Orthopedic Surgery, Aalborg University Hospital, Aalborg & Farsø
Corresponding author: Anne Mørup-Petersen, email: morup.anne@gmail.com

Abstract

Background and purpose: UCLA Activity Scale (UCLA) is a brief questionnaire assessing physical activity level from 1-10 (10 is high-level) in hip and knee arthroplasty (HA/KA) patients. We translated and culturally adapted UCLA into Danish and validated its measurement properties.

Patients and methods: methods: Translation followed the dual-panel method and cultural adaption was performed by cognitive interviews with 55 HA/KA patients. Measurement properties were validated from applying the Danish version to 131 HA and 134 KA patients preoperatively and 1-year postoperatively. Additionally, based on short patient interviews, an orthopedic surgeon and a physiotherapist proposed UCLA levels of 65 KA patients.

Results: To suit Danish patients of today, several cultural adaptations of activity examples were needed. HA/KA patients (mean age 71/68 y.) changed from UCLA 4.3 ± 1.9/ 4.5 ± 1.8 preoperatively to 6.6 ± 1.8/ 6.2 ± 1.0 at 1-year follow-up, and 79% of hip patients, and 66% of knee patients reported an increase in physical activity. Effect sizes were large (1.2/0.96). Knee (but not hip) patients who gained 8 points in Oxford Hip/Knee Score (minimal important change) had higher 1-year UCLA (and change) scores than others. In interview studies, 11 of 76 patients misinterpreted the scoring system. Examiners rated the remaining 65 patients' (mean 67 y.) UCLA levels 0.2-1.6 points lower than patients themselves.

Interpretation: UCLA has an inherent problem in lack of documented scale development and therefore has no proven content validity. Mixing frequency, intensity and duration of physical activity into one single item limits the scale's ability to discriminate between patients. In spite this, it may be valuable for measuring change in physical activity.

Introduction

Osteoarthritis (OA) of the hip and knee strongly affects a persons' ability to be physically active. When OA becomes severe and joint replacement is considered, it is in the interest of both patients and health care providers to know what change in physical activity to expect after surgery. Yet, measuring levels of physical activity is not clear-cut. Accelerometers can be regarded as the gold standard to measure non-specific physical activity, but they are resource demanding and therefore not a feasible option in many studies [22]. Also, accelerometer measurements do not necessarily reflect the difficulty and intensity of various daily activities that may be important to patients. As an alternative, physical activity assessments can be made by use of patient-reported physical activity scales like the UCLA Activity Scale (UCLA) named from University of California, Los Angeles [1, 29]. UCLA is a single-item 10-level-scale ranging from level 10, representing a highly physically active patient, to level 1, a patient who is dependent on others and unable to leave home.

UCLA is used in several languages, yet a description of the development of UCLA has to our knowledge never been published [1]. Originally, it appears to have been made for surgeons to assess the activity level of hip and knee arthroplasty patients. Now, UCLA is used as a patient-reported outcome measure (PROM), though probably not developed or originally intended to be used as such. A study comparing UCLA scores to accelerometer measurements of walking activity revealed a strong correlation but large measurement errors; patients reporting the same level of activity in UCLA varied up to a factor 15 in average number of steps per day [29]. Despite this, UCLA is widely used and has been recommended as one of the more useful physical activity PROMs in hip and knee arthroplasty (HA/KA) patient populations, mainly based on positive rating of construct validity and high completion rates compared to similar instruments [18, 20, 27]. Its brevity and simplicity make it an attractive choice when combined with other questionnaires that might be time-consuming to fill out.

In Denmark, UCLA has been used in at least two different unpublished versions. One of them, which has not undergone proper psychometric testing, demonstrated a bimodal distribution of answers, potentially due to intercultural differences in activities [23]. This called for further investigation of the measurement properties of the UCLA questionnaire. The aim of this study was to develop a new Danish version of the UCLA through formal translation and cultural adaptation and further to test the validity, reliability and interpretability of the questionnaire in relevant groups of hip and knee osteoarthritis (OA) and arthroplasty patients.

Patients and methods

The study was conducted in 4 parts: 1) translation and cultural adaptation, 2) validation in KA patients, 3) test-retest reliability in the same population, and 4) responsiveness and interpretability in a routine cohort of KA and HA patients before and 1 year after surgery.

The design and reporting of the study were guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines [16] and the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [12].

Translation and cultural adaptation

A modified dual panel translation approach was used [6, 15, 25]. The translation was made by a professional translator, a physiotherapist and an orthopedic surgeon (senior house officer) in collaboration (all were bilingual with Danish mother-tongue). First, each of the three prepared a Danish translation from the original American version (figure 1), and then all three met to agree upon a common version.

	Activity Level
1	Wholly inactive: dependent on others; cannot leave residence
2	Mostly inactive: very restricted to minimum activities of daily living
3	Sometimes participates in mild activities such as walking, limited housework, and limited shopping
4	Regularly participates in mild activities
5	Sometimes participates in moderate activities such as swimming and can do unlimited housework or shopping
6	Regularly participates in moderate activities
7	Regularly participates in active events such as bicycling
8	Regularly participates in very active events such as bowling or golf
9	Sometimes participates in impact sports such as jogging, tennis, skiing, acrobatics, ballet, heavy labor, or backpacking
10	Regularly participates in impact sports

Figure 1. *UCLA Activity Scale as first published by Amstutz et. al. in 1984. In the 1998 version (Zahiri et al.), level 10 was presented first and a patient instruction was added: "Of the following options, which statement best describes your activity level?"*

The translated version was presented to three laymen panels consisting of total 22 heart and lung patients, recruited at physiotherapy training sessions (mean age 72 ± (SD) 9 years, 10 males) to ensure wording and cultural adaption. First, all participants were asked to complete the questionnaire while "thinking out loud", and afterwards, in plenary sessions, they were encouraged to comment on any doubts and wonderings and to express their general perception of the questionnaire. During these 3 interview sessions, the questionnaire was changed in layout, instructions, wordings and activity examples.

Subsequently, to ensure cultural adaption and relevance for target patients, the revised version was presented in the same manner to target group patients, either awaiting or just having undergone HA or KA at Copenhagen University Hospital Gentofte. During meetings with 55 patients (38 pre- and 17 postoperative (21 males), mean age 69.5 ± 8 y.) in 6 rounds the questionnaire was adjusted, and when the evaluations led to no further revisions, the adaption process was ended.

Correlation with external assessment of physical activity level

As UCLA is regarded as a patient-reported outcome measure, each patient is, by definition, the only one who can properly estimate the UCLA level. However, as a consequence of the lack of documentation behind the development of UCLA and the fact that the scale was originally completed by surgeons, we sought to determine the degree of common understanding of overall content and difficulty of activity levels among knee patients and health professionals, based on interviews.

From Dec 2016 to Feb 2017, KA patients above 40 years of age were recruited at Naestved Hospital, Region Zealand, both in the outpatient clinic and in the ward, either before or at least one day after surgery. Besides primary KA patients, also patients with post-surgery complications and patients undergoing revision KA were invited. We excluded patients unable to read and understand Danish language and patients with signs of dementia who failed a clock-drawing test [14]. In addition to the UCLA score, patients reported age, sex, height, weight, and current knee pain level on a visual analogue scale (VAS) ranging from 0-10 (10 most painful).

As part of the written instructions, patients were asked to consider their level of physical activity in the preceding 4 weeks. This instruction had to be modified in some cases; for example, patients who were included just after surgery were instead asked to report their physical activity level in the 4 weeks leading up to surgery. Patients were to complete the questionnaire without help from others, yet, accompanying relatives were not asked to leave the room. Immediately after completion, patients met with either of two orthopedic surgeons (one senior house officer, one specialist registrar) and one physiotherapist (in random order) for short, separate interviews about daily physical activities, mimicking the normal clinical setting (approximately 5 minutes duration). Patients were instructed not to reveal their UCLA answer to the interviewer. Based on these interviews, examiners noted their own assessment of each patient's UCLA level, blinded to the other examiner's estimate.

Test-retest reliability

Following the interviews, participants were given a blank UCLA questionnaire to fill out 7-10 days later and return in a prepaid envelope. Patients scheduled for surgery within this period and patients recovering from surgery (< 6 weeks) were not included in retest studies because their physical activity levels could rapidly change due to surgery or recovery. On the front page of the retest questionnaire, patients answered whether their level of physical activity had changed since the first test, and if so, in which direction.

Construct validity and responsiveness

Distribution of UCLA scores in pre- and postoperative HA/KA patients and responsiveness (validity of change scores) were prospectively evaluated in Vejle Hospital, Region of Southern Denmark, with patient inclusion from March 2018 and 4 months forward. As part of the normal clinical routine, all patients scheduled for total HA total due to hip OA (not dysplasia) (n=130) and patients scheduled for total or medial unicompartmental KA due to primary (n=119) or secondary (n=15) knee OA were asked to complete PROM sets, including UCLA, before and 1 year after surgery. Most patients filled out PROMs electronically, but paper versions were available for patients with no email address. Non-responders were sent a reminder and, if necessary, were contacted by phone. Completeness at 1 year reached 96% (HA) and 95% (KA). Patients with incomplete answers and patients who went through revision surgery during the study period were excluded. Besides UCLA, PROM-sets included the joint-specific Oxford Hip or Knee Score (OHS, OKS) [4, 5, 17], the generic health PROM EQ-5D-5L and EQ-5D VAS [11] and an overall patient satisfaction question with answers ranging from "very unsatisfied" to "very satisfied" (5-point Likert scale, one neutral answer).

Since there is no gold standard approach available when assessing a true change in the construct "activity" [18], responsiveness was evaluated by use of the construct approach [28], i.e. correlation of UCLA change with other PROM change scores and overall satisfaction. We did, however, expect only fair to moderate correlations for several reasons [18]. Firstly, a successful operation leading to pain relief and increased joint function should increase a patients' ability to be physically active, but whether a change in physical activity actually occurs depends on many other factors, e.g. general health, surroundings as well as personal goals and interests – patients do not necessarily have an intention to become more physically active after surgery. Secondly, the amount of perceived change may be influenced by recall bias and preoperative expectations to change. Thirdly, the outcomes used to assess change are influenced by other matters than physical activity; OHS and OKS assess joint pain, joint function and mobility while the generic EQ-5D measures overall health and the satisfaction question is affected by the overall hospital care provided.

Based on previous reports, we expected a mean 1-year increase of 1-3 UCLA levels [7, 21, 24] and a 2-fold increase in the proportion of patients with UCLA score ≥ 6 in both HA and KA patients [21]. Besides these anchor-based methods, we also calculated the Effect Size (ES), a traditional distribution-based measure of responsiveness [2].

Statistics

UCLA scores were not expected to be equidistant or normally distributed, so the score was treated as an ordinal variable and analyzed using nonparametric statistical methods (Wilcoxon rank sum and Kruskal-Wallis test). However, for a full overview of variations in results, means and standard deviations (SD) were reported as well, and univariate linear regression analyses were performed to check for score dependence of age and BMI. Paired tests (paired t-test in OHS/OKS and Wilcoxon signed rank test in UCLA scores) were used to calculate within-patient differences. Dichotomization of OHS/OKS change scores was done to study associations between 1-year UCLA (changes) and reaching minimal important change (MIC). In both scores, the MIC was set at 8 points [3, 10].

For correlation with external (physiotherapist and surgeons') assessment of physical activity level, agreement was estimated by mean difference, limits of agreement (LoA) and Spearman's correlation coefficient. Weighted Kappa coefficient [13] was used to calculate an alternative measure of reliability (examiner vs. patient rating). In responsiveness analyses, effect size (ES) was calculated as the mean UCLA improvement divided by SD of UCLA at baseline [2, 28]. Floor or ceiling effects were considered present if more than 15% of patients marked the lowest (1) or highest score (10), respectively [26].

Sample size for the interview study was based on recommendations of 50-100 patients [26, 28] balanced with feasibility. For responsiveness and construct validity studies, we aimed for more than 100 hip and 100 knee patients and used all data available at the time. Statistical significance level was set at alpha level 0.05 (2-sided) and 95% confidence intervals (CI) were reported when relevant. All analyses were carried out in R (Rstudio) [19].

Ethics, registration, data sharing plan, funding and potential conflicts of interest

The study was funded by The Health Research Fund of the Capital Region of Denmark and ethically approved by the National Committee of Health Research Ethics. Data management approval was provided by the Danish Data Protection Agency (Jr. no. 2012-58-0004). No funds or authors had relevant conflicts of interest. Raw data is available upon request.

Results

Translation and cultural adaptation

Through the cross-cultural adaption process it quickly became evident that the original American version of UCLA had to be changed in wording to become a valid and relevant measure of physical activity in Danish patients. For example, bicycling for transportation is more common in Denmark, also among the elderly, than in USA. To prevent a bimodal distribution of scores as in a previous version with maximum counts at levels 4 and 7 [23], the bicycling activity was split by intensity and frequency to cover levels 5-8 in the new Danish version (appendix B).

As defined by the American version, physical activity was verbally graded by use of adjectives “moderate” or “high” in the Danish translation. To further clarify the content, we added examples of popular Danish sports and activities, e.g. badminton and gymnastics/fitness. Examples were mentioned only once in order to keep the text as short as possible and offer patients a quick overview of the questionnaire. Instead we let curly brackets illustrate how each group of examples referred to two levels that differed only by frequency of activities (“regularly” or “once in a while”).

The questionnaire had to be self-explanatory. Hence, efforts were made to give patients a sufficient written introduction to the task. During test rounds we learned that the instruction had to be extremely short for patients not to skip the whole introduction and misinterpret the scale. Some patients, for example, ticked several boxes (e.g. 2, 4, 6 and 8) or wrote numbers from 1-10 in every box. Eventually, instructions were cut down to a minimum, and those misunderstandings became less frequent but were not eliminated.

Correlation with external assessment of physical activity level

We invited 80 KA and knee OA patients for interviews. 2 were excluded due to poor language skills, 2 because their scores were lost, and 11 (14.5%, 67.3 y.) due to marking more than one UCLA level. In the remaining 65 patients (table 1), results were aggregated (table 2) and related UCLA scores were displayed graphically (fig. 2).

Table 1. Characteristics of interview participants

n	65
Age (mean [median] (SD))	66.4 [67] (10.5)
Male sex (n (%))	29 (44.6)
Knee arthroplasty/ knee OA (n)	34/31
BMI (kg/m²) (mean (SD)[1]	30.4 (5.8) [1]
Pain (VAS 0-10) (mean [median] (SD)) [2]	4.5 [5] (2.8) [2]

1) n=60, 2) n= 63. BMI = Body Mass Index.

Table 2. Results of interviews: Correlation with external assessment of physical activity level

Absolute scores (total sample)	**UCLA Activity scale**				
	Mean	SD	Median	IQR	Range
Patient	5.0	1.7	5	[4, 6]	[2, 10]
Surgeon	4.4	1.6	4	[3, 6]	[1, 8]
Physiotherapist	3.8	1.3	4	[3, 4]	[2, 8]

Differences (per individual patient)					**Reliability**	**Agreement**	**Correlation**
	Mean $_{diff}$	CI $_{diff}$	Median $_{diff}$	Range $_{diff}$	Weighted Kappa	LoA	Spearmans rho
Patient minus surgeon	0.6	0.2 - 0.9	0	[-3; 3]	0.63	-2.0 – 3.1	0.65
Patient minus physiotherapist	1.2	0.8 - 1.6	1	[-2, 6]	0.31	-2.0 – 4.4	0.47

IQR: Interquartile range [25%, 75%]. Spearmans rho correlation coefficient (-1 to 1) indicates the degree of linearity between measurement ranks. LoA= Limits of agreement (mean ± 2 SD). Differences are based on assessments within each patient.

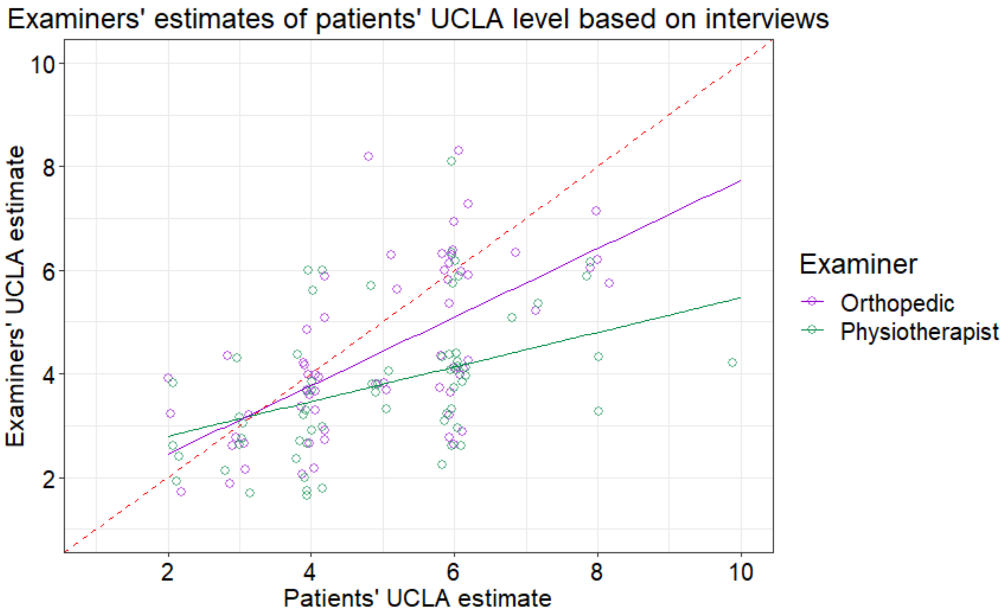


Figure 2. *Correlation with external assessment of physical activity level: Orthopedic surgeons' and physiotherapists' estimates of UCLA plotted against patients' own estimates. The red dotted line indicates perfect agreement. Random variance (jitter) is added to prevent over-plotting.*

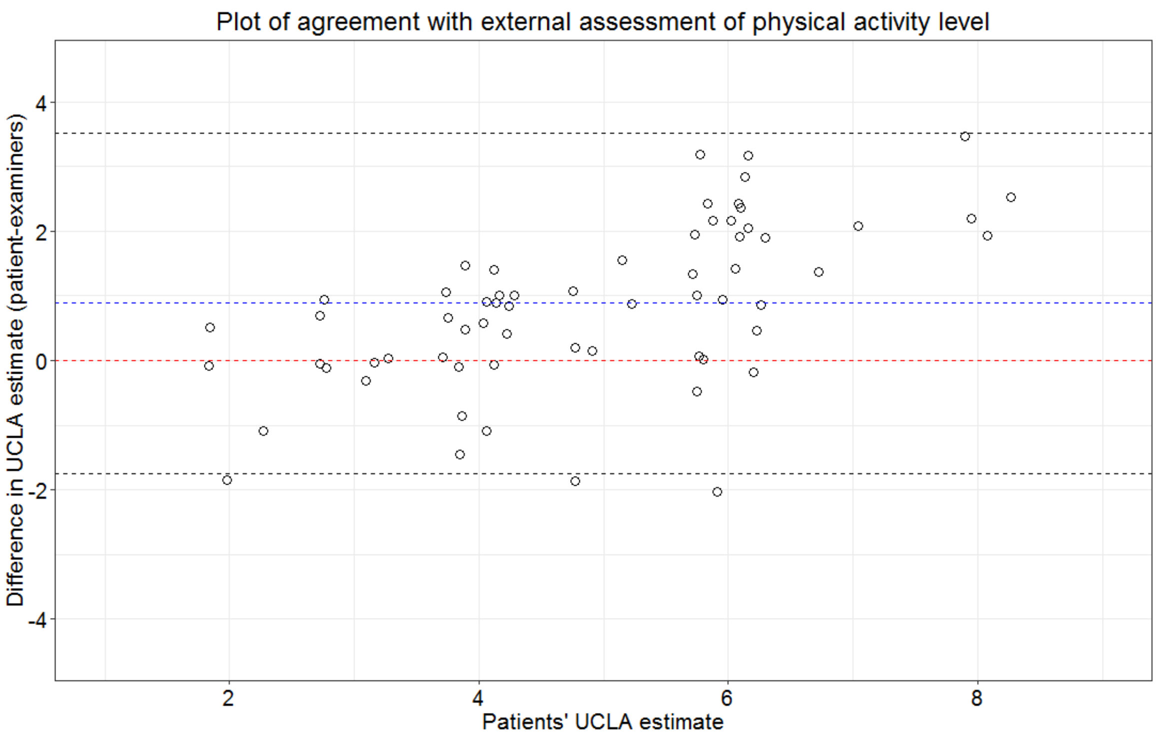


Figure 3. *Patients' UCLA estimate plotted against the difference between patient's and examiners' assessments (patient score minus mean of surgeons' and physiotherapists' scores) (modified Bland-Altman plot). The dotted lines indicate mean difference (blue) ± 2 SD (limits of agreement, black) and hypothetical perfect agreement (red). Random variance (jitter) is added to prevent over-plotting.*

Patients rated their UCLA level significantly higher than the orthopedic resident and physiotherapist did on their behalf and differences increased at higher levels of UCLA (fig. 3). In 32 cases (49%), one or both examiners marked the same level as the patient, or the patient level was in-between examiners' levels. The reliability of external interview-based assessment of patients' activity levels (table 2) was "substantial" for surgeons and "fair" for the physiotherapist, respectively. The corresponding correlations were a sign of "strong", respectively "moderate" degrees of linearity between assessments.

Univariate regression analyses revealed that patients did not report UCLA significantly different depending on sex (females 4.8 (SD 1.7), males 5.3 (1.6), CI -0.3 – 1.4, p = 0.2 (Wilcoxon test)). Neither was patient UCLA associated with age (-0.008 per year, CI -0.05 – 0.03, p = 0.7) or current knee pain (-0.1 per increase in VAS, CI -0.3 – 0.03, p = 0.1), but it was somewhat negatively associated with BMI (-0.08 per BMI unit, CI -0.15 – (-0.01), p = 0.021). Results were similar in the multiple regression analysis. In examiners' assessments, none of these factors were independently associated with activity level.

Test-retest reliability

Retest questionnaires were given to 53 patients and 43 (81%) were returned. Two patients had dated the retest form at the day of the first test and were excluded. The remaining 41 retests were filled out 8.3 days (range 1-25) after the initial test. One patient sent a blank sheet, and two patients had marked more than one answer. Only 21 of the 38 patients with complete answers reported "no change in physical activity since the first test". In this group, 13 (of 21) had perfect agreement with their initial score, 5 were 1 level apart and 1 was 2 apart. Of the 17 patients reporting "change in physical activity", 4 reported a UCLA change in the corresponding direction, 6 in the opposite direction and 7 reported the same UCLA score as before.

Construct validity and responsiveness

UCLA distributions peaked at level 4 at baseline and at level 6 one year postoperatively in both hip and knee patients (figure 4a+b). The majority of patients (79% of hip and 66% of knee patients) reported an increase in physical activity (UCLA change > 0) following joint replacement (figure 5a+b). Absolute and change UCLA scores for HA and KA patients were aggregated and grouped by overall patient satisfaction levels at 1-year (table 3).

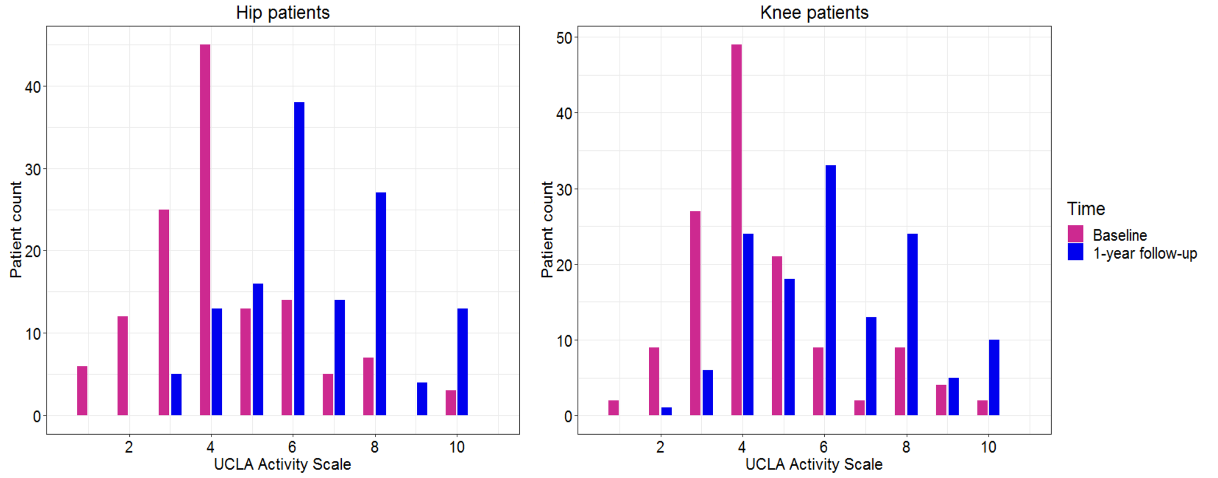


Figure 4a+b. *Distribution of pre- and 1-year postoperative UCLA scores in a) hip and b) knee replacement patients*

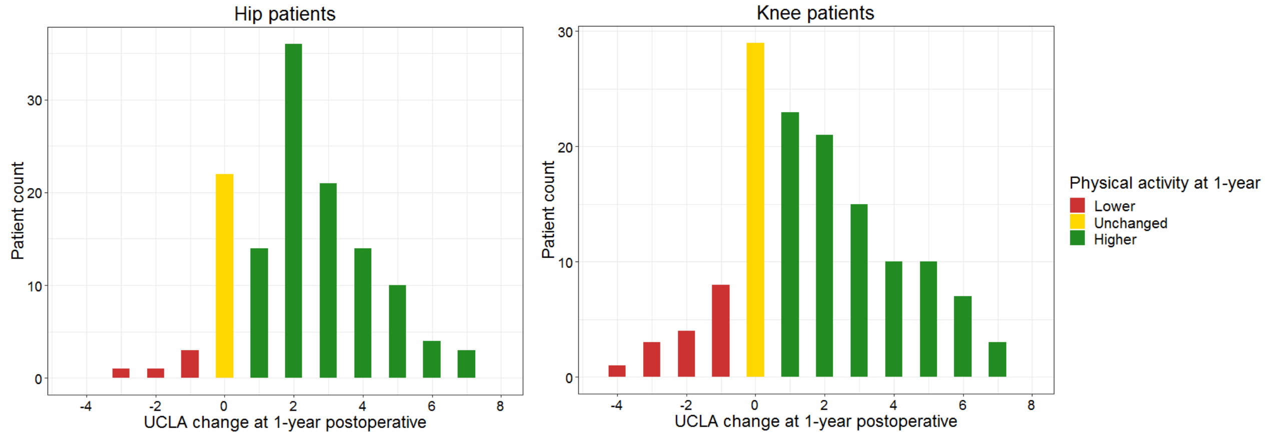


Figure 5a+b. *Distribution of UCLA change scores (1-y. follow-up minus preoperative score) in a) hip and b) knee replacement patients*

Table 3. UCLA values in Danish hip and knee arthroplasty patients

	Overall values		UCLA grouped by 1-year patient satisfaction				
	Baseline	1-y. postop.	Very dissatisfied	Dissatisfied	Neither	Satisfied	Very satisfied
Hip							
n (%)	130 (100)	130 (100)	1 (1)	0 (0)	4 (3)	16 (12)	109 (84)
M/F (n (%))	62/68 (52/48)	-	0/1	0/0	2/2	9/7	51/58
Age (y) (mean ± SD))	71 ±9	-	*	-	75 ± 6	76 ± 8	70 ± 9
UCLA (mean [median] ± SD)	4.3 [4] ± 1.9	6.6 [6] ± 1.8	*	-	6.3 [7] ± 2.1	6.2 [6] ± 2.3	6.7 [6] ± 1.8
UCLA improvement	-	2.3 [2] ± 2.0	*	-	2.5 [3] ± 1.9	1.8 [2] ± 1.8	2.4 [2] ± 2.0
UCLA ≥ 6 (n (%))	29 (22)	96 (74)	*	-	2 (50)	8 (50)	86 (79)
Floor/Ceiling (n (%))	6/3 (5/2)	0/13 (0/10)	*	-	0/0 (0/0)	0/2 (0/13)	0/11 (0/10)
Knee							
n (%)	134 (100)	134 (100)	1 (1)	11 (8)	11 (8)	40 (30)	71 (53)
M/F (n (%))	61/73 (46/54)	-	1/0	5/6	6/5	21/19	28/43
Age (y) (mean ± SD))	68 ± 9	-	*	72 ± 6	71 ± 8	66 ± 10	68 ± 9
UCLA (mean [median] ± SD)	4.5 [4] ± 1.8	6.2 [6] ± 1.0	*	4.6 [4] ± 1.7	5.1 [4] ± 2.2	6.0 [4] ± 1.6	6.8 [4] ± 1.9
UCLA improvement	-	1.7 [1] ± 2.3	*	0.0 [0] ± 1.6	0.0 [0] ± 2.0	1.3 [1] ± 1.9	2.6 [2] ± 2.3
UCLA ≥ 6 (n (%))	26 (19)	85 (63)	*	3 (27)	3 (27)	24 (60)	54 (76)
Floor/Ceiling (n (%))	2/2 (1/1)	0/10 (0/7)	*	0/0 (0/0)	0/1 (0/10)	0/1 (0/3)	0/8 (0/11)

"": Values not shown (1 patient only). Age is assessed operation date. M/F: Male/Female. Floor/Ceiling denotes the number (and percentage) of patients reporting level 1 or 10.*

With patients grouped by their individual UCLA change score after joint replacement, accompanying changes in other relevant PROMs were calculated (table 4), demonstrating weak to moderate correlations with UCLA change scores (strongest in KA patients). Knee patients who had reached the MIC of 8 OKS points reported a mean 1-year UCLA level of 6.4 compared to 5.2 in patients not reaching MIC ($p<0.036$) and UCLA change scores of 2.1 compared to -0.2 ($p<0.001$). In hip patients, the according 1-year UCLA levels were 6.6 vs. 5.7 ($p=0.2$) and according UCLA change scores were 2.4 vs. 1.5 ($p=0.3$).

Effect size (ES) was 1.2 in the HA, and 0.96 in the KA group; both ≥ 0.8 and thus "large". In these samples, there were no gender differences in UCLA levels, either ($p=0.7-1.0$). Even though the question about satisfaction could be regarded as an anchor question from which MIC could be calculated, it was not related specifically to activity. Therefore, MIC was not calculated.

Table 4. Change in relevant PROMs grouped by 1-year UCLA improvement

	UCLA improvement					Correlation
	≤ -1	0	1-2	3-4	≥ 5	Spearmans rho
Hip						
n (%)	5 (4)	22 (17)	50 (38)	35 (27)	18 (14)	
1 y. OHS (mean (SD))	45 (2)	43 (6)	44 (5)	44 (5)	44 (6)	0.09
Δ OHS (mean (SD))	24 (5)	20 (7)	22 (8)	25 (8)	27 (9)	0.21
Δ EQ-VAS (mean (SD))	13 (26)	20 (15)	24 (27)	33 (22)	38 (27)	0.29
Satisfied or very satisfied (n (%))	5 (100)	20 (91)	49 (98)	33 (94)	18 (100)	0.09
Knee						
n (%)	16 (12)	29 (22)	44 (33)	25 (19)	20 (15)	-
1 y. OKS (mean (SD))	34 (10)	38 (9)	38 (9)	39 (5)	45 (4)	0.30
Δ OKS (mean (SD))	13 (11)	14 (9)	18 (8)	19 (7)	29 (6)	0.44
Δ EQ-VAS (mean (SD))	1 (26)	16 (25)	16 (18)	26 (18)	37 (26)	0.39
Satisfied or very satisfied (n (%))	9 (56)	21 (72)	38 (86)	23 (92)	20 (100)	0.39

OHS/OKS: Oxford hip/knee Score (0-48, 48 best). Δ (Delta): change scores from baseline to 1-y. postoperatively. Correlations denote the non-parametrical correlation between the given parameter and UCLA change score (in "satisfaction", all five levels were used in correlation analyses).

Discussion

We provided a Danish translation of the UCLA Activity Scale (UCLA) for hip and knee arthroplasty patients to rate their own physical activity level on a 1-item ordinal scale ranging from 1 to 10. The process was complicated because descriptions of the original development process and purpose of the scale were not to be found in literature. When UCLA was first introduced, there was a need to determine the association between physical activity and polyethylene wear after hip and knee replacement, and the scale may have been developed for that purpose. Since then, prosthesis wear has come to play a less prominent role as cause of revisions, and the purpose of registering patient's activity levels has shifted towards evaluating the health benefits of joint replacement surgery and compare health-economic values of treatments. Nonetheless, UCLA was not redefined or redeveloped for this purpose. Patients were (probably) not involved in the original development process which remains problematic despite involvement of patients in the current translation process. Poor content validity cannot be compensated for even by otherwise good measurement properties [16]. UCLA, therefore, has no proven face- or content validity.

UCLA does have other obvious challenges: It appears to be a unidimensional scale, yet it encompasses several dimensions into one item: intensity, frequency, type, difficulty and duration of activities. This may be the price payed for brevity, but it leads to large variation in individual perception of levels. For example, the scale relies on each patient to judge whether "fitness performed once in a while" or "heavy housework performed regularly" should decide their level of activity. Since this scale was originally filled out by surgeons on behalf of patients, we found it valuable to test the agreement of concepts and levels between patients and professionals. To the best of our knowledge, such comparisons have not been made for UCLA before. The finding of systematic differences (0.2-1.6 points' lower examiner estimates) and wide limits of agreement (95% LoA -2.2 - 4.4) between patients and examiners' estimates underline that patient-reported outcomes cannot be replaced by surgeon or physiotherapist's estimates. This, along with the previous findings [29], suggests that comparison of UCLA levels among patients should be made with great caution.

Based on patient feedback, we found it necessary to make comprehensive changes in layout, wording and activity examples, and it was clear that physical activities of Danish hip and knee arthroplasty patients today differ from American standards more than 3 decades ago. As the scale was originally filled out by surgeons, we added a short instruction text to make the scale self-explanatory to patients. Despite several attempts to clarify the meaning (8 iterative rounds of changes), 11 patients (15%) still misunderstood the task. With an electronic version allowing only one response, this problem would be overcome and patients might even be guided towards a better understanding of the scale [8]. The layout used in our version, where pairs of levels refer to the same set of example activities, is likely to be responsible for some of the misunderstandings; patients may think they need to give an answer for each set of examples. The score distributions reveal how patients (and examiners, for that matter) were more likely to choose options 4, 6, 8 and 10, where activities were performed "regularly", as opposed to "once in a while" in adjacent levels 3, 5, 7 and 9, maybe because people tend to have regularity in their daily life and leisure time. Thus, we may have failed to indicate that frequency, not regularity, was in focus here. We did not find it fruitful to define frequency in detail by writing e.g. "1-2 times per week", etc., considering the vagueness of the other concepts of the scale. Future studies might provide more even distributions of scores if the term "regularly" is replaced by "often". Publications based on other language versions of UCLA have not reported histograms of score distributions, but we find it likely that other versions have had similar problems.

Limitations and strengths

Beside the already mentioned limitations, we are not sure of the magnitude of measurement error within patients as we can hardly make conclusions based on the 21 patients in test-retest analyses. Preferably, future studies of measurement properties of the Danish UCLA activity should include the suggested changes in wording, repetition of test-retest studies in a larger group of patients and validation against other and more comprehensive patient-reported activity scales, accelerometers or performance-based measures, recognizing though, that the underlying construct may be very different across these measurement methods [9, 22]. Ideally, patients who had revision surgery during the observation period, should also be included in analyses in order to cover the true spectrum of outcomes. Finally, responsiveness studies should include an anchor question regarding change in physical activity alone to allow for an anchor-based calculation of MIC.

Interpretation and generalizability

Despite several changes from the original version, UCLA scores in Danish hip and knee patients were well in line with findings from other international versions: baseline UCLA in HA/KA patients were 4.3/4.5 in our study, corresponding to 4.3/4.2 in California [24]. As hypothesized, change UCLA scores 1 year after joint replacement were between 1 and 3 (2.3 HA/1.7 KA). 4 out of 5 hip patients and 2 out of 3 knee patients reported an improvement in physical activity level at 1-year follow-up. The median UCLA changed from level 4 preoperatively to 6 at 1-year follow-up in both groups, which was in accordance with a British study of 261 patients of mean age 59 years [21]. In Scott's sample, the number of KA-patients reporting to be very physically active (UCLA scores dichotomized ≥ level 6) increased from 37 to 72%. In our sample, the according proportions changed from 19 to 63% (KA) and from 22 to 74% (HA), thus the share of very active patients more than tripled in each group. It must be kept in mind that with patient-reported levels of physical activity, we cannot be sure that *ability* to be physically active was not confused with *actual* physical activity performed. This would also be the case at baseline, but particularly after successful surgery, patients may feel obliged to report higher levels, leading to overestimation of change.

In accordance with previous studies [18], we found no floor or ceiling effects. As expected, there were only poor to moderate correlations with other PROMs and overall patient satisfaction. In knee patients, but not in hip patients, absolute UCLA and UCLA change score at 1-year differed between patients who had achieved an 8 points' gain in Oxford Knee/Hip Score (MIC).

Conclusion

No questionnaire should be sent out to patients without a clear intention of how to use the responses. Without knowledge of the original intention behind UCLA Activity Scale, quantification of content validity was difficult. Thus, interpretation of scale results remains problematic. The brevity and feasibility of the UCLA Activity Scale is an obvious and important advantage but compromises the interpretability further. Based on the findings of this study, we do not recommend UCLA Activity Scale for discrimination among individual patients. Rather, the scale has its primary advantage in documenting subjective change in physical activity within individuals during the course of surgery: 4 in 5 hip patients and 2 in 3 knee patients reported to be more physically active 1 year after joint replacement surgery. Presentation of such information may be valuable to future osteoarthritis patients who consider joint replacement surgery and want to know what to expect.

Author contributions

All authors critically revised the final manuscript. AM was the main responsible and took part in all parts of the project. MK, ML and AO took part in planning the study. SS participated in translation. CH and PH were involved in revision of the primary version and conducted to validity studies, while CV conducted responsiveness studies. AM performed all statistical analysis.

Acknowledgments

Authors thank all involved patients for their time and valuable contributions. We thank Henrik M. Schrøder and colleagues (nurses, secretaries and surgeons) at Naestved Hospital and nurses and secretaries at Gentofte Hospital for good cooperation, Carsten Bogh Juhl for helpful advice during study planning and Tobias Wirenfeldt-Klausen for statistical consulting.

References

1. Amstutz HC, Thomas BJ, Jinnah R, Kim W, Grogan T, Yale C. Treatment of primary osteoarthritis of the hip. A comparison of total joint and surface replacement arthroplasty. *J Bone Jt. Surg Am*. 1984;66:228–241.

2. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Med Res Methodol*. 2011;11:152; author reply 152.

3. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol*. 2015;68:73–79.

4. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J. Bone Jt. Surg. - Ser. B*. 1996;78:185–190.

5. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Jt. Surg Br*. 1998;80:63–69.

6. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J. Clin. Epidemiol.* 2015;68:435–441.

7. Ghomrawi HMK, Lee Y, Herrero C, Joseph A, Padgett D, Westrich G, Parks M, Lyman S. A Crosswalk Between UCLA and Lower Extremity Activity Scales. *Clin. Orthop. Relat. Res.* 2017;475:542–548.

8. Gudbergsen H, Bartels EM, Krusager P, Waehrens EE, Christensen R, Danneskiold-Samsoe B, Bliddal H. Test-retest of computerized health status questionnaires frequently used in the monitoring of knee osteoarthritis: a randomized crossover trial. *BMC Musculoskelet Disord*. 2011;12:190.

9. Hossain FS, Konan S, Patel S, Rodriguez-Merchan EC, Haddad FS. The assessment of outcome after total knee arthroplasty: are we there yet? *Bone Jt. J*. 2015;97–b:3–9.

10. Ingelsrud LH, Roos EM, Terluin B, Gromov K, Husted H, Troelsen A. Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. *Acta Orthop*. 2018;89:541–547.

11. Jin X, Al Sayah F, Ohinmaa A, Marshall DA, Smith C, Johnson JA. The EQ-5D-5L Is Superior to the -3L Version in Measuring Health-related Quality of Life in Patients Awaiting THA or TKA. *Clin Orthop Relat Res*. 2019;477:1632–1644.

12. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96–106.

13. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33:159–174.

14. Mainland BJ, Amodeo S, Shulman KI. Multiple clock drawing scoring systems: simpler is better. *Int J Geriatr Psychiatry*. 2014;29:127–136.

15. McKenna SP, Doward LC. The translation and cultural adaptation of patient-reported outcome measures. *Value Heal.* 2005;8:89–91.

16. Mokkink LB, Prinsen CAC, Patrick D, Alonso J, Bouter L, de Vet HCW, Terwee CB. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). User manual, version 1.0. 2018.

17. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. *J Bone Jt. Surg Br*. 2007;89:1010–1014.

18. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? *Clin Orthop Relat Res*. 2009;467:958–965.

19. RCoreTeam. RStudio version 1.1.463. https://www.R-project.org/. R: A Language and Environment for Statistical Computing.

20. Rolfson O, Bohm E, Franklin P, Lyman S, Denissen G, Dawson J, Dunn J, Eresian Chenok K, Dunbar M, Overgaard S, Garellick G, Lubbeke A, Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty R. Patient-reported outcome measures in arthroplasty registries Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for selection, administration, and analysis. *Acta Orthop*. 2016;87 Suppl 1:9–23.

21. Scott CEH, Turnbull GS, MacDonald D, Breusch SJ. Activity levels and return to work following total knee arthroplasty in patients under 65 years of age. *Bone Jt. J*. 2017;99–b:1037–1046.

22. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sport. Med*. 2003;37:197–206; discussion 206.

23. Skou ST, Roos E. GLA:D (Good Life with osteoArthritis in Denmark) - Annual Report. 2014:28.

24. SooHoo NF, Li Z, Chenok KE, Bozic KJ. Responsiveness of patient reported outcome measures in total joint arthroplasty patients. *J Arthroplast.* 2015;30:176–191.

25. Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting quality of life instruments. *Value Heal.* 2004;7 Suppl 1:S27-30.

26. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.

27. Terwee CB, Bouwmeester W, van Elsland SL, de Vet HC, Dekker J. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthr. Cartil.* 2011;19:620–633.

28. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011.

29. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. *J Arthroplast.* 1998;13:890–895.

Study IV: ROM

The Journal of Arthroplasty 33 (2018) 2875–2883

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Primary Arthroplasty

Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale

Anne Mørup-Petersen, MD [a,*], Pætur M. Holm, PT, MSc [b,c], Christina E. Holm, MD [d], Tobias W. Klausen, MSc [e], Søren T. Skou, PT, PhD [b,c], Michael R. Krogsgaard, MD, PhD [f], Mogens B. Laursen, MD, PhD [g], Anders Odgaard, MD, DMSc [a]

[a] Department of Orthopedic Surgery, Copenhagen University Hospital Gentofte, Copenhagen, Denmark
[b] Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Slagelse, Denmark
[c] Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark
[d] Department of Orthopedic Surgery, Copenhagen University Hospital, Rigshospitalet, København Ø, Denmark
[e] Department of Hematology, Copenhagen University Hospital, Herlev, Denmark
[f] Department of Orthopedic Surgery, Copenhagen University Hospital, København NV, Denmark
[g] Department of Orthopedic Surgery, Aalborg University Hospital, Farsø, Denmark

ARTICLE INFO

Article history:
Received 5 March 2018
Received in revised form
17 April 2018
Accepted 2 May 2018
Available online 7 June 2018

Keywords:
range of motion
patient-reported
patient-assessed
validation
knee arthroplasty
knee osteoarthritis

ABSTRACT

Background: Knee arthroplasty does not always require extensive long-term follow-up. If knee range of motion (ROM) could be assessed reliably by patients, some follow-up visits might be replaced by patient-reported outcome measures, and this additional information could be reported directly to registers. We developed and tested the validity and reliability of a simple scale for patients to self-report their passive knee ROM.

Methods: Through an iterative process, we created a 2-item scale with 11 illustrations of knee motion in 15° increments. The validity and reliability was tested in knee osteoarthritis and arthroplasty patients at different treatment stages, many with poor ROM. Patient estimates were compared to passive goniometer measurements performed blindly by a physiotherapist and a junior orthopedic surgeon.

Results: The mean difference between 100 patients' (70.9 years) estimates and goniometer measurements was −0.7° (standard deviation, 12.3°) for flexion and 1.1° (standard deviation, 11.6°) for extension, both not significant. Correlation was 0.79 and 0.63, and kappa values at retest were 0.84 and 0.66. For flexion < 110°, sensitivity of patient estimates was 88% and specificity was 88%. For a limit of 100°, values were 95% and 81%. For extension deficits >10°, sensitivity was 78% and specificity 70%. Values were 100% and 66% for a 15° limit.

Conclusion: The Copenhagen Knee ROM Scale is a patient-friendly and feasible alternative to passive ROM measurement for registers, research, and selected clinical use. This scale appears reliable and valid compared to reports of similar tools, and patient estimates are better correlated to goniometer measurements.

With increasing attention to the advantages of the use of patient-reported outcome measures in knee arthroplasty, it has been suggested that patient-reported outcome measures replace some postoperative clinical follow-up visits in uncomplicated cases [1–4]. An important barrier, however, is that information about range of motion (ROM) is not available if patients do not attend a healthcare clinic in person.

Attempts have been made to have patients self-report ROM, and the need for a tool to make this possible has been recognized [1,2,4–6]. For surgeons to rely on patient-reported ROM to replace a

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to https://doi.org/10.1016/j.arth.2018.05.011.

Funding: This work was indirectly supported by the Health Research Fund of the Capital Region of Denmark, grant 2015, part one. The funding source had no influence on the content of the study.

* Reprint requests: Anne Mørup-Petersen, MD, Department of Orthopedic Surgery, Herlev and Gentofte Hospital, Gentofte Matriklen, Kildegaardsvej 28, Hellerup 2900, Denmark.

https://doi.org/10.1016/j.arth.2018.05.011
0883-5403/

clinical follow-up visit, the tool must be valid and sensitive. The same applies for use in research and registries. Previous studies have reported promising results [3–6], but we sought to explore whether a new, simple patient-reported ROM tool, in which patient-friendliness was highly prioritized, could provide satisfactory accuracy and sensitivity. The purpose of this study was to develop an illustration-based scale for patients to report passive knee ROM and to test the validity and reliability among knee osteoarthritis (OA) and knee arthroplasty patients.

Materials and Methods

Development Process

Our first focus was to design a questionnaire, based on drawings, that patients of any adult age would easily understand and be able to complete unassisted at home. The process was guided by 3 relevant guidelines: (1) Guidelines for Reporting Reliability and Agreement Studies (GRRAS); (2) The STARD Statement for reporting studies of diagnostic accuracy; and (3) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties of health status measurement instruments [7–12].

We met with 18 individual knee OA patients (7 men, 11 women; mean age, 69.9 years) who were facing or had just undergone knee arthroplasty. They were asked to show in their own preferable way how much they could bend and straighten their affected knee. We observed that the majority of patients got up from the chair or bed to show their extension, although some remained seated with their leg stretched out with the heel on the floor, or balancing the leg in the air in front of themselves. To show flexion, most patients sat on a chair or remained in bed. Many patients used their hands to pull the ankle backward.

Through an iterative process of testing and improving our draft illustrations, a total of 34 knee arthroplasty patients (23 preoperative and 11 postoperative patients: 13 men, 21 women; mean age, 70.4 years) were shown several drafts of knees in different positions from a lateral view. Patients were asked to describe what they saw in the illustrations and mark the option that fit their knee motion. Some drafts had dotted horizontal and vertical lines to aid estimation of angles. However, with exception of an engineer and a carpenter, most people found the lines more confusing than helpful. Adding a seat and the contralateral leg as navigation points and adding arrows to show the direction of force gradually enhanced the patients' understanding of the intentions of this tool (Fig. 1). Instructions were made short, here regarding flexion: "How much can you bend your knee? Please push your lower leg as far back as possible. You can use your hand to pull your lower leg in the direction of the arrow. Tick the box that fits your situation."

For flexion, we found 6 pictures to be appropriate: 60°, 75°, 90°, 105°, 120°, and 135°. For extension, 5 illustrations of 45°, 30°, 15°, 0°, and −15° were found suitable. We chose 15° increments between the pictures for 3 reasons: Firstly, only differences above 5°-10° represent a true difference in ROM [13,14]. Secondly, with 10° intervals even the authors of this article were unable to tell the difference between neighbor illustrations. In the development phase, we noted that many patients exaggerated their ROM, both in terms of good and bad results. Particularly flexion contractures were overestimated. Collins et al [5] reported the same tendency when using 5°-10° intervals. Therefore, we considered more options to be redundant, as patients would use the scale widely no matter the underlying intervals between measures. Thirdly, our goal of making the questionnaire very easy to overview would be compromised with a higher number of illustrations.

We deliberately chose not to write the underlying angle on each picture because we wanted patients to report their unbiased perception of pictures. If angles had been shown, there would be a risk of priming patients to a certain answer, in case they had recently been told their exact ROM measure, for example, by their physiotherapist.

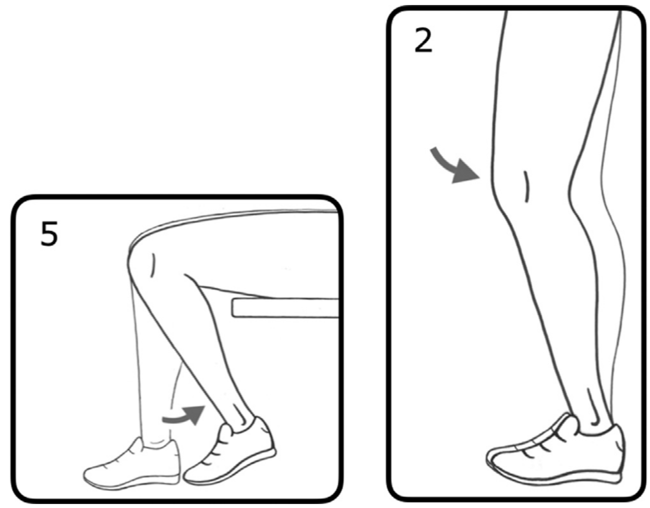


Fig. 1. Examples of illustrations of flexion and extension, respectively. Full questionnaires are available in Appendix A.

Options were placed in one row with the best score last to show a logical direction of motion. To meet patients with locked bandages or extremely limited motion, we made extra options named (for flexion) "Impossible. I am not able to bend my knee as much as in picture no. 1".

Although all illustrations show left-side knees, no patients were in doubt of which knee to think of. Some patients asked how much it was meant to hurt during testing. However, because pain level varies greatly, this subject could not be fit into instructions in a sensible manner.

The development process ended when there were no longer any new comments to facilitate meaningful changes in layout or wording, and patients understood the task without further explanation. The final version, Copenhagen Knee ROM Scale (CKRS) can be viewed in an English version in Appendix A and is available free of charge (with English or Danish text) at www.procordo.com/docs/copenhagen_rom.

Translation

Questionnaires were evaluated and tested in the original Danish version. The wording was translated into English for publication, independently by 3 bilingual persons: 1 native English layperson and 2 native Danish doctors (1 resident and 1 orthopedic knee arthroplasty surgeon with 4 years of experience from English hospitals). The 3 versions were combined to a final version by the first author. When in doubt, the native English layperson had the final say. The resulting English version of CKRS has not been evaluated among English-speaking OA patients.

Clinical Testing

A patient-reported ROM tool is probably of most clinical value in the follow-up period after knee arthroplasty. However, it may be of great value for patients considering arthroplasty to be informed of what knee motion to expect in the months and years following surgery, compared to the knee motion they have with OA. Therefore, we found the whole spectrum of patients, from the first visit in the arthroplasty clinic to the possible complications several years after arthroplasty, to be of interest when validating this tool. For clinical reliability testing, we aimed for a test group of patients reflecting this diversity. Mixing preoperative and postoperative patients also gave the opportunity of selecting patients so that all degrees of knee motion were present, and thereby be able to test the validity of CKRS in the entire range of the ROM scale.

Patients were included in both the orthopedic arthroplasty ward and the arthroplasty outpatient clinic over a period of 5 days. Inclusion criteria were age $\geq$ 40 years and clinical visit due to knee OA or hospital stay due to knee arthroplasty operation of any kind. Patients with bilateral knee OA were instructed to answer CKRS for the knee with the most restricted motion. Patients showing signs of dementia or confusion were excluded if they failed a "clock drawing test" [15], as were patients with poor Danish language skills without someone to translate for them. We also excluded hospitalized patients who were unable to get out of bed and stand on their own (walking aids were allowed, however).

First, patients filled in the CKRS paper version without the opportunity to ask any questions. Visiting relatives were allowed to stay in order to mimic the situation at home. Completion time was not measured (to avoid stressing the patients), but it was our impression that the far majority of patients completed the form within 1-3 minutes. They were, however, allowed as much time as they needed. Immediately after completion, patients met a junior orthopedic registrar and 1 of 2 experienced physiotherapists. Patients were instructed not to reveal their answers, which all obeyed. Sitting on a normal chair, the patient demonstrated his or her maximal flexion once for each examiner, who then filled in the CKRS while the other examiner turned his or her back for blinding purposes. This was repeated for extension with the patient standing up. Patients were told to press on the knee or pull the lower leg with their own hands. Examiners were only permitted to palpate for bony landmarks.

Subsequently, goniometer ROM measurements were made using the same blinding strategy with the patient lying on an examination table wearing only underwear on the lower body. We used a long goniometer (30 cm/12 inches, 1° increments) and navigated for bony landmarks: the greater trochanter, the lateral epicondyle of the femur, and the lateral malleolus [16]. External hand pressure was applied by the examiner and the patient was told to say stop when it was enough. For extension measurements, we placed a firm cylinder back roll under the Achilles tendon.

Between each examination, the knee was left in a relaxed position. The order of surgeon and physiotherapist examination was random. After all measurements were completed by the examiners individually, a consensus measurement was made by both examiners in collaboration.

Reproducibility

Examiners' CKRS estimates and ROM measures were kept secret to patients. Patients were given a retest questionnaire together with a prepaid envelope and were instructed to fill in the forms 7-10 days after the first session. Patients who participated during the first days after surgery were asked to perform the retest 1 or 2 days later because fast improvement was expected. Patients who were scheduled for surgery between test and retest were omitted from this part of the study.

Before filling in the CKRS again, patients were asked to confirm the affected side (left/right) and answer whether they had experienced any change in knee motion since the first examination. Only patients reporting "no change in knee motion" were included in retest analysis. No goniometer measurement was made on this occasion because the subject of interest was retesting patients' perception of the scale. Retest questionnaires received later than 6 weeks after the first testing and retests dated on the day of the first examination were excluded from retest analysis.

Statistical Analysis

Based on reports from similar studies, we proposed a sample size of 100 patients [4–6,17,18]. Sample size calculation aiming for a power of 0.8 based on an expected Pearson correlation coefficient of 0.7, a null correlation coefficient of 0.5, and 2-sided alpha 0.05 suggested a sample size of 80. Because we could not expect a normal distribution of answers, more patients were needed so we included 108 patients.

Descriptive statistics was made for all continuous variables including mean differences (mean goniometer measurement minus patient estimate) and 95% confidence intervals [CI]. Paired *t*-tests were used for comparisons.

Goniometer measurement was regarded as a gold standard in our calculations [18]. To describe the measurement error of the CKRS tool, we calculated overall 95% limits of agreement (LoA) as mean difference $\pm$ 1.96 $\times$ standard deviations (SD). With patients grouped by their CKRS answer, also group mean, SD, range, and LoA were calculated (LoA only for groups larger than 15 patients) to ensure clinical applicability, because measurement error was expected to vary with ROM measures.

Sensitivity and specificity for clinically relevant limits were calculated with special consideration to comparability to previously published methods. For the same reason, also Pearson correlation coefficients between methods were calculated. These, however, require equal intervals between answer options. Because we could not guarantee that patients perceived intervals between illustrations to be equal, we also calculated Spearman rank correlation coefficient, which compares only the ranking of subjects.

From the mean goniometer measurement, we calculated the "correct" CKRS answer that patients ideally should give. For example, flexion option 5 (120°) should cover the range from 112.5° to 127.5°. Absolute ROM measures (flexion minus extension) were not calculated due to their limited clinical relevance. Test-retest reliability was based on weighted kappa, paired *t*-tests, and percentage agreement between patients' first estimates and their retest estimates.

P values below .05 were considered significant and were reported when relevant. All *P* values were 2-sided. Statistical analyses were made in SAS Statistical software (SAS University Edition, version 3.6, Cary, NC). Ethical approval was provided by the National Committee of Health Research Ethics (Jr. no. 16030260) and data management was approved by the Danish Data Protection Agency (Jr. no. 2012-58-0004). Raw data for the primary tests are available in Appendix B.

Results

A total of 113 patients were asked to participate, but 5 declined (excused by business or tiredness), so 108 knee OA patients (108 knees) were included (Fig. 2). Three were excluded before testing: 1 had dementia and 2 were unable to get out of bed on their own. After testing, 5 more patients were excluded: one because goniometer measurements were performed on the contralateral knee by mistake, 1 patient failed to answer page 2, and 3 patients had marked more than one option. The final test group ready for data analysis consisted of 100 patients: 59 patients with knee arthroplasty and 41 patients with knee OA. Patient characteristics are presented in Table 1.

The goniometer measurements of the surgeon and the physiotherapists were well aligned; mean difference was 0.8° (SD, 4.2°;

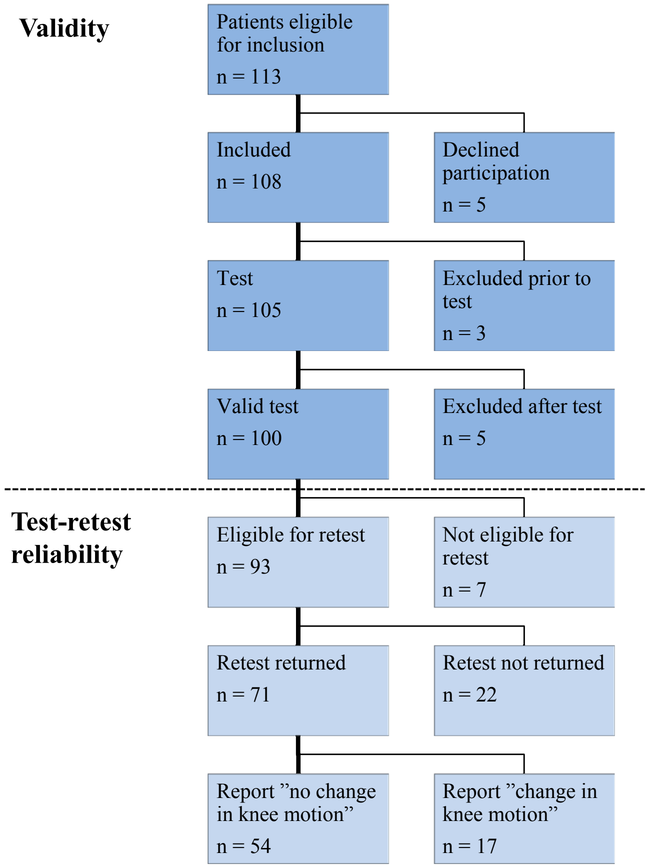


Fig. 2. Flow diagram of patients in validity and reliability studies.

range, −13° to 9°; *P* = .06) for flexion and 1.1° (SD, 3.0°; range, −7° to 7°; *P* < .001) for extension. Therefore, the consensus measurements were deemed unnecessary and left out of analysis as they are not typical of everyday practice. In the following, "goniometer measurement" refers to the mean of the 2 examiners' measurements.

Flexion

Goniometer measurements of flexion ranged from 62.5° to 150.5° (mean, 115.7°; SD, 19.6°). In CKRS, only 1 examiner and no patients made use of option 1, and no one marked option 0.

Table 1
Characteristics of Patients (N = 100) in the Clinical Test Group.

Patient Characteristics				
(N = 100)	Mean	Mean CI	Median	Range (Min-Max)
Age (y)	70.9	69.0-72.8	72.6	44.3-89.3
BMI (kg/m^2; N = 99)	28.7	27.6-29.9	27.5	19.6-46.7
Sex	Men	N = 35		
	Women	N = 65		
Knee arthroplasty	No	N = 41		
	Yes	N = 59	→	45 TKA, 13 mUKA, and 1 revision TKA Time after surgery: mean, 14.8 mo (median, 3 mo; range, 0 d to 16 y)

CI, confidence interval; BMI, body mass index; TKA, total knee arthroplasty; mUKA, medial unicompartmental knee arthroplasty.

Fifty-five percent of patients had chosen the "correct" picture and 94% were within 1 adjacent option. No one was further than 2 options away from the correct answer.

The mean difference between patient estimates in 15° intervals and goniometer measurement was −0.7° (CI, −3.2° to 1.7°; *P* = .56). Differences were normally distributed with overall SD 12.3° and total range from −32° to 28°. Hence, overall 95% LoA was 0.7° ± 24.0°.

Patient-reported flexion on CKRS had a strong Pearson correlation of 0.79 (CI, 0.70-0.85) to goniometer measurements. The according Spearman rank correlation was 0.80 (CI, 0.71-0.86). Figure 3 shows boxplots of goniometer measurements for patients grouped by their own ROM estimates on CKRS. Measurements, SD, and LoA for each group are listed for clinical applicability in Table 2.

Sensitivity and specificity was calculated for clinically relevant values (Table 3). For example, if flexion ≥ 100° is considered acceptable, CKRS is able to detect 95% of patients with an unsatisfactory flexion using cutoff between illustrations 4 and 5. The specificity in this case is 81%.

In this population, where many patients had poor knee motion, 64% of patients marked ≥ option 5. The negative predictive value, that is, the probability that a patient who marked ≥ option 5 did in fact have flexion ≥ 100° was 98%. Correspondingly, the positive predictive value of having flexion < 100° was 100% for patients marking ≤ option 2. Similar calculations for other relevant values are listed in Table 3.

Extension

Passive extension measurements ranged from −8.5° to 35° (mean, 5.8°; SD, 6.5°). All CKRS illustrations except option 0 (>45°) were used by the patients. The correct illustration was chosen by 45% of patients, 99% were within 1 option from the correct and 1 patient was 2 from the correct answer.

The mean difference between patient estimate and goniometer measurement was 1.1° (CI, −1.2° to 3.4°), *P* = .35. Overall differences were normally distributed with an SD of 11.6° and LoA 1.1° ± 22.8° (total range, −34° to 22.5°). However, goniometer measurements for each CKRS group reveal how patients perceive only 2°-9° intervals between pictures instead of the actual 15° intervals that drawings are measured by (Table 4). For example, the mean goniometer measurement for patients marking picture 5 (−15°) was 0.7°, and for patients marking picture 4 (0°) it was 4.9°. Boxplots of patient estimates against their goniometer measurements illustrate the same phenomenon: the slope is not as steep as if there was perfect agreement (Fig. 4).

On group level, variation was far lower than the overall variation; SD was 5.1°, 3.9°, and 4.6°, respectively, for the 3 pictures covering 89% of patients tested.

Pearson's correlation coefficient was 0.63 (CI, 0.49-0.73; moderate) between patient estimates and goniometer measurements and 0.57 (CI, 0.42-0.69) using Spearman's rank correlation coefficient.

Sensitivity and specificity values are listed in Table 3. If 15° is considered acceptable passive extension, a cutoff between option 3 and 4 offers a sensitivity of 100% at the cost of a specificity of 66%. If extension limit is lowered to 10°, the according values are 78% and 70%, respectively. In this population, the negative predictive value, that is, the chance of not having an extension deficit > 10° when answering ≥ option 4 was 93%. By contrast, the positive predictive value of having an extension deficit > 10° was 82% for patients marking ≤ option 2.

Reliability

Retest questionnaires were handed out to the 93 patients who were not awaiting surgery within a week (45 of patients already

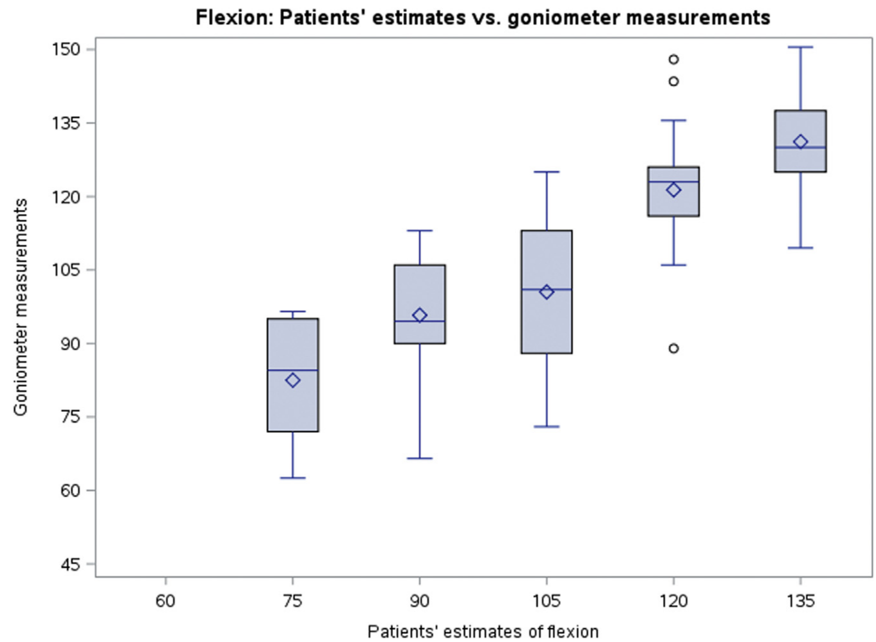


Fig. 3. Boxplot of goniometer measurements of flexion grouped by patients' estimates on Copenhagen Knee ROM Scale.

had a knee arthroplasty). We received 71 answers dated mean 8.7 days later (median, 8 days; range, 1-24 days). Fifty-four patients had replied "no change in knee motion" and so were eligible for retest analysis (Fig. 2). Of these, 32 were arthroplasty patients and 22 were OA patients.

In both flexion and extension, there were no overall differences between values in test and retest (P = .25 and .35, respectively). Forty-five patients (83%) gave the exact same answer regarding flexion as in their first test. For extension, the number was 36 (68%) and perfect agreement on both parameters was reached in 32 (59%) of the cases. Weighted kappa value for flexion was 0.84 (CI, 0.74-0.94) which represents "almost perfect" test-retest reliability [19]. For extension, weighted kappa reached 0.66 (CI, 0.52-0.80) representing "substantial" reliability.

The Role of Age, BMI, and Arthroplasty Status

We found no correlation between body mass index (BMI) and measurement error (= absolute difference between patient estimate and goniometer measurement). Nor was there any correlation between increasing age and measurement error. On the contrary, the only significant outcome was for flexion, where a weak Pearson correlation of −0.20 (CI, −0.38 to −0.01; P = .04) indicated that older patients made more accurate estimates than younger patients. A comparison of measurement error between the 41 OA patients and the 59 arthroplasty patients revealed no difference in their ability to estimate ROM using this scale (P = .75 for flexion and 0.68 for extension using unpaired t-test of unequal variances).

Examiner's Estimates of ROM

Both examiners were aware of the underlying angles behind CKRS illustrations. Their estimates of CKRS before measuring agreed well with passive goniometer measurements: the mean difference was 1.6° for flexion (SD, 6.7°; Pearson's r, 0.94) and 1.1° for extension (SD, 4.6°; r, 0.84). Examiners appointed the correct flexion option in 67% of cases, the adjacent option in 32.5%, and were 2 apart in 0.5% of cases. For extension, 77% of estimates were correct and 23% were 1 away.

Table 2
Goniometer Measurements of Flexion. Patients Are Grouped (1-6) by Their Own Estimate of Flexion on Copenhagen Knee ROM Scale.

Measurements of Flexion						
Patient Estimate (picture no.)	1	2	3	4	5	6
Underlying (hidden) Measure	60°	75°	90°	105°	120°	135°
Patients (N = 100)	0	6	11	19	29	35
Goniometer measurements (°)						
Mean	—	82.5	95.8	100.5	121.4	131.2
95% CI of the mean		71.7-93.3	88.0-103.5	93.5-107.5	117.2-125.6	128.0-134.4
Median		84.5	94.5	101.0	123.0	130.0
SD	—	(13.2)	(12.9)	15.2	11.3	9.5
95% LoA (mean ± 1.96 SD)[a]	—	[a]	[a]	70.7-130.3	99.2-143.5	112.5-149.8
Total range (min-max)	—	62.5-96.5	66.5-113.0	73.0-125.0	89.0-148.0	109.5-150.5

LoA is only calculated for groups larger than 15 patients. SDs are in parentheses for the same reason.
CI, confidence interval; SD, standard deviation; LoA, limits of agreement.
[a] 95% LoA, limits of agreement based on the SD for each group (column).

Table 3
Limits of Goniometer Measurements of Knee Motion and Their Respective Sensitivities and Specificities for Detecting Problematic Knee Motion Using Different Thresholds (Cutoff Values) in Copenhagen Knee ROM Scale (CKRS).

Limits of Knee Motion					
Limit of Knee Motion	CKRS Threshold (Between Pictures no.)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%) (Population Specific)	Negative Predictive Value (%) (Population Specific)
Flexion < 90°	2-3	31	98	67	90
	3-4	46	87	35	92
	4-5	**92**	**72**	**33**	**98**
	5-6	92	72	33	98
Flexion < 100°	2-3	27	100	100	83
	3-4	55	94	71	88
	4-5	**95**	**81**	**58**	**98**
	5-6	100	45	34	100
Flexion < 110°	2-3	19	100	100	72
	3-4	47	97	88	80
	4-5	**88**	**88**	**78**	**94**
	5-6	97	50	48	97
Flexion < 120°	4-5	65	96	94	72
	5-6	**90**	**63**	**72**	**86**
Extension ≥ 15°	2-3	57	92	36	97
	3-4	**100**	**66**	**18**	**100**
Extension ≥ 10°	2-3	50	98	82	90
	3-4	**78**	**70**	**36**	**93**
	4-5	**100**	**26**	**23**	**100**
Extension ≥ 5°	2-3	22	100	100	57
	3-4	57	78	72	66
	4-5	90	31	55	76

Guide to interpretation of values: Sensitivity: The chance that a patient with a knee motion worse than the specified limit is identified as having a knee motion problem using this CKRS threshold.
Specificity: The chance that a patient with better knee motion than the specified limit is identified as having acceptable knee motion using this CKRS threshold.
Positive predictive value: The chance that a patient reporting knee motion worse than the specified CKRS threshold does have a knee motion worse than the specified limit (note that this value is population specific).
Negative predictive value: The chance that a patient reporting better knee motion than the specified CKRS threshold does have a knee motion better than the specified limit (note that this value is population specific).
Negative and positive predictive values in the test population are listed. Authors find the numbers in bold to be of largest clinical relevance.

Discussion

Validity

Our aim to develop an easily understandable questionnaire was obtained. The scale measures the intended items and through the whole ROM there were no severe outliers. Furthermore, measurement error was unaffected by BMI and age.

We consider it a strength that CKRS was tested in a diverse group of knee OA and arthroplasty patients. Our inclusion of many patients with poor ROM has furthermore confirmed the validity of the scale in the whole range of both flexion and extension.

It is an advantage that the whole leg and the contralateral leg are both visible. Arrows clearly indicate which motion is requested and instructions are condensed to a minimum. To enhance patient relevance, we have let patient positioning be directed by patient preferences. Drawings are simplistic and, as opposed to photographs, they are neutral in terms of race, sex, and age.

Although option 0 was not used at all and flexion option 1 was only used once in the test setting, we have kept both options in CKRS as they were suggested by patients in the development phase and therefore relevant. In a prospective study of 1600 knee arthroplasty patients, where CKRS was applied 5 times during the first year after surgery (unreported, personal information), patients marked option 0 in 0.3% and option 1 in 2.5% of cases.

Table 4
Goniometer Measurements of Extension.

Measurements of Extension					
Patient Estimate (picture no.)	1	2	3	4	5
Underlying (hidden) Measure	45°	30°	15°	0°	−15°
Patients (N = 100)	2	9	28	40	21
Goniometer measurements (°)					
Mean	23.0	14.4	6.8	4.9	0.7
95% CI of the mean	−1.0 to 47.0	10.0 to 18.9	4.8 to 8.7	3.7 to 6.1	−1.3 to 2.7
Median	23.0	14.5	6.5	4.5	0.5
SD	(17.0)	(6.7)	5.1	3.9	4.6
95% LoA (mean ± 1.96 SD)[a]	[a]	[a]	−3.3 to 16.8	−2.7 to 12.5	−8.4 to 9.8
Total range (min-max)	11.0-35.0	8.0-30.0	−6.0 to 17.5	−5.5 to 13.5	−8.5 to 7.5

Patients are grouped (1-5) by their own estimate of extension on Copenhagen Knee ROM Scale.
LoA is only calculated for groups larger than 15 patients. Standard deviations are in parentheses for the same reason.
CI, confidence interval; SD, standard deviation; LoA, limits of agreement.
[a] 95% LoA, limits of agreement based on the SD for each group (column).

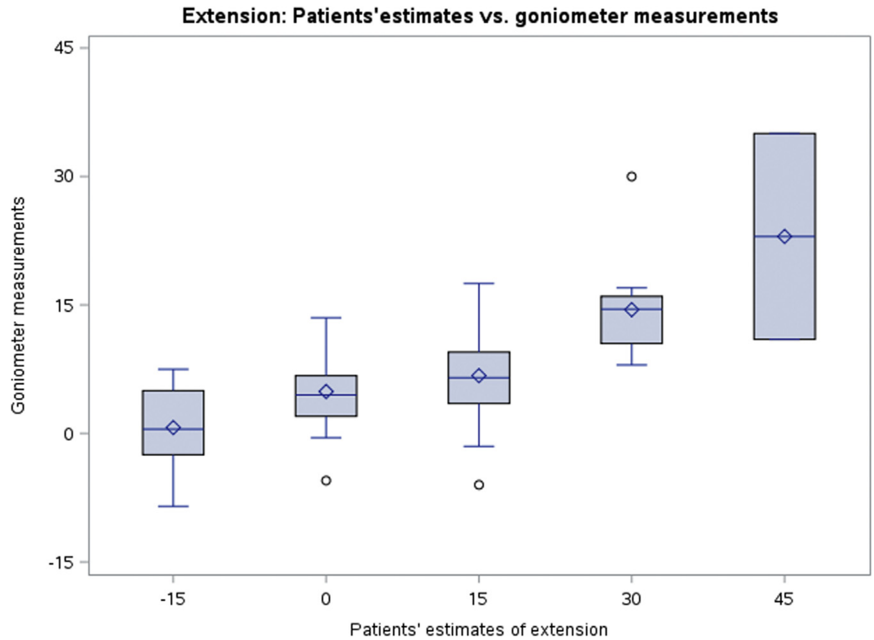


Fig. 4. Boxplot of goniometer measurements of extension grouped by patients' estimates on Copenhagen Knee ROM Scale.

It can be argued that the 15° increments lower the precision of this tool. But unlike other published methods that use 5°-10° intervals [3,4,6], we found no overall systematic difference between patient estimates and goniometer measurements for either flexion or extension. This supports our hypothesis that simplicity should be prioritized over small increments between options. In our study, Pearson's correlation coefficient was 0.79 for flexion, which is remarkably higher than the 0.44 reported by Gioe et al and the 0.35 reported by Khanna et al [3,4]. For extension, Gioe found a correlation of 0.31 between patient estimate and measurement; this was 0.13 in Khanna's study and 0.63 in the present study. These differences suggest that CKRS has an advantage over other tools in measuring the item of interest; although we examined passive ROM, which is normally considered to require help from an examiner, we did not find any overall difference between patient estimates and measurements.

Measurement Error

In flexion, obviously the variance was larger for patient estimates vs goniometer measurements (SD, 12.3°) than between goniometer measurements (SD, 4.2°) and also larger than the variance of examiners' CKRS estimates vs goniometer measurements (SD, 6.7°). It was, however, lower than the 20.6° reported by Borgbjerg et al [6] and comparable to the 12.4° and 12.8° reported by Gioe and Khanna, respectively [3,4].

In extension, although we found no systematic overall mean difference between methods, patients did exaggerate both extension deficits and hyperextension; they used the scale more widely than actual measurements justified. This was expected from our experiences in the development phase and from reports from similar studies [5,6,17]. A patient bothered by an extension deficit of 10° may well feel like option no. 2 (30°) illustrates. So, when extension measurements were grouped by CKRS answers, there were systematic differences, but this can be compensated for by using the grouped values (Table 4) in the interpretation of clinical results. Here, we have taken the consequence and let patients be the judge of what they see in ROM illustrations. For example, 95% of patients marking picture 3 (15°) can be expected to have passive extension between −8.4° and 9.8°. Collins et al [5] reported a quite similar distribution of extension estimates using 5°-10° increments (Gioe's method), regarding active ROM, however. Except for Borgbjerg, who presented Bland-Altman plots, none of the other previously mentioned publications provided information about the distribution of measurement errors in relation to absolute ROM measures [3–6].

The finding of considerable overlap between groups of answers lowers the precision of the patient estimates of extension. The overall SD for patient estimates in CKRS vs goniometer measurements was 11.6°, whereas SD for examiners' CKRS estimates vs goniometer measurement was 4.6° and between goniometer measurements only 3.0°. On group levels, however, SDs for patient-reported extension on CKRS were only 5.1°, 3.9°, and 4.6° for the 3 most used options (Table 4). This is lower than Borgbjerg's SD of 10.6° and quite similar to the 5.0°, 4.4°, and 4.2°-6.7° reported by Gioe, Khanna, and Collins respectively [3–5]. However, patient groups are difficult to compare, because Gioe and Khanna included >1 year postoperative patients only, with absolute extension measures ranging 1.4° ± SD 4.3° and 0.5° ± 2.5°, respectively. Our mix of patients had a wider extension range of 5.8° ± SD 6.5°, which was comparable to Collins' measures of active extension. To summarize, these findings together with the correlation coefficients lead us to argue that passive extension is most precisely measured with CKRS. Also, we find no reason to believe that accuracy would increase with smaller increments between illustrations.

The lack of precision in extension is caused by several factors: Even for healthcare professionals, extension measurements are more difficult than flexion measurements [20]. Estimates are affected by the extension ability of the contralateral leg. Compared to measurements of flexion, the relative difference between passive and active extension is larger given the small absolute numbers of degrees in extension. Also, instead of answering the question of passive motion, patients may be answering whether or not extension poses an actual problem for them in daily life. For example, one would expect patients answering CKRS option 4 or 5 (0 and −15°) to be satisfied with their extension. To evaluate this hypothesis would require a new study asking the additional question: “Is the extent to which you can straighten the knee a problem for you?”

This might demonstrate whether CKRS offers a better identification of patients who feel a need for treatment (for example, additional physiotherapy) than passive goniometer measurements or a simple "yes/no" question does.

Clinical Use

CKRS may be included as extra information in registries and surveys, and it may be a feasible replacement for professional goniometer measurement in some clinical settings. Whether patient-reported ROM using CKRS offers the necessary level of accuracy varies between different settings and must be evaluated for each situation [21–23].

The basis for application of CKRS as a screening tool is that the positive and negative predictive values are acceptable. A limitation to this study is that the predictive values given are not applicable to all populations of knee OA patients, as values change with the distribution of ROM measures. Values for sensitivities and specificities, however, are directly applicable to other knee OA patients. Which thresholds to use when applying CKRS to a patient group solely depends on the purpose of testing (screening, monitoring, surveillance, etc.). For example, if a clinic aims at identifying postoperative patients in need of intensive physiotherapy or manipulation under anesthesia, for example, within 3 months after surgery, a flexion limit of 95°-110° and a CKRS threshold between option 4 and 5, or even 5 and 6 may be appropriate to ensure a high sensitivity [1,2,5,22–25].

With the accelerating advancements in everyday technology, intelligent knee braces and various smartphone goniometer apps are being developed and are likely to become a natural part of future knee patients' rehabilitation programs [25–29]. One example is the "DrGoniometer" app which has recently been proven valid for knee ROM measurement [30]. A selected group of arthroplasty patients had an accompanying relative or friend take photographs of the knee in a simulated home setting. Photographs were sent to the staff who made the ROM measurements using the app. Although precision was higher than what we have found in the present study, we find that a tool such as CKRS still has some advantages over technical solutions: CKRS requires no professional intervention, it is self-explaining and quickly completed without removal of clothes, and it involves no expensive or technical equipment, factors which may be of particular importance when applied to large groups of elderly patients.

Written instructions in CKRS are brief. However, we recommend careful translation and cultural validation of text and illustrations before use in other languages. Although we have suggested an English translation, ideally, future translations should be based on the original Danish version until the English version is validated in an English-speaking patient population (both versions are available at www.procordo.com/docs/copenhagen_rom). Further validation is especially important if the tool is used for other types of knee patients, because younger or more active patients may differ in their perception of both ROM and CKRS illustrations.

In the present study, we focused on validity testing against a gold standard (goniometer measurement) in a relevant patient population. Responsiveness testing was not conducted. We do of course welcome further studies on change in CKRS estimates and ROM before and after knee arthroplasty. A mapping of patients' baseline values and postoperative results after different types of arthroplasty operations could benefit the preoperative expectation alignment, and using this instrument would provide a unique opportunity to compare ROM values across arthroplasty types, centers, and patient populations without the bias of having surgeons or physiotherapists perform measurements on their own patients.

Conclusion

CKRS is a patient-friendly and feasible tool for knee OA and arthroplasty patients to self-report their passive knee ROM for use in long-term follow-up as well as knee registries and research when professional goniometer measurement is not a feasible option, or when a virtually unbiased ROM estimate is desirable. We recommend further studies to prove responsiveness to change, for example, after knee arthroplasty operation. With 15° increments between answer options, we have reached better correlation with goniometer measurement than what was reported with similar tools using 5°-10° increments. Furthermore, we have reached at least the same level of accuracy and strong retest reliability, particularly regarding flexion. We believe this tool meets the appropriate level of ambition in the field of patient-reported passive knee ROM.

Acknowledgments

The authors would like to send special thanks to physiotherapist Hanne Hornshøj. The authors thank the following persons for their contributions to this work: Henrik Morville Schrøder, Carsten Bogh Juhl, Jonathan David Comins, Anna Pors Nielsen, Michelle Möger Andersen, Thomas Linding Jakobsen, Camilla Ryge, Karen Dyreborg, and patients and staff at the Departments of Orthopedic Surgery at Gentofte and Naestved Hospitals.

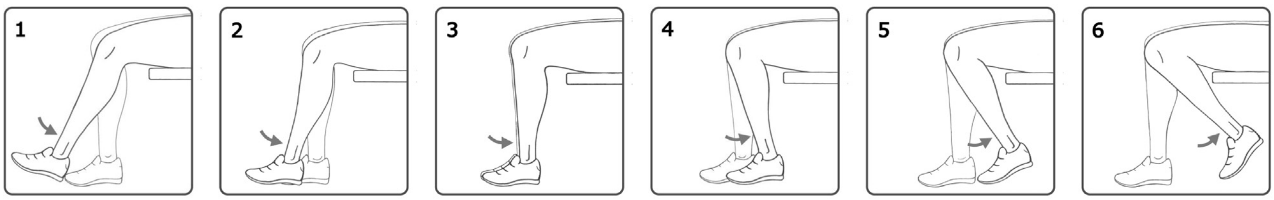
References

[1] Unver B, Nalbant A, Karatosun V. Comparison of self-reported and measured range of motion in total knee arthroplasty patients. Ann Transl Med 2015;3:192.
[2] Weick J, Bawa HS. The potential utility of patient-reported range of motion after total knee arthroplasty. Ann Transl Med 2015;3:193.
[3] Khanna GMD, Singh JAMDMPH, Pomeroy DLMD, Gioe TJMD. Comparison of patient-reported and clinician-assessed outcomes following total knee arthroplasty. J Bone Joint Surg Am 2011;93. e117(1)-(7).
[4] Gioe TJ, Pomeroy D, Suthers K, Singh JA. Can patients help with long-term total knee arthroplasty surveillance? Comparison of the American Knee Society Score self-report and surgeon assessment. Rheumatology (Oxford) 2009;48:160–4.
[5] Collins JE, Rome BN, Daigle ME, Lerner V, Katz JN, Losina E. A comparison of patient-reported and measured range of motion in a cohort of total knee arthroplasty patients. J Arthroplasty 2014;29:1378–1382.e1.
[6] Borgbjerg J, Madsen F, Odgaard A. Patient self-assessed passive range of motion of the knee cannot replace health professional measurements. J Knee Surg 2017;30:829–34.
[7] de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. practical guides to biostatistics and epidemiology. Cambridge: Cambridge University Press; 2011.
[8] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
[9] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7–18.
[10] Bossuyt PM, Cohen JF, Gatsonis CA, Korevaar DA. STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. Ann Transl Med 2016;4:85.
[11] Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96–106.
[12] Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;6:e012799.
[13] Jakobsen TL, Christensen M, Christensen SS, Olsen MM, Bandholm T. Reliability of knee joint range of motion and circumference measurements after total knee arthroplasty: does tester experience matter? Physiother Res Int 2010;15:126–34.
[14] Bovens AM, van Baak MA, Vrencken JG, Wijnen JA, Verstappen FT. Variability and reliability of joint measurements. Am J Sports Med 1990;18:58–63.
[15] Mainland BJ, Amodeo S, Shulman KI. Multiple clock drawing scoring systems: simpler is better. Int J Geriatr Psychiatry 2014;29:127–36.
[16] Norkin C. Measurement of joint motion, 5e, A guide to goniometry. In: Norkin C, White DJ, editors. 5. ed. Philadelphia: F.A. Davis; 2016.
[17] Uribe B, El Bitar Y, Wolf BR, Bollier M, Kuhn JE, Hettrich CM. Agreement between patient self-assessment and physician assessment of shoulder range of motion. J Shoulder Elbow Surg 2016;25:1649–54.

[18] Brosseau L, Balmer S, Tousignant M, O'Sullivan JP, Goudreault C, Goudreault M, et al. Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. Arch Phys Med Rehabil 2001;82:396–402.
[19] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[20] Piriyaprasarth P, Morris ME. Psychometric properties of measurement tools for quantifying knee joint position and movement: a systematic review. Knee 2007;14:2–8.
[21] McClelland JA, Feller JA, Menz HB, Webster KE. Patients with total knee arthroplasty do not use all of their available range of knee flexion during functional activities. Clin Biomech (Bristol, Avon) 2017;43:74–8.
[22] Wylde V, Lenguerrand E, Brunton L, Dieppe P, Gooberman-Hill R, Mann C, et al. Does measuring the range of motion of the hip and knee add to the assessment of disability in people undergoing joint replacement? Orthop Traumatol Surg Res 2014;100:183–6.
[23] Miner AL, Lingard EA, Wright EA, Sledge CB, Katz JN. Knee range of motion after total knee arthroplasty: how important is this as an outcome measure? J Arthroplasty 2003;18:286–94.
[24] Pua YH, Ong PH, Chong HC, Yeo W, Tan CI, Lo NN. Associations of self-report physical function with knee strength and knee range-of-motion in total knee arthroplasty possible nonlinear and threshold effects. J Arthroplasty 2013;28:1521–7.
[25] Dietz MJ, Sprando D, Hanselman AE, Regier MD, Frye BM. Smartphone assessment of knee flexion compared to radiographic standards. Knee 2017;24:224–30.
[26] Mourcou Q, Fleury A, Diot B, Franco C, Vuillerme N. Mobile phone-based joint angle measurement for functional assessment and rehabilitation of proprioception. Biomed Res Int 2015;2015:328142.
[27] Mourcou Q, Fleury A, Diot B, Vuillerme N. iProprio: a smartphone-based system to measure and improve proprioceptive function. Conf Proc IEEE Eng Med Biol Soc 2016;2016:2622–5.
[28] Stove MP, Palsson TS, Hirata RP. Smartphone-based accelerometry is a valid tool for measuring dynamic changes in knee extension range of motion. Knee 2017;25:66–72.
[29] Jenny JY, Bureggah A, Diesinger Y. Measurement of the knee flexion angle with smartphone applications: which technology is better? Knee Surg Sports Traumatol Arthrosc 2016;24:2874–7.
[30] Castle H, Kozak K, Sidhu A, Khan RJK, Haebich S, Bowden V, et al. Smartphone technology: a reliable and valid measure of knee movement in knee replacement. Int J Rehabil Res 2018;41:152–8.

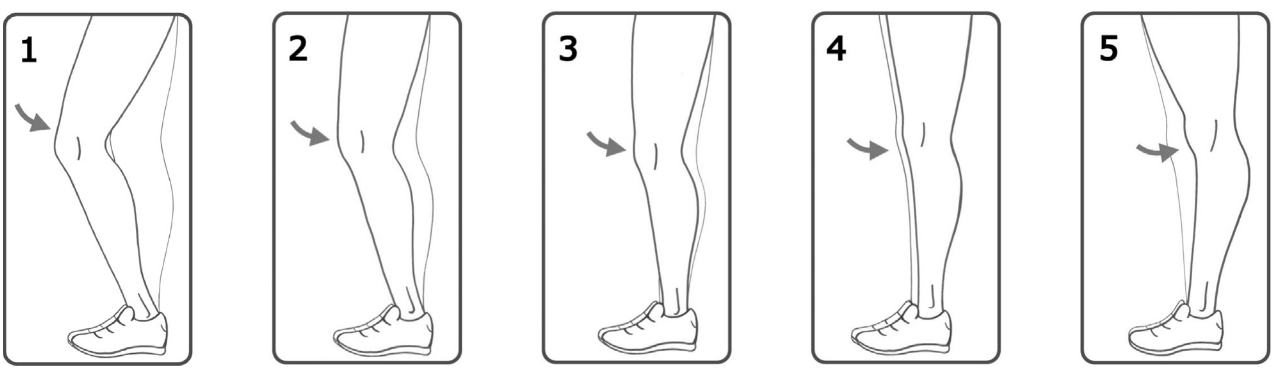
Appendix A

How much can you bend your knee?
Please push your lower leg as far back as possible.
You can use your hand to pull your lower leg in the direction of the arrow.
Tick the box that fits your situation.



0 Impossible. I am not able to bend my knee as much as in picture no. 1

How much can you straighten your knee?
Please use your hand to push your knee backward in the direction of the arrow.
Tick the box that fits your situation.



0 Impossible. I am not able to straighten my knee as much as in picture no. 1

Appendix B

Raw data for clinical testing of Copenhagen ROM Scale 2017 (Column names are explained in sheet 2)

ID	FlexPA	ExtPA	FlexPHYS	FlexPHYSgon	ExtPHYS	ExtPHYSgon	FlexSUR	FlexSURgon	ExtSUR	ExtSURgon	flexCONS	ExtCONS
1	5	3	4	106	4	7	4	106	3	9	107	7
2	5	3	6	133	3	16	5	123	3	15	127	16
3	5	2	5	127	4	8	5	125	4	8	125	10
4	4	3	4	102	4	5	4	96	4	1	95	3
5	6	4	5	126	4	6	5	128	4	1	126	8
6	6	4	6	129	5	-5	5	127	5	-6	131	-4
7	4	5	3	100	4	8	4	102	4	7	102	8
8	6	5	5	119	4	-6	6	119	5	-6	118	-7
9	3	4	3	112	4	7	4	114	3	7	118	6
10	6	5	6	131	4	-4	6	129	4	-2	130	-4
11	5	3	5	121	4	9	4	112	3	8	120	7
12	3	3	2	68	4	6	2	65	3	9	69	8
13	5	1	3	89	1	35	3	89	1	35	83	34
14	3	2	3	96	3	17	3	90	3	17	94	18
15	5	4	5	122	4	6	5	116	3	6	116	7
16	5	5	5	122	4	5	5	117	4	5	117	4
17	5	3	6	146	4	-2	6	150	4	-1	148	-1
18	3	2	3	89	3	16	3	91	4	13	86	14
19	6	4	5	126	3	7	6	135	4	2	124	3
20	5	3	5	131	4	6	5	132	4	3	133	5
21	2	4	2	66	4	3	2	59	4	3	58	5
22	6	5	6	140	4	-3	6	138	4	-2	137	-5
23	5	4	5	114	4	2	6	119	4	0	111	-2
24	4	3	5	114	4	7	4	109	4	3	113	9
25	6	4	6	134	4	6	6	135	4	3	136	7
26	5	4	6	135	4	3	6	136	4	0	137	2
27	2	4	3	99	3	16	3	91	4	11	95	13
28	6	4	6	140	4	0	6	130	4	1	139	0
29	6	4	5	117	4	5	5	118	4	3	118	4
30	5	3	6	144	4	5	6	143	4	2	145	3
31	3	3	5	106	3	9	5	106	4	2	105	4
32	2	4	3	88	4	0	4	83	4	-1	86	-4
33	3	4	4	110	4	4	5	110	4	6	115	5
34	6	5	5	126	4	3	5	129	4	3	134	4
35	3	4	3	89	4	9	3	85	3	10	92	11
36	2	4	3	83	4	8	3	84	4	11	83	8
37	6	4	5	126	3	9	5	128	3	9	129	10
38	6	5	5	120	4	7	5	122	4	4	123	6
39	4	4	3	79	4	3	2	76	4	-3	79	1
40	6	5	6	147	5	-10	5	144	4	-7	148	-4
41	2	3	3	72	4	4	2	72	4	-1	72	-1
42	4	5	4	118	3	6	4	116	3	7	114	7
43	3	2	3	96	3	17	3	93	2	15	97	16
44	4	3	4	111	3	12	4	115	4	6	121	10
45	6	4	6	140	4	0	6	137	4	2	142	0
46	5	4	4	124	4	7	6	124	4	3	122	5
47	4	3	4	102	3	11	4	106	3	9	109	11
48	6	4	6	152	4	1	6	149	4	2	152	0
49	6	5	6	146	4	-2	6	150	4	-2	149	-4
50	4	4	3	88	4	6	3	88	4	4	88	4
51	3	4	3	105	3	6	3	98	4	3	100	2
52	6	4	6	130	4	5	5	130	4	2	128	5
53	3	3	4	102	3	19	4	100	3	16	103	16
54	6	5	5	135	4	4	6	135	4	2	134	5
55	5	3	5	130	3	9	6	133	4	10	135	11
56	5	4	5	124	3	12	5	123	3	9	122	8
57	6	5	6	145	5	5	6	141	5	0	140	-4
58	5	3	5	120	4	-5	5	122	5	-7	118	-8
59	6	4	4	116	3	13	5	117	3	14	121	13
60	5	4	4	112	4	8	5	112	4	6	108	7

61	5	5	6	128	4	0	5	118	5	-7	122	-10
62	5	4	5	123	3	5	6	126	4	1	124	5
63	6	3	5	120	3	6	6	128	4	7	121	6
64	3	3	3	93	4	4	3	89	3	7	89	4
65	6	5	5	125	4	1	6	126	4	-5	123	-1
66	6	5	6	142	4	-4	6	149	4	-6	145	-5
67	5	3	5	124	4	2	5	127	4	5	127	4
68	6	5	5	125	4	3	6	124	4	-2	125	1
69	5	4	4	110	4	7	5	117	4	6	112	3
70	2	2	3	96	3	11	3	97	3	18	91	15
71	5	4	4	112	4	2	6	115	4	2	116	2
72	4	4	4	95	4	4	3	92	4	4	94	4
73	6	5	5	137	4	5	6	138	4	-1	134	1
74	6	5	6	142	4	-2	6	142	4	-1	143	1
75	6	5	4	108	4	6	5	111	3	6	114	5
76	6	4	5	130	4	4	5	130	4	0	131	4
77	6	3	6	135	3	15	6	133	3	11	137	14
78	5	3	6	125	4	14	5	121	3	17	124	15
79	6	4	6	131	4	0	5	127	3	4	134	0
80	4	2	3	83	4	7	3	84	3	9	81	7
81	4	1	3	110	3	10	4	107	3	12	115	11
82	4	3	3	106	4	5	4	103	4	1	104	7
83	4	4	3	88	4	8	3	90	4	4	85	5
84	5	3	4	108	4	6	4	109	4	0	110	5
85	5	4	5	127	4	9	6	126	3	8	126	9
86	4	3	5	124	4	8	5	121	4	7	124	8
87	5	4	5	115	3	12	5	111	3	11	115	11
88	6	5	6	135	4	-1	6	139	4	1	140	0
89	5	3	4	115	4	9	5	117	3	13	114	11
90	4	3	4	100	3	9	4	96	3	10	96	10
91	6	2	5	131	2	33	5	128	2	27	130	33
92	4	4	5	124	4	5	5	126	4	8	120	8
93	5	2	5	122	4	9	5	124	3	12	121	13
94	4	2	3	91	3	12	3	78	2	11	86	11
95	5	4	5	114	4	6	4	123	4	3	117	4
96	4	4	4	115	3	6	4	118	3	7	112	7
97	6	3	5	121	3	10	6	118	3	3	122	9
98	6	5	6	137	4	5	5	134	4	9	135	6
99	6	4	6	129	4	6	5	121	3	5	128	7
100	4	3	1	72	4	3	2	74	4	4	72	3

Column names

ID	Patient ID (new)
FlexPA	Flexion on CKRS by patient (picture no.)
ExtPA	Extension on CKRS by patient (picture no.)
FlexPHYS	Flexion on CKRS by physiotherapist (picture no.)
FlexPHYSgon	Flexion measured with goniometer by physiotherapist (degrees)
ExtPHYS	Extension on CKRS by physiotherapist (picture no.)
ExtPHYSgon	Extension measured with goniometer by physiotherapist (degrees)
FlexSUR	Flexion on CKRS by surgeon (picture no.)
FlexSURgon	Flexion measured with goniometer by surgeon (degrees)
ExtSUR	Extension on CKRS by surgeon (picture no.)
ExtSURgon	Extension measured with goniometer by surgeon (degrees)
FlexCONS	Flexion measured with goniometer by physiotherapist and surgeon together (not used in data analysis) (degrees)
ExtCONS	Extension measured with goniometer by physiotherapist and surgeon together (not used in data analysis) (degrees)

ARTICLE IN PRESS

The Journal of Arthroplasty xxx (2019) 1–2

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Corrigendum

Corrigendum to 'Knee Osteoarthritis Patients Can Provide Useful Estimates of Passive Knee Range of Motion: Development and Validation of the Copenhagen Knee ROM Scale' [The Journal of Arthroplasty 33 (2018) 2875-2883]

Anne Mørup-Petersen, MD [a,*], Pætur M. Holm, PT, MSc [b,c], Christina E. Holm, MD [d], Tobias W. Klausen, MSc [e], Søren T. Skou, PT, PhD [b,c], Michael R. Krogsgaard, MD, PhD [f], Mogens B. Laursen, MD, PhD [g], Anders Odgaard, MD, DMSc [a]

[a] *Department of Orthopedic Surgery, Copenhagen University Hospital Gentofte, Copenhagen, Denmark*
[b] *Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Slagelse, Denmark*
[c] *Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark*
[d] *Department of Orthopedic Surgery, Copenhagen University Hospital, Rigshospitalet, København Ø, Denmark*
[e] *Department of Hematology, Copenhagen University Hospital, Herlev, Denmark*
[f] *Department of Orthopedic Surgery, Copenhagen University Hospital, København NV, Denmark*
[g] *Department of Orthopedic Surgery, Aalborg University Hospital, Farsø, Denmark*

The authors regret that there are errors in Table 3, which appears on page 2880 of their article.

The authors would like to apologise for any inconvenience caused and provide the corrected Table below.

DOI of original article: https://doi.org/10.1016/j.arth.2018.05.011.

* Reprint requests: Anne Mørup-Petersen, MD, Department of Orthopedic Surgery, Herlev and Gentofte Hospital, Gentofte Matriklen, Kildegaardsvej 28, Hellerup 2900, Denmark.

https://doi.org/10.1016/j.arth.2019.04.054

Table 3
Limits of Goniometer Measurements of Knee Motion and Their Respective Sensitivities and Specificities for Detecting Problematic Knee Motion Using Different Thresholds (Cutoff Values) in Copenhagen Knee ROM Scale (CKRS).

Limits of Knee Motion					
Limit of Knee Motion	CKRS Threshold (Between Pictures No.)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%) (Population Specific)	Negative Predictive Value (%) (Population Specific)
Flexion < 90°	2-3	31	98	67	90
	3-4	46	87	35	92
	4-5	**92**	**72**	**33**	**98**
	5-6	100	40	20	100
Flexion < 100°	2-3	27	100	100	83
	3-4	55	94	71	88
	4-5	**95**	**81**	**58**	**98**
	5-6	100	45	34	100
Flexion < 110°	2-3	19	100	100	72
	3-4	47	97	88	80
	4-5	**88**	**88**	**78**	**94**
	5-6	97	50	48	97
Flexion < 120°	4-5	65	96	94	72
	5-6	**90**	**63**	**72**	**86**
Extension ≥ 15°	2-3	57	92	36	97
	3-4	**100**	**66**	**18**	**100**
Extension ≥ 10°	2-3	50	98	82	90
	3-4	**78**	**70**	**36**	**93**
	4-5	**100**	**26**	**23**	**100**
Extension ≥ 5°	2-3	22	100	100	57
	3-4	57	78	72	66
	4-5	90	31	55	76

Guide to interpretation of values:
Sensitivity: The chance that a patient with a knee motion worse than the specified limit is identified as having a knee motion problem using this Copenhagen Knee ROM Scale (CKRS) threshold.
Specificity: The chance that a patient with better knee motion than the specified limit is identified as having acceptable knee motion using this CKRS threshold.
Positive predictive value: The chance that a patient reporting knee motion worse than the specified CKRS threshold does have a knee motion worse than the specified limit. (Note that this value is population specific).
Negative predictive value: The chance that a patient reporting better knee motion than the specified CKRS threshold does have a knee motion better than the specified limit. (Note that this value is population specific).
Negative and positive predictive values in the test population are listed. Authors find the numbers in bold to be of largest clinical relevance.

Appendix

A: Danish regional revision rates after primary knee arthroplasty

B: SPARK patient information

C: Copenhagen Knee ROM Scale (Danish)

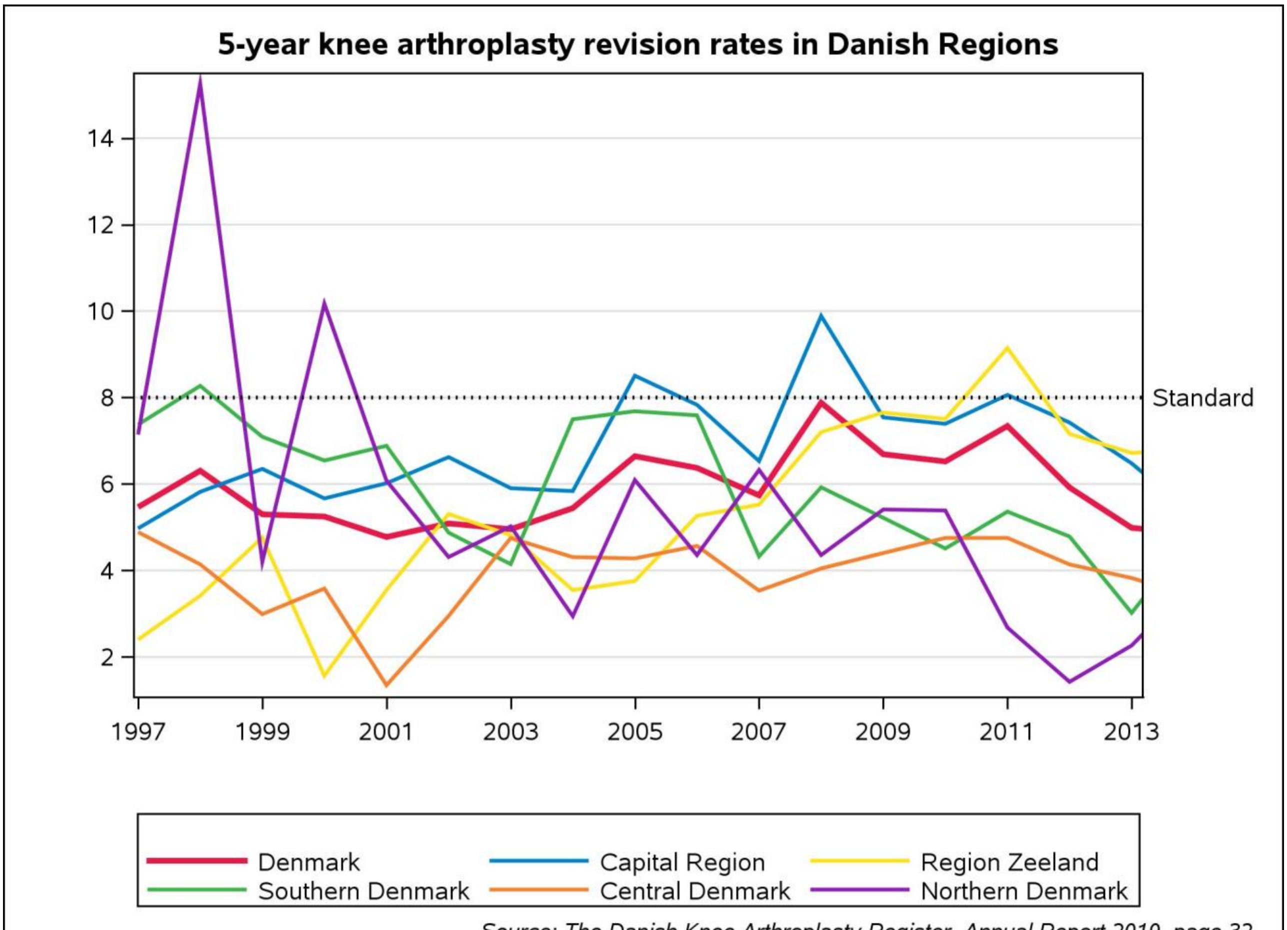
D: Copenhagen Knee ROM Scale (English)

E: UCLA Activity Scale, Danish version 2017

Appendix A

Knee arthroplasty revision rates in Danish Regions per operation year

	1-y. revision rates						2-y. revision rates						5-y. revision rates					
Region	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Denmark	2.9	2.0	1.8	1.8	1.9	1.9	3.8	4.7	3.9	3.3	3.3	3.1	5.2	7.2	6.6	6.4	7.3	5.9
Capital Region	3.0	2.6	2.2	2.0	2.2	1.8	4.2	5.1	5.0	4.1	3.3	3.7	5.8	9.0	7.4	7.3	8.0	7.4
Region Zeeland	2.7	2.6	1.9	2.4	1.5	2.4	4.9	5.5	4.8	4.8	5.0	2.8	4.9	6.3	7.5	7.5	9.3	7.2
Southern Denmark	2.0	1.6	1.3	1.7	1.8	2.2	2.8	3.1	2.9	2.0	3.1	2.5	4.0	5.6	5.2	4.4	5.4	4.8
Central Denmark	1.5	1.4	1.5	1.5	2.1	1.5	2.3	2.8	2.2	2.4	2.7	3.1	3.2	3.7	4.3	4.7	4.5	3.8
Northern Denmark	1.8	0.5	1.1	0.8	1.7	1.4	3.2	1.8	1.0	1.7	2.1	3.2	5.8	4.0	5.2	5.0	2.7	1.4



Source: The Danish Knee Arthroplasty Register, Annual Report 2019, page 32

Appendix B

INFORMATION OM "SPARK"-UNDERSØGELSEN

Tag dette med hjem!

Til dig, der skal opereres med protese i knæet

Her på Ortopædkirurgisk Afdeling vil vi i det kommende år følge, hvordan du har det både før og efter din operation med kunstigt knæ.

SPARK-undersøgelsens formål er

- at kirurger i hele landet bliver endnu bedre til at forudse, hvem der har gavn af knæoperation
- at patienterne oplever færre komplikationer
- at eventuelle komplikationer behandles bedst muligt

Hvad betyder det for dig?

Du vil få tilsendt et spørgeskema via mail eller brev fem gange i forløbet. Det tager 10-15 minutter at udfylde første gang og de næste gange 5-10 minutter.

For at undersøgelsen skal lykkes, er det vigtigt, vi får svar fra alle patienter - både fra dem, hvor det hele går glat og fra dem, der er mindre tilfredse. Vi håber altså, at du vil afsætte sammenlagt en time derhjemme i løbet af det næste år til at gøre en stor forskel for patienter med samme knæproblem som dig selv.

Når du har underskrevet samtykkeerklæringen og afleveret den i ambulatoriet, behøver du ikke foretage dig noget, før vi kontakter dig.

På forhånd mange tak for din hjælp!

Med venlig hilsen **SPARK-gruppen**

Anne Mørup-Petersen
Reservelæge, Ph.D. studerende

Frank Madsen
Overlæge,
Sektorchef,
Aarhus Universitets-hospital

Mogens Berg Laursen
Overlæge, Ph.D., lektor
Aalborg Universitetshospital,
Aalborg og Farsø

Anders Odgaard
Overlæge, dr.med.
Gentofte Hospital

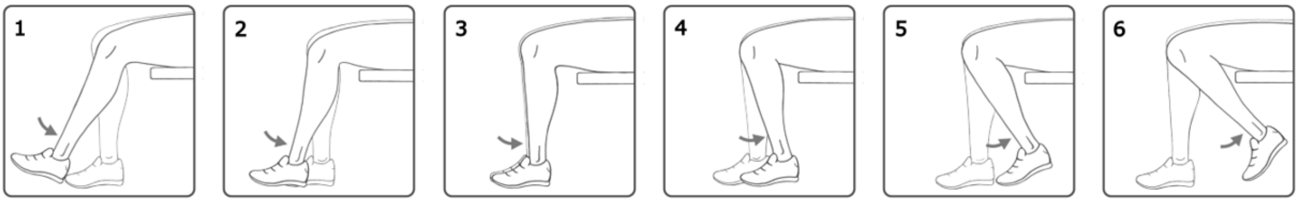
SPARK er en forkortelse af engelsk for "Variation i patienttilfredshed, patient-rapporterede oplysninger, radiologisk artrose og genoperationshyppighed blandt patienter med knæartrose på hospitaler i tre danske regioner". Eventuelle spørgsmål rettes venligst til den projektansvarlige læge, Ph.D.-studerende Anne Mørup-Petersen på mail: anne.moerup-petersen.02@regionh.dk eller tlf: 38 67 38 40.

Appendix C: Copenhagen Knee ROM Scale (Danish version)

Hvor meget kan du bøje dit knæ?

Pres underbenet længst muligt bagud, f.eks. ved at trække med hånden i pilens retning.

Markér den boks, der passer bedst.



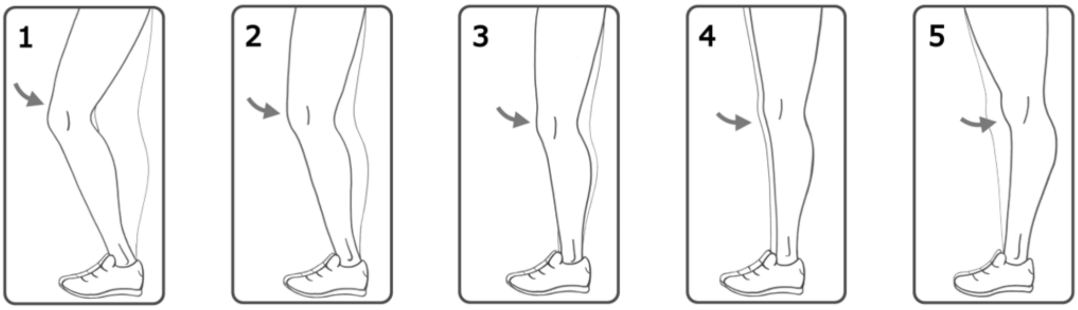
0 Umuligt. Jeg kan slet ikke bøje mit knæ så meget, som det første billede viser

Copenhagen Knee ROM Scale (side 1 af 2)

Hvor meget kan du strække dit knæ?

Brug hånden til at presse knæet bagud i pilens retning.

Markér den boks, der passer bedst.



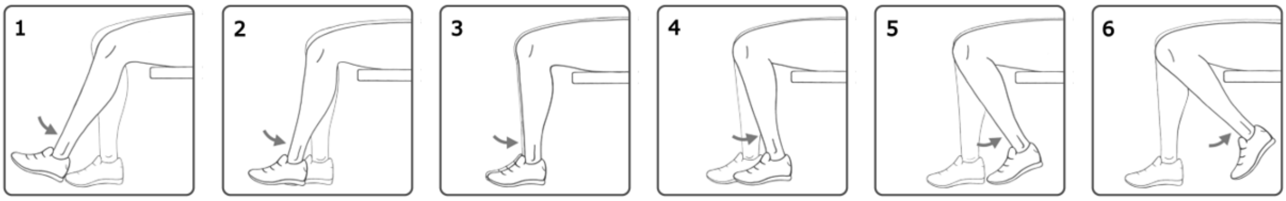
0 Umuligt. Jeg kan slet ikke strække mit knæ så meget, som det første billede viser

Copenhagen Knee ROM Scale (side 2 af 2)

Appendix D: Copenhagen Knee ROM Scale (English version)

How much can you bend your knee?

Please push your lower leg as far back as possible.
You can use your hand to pull your lower leg in the direction of the arrow.
Tick the box that fits your situation.

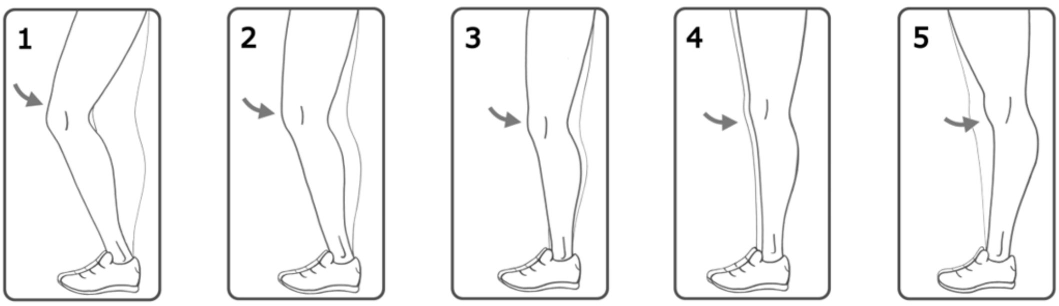


0 Impossible. I am not able to bend my knee as much as in picture no. 1

Copenhagen Knee ROM Scale (page 1 of 2)

How much can you straighten your knee?

Please use your hand to push your knee backwards in the direction of the arrow.
Tick the box that fits your situation.



0 Impossible. I am not able to straighten my knee as much as in picture no. 1

Copenhagen Knee ROM Scale (page 2 of 2)

Appendix E: UCLA Activity Scale (Danish version)

UCLA Aktivitetsskala

Hvor højt er dit aktivitetsniveau lige nu?

Tænk på dit aktivitetsniveau i de seneste 4 uger.

Niveau 10 er meget højt, og niveau 1 er meget lavt.

Sæt kun ét kryds ved én af de ti muligheder.

Sæt kun ét X

10	**Jeg er regelmæssigt fysisk aktiv med høj belastning**	*f.eks. løb, fodbold, håndbold, badminton, tennis, skiløb, tungt fysisk arbejde, bjergvandring eller lignende*
9	**Jeg er af og til fysisk aktiv med høj belastning**	
8	**Jeg er regelmæssigt meget fysisk aktiv**	*f.eks. lange cykelture, golf, krævende gymnastik/fitness eller lignende*
7	**Jeg er af og til meget fysisk aktiv**	
6	**Jeg er regelmæssigt fysisk aktiv**	*f.eks. svømning, cykelture, lange gåture, krævende husligt arbejde, større indkøb eller lignende*
5	**Jeg er af og til fysisk aktiv**	
4	**Jeg er regelmæssigt lettere fysisk aktiv**	*f.eks. gang, let husligt arbejde, mindre indkøb eller lignende*
3	**Jeg er af og til lettere fysisk aktiv**	
2	**Jeg er nærmest aldrig fysisk aktiv**	~ *begrænset til et minimum af dagligdags aktiviteter*
1	**Jeg er slet ikke fysisk aktiv**	~ *afhængig af andre, kan ikke forlade hjemmet*