The Effects of Mental Health on Knee Arthroplasty Outcomes

Technical Report

Queen Mary University London

Version Number: V1.1

Version Date: 13.6.25

Author: Ben Gabbott

Project Contributors:
Principle Investigator – Ben Gabbott
Lead Supervisor – Prof Xavier Griffin
Data Advisory Group Lead – Dr John Robson

Table of Contents

Acronyms

APC – Admitted Patient Care
CEG – Clinical Effectiveness Group
DAG – Data Access Group
DDS – Discovery Data Service
DSH – Data Safe Haven
ED – Emergency Department
EQ5D – Euroqol 5 Dimension
EQVAS – Euroqol Visual Analogue Scale
GP – General Practitioner
HES – Hospital Episode Statistics
HQIP – Health Quality Improvement Partnership
K1 – Knee form 1
K2 – Knee form 2
MH – Mental Health
MHD – Mental Health Disorder
NEL – Northeast London
NHS – National Health Service
NJR – National Joint Registry
OKS – Oxford Knee Score
OP - Osteoporosis
PI – Principal Investigator
PROMS – Patient reported Outcome measure Score
QMUL – Queen Mary University London
SKID – Salt Key Identifier
TKR – Total Knee Replacement

1. Introduction

1.1 Executive Summary

Background

Patients with Mental Health (MH) disorders (MHD) are at higher risk of dissatisfaction, complications and chronic pain following knee arthroplasty (TKR). Chronic pain following knee arthroplasty alone costs the NHS £33,000,000 per Anum in direct care costs. Despite MH disorders being common problem, they are often underrepresented and overlooked in perioperative guidance.

Extensive literature has identified MHD as a key risk factor for poor outcome, the next translational step, patient level prediction of individuals at high risk has proven to be difficult, predominantly due to lack of granular MH data. Little research has considered the importance of specific diagnosis, severity and chronicity of the MHD.

Aim

Our overall aim is to explore mental health related risk factors which may place patients at risk of poor outcomes following knee replacement surgery.

Methodology

To achieve our aim, we will link and subsequently analyse patient data from 3 routine NHS data sources.

1.Discovery Data Service (DDS) - Holds linked patient data combining GP, Social service and Mental health records. It covers the population of North-East London (approximately 2 million people).

2.National Joint Registry (NJR) – A database for patient undergoing arthroplasty in the UK. Details include arthritis, surgery, and hospital data. It also has a linked Patient Reported Outcomes Score (PROMS) database, controlled by NHS England. Patients answer questionnaires about their general health and knee function before surgery and at approximately 6 months post operatively.

3.Hospital Episode Statistics (HES) - Is linked to the NJR. Provides details about patients' engagement with secondary health providers, and further inpatient data.

Outcomes

Knee Replacement Outcomes include

Patient knee function – As recorded using the Oxford Knee score at 6 months
Surgical complication – Any re-admission to hospital during the first 12 weeks after surgery
Chronic pain – As recorded in the EQ5D Patient questionnaire at 6 months

Primary Objective

1. To estimate the prevalence and distribution of different mental health disorders diagnoses, in patients before they undertake a knee replacement within the NHS in Northeast London from January 2014 to January 2024.

2. Estimate the association between different mental health disorder diagnoses and knee replacement outcomes within the NHS in Northeast London from January 2014 to January 2024.

Secondary Objective

3. To estimate the level of association between severity of disease (for depression, anxiety, mixed anxiety and depression only), and timing of diagnosis on the three outcomes of interest following knee replacements within the NHS in Northeast London from January 2014 to January 2024.

Analysis Plan

Initial analysis will describe the distribution and prevalence of mental health disorders diagnosis at the time of surgery, for patients undergoing knee arthroplasty.

Following this regression analysis will be conducted to investigate the effects of mental health disorder diagnosis, severity and timing on the outcomes of interest following knee replacement. Outcomes will be measured at 6 months considering baseline score and other health confounder variables.

1.2 Data Source Description

1.2.1 National Joint Registry

The National Joint Registry (NJR) is a comprehensive dataset that collects and monitors information on joint replacement procedures in the United Kingdom. Established in 2003, the NJR is one of the largest arthroplasty registries globally, providing detailed records on implants, surgical techniques, patient demographics, and clinical outcomes. The dataset includes data from primary and revision procedures for hip, knee, shoulder, elbow, and ankle replacements.

The Health Quality Improvement Partnership (HQIP) acts as the data controller for the NJR. Data management processors are covered by a number of organisations such as the NEC software solutions.

As of the most recent update, the registry comprises over 3 million procedures. The NJR boasts a compliance rate nationally of 89'\%', patient consent rate of 92'\%' and has previously been show to have a successful linkage rate of 96'\%' to other major national datasets.

1.2.2 Discovery Data Service

The Discovery Data Service (DDS) is a secure and centralised platform designed to support health data research by integrating and standardising large-scale datasets from diverse sources across the United Kingdom. It facilitates longitudinal analyses by linking primary care, social care, and other data sources while ensuring patient confidentiality. The DDS is controlled by the Discovery Data Clinical Effectiveness Group (CEG) on behalf of GPs in East London. The data processor and provider is North East London (NEL) Integrated Care Strategy

It currently stores 6 million GP patient records, in NE London alone. It has 11 million MH records nationally. Its main goal is to allow for robust analysis of population level outcomes and is focussed on combining all sources of patient data held by different providers.

Data are held in a single common data model with data values represented by a super ontology, composed of the world's leading health ontologies (EMIS, System1, Cerner). Its data completeness is 95'\%'. The data included is; demographics, diagnosis, health resource use, prescribing data, investigations and referrals. The primary data vocabulary is SNOMED CT codes. The database has never been linked to NJR previously.

1.2.3 NHS England

NHS England acts as data controller for the two below databases. The data processor is NHS Digital.

Hospital Episode Statistics

The Hospital Episode Statistics (HES) dataset, managed by NHS England, is a comprehensive administrative database that captures details of all hospital admissions, outpatient appointments, and emergency care visits in England. It includes data on patient admission details, admission diagnoses, treatments and procedural codes. It offers a robust foundation for healthcare research and service evaluation. HES comprises several datasets: Admitted Patient Care (APC), Outpatient (OP), and Accident and Emergency (AE), collectively covering over 125 million episodes annually. It is considered 'gold

standard' standard data monitoring all hospital admissions. The NJR has previously been linked to EHRs for a number of projects, resulting in multiple peer review publications.

National Patient Reported Outcome Measures

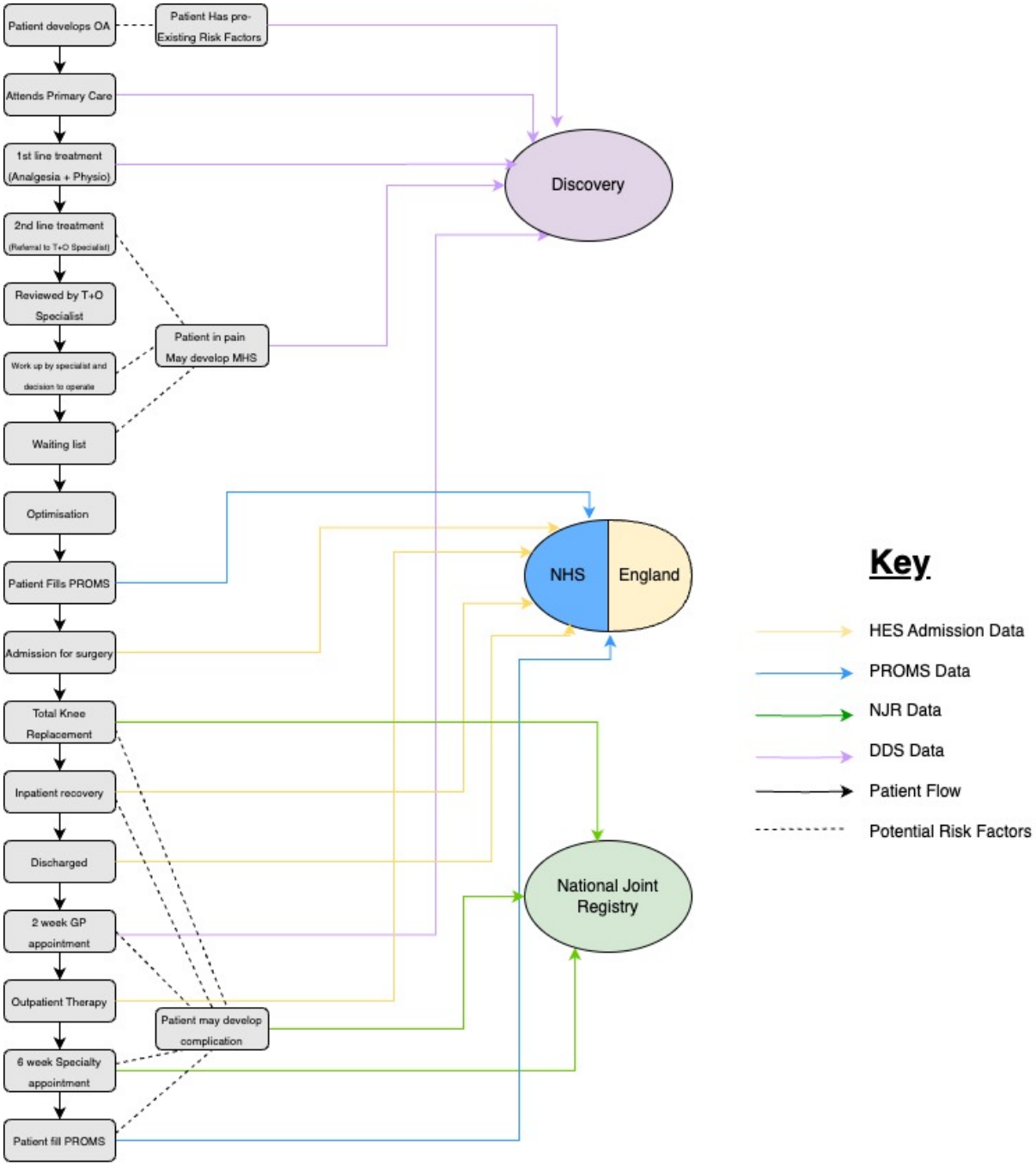
The National Patient Reported Outcome Measures (PROMs) database, administered by NHS England, captures patients' self-reported health outcomes before and after elective procedures, such as hip and knee replacements. PROMs assess health-related quality of life using standardised tools. PROMs data provide a unique perspective on the effectiveness of care from the patient's viewpoint. Participation is voluntary.

The 'Knee data set', monitoring the success of Knee arthroplasty, includes the EQ5D, EQ visual analogue scale (EQVAS) and Oxford Knee Score (OKS). The EQ-5D is a standardised measurement of health-related quality of life, with 5 questions on a 3 point Likert scale. The EQVAS is a standardised scale, where patients rate their overall health out of 100. The Oxford Knee Score (OKS) is a commonly used outcome measure, where patients self-report 12 items relating to their function and pain, giving a score out of 40.

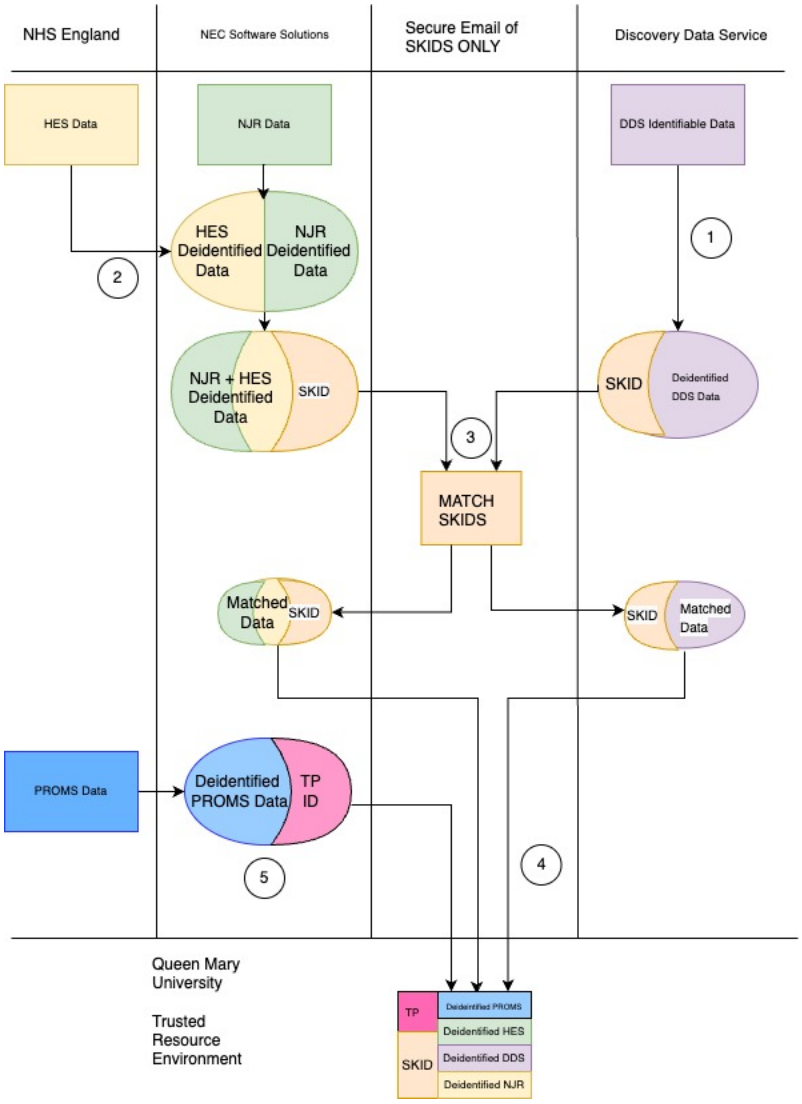
EQVAS, EQ5D and OKS have all been previously validated and are widely used to monitor knee arthroplasty. The minimal clinically important difference for TKR using the OKS is 5.

The 'Knee 1' (K1) question pack is answered pre-operatively, with the 'Knee 2' (K2) question pack being answered at 6 months after surgery. NHS England does not routinely report response rate for K1 and K2. Previous studies have reported response rate of 60'\%' and 40'\%' respectively (ref), however no study has been conducted after the COVID pandemic, which may have decreased these rates further.

1.3 Example Arthroplasty Pathway



1.4 Data Flow Diagram and Summary



Summary of the 2 stage pseudonymised linkage process between DDS, NJR and HES data sets is included below. Pseudonymisation is completed using a free and validated online secure software, open P (ref). The 2 stage linkage process involves an external linkage performed outside of the host institution, followed by an internal link within host institution Data Safe Haven (DSH). The benefits of this method include reduction of the amount of data transferred (eg - national data sets are not transferred), maintains highest levels of confidentiality (eg - QMUL never receives identifiable data, or the key to decrypt the pseudonymisation process).

1. DDS : The OpenP secure algorithm has already been run in the identifiable data in DDS. Every patient in NE London has had this new field added to their record with new unique study specific number – SKID

2. NJR and HES data: Please note that the NJR and HES have been previously linked via NEC and are linked via a unique 'Primary NJR Index Number'. Following this, the same OpenP salt identifier algorithm will be sent to the NEC. In the same way as above, NEC staff will encrypt the NHS number of all patients that fit the inclusion criteria.

3. External Link: The encrypted SKIDs of the NJR/HES and DDS cohort are then exchanged via encrypted email between the CEG and the NEC, who will match patients in both. Please note SKIDs only will be transferred at this stage, no other data.

4. Discovery/NJR Data extraction. Once the SKID exchange has occurred individuals across the 2 databases are considered 'matched' and therefore included in the project. Relevant data is then extracted for each matched person in the NJR/HES and DDS (with no identifiable data included).

5. Data transfer and internal linkage: The matched data, with SKID ID, is then surely transferred to the researcher's secure DSH via a secure file transfer. The DSH has been established, maintained and run by Queen Mary University London. All data is stored in line with GDPR and the NHS standard security model. Once within the DSH, the research team can then link the data, via any data analysis software, using the previously mentioned pseudonymous IDs.

6. PROMs Transfer: Due to administrative difficulties, the same two stage link process was not performed for the PROMS data set. A different transfer method was co-ordinated with the data controllers. The whole national knee PROMS dataset was transferred to the host institution under a different pseudonymisation number. This number could be linked to the HES data set and therefore patients linked across all datasets

2. Data Preparation

A data access group (DAG) was formed to facilitate data preparation. It contained key research representatives from the research team and data scientists from the DDS, CEG and NEC. \\
In accordance with the criteria stipulated by the DDS information governance group, primary care data was transferred for only the patients that 'matched' SKIDS. Furthermore, only a partial dataset of their relevant data was transferred (e.g, non-relevant past medical history was not transferred). Larger, complete datasets were transferred for the NJR, HES and PROMS datasets.

The patient identification criteria, geographical area's and relevant data identification is discussed in this section, followed by the linking process.

2.1 Patient Identification

2.1.1 Inclusion Criteria

- Patients are identifiable in the Discovery Data Service. EG – Registered to a GP practice in the NE London Area (see section 2.1.3).
- Patients are identifiable in the NJR data set from January 2014 to January 2024. EG – Received a TKR in any NE London NHS hospital (see below).

2.1.2 Exclusion Criteria

- Revision Knee Arthroplasty
- Uni / Partial Knee Arthroplasty
- Knee Replacements performed privately
- Data not linked across more than 1 database

2.1.3 Geographical Dataset Descriptions

Discovery Data Service

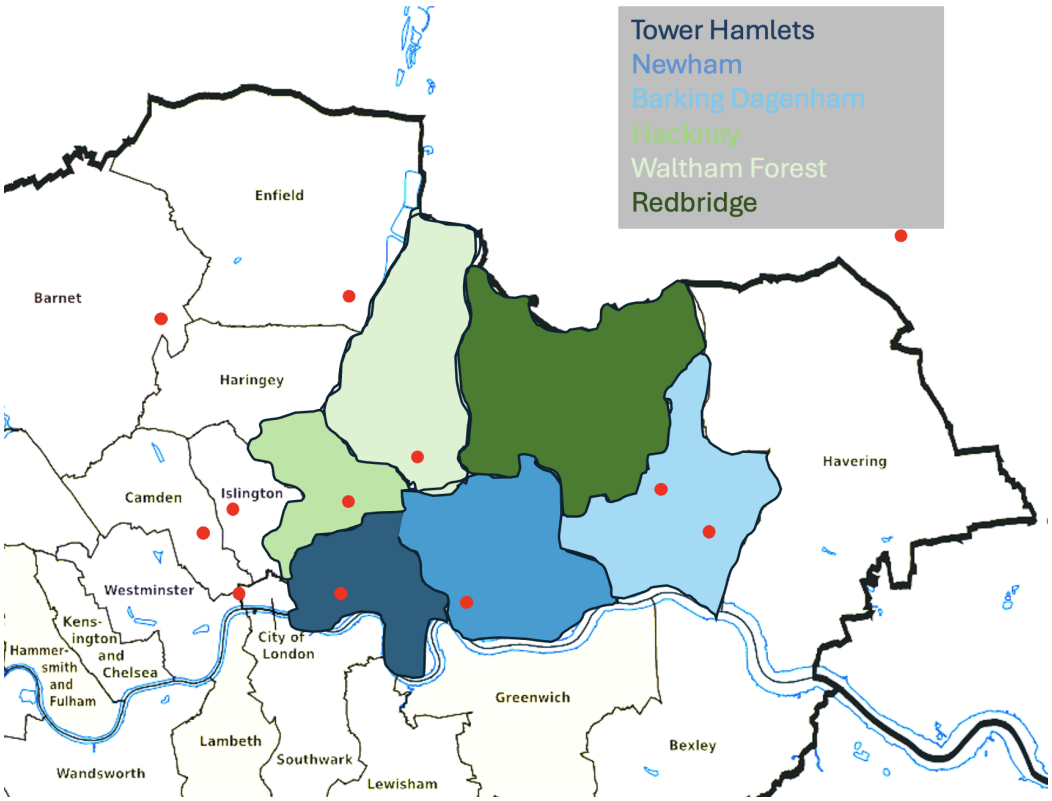
The DDS covers the 'North East London Integrated Care Board', which approximately equates to the 6 london boroughs of Barking and Dagenham, Hackney, Tower Hamlets, Newham, Redbridge and Waltham Forest(see figure 4.3a).

National Joint Registry / Hospital Episode Statistic

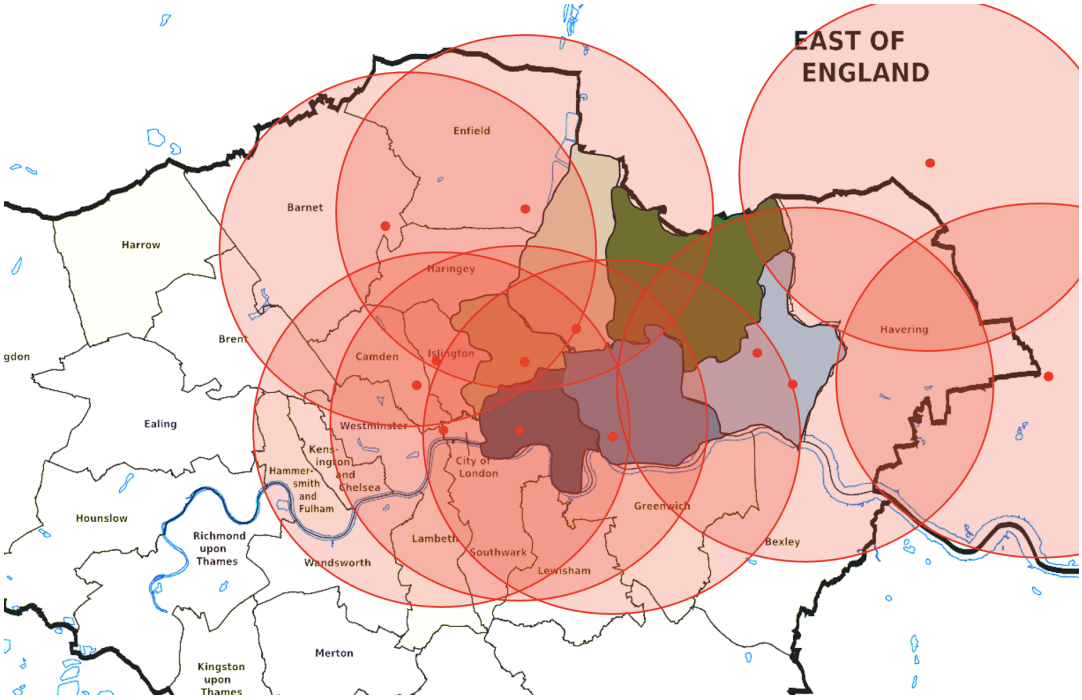
The NJR and HES datasets are organized by hospital, as indicated by the red dots in Figure 2a. Within the NEL ICB, there are six acute hospitals with orthopaedic departments. While it is reasonable to assume that the majority of patients within the North East London area will attend these hospitals, additional hospitals located outside the North East London area but within a short travel time have also been included. This inclusion is based on two primary considerations. First, the geographical coverage areas of GPs and hospitals do not precisely align. A patient may have a GP with an address local to North East London but be referred to a hospital outside of this area. Second, in accordance with the UK government's 'choose and book' policy, patients with a GP address local to North East London may opt to receive treatment at a hospital outside of this area.
The approximated hospital coverage areas are depicted in red in Figure 2b. The hospitals included are Gateway Surgical Centre, King George Hospital, North Middlesex Hospital, Royal Free Hospital, Whipps Cross University Hospital, Homerton Hospital, Royal London Hospital, Whittington Hospital, University College London Hospital, Barnet Hospital, Chase Farm Hospital, Queens Hospital, Basildon Hospital, Broomfield Hospital, and Braintree Hospital. This approach is likely to result in larger datasets being transferred for the NJR and HES, however these are likely to include a high portion of unmatched patients who will not be included in the analysis.

PROMS Data

Please note - due to administrative difficulties outside of our control, no attempts were made to reduce the size of the PROMS dataset. This resulted in the whole national knee PROMS dataset being transferred to the host institution under a previously linked “Token person ID”. The “Token person ID” retains the appropriate level of pseudonymisation, and is still in keeping with our ethical approval. This number can be linked to the HES data set and therefore patients linked across all datasets.2.2 Relevant Data Set Identification



(a) Discovery Data Service Coverage (Blue) with Local Hospitals identified in Red



b) Discovery Data Service Coverage (Blue) with Approximated Hospital Coverage Areas (red)
Figure 2: Diagrams displaying the DDS and the HES/NJR Data Coverage Areas

2.2 Data Identification

Dataset Summary

As mentioned previously all data within the NJR, HES and PROMS datasets were transferred, as it was considered relevant to the project by the DAG. However in agreement with the DDS information governance, only 'relevant data' from the DDS was transferred (see below).

DDS data was divided into demographic data, diagnostic data and prescription data.

Demographic Data

All demographic data was transferred.

Diagnostic / Appointment Data

Any snowmed code relevant to MHD were identified and included. Snowmed codes lists were created by the DAG. They were constructed in the majority from previous code lists from either related research, or related audit projects. If no previously used code list was available, a new code list was created by the data analyst within the DAG. This was completed via the use of the snomed CT browser.

Snowmed codes for related co-variates (eg – significant past medical history) were created using the same method. Related co-variates were identified via systematic review. Any medical risk factor, identified by previous published research, for poor outcome was included.

Prescription Data

All prescription data for MHD related medications and pain medications were identified and included. Related drug lists were created by the DAG, from previous published research or related audit projects.

Lists of all relevant snowmed codes and prescription codes are available in the appendix.

2.3 External Link Process

2.3.2 Encryption

For the project link to be performed without requiring the transfer of patient identifiable information to QMUL, patients must be pseudonymised, via a repeatable encryption process, by the original data controllers. This was performed via the use of an open source, trusted encryption website, produced by the University of Nottingham called Open P (*https://www.openpseudonymiser.org/)*.

Open P securely generates a random string of data (or SALT), associated to each NHS number. Each SALT is unique to each NHS number. The study Salt key ID (SKID) algorithm is kept by a member of DDS staff (Kelvin Smith) on behalf of the project and is the trusted third-party key holder for the research project. The key is a one-way hash and cannot be used to recreate the NHS number by QMUL. The key is NOT shared with other data controllers, with the research team or the university.

2.3.2 External Link Method

The below methodology was created by the DAG to allow for an efficient, specific link of patients while avoiding the encryption of excess data and national databases.

- Discovery Data Service creates SKID via Open P software for its NE London data base. (eg – the population of patients registered at NE London GP, approximately 2 million people)
- The NEC Staff run query in their national NJR data base, extracting patients in the identified hospitals over the prescribed time period.
- All identified NJR patients are then run through the Open P software to create SKID IDs.
- SKID ID's (only) are then shared via encrypted email between the DDS CEG and NJR NEC. Patients who's SKIDS are matched between the two SKID lists are then included as the study population.

- Complete NJR / HES / PROMS data transferred to QMUL for study population under SKID IDs. All identifiable details removed
- Relevant data only for study population transferred by DDS to QMUL under SKID IDs. All identifiable details removed.

Total number of patients identified by external linking process = 8539

3. Internal Linking Process

3.1 Received Data Sets

Data Set	Original Data Controller	Description	Number of Cases	Data Type + Size	Date Transferred
NJR Data	NEC Software Solutions	Complete NJR dataset transferred for patients matched during external link process. Details for the updated minimum NJR dataset (MDSv8) can be found online; https://www.njrcentre.org.uk/healthcare-providers/update-of-njr-minimum-data-set-forms/	33,241 TKR's in 28,272 patients.	1x Text Document (32,333KB)	26/6/24
HES Data	NHSE	Complete HES Admitted Patient Care (APC) dataset transferred for patients matched during external link process. Details for the APC dataset can be found online; https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity#past-publications with a published data dictionary available; https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary	374,491 Hospital Episodes in 27,809 patients	1x Text Document (159,721KB)	26/6/24
PROMS Data	NHSE	Complete Knee PROMS data set transferred. The national dataset transferred for reasons described previously. Knee PROMS data set includes the 'K1' and 'K2' datasets (including OKS, EQ5D, EQVAS). Details for the K1 and K2 dataset can again be found online with the updated NJR minimum dataset description; https://www.njrcentre.org.uk/healthcare-providers/njr-data-collection-forms-mdsv8/	1,941,980 responses in 1,445,514 patients.	1x Text document (372,916KB)	26/6/24
DDS Data	NEL ICB	Transferred data included demographics, predictor variable, potential covariates, and related prescriptions transferred (see section 3.2.2). Data transferred for patients matched during external link process only. Data transferred – - Demographic dataset (study ID, age, sex, area code, ethnicity, deprivation decile) as one CSV file.	Patients identified - 8539 Total Codes Transferred - 35,267 MH Snowmed Codes	30 x CSV Files Size range (24KB – 156,535 KB) Average size 8,836 KB	5/11/24

		- Predictor variables (anxiety, chronic pain, depression, chronic fatigue, low mood, serious mental illness) all transferred as individual CSV files. - Covariate variables (asthma, coronary heart disease, chronic kidney disease, chronic obstructive pulmonary disease, diabetes, frailty, osteoporosis, rheumatoid arthritis, cerebral vascular accident, anaemia, alcohol, obesity (BMI), smoking, fit note use) all transferred as individual CSV files. - Prescription data (antidepressants / anxiolytics – including amytriptyline, opioids weak, opioid strong, pregabalin) all transferred as individual CSV files. Each CSV file includes the exact snowmed/prescription code used, it's 'free text' title, the date of code and the person ID it was associated to. Lists of all included snowmed codes for each variable / prescription can be found in the appendix.	23,002 Co-variate Snowmed Codes 901,387 Presciptions snowmed codes.		

3.2 Internal Link Method

To ensure a reliable and repeatable linkage process across the four datasets, the following methodology was developed and conducted;

- **Linking HES to NJR**. The HES dataset has been previously linked and validated to the NJR dataset via the 'Primary NJR Index Number' (ref), which uniquely identifies each knee replacement surgery. The NJR dataset records each total knee replacement (TKR) under this index, while HES includes both the 'Primary NJR Index Number' and a specific person ID ('Token Person ID'). By matching these identifiers, patient-level data can be linked across NJR and all patient hospital admissions in HES.

- **Linking NJR/HES to PROMS**. The PROMS dataset has been previously validated against HES using the same 'Token Person ID'. This identifier allows PROMS data to be integrated with the linked NJR/HES dataset.

- **DDS Internal Linkage**. Multiple DDS CSV files were linked to the DDS demographic file via a specific 'person ID', already included in DDS. This created a unified DDS database (RDS file) where all relevant data is consolidated under each unique person ID.

- **Linking NJR/HES/PROMS to DDS.** Both the NJR dataset and the unified DDS database retained the SKID identifier (see section 2.3.2), enabling linkage between DDS and the combined NJR/HES/PROMS dataset.

This process results in a fully integrated database containing primary/community care data, hospital/secondary care data, and arthroplasty registry data. All linkages are conducted using encrypted IDs, ensuring that no identifiable patient data is transferred to the host institution. This approach provides a robust framework for cross-database linkage while maintaining data security and integrity.

3.3 Linking Assumptions

Please note – the PROMs database can be filled out multiple times by one patient, and does not identify which side the patient is reporting their symptoms to. Due to this a number of assumptions were forced in the linking process. These assumptions were decided prior to the linking process and before any subsequent analysis of the database, to avoid introduction of bias.

- In the incidences that there are multiple PROMS scores pre/post op for one "NJR Index Number". The lowest scores were used as the patients' response for analysis.
- In the incidences where patients had bilateral knee replacements (at different times), any K1 score prior to the first operation date was assumed to be about the first operated knee. Any K1 score following the first operation date was assumed to be about the contra-lateral (or second) knee placement. Equally any K2 score before the second operation date was assumed to be about the first operated knee. Any K2 score after the second operation date was assumed to be about the second operated knee.
- In the rare incidences that both knees are replaced in the same operation they are given one 'Primary NJR index Number' and two 'Procedure ID's'. In these cases they are considered as only 1 case. (Number identified = 52, 0.052% of cases)

4. Dataset Formation

4.1 Final Dataset Description

Data Set	Data Controller	Description and Successful Link Rate	Number of Cases	Data Type + Size	Date Completed
"mh_in_tkr_data" dataset	Queen Mary University London	Complete linked dataset transferred for patients matched during external link process. **Summary** Externally linked patients – Total 8539 patients Internally linked patients - Total 8539 patients (100%) **Individual Database Rate** Successfully linked patients (NJR to HES) – 7129/8539 (83.5%) Successfully linked patients (NJR to PROMS) - K1 EQ5D form = 4647/10,005 (46.4%) - K2 EQ5D form = 3074 /10,005 (30.7%) -K1 OKS form = 4814/10,005 (48.1%) -K2 OKS form = 3173/10,005 (31.7%) Successfully linked patients (NJR to DDS) - 8539/8539 (100%) Duplicates identified – 44 (removed at final formatting process) For a complete analysis of missing data please see chapter 6 of project report.	10,005 TKR's in 8,539 patients.	1x RDS File (KB)	26/6/24

4.2 Cleaning

A secondary review of the transferred diagnostic / prescription snowmed codes was performed by the PI. A secondary removal of codes deemed to be irrelevant to the project was conducted. A list of all excluded codes can be found in the appendix.

Subsequent to this all data cleaning and formatting was conducted using R Studio version 4.2. Subsequent versions of the linked database were saved in R Data Format (RDS file).

All R scripts used for cleaning and formatting are publicly available for download on the PI's git hub page (ref). Please note that due to administrative constraints of the project, data cleaning procedures were performed both before and after the internal linking process.

4.2 Dictionary

Below is the data dictionary for the linked dataset, all assumptions for each covariate is described.

Column Title	Description	Type (R code)	Values	Data assumptions
"njr_index"	NJR Index Number – used as study ID.	Continuous (Integer)	7 digit ID Number	Nil
"sex"	Sex of participant - as identified by their GP data	Categorical (factor)	male female other (NA)	Nil
"nhs_5"	Ethnicity of the patient – Self declared, by the previously defined 5 major groups	Categorical (factor)	white mixed black South Asian Other	Nil
"ethnicity_18"	Ethnicity of the patient – Self declared, by the previously defined 18 major group	Categorical (factor)	British Bangladeshi Unclassified Indian Caribbean Pakistani Any other Black Any Other Asian Any other White Irish African White and Black Caribbean Chinese Not Stated Any other ethnic group White and Black African Any other Mixed White and Asian Gypsy or Irish Traveller Arab	Nil
"lsoa_2011"	Lower Super Output Area (2011) – as defined by ONS 2011 survery. An area code defining a population of approximately 1000 – 3000 patients	Categorical (factor)	Please see HES data dictionary / or ONS website for detailed locations.	Nil
"imd_2019"	Index of Multiple Deprivation – Score, as defined by the ONS, related to the patients "lsoa_2011".	Continuous (numerical)	Range 2.3 – 56.2	If multiple values score, highest score used
"nel_imd_quintile"	NE London IMD quintile – Ranking of IMD_2019 scores within NE London, divided into 5 groups.	Categorical, Ordinal Ordered (factor)	1 2 3 4 5	If multiple values score, lowest score used
"age_at_primary"	Age of participant at time of TKR	Continuous (numerical)	Range 0 – inf	Nil
"bmi"	Body Mass Index – As scored by GP or by NJR	Continuous (numerical)	Range 0 - 100	BMI code must be before "primary_op_date". If multiple values score, closet to op date used.
"asthma"	Asthma – Active diagnosis in the patients GP past medical history	Categorical, Dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"chd"	Coronary Heart Disease – Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"copd"	Chronic Obstructive Pulmonary Disease– Active diagnosis in the patients past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date"

				If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"dm"	Diabetes Mellitus – Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"op"	Osteoporosis – Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"oa"	Osteoarthritis – Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"ra"	Rheumatoid Arthritis – Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"cva"	Cerebral Vascular Accident – Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"ckd"	Chronic Kidney Disease – Active diagnosis in the patients GP past medical history	Categorical (factor)	None Stage 1 Stage 2 Stage 3 Stage 4 Stage 5	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – included as original group.
"frailty"	Frailty - Active diagnosis in the patients GP past medical history	Categorical (factor)	None Mild Moderate Severe	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "None"
"asa"	American Association of Anaesthics Grade – Functional score given to patient pre – op. Recorded in NJR data set.	Categorical (factor)	P1 - Fit and healthy P2 - Mild disease not incapa citating P3 - Incapacitating systemic disease P4 - Life threatening diseas e	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date - included as original group
"cpain"	Chronic Pain - Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"

"smoke"	Smoking - Active diagnosis in the patients GP past medical history	Categorical (factor)	non smoker smoker ex smoker	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – included as original group
"alcohol"	Alcohol - Active diagnosis in the patients GP past medical history	Categorical, dichotomous (factor)	Yes No	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "Yes"
"fnote_12_pre"	Fit note pre op – Has the patient required a fit note in the 12 months prior to the "primary_op_date". As documented in DDS.	Categorical, dichotomous (factor)	Yes No	Code date must be within 12 months preceeding "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "No"
"diag_group"	Mental Health Diagnosis – Divided into groups guided by DSM classification	Categorical (factor)	Control Group Depression Anxiety Depression and Anxiety Serious Mental Illness	Code date must preceed "primary_op_date" If no snowmed code present – patient coded as "Control". For more details on SMI grouped diagnoses please see the appendix – "Snowmed code lists". If no time stamp present – patient coded as original group
"severity_group"	Mental Health Severity – Divided into groups guided by NICE treatment pathways	Categorical (factor)	Control Group Mild Symptoms Moderate Symptoms Severe Symptoms	SMI Group excluded (NA) Code date must preceed "primary_op_date" Group definitions - *Control Group – No MHD (as coded in diag_group)* *Mild Symptoms – <2 attendances to GP for MH issue (ie - <2 Snowmed codes)* *Moderate - >2 attendances to GP for MH issue (ie - >2 Snowmed codes)* *Severe – Required further treatment for MH in 12 months preceding "primary_op_date". (treatment includes onward referral or MH prescription).*
"chron_group"	Mental Health Chronicity – Divided into groups guided by time from GP engagement and "primary_op_date".	Categorical (factor)	Control Group Long Term Symptoms Moderate Term Symptoms Short Term Symptoms	SMI Group excluded (NA) Code date must precede "primary_op_date" Group definitions - *Control Group – No MHD (as coded in diag_group)* *Long Term – has NOT attended GP for MH issue within 1 year of "primary_op_date".* *Moderate - has attended GP for MH issue within 1 year of "primary_op_date".* *Severe – has attended GP for MH -issue within 3 months of "primary_op_date".*

"pre_op_opioid"	Opioids Pre Op – Patient prescription for opioids at time of "primary_op_date"	Categorical (factor)	No Opioid Weak Opioid Moderate Opioid Strong Opioid	Code date must precede AND be within 3 months. "primary_op_date" Group Definitions No opioid – no prescription code for opioid Weak Opioid – prescription code for 'codeine / dihydrocodeine' related analgesia. Moderate Opioid – prescription code for an opioid stronger than 'codeine / dihydrocodeine' related analgesia. Strong Opioid – Prescription code for opioid with OME greater than 90 (ie – "dangerous" opioid level). If no time stamp present – snowmed code NOT included in analysis
"npp_3_pre"	Neuropathic Analgesia Pre Op – Patient prescription for neuropathic pain medication at the time of "primary_op_date"	Categorical, dichotomous (factor)	Yes No	Code date must precede AND be within 3 months. "primary_op_date" Any neuropathic analgesia included (variations of pregabalin, amitriptyline and gabapentin)
"no_appointments"	Number of MH appointments – How many times the patient has required an appointment with their GP for MH related issues.	Continuous (Integer)	Range 0 - inf	Code date must preceed "primary_op_date" If no time stamp present – snowmed code NOT included in analysis
"prescrip12"	12 Month Prescription for MH medication – Has the patient required an antidepressant / anxiolytic in the 12 months prior to "primary_op_date".	Categorical, dichotomous (factor)	Yes No	Code date must precede AND be within 12 months. "primary_op_date" If no time stamp present – snowmed code NOT included in analysis
"prescrip6"	6 Month Prescription for MH medication – Has the patient required an antidepressant / anxiolytic in the 6 months prior to "primary_op_date".	Categorical, dichotomous (factor)	Yes No	Code date must precede AND be within 6 months. "primary_op_date" If no time stamp present – snowmed code NOT included in analysis
"prescrip3"	3 Month Prescription for MH medication – Has the patient required an antidepressant / anxiolytic in the 3 months prior to "primary_op_date".	Categorical, dichotomous (factor)	Yes No	Code date must precede AND be within 3 months. "primary_op_date" If no time stamp present – snowmed code NOT included in analysis
"psych12"	Specialist referral – The patient has been referred on for further MH assessment.	Categorical, dichotomous (factor)	Yes No	Code date must precede AND be within 12 months. "primary_op_date" If no time stamp present – snowmed code NOT included in analysis
"chron_days_min"	MH timing - the number of days between GP MH Diagnosis code and "primary_op_date".	Continuous (Integer)	Range 0 – inf	Code date must preceed "primary_op_date" If no time stamp present – snowmed code NOT included in analysis
"ome_pre"	Oral Morphine Equivalent (OME) pre op – The strength of opioids prescription (pre_op_opioid), as guided by OME calculations.	Continuous (numerical)	Range 0 – 1000	Code date must precede AND be within 12 months. "primary_op_date" If multiple prescriptions used per patient, strongest opioid used. Prescription then matched to it's OME dose. For

				dosing table please see appendix. If no time stamp present – snowmed code NOT included in analysis
"primary_op_date"	Primary Op Date – Date of the primary TKR operation. As documented by NJR	Date (date)	Date	Nil
"fixed_flexion_deformity"	Fixed Flexion Deformity - As documented by NJR	Categorical (factor)	Not Available Less than 10 10 to 30 More than 30	Nil
"symp_period"	Length of symptoms pre operation – Patient reported in PROMS data set (K1).	Categorical (factor)	Less than 1 year 1 to 5 years 6 to 10 years More than 10 years	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"prev_surg"	Previous surgery to knee - Patient reported in PROMS data set (K1).	Categorical, dichotomous (factor)	Yes No	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"pre_night_pain"	Pre-operative knee pain at night - Patient reported in PROMS data set (K1).	Categorical (factor)	Every Night Most Nights Some Nights 1 or 2 Nights No Nights	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"pre_pain"	Pre-operative knee pain - Patient reported in PROMS data set (K1).	Categorical (factor)	Severe Moderate Mild Very Mild None	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"pre_oks"	Pre-operative OKS - Patient reported in PROMS data set (OKS).	Continuous (Integer)	Range 0 - 40	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"pre_mhs"	Pre-operative Mental Health Symptoms- Patient reported in PROMS data set (EQ5D).	Categorical (factor)	No symptoms Moderate Symptoms Severe Symptoms	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"pre_gen_pain"	Pre-operative general pain - Patient reported in PROMS data set (EQ5D)	Categorical (factor)	No symptoms Moderate Symptoms Severe Symptoms	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"pre_eqvas"	Pre-operative EQVAS - Patient reported in PROMS data set (EQVAS)	Continuous (numerical)	Range 0 - 100	In the case of multiple responses – K1 form with lowest OKS or EQ5D score included
"q1_eq5d_index"	Pre-operative EQ5D - Patient reported in PROMS data set (EQ5D)	Continuous (numerical)	Range 0 - 1	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"outcome"	Outcome – NJR documented outcome	Categorical (factor)	Unrevised Revised Death	Nil
"post_op_stay"	Length of Post Op stay – As documented by the HES dataset.	Continuous (Integer)	Range 0 - Inf	Nil
"post_bleeding"	Post op bleeding - Patient reported in PROMS data set (K2)	Categorical, dichotomous (factor)	Yes No	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_reop"	Required a re-operation post op – Patient reported in PROMS data set (K2)	Categorical, dichotomous (factor)	Yes No	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_readmission"	Required a re-admission post op – Patient reported in PROMS data set (K2)	Categorical, dichotomous (factor)	Yes No	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_wound"	Had wound issues post op - Patient reported in PROMS data set (K2)	Categorical, dichotomous (factor)	Yes No	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"complication"	Had a complication post op - Patient reported in PROMS data set (K2)	Categorical, dichotomous (factor)	Yes No	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"revision_njr_no"	Required Revision TKR – NJR documentation of revision NJR index no.	Continuous (Integer)	7 digit ID Number	Nil
"revision_date"	Date of Revision TKR – As per NJR dataset	Date (date)	Date	Nil

"revision_for_pain"	Pain as key indication for revision – as documented by NJR	Categorical, dichotomous (factor)	Yes No	Nil
"revision_for_infection"	Infection as key indication for revision – as documented by NJR	Categorical, dichotomous (factor)	Yes No	Nil
"post_night_pain"	Post-operative knee pain at night - Patient reported in PROMS data set (K2).	Categorical (factor)	Every Night Most Nights Some Nights 1 or 2 Nights No Nights	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_pain"	Post - operative knee pain - Patient reported in PROMS data set (K2).	Categorical (factor)	Severe Moderate Mild Very Mild None	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_oks"	Post-operative OKS - Patient reported in PROMS data set (OKS).	Categorical (factor)	Range 0 - 40	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_mhs"	Post-operative Mental Health Symptoms- Patient reported in PROMS data set (EQ5D).	Categorical (factor)	No symptoms Moderate Symptoms Severe Symptoms	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_gen_pain"	Post-operative general pain - Patient reported in PROMS data set (EQ5D)	Categorical (factor)	No symptoms Moderate Symptoms Severe Symptoms	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"post_eqvas"	Post-operative EQVAS - Patient reported in PROMS data set (EQVAS)	Continuous (numerical)	Range 0 - 100	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"q2_eq5d_index"	Post-operative EQ5D - Patient reported in PROMS data set (EQ5D)	Continuous (numerical)	Range -1 - 1	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"surg_success"	Surgical Success - Patient reported in PROMS data set (K2).	Categorical, dichotomous (factor)	Yes No	In the case of multiple responses – K2 form with lowest OKS or EQ5D score included
"readmis_1"	Patient re-admission within 1 week of discharge from TKR episode– As documented in the HES dataset	Categorical, dichotomous (factor)	Yes No	All re-admissions/attendances included (including under different specialties)
"readmis_6"	Patient re-admission within 6 weeks of discharge from TKR episode– As documented in the HES dataset	Categorical, dichotomous (factor)	Yes No	All re-admissions/attendances included (including under different specialties)
"readmis_12"	Patient re-admission within 12 weeks of discharge from TKR episode– As documented in the HES dataset	Categorical, dichotomous (factor)	Yes No	All re-admissions/attendances included (including under different specialties)
"readmi_team"	Primary re-admission team - As documented in the HES dataset	Categorical (factor)	Please see HES data dictionary fpr team codes.	Nil
"readmi_diag"	Primary re-admission diagnosis - As documented in the HES dataset	Categorical (factor)	Please see HES data dictionary for diagnosis codes.	Nil
"fnote_12_post"	Fit note – Has the patient required a fit note in the 12 months post "primary_op_date". As documented in DDS.	Categorical, dichotomous (factor)	Yes No	Code date must be within 12 months following "primary_op_date" If no snowmed code present – patient coded as "No" If no snowmed code date – coded as "No"
"post_op_opioid"	Opioids Post Op – Patient required prescription for opioids following "primary_op_date"	Categorical (factor)	No Opioid Weak Opioid Moderate Opioid Strong Opioid	Code date must follow AND be within 6 - 18 month time window of "primary_op_date" . (This is to EXCLUDE standard post operative pain prescriptions) Group Definitions No opioid – no prescription code for opioid Weak Opioid – prescription code for 'codeine /

				dihydrocodeine' related analgesia. Moderate Opioid – prescription code for an opioid stronger than 'codeine / dihydrocodeine' related analgesia. Strong Opioid – Prescription code for opioid with OME greater than 90 (ie – "dangerous" opioid level). If no time stamp present – snowmed code NOT included in analysis
"npp_6_to_18_post"	Neuropathic Analgesia Pre Op – Patient required prescription for neuropathic pain medication following "primary_op_date"	Categorical, dichotomous (factor)	Yes No	Code date must follow AND be within 6 - 18 month time window of "primary_op_date" . (This is to EXCLUDE standard post operative pain prescriptions) Any neuropathic analgesia included (variations of pregabalin, amitriptyline and gabapentin)
"ome_post"	Oral Morphine Equivalent (OME) post op – The strength of opioids prescription (post_op_opioid), as guided by OME calculations.	Continuous (numerical)	Range 0 – 1000	Code date must precede AND be within 12 months. "primary_op_date" If multiple prescriptions used per patient, strongest opioid used. Prescription then matched to it's OME dose. For dosing table please see appendix. If no time stamp present – snowmed code NOT included in analysis

4.4 Validation

For validation of the linked dataset please see the final project report.

4.5 Missing Data

For missing data description and analysis of the linked dataset please see the final project report.

5. Data Management

5.1 Consent for Data Use

Patients who have declared a "Type 1 opt-out" have been excluded. Inclusion of all other patients, without direct consent, aligns with established legal frameworks. Following legal review by King's Counsel, de-identified data usage for research purposes under secure ICO-approved protocols is permitted. See ethical considerations for more details.

5.2 Data Storage and Security

Data are securely encrypted and transferred using open-source servers (Open P). All de-identified records are stored within QMUL's ISO 27002-approved Data Safe Haven, ensuring compliance with published legal security standards.

5.3 Metadata Standards and Documentation

Internal validation, script checking and final analysis will be conducted by the Bone and Joint Health "Data Group".

5.4 Risks to Data Security

Theoretical risks of re-identification exist in pseudonymized datasets, particularly if combined with prior patient knowledge. However, ethical review boards deem this risk negligible given strict access controls and de-identification protocols (see ethics section for more details).

5.5 Data Sharing and Governance

5.5.1 Data Sharing

Raw data access is restricted to the immediate research team. Metadata and summarized findings will be published on the Bone and Joint Health Department webpage and advertised via the HDRUK Gateway.

5.5.2 Governance

Data governance responsibilities rest with the Queen Mary University of London (QMUL) and the Bone and Joint Health Department, ensuring compliance with institutional, legal, and ethical frameworks.