Health Technology
Assessment Programme

National Institute for Health Research

NIHR HTA PROGRAMME

HTA project 15/80/40

This project was funded by the National Institute for Health Research Health Technology Assessment Programme (project number 15/80/40). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Health Technology Assessment Programme, NIHR, NHS or the Department of Health.

Protocol version 2.0

02 August 2019

UTMOST study

Risk-benefit and costs of unicompartmental (compared to total) knee replacement for patients with multiple co-morbidities: a non-randomised study, and different novel approaches to minimize confounding

Version	Date
Version 1.0	26/01/2017
Version 2.0	02/08/2019

Principal Investigator	Associate Professor Daniel Prieto-Alhambra Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences University of Oxford daniel.prietoalhambra@ndorms.ox.ac.uk 01865223401
Co-Investigators	Victoria Strauss, University of Oxford Albert Prats-Uribe, University of Oxford Spyros Kolovos, University of Oxford Andrew Judge, University of Oxford Nigel Arden, University of Oxford Mark Wilkinson, University of Sheffield Ian Douglas, London School of Hygiene and Tropical Medicine Susan Thwaite, National Rheumatoid Arthritis Society Alan Silman, University of Oxford David Murray, University of Oxford Irene Petersen, University College London Jose Maria Valderas, University of Exeter Andrew Carr, University of Oxford David Beard, University of Oxford Rafael Pinedo-Villanueva, University of Oxford Sarah Lamb, University of Oxford Antonella Delmestri, University of Oxford Klara Berencsi, University of Oxford
Sponsor	University of Oxford
Funder	NIHR Health Technology Assessment Programme

TABLE OF CONTENTS

SUMMARY OF RESEARCH

RATIONALE

Although an NIHR HTA-funded RCT (TOPKAT) is ongoing to compare unicompartemental (UKR) and total knee replacement (TKR), limited follow-up and restrictive eligibility criteria will limit external validity to the large number of patients with multiple co-morbidities (1 in 6 according to the National Joint Registry data).

AIMS

- STAGE 1. To validate a number of novel analytical methods to minimise confounding: we will replicate TOPKAT by analysing the association between UKR (compared to TKR) and post-operative patient reported outcomes (PROMs) amongst participants in the National Joint Registry (NJR) eligible for TOPKAT (ASA grade <3) using different methods, and then test for a difference between the obtained estimates and TOPKAT.
- STAGE 2. To study the benefits (PROMs), risks (revision, complications), mortality, costs and cost-effectiveness of UKR (compared to TKR) amongst NJR participants not eligible for TOPKAT (ASA 3+). Methods previously validated (in STAGE 1) will be applied for this second Aim.

METHODS

Setting and design

We will conduct a cohort analysis using routinely collected data from the NJR linked to hospital admission records (HES) and the National PROMs Database.

Participants

Two cohorts: 1.NJR participants undergoing UKR/TKR with ASA 1 or 2, eligible for TOPKAT (*comparison cohort*); AND 2.NJR participants undergoing UKR/TKR with ASA 3+ (*co-morbidity cohort*).

Health technology being assessed: UKR versus TKR.

Outcomes

- Primary: post-operative Oxford Knee Score (PROMs).
- Secondary: 90 days risks of surgery complications (prosthetic joint infection, cardiovascular disease and venous thromboembolism) and 5-year risks of revision , mortality, health-related quality of life (EQ-5D), NHS hospital costs (as identified in HES).

Power

Based on published data(11), >720 UKR and 8,400 TKR recipients in the smaller (*co-morbidity*) cohort will have linked PROMs. With an expected standard deviation of 8, power will be 90% to detect a minimally clinically important difference of 2+ points in Oxford Knee Score(12).

Statistics

Linear regression models will be used to study the association between surgery performed (UKR vs TKR) and post-operative PROMs, account for clustering on surgeons. Survival models will be fitted to study time-to-event (one model for each of the proposed secondary outcomes) according to UKR/TKR (in Stage 2). Poisson models will be fitted to study secondary outcomes in Stage 1 because it would allow to have a direct comparison with TOPKAT findings. Generalized linear models (GLMs) will be used for the study of costs and their relationship with surgery type (UKR vs TKR).

In the first stage, different methods will be tested to evaluate the association between KR type and both primary and secondary outcomes in the comparison cohort: 1.Propensity score (PS) methods; 2.High-dimensional PS; and 3.Instrumental variable analyses. A chi square test for heterogeneity will be used to formally test for differences between the estimates obtained in TOPKAT compared to the different observational analyses.

In a second stage, those methods able to obtain results equivalent (i.e. not significantly different) to the TOPKAT post-operative PROMs findings will be applied to the analysis of the association between UKR (compared to TKR) and all study outcomes (risk/s, revision, benefits, mortality, costs and cost-effectiveness) in the co-morbidity cohort.

TIME TABLE

Stage 1

Months 1-6: Submission of application form, approval by the NJR Research Sub-committee, and data linkage/extraction by the NJR; Months 7-10: Data management; Months 11-14: Stage 1 data analyses; Month 15: Steering Committee meeting to review Stage 1 results and to decide on study progress (Go/No-Go);

Stage 2

Months 16 to 22: Data Analyses (including health economics); Months 23-25: Writing of study report (PI + PC), dissemination, and study closure.

EXPERTISE

Multidisciplinary team including academic general practitioners, rheumatologists, orthopaedic surgeons, allied healthcare professionals, and experts in orthopaedics, rheumatology, epidemiology, statistics, health economics, clinical trials, and analysis of routinely collected data, as well as a patient representative.

BACKGROUND AND RATIONALE

Surgical randomised controlled trials (RCTs) generate gold standard evidence. However, despite recent evidence suggesting that surgical RCTs are both safe and useful(13), they remain uncommon for a number of reasons, including costs, time, ethical concerns, surgeon equipoise, and feasibility(14, 15). Non-randomised studies relying on routinely collected data could offer an efficient alternative for the comparative assessment of surgical interventions in the National Health Service (NHS). In addition, these studies offer results potentially generalizable to the whole population of real world NHS patients (regardless of comorbidities or age) including patients who would have been excluded in RCTs, and they can be conducted at a much lower cost as well as within a shorter time. However, observational studies are limited by confounding and related bias due to the non-random allocation of treatment alternatives.

One therapeutic area where this is a live and highly relevant issue is the choice of Unicompartmental Knee Replacement (UKR) or Total Knee Replacement (TKR), which has led to the funding of TOPKAT (NIHR HTA – 08/14/08: Total or Partial Knee Arthroplasty Trial), an ongoing Health Technology Assessment (HTA)-awarded, £2.7 million multi-centre RCT. TOPKAT has successfully recruited the target number of patients, and 1-year follow-up data on patient reported outcomes (PROMs) will be disseminated in the coming weeks. Despite the unquestionable quality and internal validity of RCTs such as TOPKAT, two key issues limit their usefulness for determining the comparative cost and effects (risk-benefit) of the studied NHS interventions in actual practice conditions: 1. Limited external validity: only patients with ASA (American Society of Anesthesiologists) grades 1-2 were eligible for enrolment in TOPKAT thus excluding patients with multiple comorbidities; and 2. Length of follow-up and power: due to the cost and difficulty of primary data collection, surgical RCTs are underpowered to detect rare events, and short in length of follow-up. This limits the availability of data on key (usually rare and long-term) safety outcomes including complications (revision, systemic infection, wound infection, cardiovascular disease, and venous thromboembolism) and mortality.

National Joint Registry (NJR) reports suggest that about 1 in 6 candidates (16.7%) for Knee Replacement (KR) surgery are ASA grade 3 or worse (16), therefore not eligible for TOPKAT. There is thus an urgent need for data on the performance of such different surgical approaches for multi-morbid patients requiring knee surgery, which TOPKAT will not provide.

Observational data from the NJR could, conversely, be helpful: in a recent Lancet paper, the authors used one of the most widely extended methods (propensity score matching) to minimise bias(5, 6, 11, 17). In this manuscript, the authors acknowledged that unmeasured confounders (such as unrecorded conditions, disease severity, or drug use) could at least partially explain the study findings since propensity score matching could only account for

measured confounders. Such unresolved bias can however be minimised with alternative (more novel and robust) pharmaco-epidemiological analytical methods, such as instrumental variables(18) or high-dimensional propensity scores(19): these methods have recently been applied in observational comparative safety and/or effectiveness research. These methods have been developed and tested in drug and vaccine studies but – to our knowledge - they have not been used to compare the performance of different surgical procedures. Hence, there is a need for a better understanding of the performance of the different methods listed above for comparative effectiveness/safety studies for the evaluation of surgical and device alternatives using (observational) routinely collected data.

EVIDENCE EXPLAINING WHY THIS RESEARCH IS NEEDED NOW

Although the ongoing multi-centred RCT (TOPKAT) will soon provide top quality evidence on the clinical- and cost-effectiveness of unicompartmental (UKR) compared to total knee replacement (TKR) for medial compartmental knee osteoarthritis, the result is not generalizable to patients with ASA grade 3 or worse (equivalent to severe or very severe systemic disease)(16), who were not eligible for TOPKAT: as reported in recent NJR reports(16), there are differences in PROMs according to ASA grade, and there are also known associations between co-morbidities and both post-operative complications and mortality(4).

This group is not to be underestimated, representing almost 17% of the people undergoing KR surgery according to National Joint Registry (NJR) report, and probably –given their baseline medical history and risk factors- a much higher proportion of the NHS expenditure in knee replacement surgery and related hospital admissions. However, the difficulty to recruit elderly as well as patients with multiple co-morbidities for surgical RCTs is well known, requiring alternative solutions to generate evidence for this substantial group of patients. We expect that our proposed study will provide an efficient answer on the risks, benefits, and cost-effectiveness of UKR compared to TKR for patients with multiple co-morbidities (classified as grade 3 or worse according to the ASA system), equivalent to severe or very severe systemic disease (16). This research is in line with the recently published NIHR themed call on the evaluation of interventions or services delivered for older people with multi-morbidity.

AIMS AND OBJECTIVES

AIMS

In line with the NIHR themed call on "Multimorbidities in older people", our overarching aims are:

1. To study the validity of different epidemiology analytical methods -used in drug and vaccine studies to minimise confounding- for the assessment of alternative surgical procedures. We will use knee replacement (UKR compared to TKR) amongst ASA grade 1-2 (eligible for TOPKAT) as a use case for demonstration, where the previously mentioned TOPKAT RCT will be used as a "gold standard" for comparison.
2. To apply "valid" methods (as identified/validated in Aim 1) to the analysis of risks, benefits, costs and cost-effectiveness of surgical alternatives for knee replacement (UKR versus TKR) for patients with multiple comorbidities, ineligible for the TOPKAT RCT.

SPECIFIC OBJECTIVES

1. To assess whether any of a number of proposed pharmaco-epidemiological methods offer comparable findings to those obtained from the TOPKAT RCT amongst participants in the NJR eligible for TOPKAT (ASA grade <3). The proposed different methods (propensity score methods, high-dimensional propensity scores, and instrumental variables) will be applied in separate analyses, and those offering results comparable to TOPKAT will be deemed valid, and applied to Objective 2.

2. To apply the "validated" methods (as in Objective 1) to the comparative study of benefits (PROMs), risks (revision surgery, complications, and mortality), hospital costs and cost-effectiveness of UKR (compared to TKR) amongst NJR participants with multiple and severe co-morbidities (ASA 3 or above) and therefore not eligible for TOPKAT.

RESEARCH PLAN

METHODS

Study Stages/Structure

We propose a 2-staged approach:

Stage 1: Validation of pharmaco-epidemiological methods for the comparison of knee surgery alternatives.

Firstly, we will establish the validity of different novel statistical methods for the non-randomised evaluation of surgical interventions by attempting to replicate the 1-year results from the TOPKAT RCT: we will study the association between unicompartmental (compared to total) KR and post-operative PROMs (primary outcome in TOPKAT) amongst participants in the observational dataset deemed eligible for TOPKAT (ASA 1 or 2 in the NJR dataset). The different analytical methods described above will be tested, and the one/s successful to obtain equivalent results to those from the RCT will be applied in a second stage (below).

We will compare post-operative PROMs (Oxford Knee Scores, OKS) from TOPKAT with those obtained from the NJR-HES-PROMS dataset after each of the proposed analyses. We will report on the difference in estimates, as well as on the overlap in the OKS 95% confidence intervals (95% CIs), to then formally test for significant differences between the TOPKAT findings and each of the database analyses using a chi-square test for heterogeneity (Tau squared) as proposed by M Egger et al(10). Only those observational methods providing estimates not (statistically) significantly different from TOPKAT will be considered "valid" in Stage 1. These will then be applied to the "co-morbidity cohort" in Stage 2.

The following methods will be tested as part of Stage 1:

1. Propensity score methods: previously used in similar studies(6), propensity scores (PS) represent the probability that a specific patient receives an intervention (i.e. here UKR) based on his/her baseline characteristics. In PS analyses, PS are calculated using logistic regression modeling where treatment group (UKR vs TKR) is the outcome of interest. Once PS are estimated for all participants, a number of "exposed" subjects (those undergoing UKR in our case) are matched to non-exposed (i.e. TKR recipients in our proposal) ones, therefore providing with comparable groups (PSM). Despite existing evidence of their usefulness in pharmaco-epidemiology, recent concerns have been raised on their limitations(17), which do not account for unobserved patient characteristics/variables, often leading to residual confounding and bias. We will also use other propensity score models including propensity score stratification and propensity score adjustment. Unlike the PS matched analyses, both methods use all

eligible participants in the NJR-HES. In Propensity score stratification, outcome analyses, i.e. linear regression/Poisson models, will be done by per strata before deriving into a combined estimate. Propensity score adjustment means that outcome analyses will be adjusted for the propensity score.

2. High-dimensional propensity scores (HDPS)(20) use all information available in electronic medical records (EMR), including not only patient characteristics (like PS) but also physician-based (surgeon in this case) and health care system features (type of hospital, surgical volume, etc) in an attempt to account for unmeasured confounding. Recent research suggests that HDPS may reduce bias compared with 'traditional' PS(19). This is based on the assumption that if we could measure a battery of proxies, we would increase the likelihood that in combination they are a good overall proxy for relevant unobserved confounders. This approach was not implemented during the analyses because this method cannot apply for continuous outcomes (OKS).

3. Instrumental variable analyses (IV) rely on the existence of an 'instrument', a variable that is related to the treatment but not to the study outcome (other than through treatment effects). Common examples of instruments identified in healthcare research include: 1.distance to hospital providing specialized treatment/s(21); and 2.physician (prescription or surgery) preference (based on the assumption that different practitioners have different preferences for one treatment versus another)(22). IV methods have been compared to other techniques both in clinical and simulated datasets(9, 23, 24), and shown to be of value particularly in comparative effectiveness research using observational data(18). An important requirement of preference-based instruments is that there should be variability among physicians in their preference for the different treatments under study. We will start from a number of potential instruments to include 1.surgeon preference, 2.surgical experience, 3.volume of surgical procedures, 4.geographical location, and 5. Calendar time. We will first test each of these for the underlying IV assumptions explained above. We will then conduct the proposed analyses using only those instruments proved to fulfill the mentioned assumptions. This strategy is recommended by experts in the field (M Hernan et al. Epidemiology 2006) and supported by previous literature(9).

Stage 2: Risks, benefits, costs, and cost-effectiveness of UKR versus TKR for patients with multiple co-morbidities.

Secondly, we will use the previously validated methods (as in Stage 1) to assess -amongst patients with an ASA grade 3 or above- the effect of unicompartmental (compared to total) KR on: 1.PROMs, 2. complications (revision surgery, prosthetic joint infection, cardiovascular

disease, and venous thromboembolism), 3. Mortality, and 4. Incremental costs, health-related quality of life (based on EuroQoL 5-D provided as part of the PROMs database), and cost-effectiveness.

Study Period and Follow-up

The NJR for England and Wales (and subsequently Northern Ireland) was established in 2003 and collects information on all hip and knee replacements nationwide. The HES database contains inpatient data from 1992, while the linked National PROMs database has been collecting data on all NHS-funded hip and knee replacements undertaken in England since 2009. Our study period will therefore cover from 01-01-2009 to the latest data extraction for the identified data sources (likely 31/12/2016) for Stage 1 and for some of the analyses of Stage 2 (where PROMs data are needed). The analysis of revision, complications, costs, and mortality to be conducted for Stage 2 will also include earlier data potentially from 2003 to end/2016.

Follow-up and Study Outcomes

The proposed study will have different follow-up periods for the various outcomes as follows:

Primary Outcome (PROMs)

For the study of PROMs (Oxford Knee Score, OKS), patients will be followed from the start of the study period (date of UKR or TKR) until the collection of post-operative OKS (6 to 12 months post-operatively).

Secondary outcomes

Secondary outcomes include NHS hospital costs, health-related quality of life (based on pre- and post-operative EuroQoL 5-D in the PROMs database), derived cost-effectiveness, and 90-days risks of surgery complications (prosthetic joint infection, cardiovascular disease and venous thromboembolism) and 5-year risks of revision or mortality.

Patients will be studied from their index date of surgery (UKR or TKR) until the earliest of the following:

- Date of last data update (likely end/2016)
- Date of revision surgery (not applicable for costs or cost-effectiveness)
- Death
- End of the study: up to five years after index date (UKR/TKR surgery date as recorded in NJR)

TARGET ORGANIZATIONS, PROFESSIONS, AND PATIENT CARE GROUPS

The proposed study is targeted for regulators (MHRA and alike), policy-makers (NICE and similar), NHS healthcare professionals (specifically, general practitioners, rheumatologists, orthopaedic surgeons, and allied healthcare professionals), and patients suffering from severe knee osteoarthritis requiring knee replacement.

DISEASE AREA

Our proposal studies the most widely recommended surgical treatment for patients with severe knee osteoarthritis, disease responsible for >95% of the knee replacements performed in the UK(16).

RESEARCH TEAM

We have constituted a multi-disciplinary research team, with expertise in the following areas:

- Musculoskeletal Epidemiology: Prof Nigel Arden, Prof Alan Silman, and Associate Professors Daniel Prieto-Alhambra (lead applicant) and Andrew Judge have a long track record in the use of routinely collected data for musculoskeletal epidemiology research.
- Pharmaco and device epidemiology: DPA (lead applicant) and AJ have experience in the use of advanced pharmaco-epidemiological methods, and they are co-chairs of the departmental 'Big Health Data User Group'.
  In addition, two world experts in the field (Dr Irene Petersen and Dr Ian Douglas) have also been incorporated to collaborate with the proposed analyses.
- Orthopaedic Surgery and National Joint Registry data: Prof Andy Carr, Prof David Murray, and Prof Mark Wilkinson are recognised leaders in the field of orthopaedic surgery research. They have a strong track record of publications using NJR data in high impact factor journals (BMJ, Lancet), and will be key for the interpretation of our findings from observational (NJR linked to HES-PROMS) data (Stage 2).
- Primary Care, Health Delivery, and Co-morbidity research: Prof JM Valderas is a world lead in the research of management of patients with multiple co-morbidities. He is also a practicing GP, and will be providing key feedback on the consequences of our findings from both angles.
- Randomised Controlled Trials: Prof Sallie Lamb is the Co-Director of the Oxford Clinical Trials Research Unit; Prof David Beard has extensive expertise in the conducting of surgical RCTs, he is the lead investigator for TOPKAT as well as Co-Director of the Royal College of Surgeons (RCS) Surgical Intervention Trials Unit.

- Health Economics: Dr Pinedo-Villanueva is a Senior Health Economist within the NDORMS Musculoskeletal Epidemiology group. He has extensive experience using routinely-collected data for the economic assessment of joint replacement interventions.
- Patient and Public Involvement: Ms Sue Thwaite is a patient representative for the National Rheumatoid Arthritis Society. She underwent knee replacement surgery (two primary procedures and one revision) for severe osteoarthritis and rheumatoid arthritis, and will be instrumental for the interpretation and dissemination of our findings from a lay audience/patient perspective.

HEALTH TECHNOLOGY BEING ASSESSED

There are two common different approaches to replacing a knee joint once severely damaged by knee osteoarthritis. Some surgeons feel that it is always best to replace the entire knee with a Total Knee Replacement (TKR) while others feel it is best to preserve –as much as possible- the original anatomy and replace only the damaged component of the knee with a Unicompartmental Replacement (UKR). There is little agreement amongst knee surgeons on the best procedure although both interventions are well established in the NHS. As a result, there is high variation in the uptake of UKR nationally.

The current proposal will complement the results of the HTA-funded TOPKAT RCT, where unicompartmental (UKR) is compared to total Knee Replacement (TKR). Our study will –once observational methods are validated in Stage 1- assess the effects (risks and benefits) and costs of these two alternative surgical procedures amongst NHS patients with multiple co-morbidities undergoing any of both, who were not eligible for TOPKAT according to the listed inclusion/exclusion criteria.

DESIGN AND THEORETICAL/CONCEPTUAL FRAMEWORK

CONCEPTUAL FRAMEWORK

-Stage 1: An observational cohort study using routinely collected data (NJR linked to HES-PROMS) will be performed to evaluate the validity of different analytical methods – discussed above- for the assessment of two alternative surgical treatments for severe knee osteoarthritis: UKR and TKR. The previously mentioned TOPKAT RCT will be used as a "gold standard" for comparison as a reference for validation.

-Stage 2: As a next step, the "validated" methods (in Stage 1) will be applied to patients in the linked dataset (NJR linked to HES and PROMS) with an ASA 3+ to study the benefits (PROMs), risks (complications and mortality), costs, and cost-effectiveness of UKR (compared to TKR) amongst NJR participants not eligible for TOPKAT (ASA 3 or above).

STUDY DESIGN

The chosen study design is a retrospective cohort study based on routinely collected data from the NJR linked to hospital admissions (HES) and national patient reported outcomes (PROMs). This approach will ensure the inclusion of potentially all NHS patients and treatment centres providing knee replacement surgery in England, therefore maximizing generalizability. In this way, our proposal will include patients with multiple and severe co-morbidities (ASA grade 3 or above), who represent almost 17% of the NHS patients undergoing knee replacement surgery(16) but yet have been excluded from the ongoing TOPKAT RCT.

TARGET POPULATION

-Stage 1: The target population for the validation study (Step 1) will be those patients in the NJR undergoing primary UKR or TKR who were relatively healthy at the time of surgery (ASA grade 1-2) and therefore eligible for the TOPKAT RCT.

-Stage 2: The target population for this comparative study of risks, benefits, costs and cost-effectiveness of UKR (compared to TKR) will be NJR participants undergoing primary UKR or TKR but with multiple co-morbidities (ASA grade 3 or above) at the time of surgery, and therefore not eligible for TOPKAT.

INCLUSION/EXCLUSION CRITERIA

Comparison cohort (Stage 1)

From the target population described above, we will exclude those with any of the following exclusion criteria:

- NJR participants with no linked pre-operative PROMs data available
- NJR participants with no linked post-operative PROMs data available
- NJR participants with previous cruciate ligament injury, therefore ineligible for UKR
- NJR participants with ASA 3 or above (according to NJR data)
- NJR participants with any other exclusion criteria for TOPKAT (to maximize comparability)(26):
  - Undergoing revision surgery (according to NJR)
  - History of inflammatory arthritis (either in HES or NJR data)
  - Evidence of lateral cartilage or patello-femoral injury (as coded in either HES or NJR data)
  - History of foot, hip or spinal pathology, or septic arthritis (as coded in HES or NJR)
  - Previous knee surgery other than diagnostic arthroscopy and medial meniscectomy (as coded in HES or NJR)

Co-morbidity cohort (Stage 2)

From the target population described above, we will exclude those with any of the following exclusion criteria:

- NJR participants with no possible linkage to HES (for the study of complications and costs)
- NJR participants with no linked data available on pre or post-operative PROMs (for the study of PROMs and related effectiveness evaluation)
- NJR participants with previous cruciate ligament injury or inflammatory arthritis (ineligible for UKR)
- NJR participants undergoing revision surgery

SETTING/CONTEXT

The proposed study focuses on secondary care surgical therapies for knee osteoarthritis. We will obtain data from potentially all NHS centres and patients undergoing knee replacement, as coded in the NJR.

SAMPLING

One of the advantages of routinely collected datasets is that the contained information is readily available, not needing active recruitment. Therefore, all patients eligible (see target population and exclusion criteria above) registered in the chosen data sources will be included.

SAMPLE SIZE/POWER

We have calculated sample size for the proposed analyses of primary (post-operative OKS) and secondary (revision, complications, and mortality) outcomes separately using the English version of the GranMo software, available online at: http://www.imim.cat/ofertadeserveis/software-public/granmo/

Primary outcome

Based on published data(11), over 720 UKR and 8,400 TKR recipients in the smaller (*multi morbidity*) cohort will be available in the linked dataset including NJR-HES-PROMs. With an expected standard deviation of 8 in post-operative OKS, a Group size 1/Group size 2 ratio of 11.66 (8,400/720), an alpha risk of 0.05, a 10% drop-out, and 90% power, 203 UKR and 2,366 TKR subjects are needed to detect as significant a minimally clinically important difference of 2 points in Oxford Knee Score -as demonstrated in previous studies(12) and proposed for the power calculation of TOPKAT(26) - in a two-sided test. This number is lower than the expected sample size available in our proposed study.

Secondary outcomes

Accepting an alpha risk of 0.05 and a beta risk of 0.2 (80% power) in a two-sided survival analysis (Log-Rank test), and assuming cumulative rates at 1, 3 and 5 years of 3%, 5% and 10% in the reference group (TKR patients) with 10% drop-outs, 3,850 UKR (193 with the outcome of interest), and 42,338 TKR recipients (2,104 events) would suffice to detect as significant a >20% risk reduction in UKR participants. According to previous studies, over 30,000 UKR and 350,000 TKR will be available for these analyses(6), with 16% of them (4,800 UKR and 56,000 TKR) being eligible for the smaller (*co-morbidity*) cohort according to NJR reports, a number that will provide >80% power for the study of interest. Smaller (<20%) risk reduction/s or more uncommon (<10% at 3 years) events will be underpowered but probably less clinically relevant.

DATA COLLECTION, LINKAGE, AND MANAGEMENT PLAN

As stated above, one key advantage of routinely collected data is the immediate access to large and representative samples of patients with no need for prospective data collection.

More challenging is however the workload involved in the data management required for the current study, where 3 different data sources (PROMS, HES and the UK NJR) will be linked and anonymised by a trusted third party (NHS Digital). To guide this process, we have included co-applicants with extensive experience in linkage between these data sources as described above. All data management tasks needed to produce a final working dataset will be carried out by an expert senior data manager. With supervision from the PI and the research team, she will develop ad-hoc code in Python and SQL to produce a dataset that will then be analyzed using standard statistical packages such as Stata and R.

All the events/outcomes of interest will be ascertained using ad-hoc collected data from the NJR (e.g.revision surgery) and pre-specified lists of either validated or agreed ICD/OPCS (HES) codes (e.g.venous thrombo-embolism). Pre-specified lists of codes will be used, set after following a number of steps:

1. Literature review of previous studies using or validating HES data(1, 2). Where available,

the list/s of validated or previously used codes will be pulled from the manuscript or online (supplementary) appendices and used as the proposed list of codes for our study.

2. If no such studies exist for any of the study outcomes, new lists will be developed after consensus by all the clinicians who are co-investigators in the research team (including two GPs, two orthopaedic surgeons, and two rheumatologists) following the steps recommended by Dave S et al(27).

DATA ANALYSIS

Different pharmaco-epidemiological methods will be validated in Stage 1 of the study. Those deemed 'valid' will then be applied in Stage 2. The methods under study will be:

Propensity score based methods

Propensity scores represent the probability that a patient will receive the intervention of interest (i.e. UKR) according to their baseline socio-demographics and clinical characteristics. Multivariable logistic regression equations will be used to calculate one propensity score for each of the outcomes of interest. We will explore four different PS approaches: propensity score matched analyses, propensity score stratification, inverse probability weighting, and propensity score adjustment.

PS matched analyses: the created propensity scores will be used to match UKR patients to comparable TKR patients with a caliper matching technique with a maximum caliper width of 0.02 standard deviations (SDs). In short, this means that TKR patients will only be eligible for matching if their propensity score falls within a bandwidth of 0.02 SDs of the UKR patient's propensity score. This matching method has been shown to be the most efficient to minimize confounding by indication in pharmacoepidemiological studies(20) and typically excludes the small proportion of patients with extremely high or extremely low risk for the outcome that are not present in both intervention and comparator patient samples.

PS stratification: Outcome analyses, i.e. linear regression/Poisson models/Competing risk analyses(Stage 2), will be done by per strata before deriving into a combined estimate. Propensity score adjustment means that outcome analyses will include the created PS score as a covariate.

Inverse Probability Weighting: IP weighting means that in outcome analyses will be weighted by the inverse of the probability of individuals receiving UKR based on measured confounders. In other words, outcome analyses will be weighted by the inverse of PS. With the implement of weighting, IPW creates a pseudo-population in which measured counfounders and the exposure of interests are independent of each other.

High-dimensional propensity score (HDPS)

HDPS methods are an advanced solution to the issue of residual confounding due to the lack of information on relevant co-variables or patient characteristics in PS-matched analyses. HDPS methods accomplish this by measuring proxies for important confounder constructs. The algorithm involves the following steps: 1) identifying data dimensions, eg, diagnoses, procedures, medications, surgeon and health-care system (hospital) characteristics and; 2) empirically identifying candidate covariates and assessing recurrence of codes; 3) prioritizing covariates; 4) selecting covariates for adjustment; 5) estimating the exposure propensity score; and 6) estimating an outcome model. Once HDPS are estimated, HDPS-matching with a pre-specified caliper (similar to PS-matching above) will be applied. This analysis is not implemented because this method cannot deal with continuous outcomes.

Instrumental variable analyses (IV)

IV methods rely on the existence of an 'instrument', an observed variable related to the exposure/treatment under study and to the outcome/s of interest only through the treatment/exposure effect/s. This resembles a randomized trial, in which treatment allocation typically almost perfectly coincides with the actual treatment received and (in case of a double-blinded RCT) treatment assignment only affects the outcome through the allocated treatment (hence the term pseudo-randomisation that is used for IV methods). The following instruments will be constructed and then tested against the underlying IV assumptions: 1.Preference-based instruments (physician (here 'surgeon') preference for a treatment (here UKR); surgical experience; and hospital volume); 2.Geographical location; and 3.Calendar time (ie date of surgery).

Constructing instrumental variable

For **surgeon preference**, NJR data will be sorted in an increasing order of dates of operation and, second, we will use three different approaches: 1) Surgeon preference is calculated based on the last twenty consecutive preference/procedures (UKR/TKR) or 2) the last thirty consecutive procedures (UKR/TKR) or 3) the last 50 consecutive procedures, i.e, For each patient, we will see what was the surgeon PREVIOUS (20 or 30 or 50) preference(s) (UKR/TKR) and then this proportion of UKR procedure will be used as an instrument for the patient. The choice in the number of previous consecutive preferences depend on the average proportion of UKR and TKR in surgeons.

For **surgeon's experience** and **hospital volume**, we will use the number of knee replacement procedures undertaken by/in each of the surgeons/centres identified in the NJR to create these instruments.

Regarding **geographical location**, patient region of residence as provided by HES will be used to construct the instrument.

Finally, calendar time will be constructed based on the recorded date of surgery. We will determine the secular trends of UKR surgery in the NJR data and establish an inflexion point when UKR uptake took off (if there is such) as an instrument.

Checking underlying assumptions and selection of 'valid' instruments

As noted, IV analysis must satisfy three basic assumptions(28, 29), which we will test for each of the proposed instruments before conducting such analyses as follows:

> 1) The IV must be strongly associated with the exposure. The F-statistic value from the first-stage linear regression model will be used to statistically test this assumption for each of the instruments. We will use the rule of thumb that if the F-statistic value is greater than 10, the first assumption holds (30, 31). Otherwise, the instrument will be deemed 'invalid' and not applied for the IV analyses (see 3.3).
>
> 2) The IV must not have direct effects on the outcome except through its association with the exposure
>
> AND 3) The IV is independent of confounders

The two latter assumptions are unverifiable or not directly testable as they involve unobservable variables(29). We will use circumstantial evidence to support them, and (specifically for the second one), we will work under the assumption that surgeon and hospital allocation, region of residence, and date of surgery are at random, and therefore not associated with any potential confounders.

In addition, we will use a falsification test based on the standardized difference to test for the third assumption: if the IV is associated with measured confounders then it might also be associated with unmeasured confounders. A cut-off point of 10% for the standardized difference in means or proportions of confounders between IV groups has been proposed for the formal testing of this(32). Again, if any of the proposed instruments violates this assumption it will be deemed not valid and therefore not applied for the IV analyses.

IV analyses

Once the assumptions above have been checked, we will proceed to the IV analyses. We will use the best known two-stage method for IV analysis, the 2SLS (2-stage least squares) method which is traditionally used in IV analyses(33). Unlike other estimators (e.g., ratio estimators), this method is able to adjust any possible measured confounders. The 2SLS estimator can be obtained by the following models:

1. The first stage model will estimate the effect of the IV on exposure (UKR versus TKR).
2. The second model will compare outcomes in terms of predicted exposure rather than the actual exposure. The latter model yields the estimated parameter, which is the IV

estimator.

Economic evaluation of UKR compared to TKR

To determine the cost-effectiveness of UKR vs. TKR, we will conduct a cost-utility analysis up to five years post-surgery, similar to that proposed in TOPKAT. The analysis plan will follow best practice guidelines for cost-effectiveness analysis for non-randomised individual patient level data (34). The 'validated' method/s identified in stage 1, will be applied to estimate the cost-effectiveness of these two interventions in patients with multiple co-morbidities (ASA grade 3 or above) at the time of surgery, and therefore not eligible for TOPKAT.

A UK NHS perspective will be adopted and quality-adjusted life years (QALYs) will be used as the main health outcome measure. The impact on health-related quality of life (HRQL) associated with UKR and TKR will be assessed using data from the EQ-5D instrument collected pre-operatively and 6-month post-operatively via the HES-PROMs. The UK value set to derived utilities (35) will be used to create a patient-specific utility profile to generate QALYs. These scores express the HRQL in a given time period on a scale anchored at 0 (dead) and taking values up to 1 (perfect health). Utility values will be connected using a straight-line association between follow-up points. Missing data on EQ-5D questionnaires will be imputed if necessary using multiple imputation approaches (36). A comparison will be made with mapping OKS, if not missing, to utilities using a validated algorithm (37). In the base case, the path in utilities beyond 6-month post-operative time period will be assumed to remain unchanged up to five years after surgery based on findings from unpublished work (Burn, et al). Other plausible scenarios will be explored in sensitivity analyses including a flat degradation to the pre-operative level over the average implant survival (38) and the natural degradation of HRQL observed in patients with long term conditions (39).

Hospital costs will be estimated using the patient-level data provided in HES. All hospital care associated with the knee replacement surgeries as well as selected complications will be included in the analysis for patients with ASA 3+. Hospital costs will be derived by grouping each hospital episode to a healthcare resource group (HRG), which will be valued using NHS Reference Costs (40). As the economic evaluation will be conducted based on observed patient-level data and PROMs records have been systematically collected since 2009, resource use and outcome data between this year and 2016 will be considered for the economic analysis.

Results will be produced aggregated for all patients and for relevant subgroups as permitted by the richness of the data. Cost and effects results will be reported as means with 95% confidence intervals. An incremental cost-effectiveness ratio (ICER) will be estimated by dividing the difference in costs by the difference in QALYs of the two treatments under analysis and will be depicted on the cost-effectiveness plane. This will be interpreted as the additional

costs/savings associated to the additional QALY benefits from doing UKR compared with TKR. Uncertainty around the ICER will be presented using parametric and non-parametric confidence intervals for the ICER (if appropriate), net-benefits, and plotting mean differences in costs and effects on the cost-effectiveness plane (41, 42). Cost-effectiveness acceptability curves will be constructed to derive the probability of UKR being cost-effective for different values of willingness to pay for QALY gained compared to TKR. Value of information techniques will be used to inform policy decisions about the value of further research. An annual discounting rate of 3.5%, based on current guidance (43), on both costs and health effects will be used.

Resulting estimates

Beta coefficients

Linear regression modeling will be used to study the association between knee surgery type (UKR versus TKR) and post-operative Oxford Knee Score (primary outcome), account for clustering on surgeons Beta coefficient (representing adjusted average difference in means between groups) and 95% confidence intervals will be reported for the comparative effect of UKR (versus TKR) on postoperative Oxford Knee Score.

Hazard ratios (Stage 2)

We will use proportional hazards Cox regression modelling to estimate the Hazard Ratio (HR) and 95% Confidence Intervals for each of the events of interest (implant survival, systemic infection, wound infection, cardiovascular disease, venous thromboembolism, and death) according to the type of Knee Replacement. To account for the matched cohort approach proposed, we will use Cox regression stratified by matched sets. If there were differences in mortality amongst TKR and UKR participants, Fine and Gray survival analyses(44) would be used instead, to account for a competing risk with death. Failing to do so would results in a biased estimation of the excess/reduced risk of the events of interest amongst UKR recipients (45).

Relative Risks (Stage 1)

We will use Poisson models to estimate 90-days and 5-year surgical complication risk for each of the complication outcomes (prosthetic joint infection, cardiovascular disease and venous thromboembolism, revision or mortality). Clustering on surgeons will be accounted in the multi-level analysis.

Comparing observational and randomized (TOPKAT) results [Stage 1]

We will compare post-operative PROMs (Oxford Knee Scores, OKS) from TOPKAT with those obtained in the NJR-HES-PROMS dataset after each of the proposed analyses. We will then report on the difference in effect estimates (mean differences/beta coefficients comparing OKS amongst those undergoing UKR and TKR), as well as on the overlap in the OKS 95%

confidence intervals (95% CIs). As suggested in the co-applicants meeting, it was decided to use a minimally clinical significance difference of 4 was a cut-off for a meaningful difference.

We will then formally test for significant differences between the obtained estimates (mean difference in postoperative PROMs between UKR and TKR patients) in TOPKAT and each of the database analyses using a chi-square test for heterogeneity, and finally estimate the between study variance (Tau squared) as proposed by M Egger et al(10).

Only those observational methods providing estimates not (statistically) significantly different from TOPKAT will be considered "valid" in Stage 1. These will then be applied to the "co-morbidity cohort" in Stage 2.

Comparing effects and costs of UKR versus TKR using observational data [Stage 2]

The "validated" method in stage 1, will be applied to in the comparative study of risks, benefits, costs and cost-effectiveness of UKR (compared to TKR) using NJR participant (linked to HES) undergoing primary UKR or TKR but with multiple co-morbidities (ASA grade 3 or above) at the time of surgery, and therefore not eligible for TOPKAT. The statistical analysis will be similar to what has been described in the previous Paragraphs.

Missing data

Missing data as a key issue in routinely collected data analyses, although no missing is expected for the study exposure (UKR/TKR) or outcomes, it is indeed likely that data will be missing for some of the confounders in our study. The cumulative effect of missing data in several variables would otherwise (in complete case analyses) lead to exclusion of a substantial proportion of the original sample, causing bias as well as a loss of precision and power. This bias can be overcome using multiple imputation, which allows for the uncertainty about missing data by creating several plausible imputed datasets and appropriately combining their results.

We will impute missing covariates for the propensity score (logistic) models using multiple imputation by chained equations methods. We will use the multiple imputation procedure in Stata(46), including all predictor variables in the multiple imputation process, together with the outcome variable and length of follow up time on the log scale(47) as this carries information about missing values of the predictors.

Sensitivity analyses

Three pre-defined interactions will be tested for using multiplicative terms in the above models, and if borderline or significant (p-val<0.1), stratified analyses will be reported: 1.by gender; 2.by age (younger / older than median age in the study datasets), and 3.by ASA grade.

To explore the impact of surgeons'' experiences on outcomes, several ad-hoc sensitivity

analyses were decided after the co-applicant meeting on Feb2019. All analyses will be restricted on surgeries by lead surgeons with at least a) 10 corresponding (TKR or UKR depending on patient's index procedure) surgeries in the past year; b) 30 surgeries; c) 50 surgeries. Only a) cohort will be used in the OKS analyses due to the lack of statistical power. An interaction between surgeons' experience and surgical procedure (TKR/UKR) will be tested in the outcome analyses, and if borderline or significant (p-val<0.1), stratified analyses will be reported.

DISSEMINATION AND PROJECTED OUTPUTS

OUTPUTS

Our study will have clear impact on and benefits for both the public and the NHS, as well as for clinical research funders including NIHR by:

- Providing information on the comparative risks, benefit (patient-reported outcomes), and cost of partial and total knee replacement for patients with multiple co-morbid conditions.

  If –as expected- UKR is safer, as effective, less costly, and thus more cost-effective than TKR for this specific patient group, it might become the first line surgical solution for severe knee arthritis in multi-morbid patients. We would then inform NICE and the Medicine and Healthcare products Regulatory Authority (MHRA) of our findings with the aim to impact on future guidelines for the treatment of severe knee arthritis. Depending on our study results, we would –if relevant- produce UK guidance documents and information leaflets for patients and health care professionals in both primary and secondary care involved in this area.

- Informing on the usefulness of efficient studies using routinely collected (non-randomised) data for the evaluation of surgical alternatives in the NHS to complement randomized studies.

  If some or all of the proposed pharmaco-epidemiological analytical methods are able to replicate the findings from an ongoing surgical RCT, these could be used in the future to provide information on the comparative risk-benefit and cost-effectiveness of surgical options for patients typically under-represented in (or even excluded from) randomized studies. This would typically include a growing proportion of the UK population: the elderly and multi-morbid patients.

DISSEMINATION

We will write a thorough report of the research at the end of the project to be included in the NIHR HTA Journal. In addition, we will publish at least two papers in national and/or international scientific journals to report key findings including (aiming for) the Lancet, the British Medical Journal, Journal of Trauma and Orthopaedics, Journal of Bone and Joint Surgery and/or similar. In order to increase the impact and accessibility of our findings, we will publish in open access when possible, and we are requesting funding for this as part of the current proposal.

Our results will also be presented at national (British Orthopaedic Association, British Society of Rheumatology, or similar) and international (American Association of Orthopaedic Surgeons,

American College of Rheumatology, or similar) scientific conferences, preferably in the format of oral presentation/s.

We will discuss our results (including risk-benefit and cost-effectiveness evaluation/s) with relevant panels at NICE to make them available for future health technology assessment/s.

We will also disseminate our findings to the public. Our PPI co-applicant will help to design materials such as leaflets for this purpose, which will be distributed in key places/events like surgeries, hospitals and meetings organised by charities. The PI will present the results in meetings with both local and regional patient groups (such as the NJR patient Network), and charities such as National Rheumatoid Arthritis Society (NRAS) and Arthritis Care will be involved in this stage to ensure we reach the public in an effective and respectful way.

Finally, we will disseminate our results through the media when possible – local radio, charity magazines, etc.- following advice from our departmental Outreach and Communications officers, as well as resources available through the Oxford NIHR Biomedical Research Centre network.

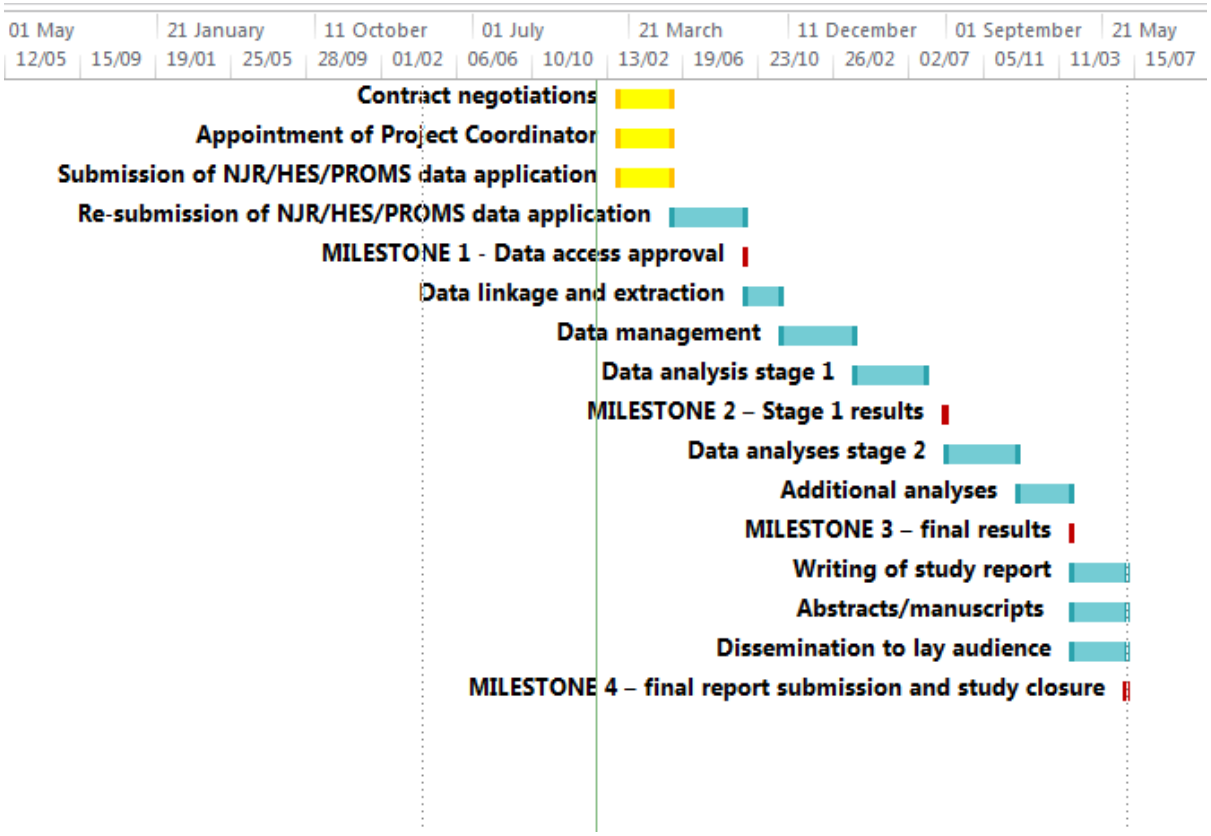
PLAN OF INVESTIGATION AND TIMETABLE

A summary of the tasks, related duration, responsible parties, and milestones (in italics) are outlined in this timetable, and graphically depicted in the following Gantt chart. Both this timetable and Gantt chart will be used for monitoring study progress and achievement of each of the listed milestones.

Task / *Milestone*	Duration (months)	Responsible
1) Contract negotiations	-3 to 0	PI, University, NIHR
2) Appointment of Project Coordinator	-3 to 0	PI
3)Submission of NJR/HES/PROMS application	-3 to 0	PI, PC
4) Kick-off meeting	1	PI, PC, SIG
5) Amendment and re-submission of NJR/HES/PROMS data application	2 to 4	PI, PC
6)Milestone 1: Data access approval	5	NJR research subcommittee
7) Steering Committee meeting: data access	5	SC
8) Data linkage and extraction	5 to 6	NJR
9) Study investigators group update meeting	7	PI, PC, SIG
10) Data management	7 to 10	DM
11) Stage 1 data analyses	11 to 14	PE
12) Study investigators meeting: internal discussion of Stage 1 results.	14	PI, PC, SIG

13) Steering Committee meeting. Stage 1 results and decision on study progress	15	SC
14) Milestone 2 (go/no-go): Reporting on Stage 1 results	16	PI, PC, CoA
15) Stage 2 data analyses	16 to 19	PE, HE
16) Study investigators meeting: internal discussion of Stage 2 results	19	PI, PC, SIG
17) Additional analyses	20 to 22	PE, HE
18) Steering Committee meeting for discussion of final results	23	SC
19) Milestone 3: final study results	23	PI, SIG
20) Writing of study report	23 to 25	PI, PC
21) Dissemination to scientific audience	23 to 25	PI, SIG
22) Dissemination to lay audience	23 to 25	PI, PPI co-applicant
23) Milestone 4: final report submitted and study closed.	25	PI, PC, CoA

PI = Principal Investigator; PC = Project Coordinator; SC = Steering Committee; CoA = Co-applicants; SIG = Study investigators group; DM = funded Data Manager; PE = funded epidemiologist; HE = senior health economist and funded health economist.



PROJECT MANAGEMENT

A part-time project coordinator will be appointed for the whole study duration, who will be responsible for the management of the study, including facilitating communication between investigators, monitoring tasks and milestones, organising steering committee meetings, assisting in writing and submitting reports and planning and ensuring dissemination of the study results.

Co-investigators will have regular email/phone communications during the 25 months of the project. In addition, there will be periodical meetings and teleconferences organized by the project coordinator to guarantee that all milestones are delivered on time and that resources can be reallocated if necessary:

- Month 1: kick-off meeting.
- Month 4: Co-investigators teleconference.
- Month 7: meeting after data extraction.
- Month 10: Co-investigators teleconference.
- Month 14: meeting to discuss on stage 1 analysis results.

- Month 16: Co-investigators teleconference.
- Month 19: meeting to discuss on stage 2 analysis results.
- Month 22: Co-investigators teleconference.
- Month 25: Co-investigators teleconference.

More meetings will be arranged according to the needs of the project. WebEx or a similar videoconferencing application will be used to improve the quality of communication between investigators. This will make sharing materials and presentations possible even when a face to face meeting is not scheduled.

A Steering committee formed by an external chair, a patient/public representative (preferably from an organization like Arthritis Care), a statistician/epidemiologist and a relevant clinician will meet three times (on months 5, 15 and 23) to evaluate progress and to decide on progression to Stage 2 (Go/No Go milestone, during month 15).

APPROVAL BY ETHICS COMMITTEES

The proposed study will only use retrospective, routinely collected data. The identified data sources (NJR, PROMs, HES) do not request ethics committee approval to access/extract their data. Instead, approval by internal independent data access committees is required.

Access to routinely collected data from the National Joint Registry (NJR) linked to hospital admissions and patient-reported outcomes data (HES-PROMs) has been provisionally approved by the NJR (please see enclosed Support Letter). A full application will be submitted to the NJR Research Subcommittee for formal evaluation and approval in the first few months of the study (see Research Timetable).

PATIENT AND PUBLIC INVOLVEMENT

Patient and public views have usually been different from researchers' views, even though research is aimed to improve their quality of life. Research objectives must match public interest, and for that reason it is necessary to involve patients and public representatives as members of research teams.

An osteoarthritis patient helped to prioritise the study questions in early stages of the application. Also, a key patient and public representative has been identified and included as co-applicant of the current proposal from its early stages: Ms S. Thwaite, committed patient representative with broad experience in organising patient support activities, who was diagnosed with rheumatoid arthritis since young age (in 1982), had a knee replacement surgery on both legs (Right Knee in 1988 and left knee in 1991) and a revision surgery (2012). This will ensure the public perspective is present in all decisions made within the team. During the planning of the project, she has contributed by reviewing the lay summary, revising the PPI section, including plans for dissemination, and reviewing and commenting on the application form and "Detailed Project Description".

We are also aiming to have an independent representative of Arthritis Care on the Steering Committee of the project to have an external opinion of the progress of the study and its relevance for society.

In addition, a group of interested patients and members of the general public will be invited to discuss the use of de-identified NHS patient records as an alternative to surgical randomised studies. We believe this increases the validity of the study results by including not only the technical comparison of both kinds of study but also the public acceptance of alternative study designs.

The proposed study affects mainly patients with arthritis, and at least two charities provide the natural environment for the dissemination of our results to the target lay audience: Arthritis Care and the National Rheumatoid Arthritis Society (NRAS). Our PPI co-applicant will supervise dissemination of the study results in order to effectively reach the public.

The overarching aims of their involvement are: 1.to assist the study investigators in identifying the most relevant study outcomes (adverse events) from a patients' perspective, 2.to collaborate in the drafting of the grant lay summary, 3.to participate in the study investigators group to monitor and discuss study progress and preliminary results, and 4.to organize dissemination to the lay audience. By doing this we expect to increase the impact of the study and to contribute to the continual increase of the acceptance of PPI representatives as research members/advisors.

REFERENCES

1. Wright FL, Green J, Canoy D, Cairns BJ, Balkwill A, Beral V. Vascular disease in women: comparison of diagnoses in hospital episode statistics and general practice records in England. BMC medical research methodology. 2012;12(1):1.

2. Holt PJ, Poloniecki JD, Thompson MM. Multicentre study of the quality of a large administrative data set and implications for comparing death rates. Br J Surg. 2012;99(1):58-65.

3. Clarke A, Pulikottil-Jacob R, Grove A, Freeman K, Mistry H, Tsertsvadze A, et al. Total hip replacement and surface replacement for the treatment of pain and disability resulting from endstage arthritis of the hip (review of technology appraisal guidance 2 and 44): systematic review and economic evaluation. Health Technol Assess. 2015;19(10):1-668, vii-viii.

4. Hunt LP, Ben-Shlomo Y, Clark EM, Dieppe P, Judge A, MacGregor AJ, et al. 45-day mortality after 467,779 knee replacements for osteoarthritis from the National Joint Registry for England and Wales: an observational study. Lancet (London, England). 2014;384(9952):1429-36.

5. Kendal AR, Prieto-Alhambra D, Arden NK, Carr A, Judge A. Mortality rates at 10 years after etal-on-metal hip resurfacing compared with total hip replacement in England: retrospective cohort analysis of hospital episode statistics. BMJ (Clinical research ed). 2013;347:f6549.

6. Liddle AD, Judge A, Pandit H, Murray DW. Adverse outcomes after total and unicompartmental knee replacement in 101,330 matched patients: a study of data from the National Joint Registry for England and Wales. Lancet (London, England). 2014;384(9952):1437-45.

7. Hunt LP, Ben-Shlomo Y, Clark EM, Dieppe P, Judge A, MacGregor AJ, et al. 90-day mortality after 409,096 total hip replacements for osteoarthritis, from the National Joint Registry for England and Wales: a retrospective analysis. Lancet (London, England). 2013;382(9898):1097-104.

8. Boef AG, Souverein PC, Vandenbroucke JP, van Hylckama Vlieg A, de Boer A, le Cessie S, et al. Instrumental variable analysis as a complementary analysis in studies of adverse effects: venous thromboembolism and second-generation versus third-generation oral contraceptives. Pharmacoepidemiol Drug Saf. 2016;25(3):317-24.

9. Uddin MJ, Groenwold RH, de Boer A, Afonso AS, Primatesta P, Becker C, et al. Evaluating different physician's prescribing preference based instrumental variables in two primary care databases: a study of inhaled long-acting beta2-agonist use and the risk of myocardial infarction.

Pharmacoepidemiol Drug Saf. 2016;25 Suppl 1:132-41. 10. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. BMJ (Clinical research ed). 1998;316(7125):140-4.

11. Liddle AD, Pandit H, Judge A, Murray DW. Patient-reported outcomes after total and unicompartmental knee arthroplasty: a study of 14,076 matched patients from the National Joint Registry for England and Wales. The bone & joint journal. 2015;97-b(6):793-801.

12. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and knee scores. The Journal of bone and joint surgery British volume. 2007;89(8):1010-4.

13. Wartolowska K, Judge A, Hopewell S, Collins GS, Dean BJ, Rombach I, et al. Use of placebo controls in the evaluation of surgery: systematic review. BMJ (Clinical research ed). 2014;348:g3253.

14. London AJ, Kadane JB. Placebos that harm: sham surgery controls in clinical trials. Statistical methods in medical research. 2002;11(5):413-27.

15. Macklin R. The ethical problems with sham surgery in clinical research. The New England journal of medicine. 1999;341(13):992-6.

16. Board TNE. 12th Annual Report 2015. National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. Surgical data to 31 December 2014. 2015.

17. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ (Clinical research ed). 2013;347:f6409.

18. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiology and Drug Safety. 2010;19(6):537-54.

19. Garbe E, Kloss S, Suling M, Pigeot I, Schneeweiss S. High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. European journal of clinical pharmacology. 2013;69(3):549-57.

20. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology (Cambridge, Mass). 2009;20(4):512-22.

21. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. Jama. 1994;272(11):859-66.

22. Wang PS, Schneeweiss S, Avorn J, Fischer MA, Mogun H, Solomon DH, et al. Risk of death in lderly users of conventional vs. atypical antipsychotic medications. The New England journal of medicine. 2005;353(22):2335-41.

23. Cai B, Hennessy S, Flory JH, Sha D, Ten Have TR, Small DS. Simulation study of instrumental variable approaches with an application to a study of the antidiabetic effect of bezafibrate. Pharmacoepidemiology and Drug Safety. 2012;21:114-20.

24. Suh HS, Hay JW, Johnson KA, Doctor JN. Comparative effectiveness of statin plus fibrate combination therapy and statin monotherapy in patients with type 2 diabetes: use of propensityscore and instrumental variable methods to adjust for treatment-selection bias. Pharmacoepidemiology and Drug Safety. 2012;21(5):470-84.

25. Beard DJ, Holt MD, Mullins MM, Malek S, Massa E, Price AJ. Decision making for knee replacement: variation in treatment choice for late stage medial compartment osteoarthritis. The Knee. 2012;19(6):886-9.

26. Beard D, Price A, Cook J, Fitzpatrick R, Carr A, Campbell M, et al. Total or Partial Knee Arthroplasty Trial - TOPKAT: study protocol for a randomised controlled trial. Trials. 2013;14:292.

27. Dave S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. Pharmacoepidemiol Drug Saf. 2009;18(8):704-7.

28. Angrist J, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. National Bureau of Economic Research, 2001.

29. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology (Cambridge, Mass). 2006;17(3):260-7.

30. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. J Clin Epidemiol. 2012;65(2):155-62.

31. Uddin M, Groenwold RH, Boer A, Belitser SV, Roes KC, Hoes AW, et al. Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. Pharmacoepidemiology and drug safety. 2014;23(2):165-77.

32. Ali MS, Uddin MJ, Groenwold RH, Pestman WR, Belitser SV, Hoes AW, et al. Quantitative falsification of instrumental variables assumption using balance measures. Epidemiology (Cambridge, Mass). 2014;25(5):770-2.

33. Groenwold RHH, Uddin MJ, Roes KCB, de Boer A, Rivero-Ferrer E, Martin E, et al. Instrumental variable analysis in randomized trials with non-compliance and observational pharmacoepidemiologic studies. OA Epidemiology. 2014;2(1):9.

34. Faria R, Hernandez Alava M, Manca A, Wailoo A. The use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data. Decision Support Unit, Sheffield SUo; 2015 Contract No.: NICE DSU Technical Support Document 17.

35. Dolan P. Modeling Valuations for EuroQol Health States. Medical Care. 1997;35(11):1095-108.

36. Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2015;24(4):805- 15.

37. Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. Quality of Life Research. 2013;22:683-94.

38. Health & Social Care Information Centre. NHS Outcomes Framework Indicators 2015 [cited 2016 April 18]. Available from: http://www.hscic.gov.uk/searchcatalogue?productid=18497&q=eq-5d&topics=0%2fIllnesses+and+conditions&sort=Relevance&size=10&page=1#top.

39.Appleby J, Poteliakhoff E, Shah K, Devlin N. Using patient-reported outcome measures to estimate cost-effectiveness of hip replacements in English hospitals. J R Soc Med. 2013;106(8):323- 31.

40. Department of Health. NHS reference costs 2014 to 2015 2015 [cited 2018 April 18]. Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015.

41. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making. 1998;18(2 Suppl):S68-80.

42. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technol Assess. 1999;3(2):1-134.

43. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013.

44. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94(446):496-509.

45. Bolland MJ, Jackson R, Gamble GD, Grey A. Discrepancies in predicted fracture risk in elderly people. BMJ (Clinical research ed). 2013;346:e8669.

46. Royston P. Multiple imputation of missing values: Update of ice. Stata Journal. 2005;5(4):527-36.

47. White IR, Royston P. Imputingmissing covariate values for the Cox model. Statistics in medicine. 2009;28(15):1982-98.

Appendix

Varaibel Name	TOPKAT	NJR	HES
Age at Surgery	X	X	X
Gender	X	X	X
HospitalName	X	X	X
Admission Date	X	X	X
Discharge Date	X		X
Admission Reason/Indication for Implanatation	X	X	X
Side of Knee replacement	X	X	X
Name of Surgeon	X	X	
Procedure Type	X	X	
Primary procedure type	X	X	
Special Instruments and detail	X	X	
Who performed the procedure	X	X	
Lead Surgeon a locum		X	
Type of Anesthetic used	X	X	
Cement yes/No	X	X	
Cement detail	X	X	
Patella replaced	X	X	X
Bearing	X		X
Bone graft Used		X	
Consultant in charge		X	
Thromboprophylaxis	X	X	X
EaseOfReplacement	X		
Patient ASA Grade		X	
Tuorniquate use		X	X
Laminar Flow Theatre		X	
Minimally Invasive Technique Used		X	
Admitted to ICU	X		X
Number of days in ICU	X		X
Discharge Destination	X		X
Septicaemia	X		X
ConfirmedMI	X		X
Admitted to ICU	X		X
PostOperative Complication	X		X
Knee Dislocation	X		X
PEconfirmed	X		X
WoundInfection	X		X
MedicalComplications	X		X
StaffingProblems	X		X
OtherComplication	X		X
OtherComplicationDetails	X		X
Alcohol related diagnosis			X
Area of residence			X
Type of patient/Source of funding		X	X
Other diagnosis Codes (ICD9)			X
Waiting time			X
Ethnic category			X
Regional Office area where patient's GP was registered			X
Primary Care Trust area where patient's GP was registered			X
Patient Identifier		X	X
Postcode of patient			X

Overlap of Covariates in the TOPKAT, NJR, and HES.
**HES will have additional covariates including list of co-morbidities. This table is not an exhaustive list for HES, the additional variables will be used for high dimensional propensity score methods.

Health Technology
Assessment Programme

National Institute for
Health Research

Protocol version 2.0

02 August 2019