Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE)
Work Package 2a (RO-HES)

Version	2.2
Date	18/03/2021
Sponsor	University of Leeds
Research Ethics (IRAS) Reference	220520
Sponsor Reference	220520
Funder Reference:	14/70/146

Principal Investigator	
Name	Professor Philip Gerard Conaghan
Department	Leeds Institute of Rheumatic and Musculoskeletal Medicine
Specialty	Rheumatology
Job Position	Professor of Musculoskeletal Medicine
Email	p.conaghan@leeds.ac.uk
Phone	0113 3924884

Administrative Overview

This document describes Work Package 2a (RO-HES) of the NIHR (HS&DR) funded research project: *Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE)*, which is sponsored by the University of Leeds.

It is intended to provide sufficient detail to enable: i) an understanding of the administration, background, rationale, objectives, population, methods, analyses, ethical considerations, and dissemination plans of the study, ii) replication of key aspects of study methods and conduct, and iii) appraisal of the study's scientific and ethical rigour from the time of ethics approval through to dissemination of the results.

Whilst every care has been taken in drafting this protocol, corrections or amendments may be necessary.

Structured Project Summary

Public Title	Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE) – Work Package 2a (RO-HES)
Scientific Title	Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE) – Work Package 2a (RO-HES)
Primary Registry and Project Identifying Number	HS&DR 14/70/146
Source of Monetary or Material Support	National Institute for Health Research (NIHR) – Health Services and Delivery (HS&DR) Programme (Funder Reference: 14/70/146)
Sponsor	University of Leeds
Contact for Public Queries	Dr Sarah Kingsbury
Contact for Scientific Queries	Professor Philip Gerard Conaghan
Countries of Recruitment	United Kingdom (England)
Areas of Enquiry	Arthroplasty in Hospitals in England
Overview of UK-SAFE **(* subject of this protocol)**	This study comprises of 4 work streams: 1) Systematic Review 2) Determining when, which and how patients present for revision surgery a. Routine Data i. CPRD-HES **ii. RO-HES*** iii. NJR-HES-PROMS b. Prospective Cohort 3) Health Economics 4) Consensus Process
Study Type	WP2a (RO-HES) is a retrospective cohort study using routinely collected (observational) NHS data from:

	• **NHS Digital:** Data will obtained from the Hospital Episode Statistics (HES) database in relation to A&E, Inpatient and Outpatient episodes in hospitals in England. • **The Phoenix Partnership:** Data will be obtained from the ResearchOne database in relation to General Practice for patients identified by NHS Digital. Data from these sources will be linked to construct longitudinal records for analysis.
Target Sample Size	All hospitals in England
Primary Outcome(s)	• Time to revision.
Key Secondary Outcomes	n/a

Roles and Responsibilities

Sponsor

UK SAFE is sponsored by the University of Leeds. University of Leeds has overall responsibility for the design and management of the study.

Funder

UK SAFE is funded by National Institute for Health Research (NIHR) – Health Services and Delivery Research (HS&DR) Programme (Funder Reference: 14/70/146). NIHR (HS&DR) has responsibility for the project design and funding.

Project Management

The Project will be conducted in compliance with the approved protocol, the Declaration of Helsinki (2008), the principles of Good Clinical Practice (GCP), the UK Data Protection Act, and the National Health Service (NHS) Research Governance Framework for Health and Social Care (RGF).

Any possible breaches of compliance will be reported immediately to the project co-coordinator by phone or email. The Principal Investigator (PI) will assess whether or not the breach is 'serious'. A 'serious breach' is one that is likely to affect to a significant degree:

- The safety or physical or mental integrity of the subjects in the Study, or
- The scientific value of the Project.

The PI will have overall responsibility for the conduct of the study in accordance with the Research Governance Framework and will oversee the financial management of the study. He will receive support in this capacity from the finance manager at the Leeds Institute of Musculoskeletal and Rheumatic Medicine (LIRMM).

Each work package will be managed by a designed lead supported by a small team of co-applicants and appointed staff. Work package leads are as follows:

Work Package	Lead Name	Affiliation
1	Ms Carolyn Czoski-Murray	University of Leeds
2a (CPRD-HES)	Dr Andrew Judge	University of Oxford
2a (RO-HES)	Dr Chris Smith	University of Leeds
2a (NJR-HES-PROMS)	Dr Andrew Judge	University of Oxford
2b	Mr Martin Stone	Leeds Teaching Hospitals Trust
3	Dr Rafael Pinedo Villanueva	University of Oxford
4	Professor Philip Conaghan	University of Leeds

Additionally, a Senior Research Fellow, reporting to PI, will have responsibility for project co-ordination, liaison between co-applicants and monitoring of project milestones and deliverables.

Study Management Group

Study supervision will be established in line with the MRC GCP guidelines and include a core team who will form a Study Management Group (SMG). The SMG will be assigned responsibility for the setup, on-going management, and promotion of the study and the interpretation of the results. The SMG will meet monthly during set-up and at least quarterly throughout the study. Membership of the SMG is as follows:

Name	Affiliation	Role
Professor Philip Conaghan	University of Leeds	Principal Investigator
Mr Martin Stone	Leeds Teaching Hospitals Trust	Co-Applicant
Professor Robert West	University of Leeds	Co-Applicant
Dr Chris Smith	University of Leeds	Co-Applicant
Dr Sarah Kingsbury	University of Leeds	Co-Applicant
Ms Carolyn Czoski-Murray	University of Leeds	Co-Applicant
Ms Judy Wright	University of Leeds	Co-Applicant
Professor Nigel Arden	University of Oxford	Co-Applicant
Associate Professor Andrew Judge	University of Oxford	Co-Applicant
Dr Rafael Pinedo Villanueva	University of Oxford	Co-Applicant
Dr Lindsay Smith	University of the West of England	Co-Applicant
Mrs Christine Thomas	NIHR Leeds Musculoskeletal Biomedical Research Unit	Co-Applicant
Dr Jamie O'Shea	University of Leeds	Co-Applicant

Independent Advisory Group

An Independent Advisory Group will be established in the set-up period to oversee, provide advice on and monitor study progress. The IAG will play a significant role in supporting the Study Management Group to deliver the study and provide strategic advice. The IAG will be consulted several times per year to ask their advice and update them on the progress of each work package. Membership of the IAG is as follows:

Name	Position	Affiliation	Role
Professor Alistair Hart	Professor of Academic Clinical Orthopaedics	University College London	Chair
Professor Peter Kay	Professor of Orthopaedic Surgery, National Clinical Director of Musculoskeletal Services	University of Manchester	Member
Mr Martyn Porter	Orthopaedic Surgeon, Medical Director	Wrightington Hospital, NJR	Member
Professor Matt Stephenson	Professor of Health Technology Assessment	University of Sheffield	Member
Professor Andrew Prevost	Professor of Medical Statistics and Clinical Trials	Imperial College London	Member
Dr Steve Laville	Head of Planned Care	Leeds West CCG	Member
Professor Yvonne Birks	Co-Director	Social Policy Research Unit, University of York	Member
Michael Nicholson	PPI Representative	n/a	Member
Valerie Thurlow	PPI Representative	n/a	Member
Barbara Hartley	PPI Representative	n/a	Member

Change Log

Version 1 (June 2017)

Initial version of the Ethics Protocol which was included in the submission to the Health Research Authority for review by an NHS Research Ethics Committee. Favourable opinion was received from an NHS Research Ethics Committee (Leeds East) on 8th August 2017, subject to Management Permission, which was received on 9th September 2017.

Version 2 (January 2021)

Incorporates changes to Version 1 which will be included in the submission to the Health Research Authority for amendment to ethical approval. A summary of the changes is provided in the table below:

Description	Author
Amendment of the data items specified in the 'Data Items' section and Appendix 1 of the Ethics Protocol due to requirements of the production team at NHS Digital. We have been advised by NHS Digital that the following data items are required to be included in the data supplied by NHS Digital: i) AEKEY (Hospital Episode Statistics - Accident and Emergency), ii) EPIKEY (Hospital Episode Statistics - Admitted Patient Care) and iii) ATTENDKEY (Hospital Episode Statistics - Outpatient). Additionally, we have been advised by NHS Digital that the following data items are required to be removed from the data requested: i) Hospital Episode Statistics - Accident and Emergency: DIAG2_NN, DIAG3_NN, INVEST2_NN, PSEUDO_HESID, TREAT2_NN and TREAT3_NN, ii) Hospital Episode Statistics - Admitted Patient Care: ADM_CFL, CHAPTER, DIAG_3_CONCAT, DIAG_3_NN, DIAG_4_CONCAT, DIAG_4_NN, DIS_CFL, DOB_CFL, ELEC_CFL, EPIE_CFL, EPIS_CFL, HESID_ORIG, CAUSE_3, CAUSE_4, OPERTN_3_01, OPERTN_3_CONCAT, OPERTN_3_NN and OPERTN_4_CONCAT, iii) Hospital Episode Statistics - Outpatient: DOB_CFL, HESID_ORIG, CHAPTER, WAITWEEKS, DIAG_3_01, DIAG_04_01 and OPERTN_3_01. Finally, the HES-specific identifier used to reference patients in the data supplied by NHS Digital is now referenced as 'ENCRYPTED_HESID'. We understand this field to supersede the following fields that were available on the DARS online service at the point of application: PSEUDO_HESID (Hospital Episode Statistics - Accident and Emergency), HESID_ORIG (Hospital Episode Statistics - Inpatient), and HESID_ORIG (Hospital Episode Statistics - Outpatient). The data items specified in the Ethics Protocol (Main Section and Appendix 1) have been updated to reflect these changes. We have also added an explicit reference to 'Patient Identifier' in the specification of data items to be included from ResearchOne that is included in the 'Data Items' section to be consistent with Appendix 1 (Data Specification). Please note, the *KEY fields detailed above are referenced as 'Record Identifier'. The Data Sharing Agreement between the University of Leeds and NHS Digital reflects these changes in the data items to be supplied by NHS Digital. We expect the planned analysis to remain feasible on the basis of the data items supplied by NHS Digital.	Chris Smith
Amendment of a field used in the inclusion/exclusion criteria for data items from Hospital Episode Statistics – Accident and Emergency due to requirements of the production team at NHS Digital. We have been advised by the production team at	Chris Smith

NHS Digital that the 'A&E Diagnosis (2 character)' field is required to be removed from the data items requested from Hospital Episode Statistics – Accident and Emergency. This field was used to define inclusion/exclusion criteria for data items from Hospital Episode Statistics – Accident and Emergency, specifically, episodes are included only if they have a value of '05' (Dislocation/fracture/joint injury/amputation) within the 'A&E Diagnosis (2 character)' field. Additional criteria is also defined over the 'A&E Diagnosis – Anatomical Area' field. As we have been advised by NHS Digital to remove the 'A&E Diagnosis (2 character)' field, we have requested that NHS Digital confirm that the criteria relating to A&E Diagnosis will be evaluated as part of the inclusion/exclusion criteria for episodes from Hospital Episode Statistics – Accident and Emergency and to confirm the specific field that will be used to evaluate this criteria in the data to be supplied. We expect that the first two characters of the 'A&E Diagnosis' field (which contains the A&E diagnosis code) will be used, but we are currently awaiting confirmation from the production team at NHS Digital.	
Amendment of the Ethics Protocol document to confirm the status of Data Processing Agreements with NHS Digital and The Phoenix Partnership. We stated in Section A6-2 of the IRAS form that University of Leeds will establish Data Processing Agreements with NHS Digital (if applicable) and ResearchOne and within the Ethics Protocol document that 'Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security'. Following correspondence with NHS Digital in relation to the NIHR-funded LP-MAESTRO project (IRAS ID: 178391), at the University of Leeds (which uses the same linkage methodology and provides a precedent for UK-SAFE WP2a (RO-HES), it was determined that a Data Processing Agreement between the University of Leeds and NHS Digital was not required (August 2017). University of Leeds confirmed that LP-MAESTRO could proceed on this basis. UK-SAFE WP2a (RO-HES) also proceeded on this basis using LP-MAESTRO as the precedent. A Data Processing Agreement between the University of Leeds and The Phoenix Partnership has been established for this work (December 2018).	Chris Smith
Amendment of the Ethics Protocol document to clarify that the ResearchOne data request form made available by The Phoenix Partnership, as Data Controller for ResearchOne, is used to specify the project-specific terms, including terms relating to information security, under which the data is made available from ResearchOne. Project-specific terms are reviewed and must be approved by the ResearchOne Project Committee prior to the release of data from ResearchOne. The Ethics Protocol document previously referenced that a Data Sharing Agreement would be established with ResearchOne. However, The Phoenix Partnership do not explicitly reference a Data Sharing Agreement. Therefore, references within the documentation have been amended to clarify that the project-specific terms under which data is made available from ResearchOne are specified in the data request form that is approved by the ResearchOne Project Committee, rather than referencing a Data Sharing Agreement between University of Leeds and ResearchOne.	Chris Smith
Amendment of the Ethics Protocol document to provide an update with respect to the requirement for Section 251 support from the Confidentiality Advisory Group (CAG). Following submission of an application to the Confidentiality Advisory Group at the Health Research Authority (22nd June 2017), and following subsequent correspondence with the Health Research Authority, NHS Digital and The Phoenix Partnership, it was determined that Section 251 support was not required for the study and the application to the Confidentiality Advisory Group with	Chris Smith

withdrawn (4th October 2017). A footnote has been added to the Approvals section of the Ethics Protocol document to document this outcome.	
Amendment of the Ethics Protocol document to clarify references to the databases (Hospital Episode Statistics and ResearchOne) and to the organisations which are the data controllers for these databases (NHS Digital and The Phoenix Partnership). References have been amended throughout all sections of the Ethics Protocol document, including within diagrams, to provide this clarification.	Chris Smith
Amendment of Appendix 4 (Data Processing) of the Ethics Protocol document to reflect changes that were required by NHS Digital to demonstrate compliance with their standards for Privacy Notices. Content and structure provided in the Ethics Protocol document was amended to comply with these standards and the information was published on the project Web site (https://leedsbrc.nihr.ac.uk/uk-safe-project/). Ethics Protocol document has been amended to reflect these changes. In addition, the Appendix 4 (Data Processing) of the Ethics Protocol document has been amended to clarify references to these databases and organisations (see Change 8) and to correct an error with the name of NHS Research Ethics Committee from which a Favourable Opinion was obtained. 'North of Scotland' is named rather than 'Leeds East'. For information, 'North of Scotland' is the NHS REC for which a Favourable Opinion was provided for another project (LP-MAESTRO) at the University of Leeds which uses the same linkage methodology. Information published on the project Web site will be updated accordingly.	Chris Smith
Amendment of the Ethics Protocol document to reflect the new version and date on the cover page; to correct textual errors, including: i) reference to NHS Digital rather than The Phoenix Partnership in 'Data specification will be formally specified in the application for data from NHS Digital' of the Privacy Impact Assessment and, ii) reference to UoL rather than The Phoenix Partnership on Page 78 the Data Linkage Methodology, iii) incorrect footnote reference to NJR report in Consultation Details section of Privacy Impact Assessment; to update references to specific people/documents/schemes, including: i) update of Contact for Public Queries in 'Administrative Overview', ii) updates of references to NHS Information Security Toolkit to NHS Data Security and Protection Toolkit, ii) update of references to data protection legislation from Data Protection Act 1998 to Data Protection Act 2018 and General Data Protection Regulation, and iii) updates of references to training course provided by the Medical Research Council from 'Research Data and Confidentiality' to 'Research, GDPR and Confidentiality'; and to provide a summary of changes made to the protocol from Version 1 to Version 2. References have been amended throughout all sections of the Ethics Protocol document to reflect these updates.	Chris Smith

Please note:

- Changes described above relate to the contents of the Ethics Protocol document. Additional changes, including a change to the research team and an extension of the study end date, are described in the Amendment Form to be submitted to the Health Research Authority and not within this document.

- A previous amendment of ethical approval was submitted to the Health Research Authority on 28th April 2020 and confirmation received on 30th April 2020 which included a previous extension of the study end dates to 30th August 2020.

Version 2.1 (March 2021)

Incorporates changes to Version 2 which will be included in the submission to the Health Research Authority for amendment to ethical approval. A summary of the changes is provided in the table below:

Description	Author
Further to the following amendment in Version 2: Amendment of a field used in the inclusion/exclusion criteria for data items from Hospital Episode Statistics – Accident and Emergency due to requirements of the production team at NHS Digital. We have been advised by the production team at NHS Digital that the 'A&E Diagnosis (2 character)' field is required to be removed from the data items requested from Hospital Episode Statistics – Accident and Emergency. This field was used to define inclusion/exclusion criteria for data items from Hospital Episode Statistics – Accident and Emergency, specifically, episodes are included only if they have a value of '05' (Dislocation/fracture/joint injury/amputation) within the 'A&E Diagnosis (2 character)' field. Additional criteria is also defined over the 'A&E Diagnosis – Anatomical Area' field. As we have been advised by NHS Digital to remove the 'A&E Diagnosis (2 character)' field, we have requested that NHS Digital confirm that the criteria relating to A&E Diagnosis will be evaluated as part of the inclusion/exclusion criteria for episodes from Hospital Episode Statistics – Accident and Emergency and to confirm the specific field that will be used to evaluate this criteria in the data to be supplied. We expect that the first two characters of the 'A&E Diagnosis' field (which contains the A&E diagnosis code) will be used, but we are currently awaiting confirmation from the production team at NHS Digital. We have now received a response from the production team at NHS Digital. Accordingly, we update the documentation of the amendment to the following: Amendment of a field used in the inclusion/exclusion criteria for data items from Hospital Episode Statistics – Accident and Emergency due to requirements of the production team at NHS Digital. We have been advised by the production team at NHS Digital that the 'A&E Diagnosis (2 character)' field is required to be removed from the data items requested from Hospital Episode Statistics – Accident and Emergency. This field was used to define inclusion/exclusion criteria for data items from Hospital Episode Statistics – Accident and Emergency, specifically, episodes are included only if they have a value of '05' (Dislocation/fracture/joint injury/amputation) within the 'A&E Diagnosis (2 character)' field. Additional criteria is also defined over the 'A&E Diagnosis – Anatomical Area' field. As we were advised by NHS Digital to remove the 'A&E Diagnosis (2 character)' field, we requested that NHS Digital confirm that the criteria relating to A&E Diagnosis will be evaluated as part of the inclusion/exclusion criteria for episodes from Hospital Episode Statistics – Accident and Emergency and to confirm the specific field that will be used to evaluate this criteria in the data to be supplied. The production team at NHS Digital have now confirmed that the first two characters of the 'A&E Diagnosis' field (which contains the A&E diagnosis code) will be used and that the 'A&E Diagnosis' field is "a full length version" of the 'A&E Diagnosis (2 character)' field.	Chris Smith

Version 2.2 (March 2021)

Incorporates changes to Version 2.1 which will be included in the submission to the Health Research Authority for amendment to ethical approval. A summary of the changes is provided in the table below:

Description	Author
Updated contact for public queries in the Structured Project Summary (Page 2), updated the version and date on the title page (Page 1) and corrected date for Version 2.1 in the Change Log (Page 10) from 'February 2021' to 'March 2021'.	Chris Smith

Introduction

Plain English Summary

Most people now understand the need for a cost-effective NHS, but seek reassurance that this will not reduce the standard of care. This is particularly true of older people, who are the group most likely to be affected by this research. This research seeks to demonstrate that good after-care is not necessarily expensive, in terms of time and money on the part of both patient and hospital staff, and that individual patient-centred follow-up can better identify potential problems in a timely fashion, to the benefit of all concerned.

Total joint replacement provides considerable improvement in quality of life to people suffering with severe joint damage and in 2013 over 150,000 total hip and knee replacements were conducted. Due to increasing ageing and obesity in the UK population, this number is likely to increase each year. Sometimes, problems can develop with the replaced joint over time and a small percentage of people require further surgery. Because joint failure is not always associated with symptoms, follow-up care is provided to ensure that problems are identified as early as possible.

Scientific Abstract

In the current economic climate there is increasing pressure to identify cost-saving measures across the NHS. Our recent work suggests that many centres are curtailing primary hip and knee arthroplasty follow-up services to deal with the growing pressure on their services. However, such disinvestment is without evidence-base and raises questions of the consequences to patients. This project will examine the requirements for arthroplasty follow-up and produce evidence and consensus-based recommendations as to how, when and on whom follow-up should be conducted.

This project will make use of routine data from 5 national datasets (National Joint Registry, Hospital Episode Statistics, Patient Reported Outcome Measures, Clinical Practice Research Database and ResearchOne) together with prospective data collected on patients presenting for revision surgery and evidence gathered through a comprehensive literature search. 1) NJR-HES-PROMS data will be used to model time to revision to determine when follow-up should occur. A smoothed Nelson-Aalen cumulative hazard rate will be used to identify any peak in the mid-long term revision risk. 2) NJR-PROMS-HES linked data, together with CPRD-HES and RO-HES to understand primary care involvement, will be used to examine which patients are most likely to require intervention.

We will use cox proportional hazards regression modelling to identify pre, peri- and postoperative predictors of mid-late term revision (5 or more years post primary surgery). 3) Prospective data collected on 675 patients presenting for elective or emergency revision across 25 NHS centres will determine how patients currently present for revision surgery. 4) Markov modelling will be used to simulate long-term costs and quality adjusted life years associated with different follow-up care models to determine how follow-up should occur. 5) Together with a systematic literature review, this work will inform a Delphi-consensus process, involving all stakeholders, to develop a policy document which includes a stratification algorithm to determine appropriate follow-up for an individual patient.

We have assembled a highly experienced multi-disciplinary team from the Leeds and Oxford NIHR Musculoskeletal Biomedical Research Units with the necessary academic expertise to tackle the methodological challenges in undertaking this work to ensure the project yields meaningful results that are accepted by key stakeholders. We have support for the conduct and dissemination of this

project from all relevant stakeholders, including major national and international orthopaedic bodies, clinical commissioning groups, general practitioners and patients. This research, by making use of large routinely collected, observational datasets and prospective data collection will establish what constitutes the most effective and cost-effective arthroplasty follow-up care pathways, thereby ending the current confusion over how, when and on whom follow-up should be carried out. Novel, evidence-based care pathways for hip and knee arthroplasty follow-up will have major immediate effect on national NHS planning and budgeting, potentially saving millions of pounds per year, and on patient well-being. The national and global impact will be to reduce the burden on patients and health services in terms of outpatient visits and clinical tests that do not add benefit, while optimising detection of potential problems.

Background and Rationale

What is the problem being addressed?

Hip and knee replacement surgery is one of the great success stories of Medicine in the 20th century, revolutionising the management of degenerative joint disease. In 2013-14, 158,541 primary hip and knee replacements and 15,534 revision procedures were carried out in the UK [19] an increase of 25% in only 4 years and 300% over the past 20 years. Orthopaedics services are already one of the worst performers across the NHS in terms of failure to meet waiting list targets, with an estimated 8,000 orthopaedic NHS breaches each month [20]. With a rapidly aging, increasingly obese population and medical advances that mean less stringent criteria for surgery eligibility [21], there is no sign that demand will recede in coming years and orthopaedic services will soon be stretched to breaking point.

The burden on orthopaedic services does not stop post-operatively; follow-up requirements for hip and knee arthroplasty are conservatively estimated at 500,000-1,000,000 outpatient attendances annually. With pressure to meet waiting list targets and maintain budgets in a difficult financial climate, there is significant pressure on orthopaedic centres to reduce the amount of follow-up provided. Our work has demonstrated that in many centres follow-up services have been curtailed or stopped entirely in order to cope with the demand on services [13]. However, there is no evidence-base that such disinvestment is safe for patients. Identification of problem patients in a timely fashion is important to avoid complex revision surgery which is considerably more costly in terms of surgical and subsequent rehabilitation costs, more traumatic to the patient and carries higher complication risk. Urgent work is required to determine the most cost-effective follow-up pathway to minimise potential harm to patients.

This timely project therefore aims to examine the consequences, if any, of disinvestment in arthroplasty follow-up.

Why is the research important in terms of improving the health of the public and/or to patients and the NHS?

The importance of appropriate arthroplasty follow-up to ensuring the health of patients was highlighted in March 2014 when the James Lind Alliance and NIHR Priority Setting Partnership (PSP) for Hip and Knee Replacement for Osteoarthritis listed ‘defining the ideal postoperative follow-up period and the best long-term care model for people with OA that have had hip/knee replacement’ amongst its top ten research priorities.

In the early days of arthroplasty all patients were followed for life with annual X-rays. As the techniques became part of mainstream practice, the need for universal follow-up passed. However the 'correct' follow-up of these patients has never been established. Gradually the orthopaedic community have reduced the universal long term follow-up due to economic and time constraints rather than for scientifically evidenced reasons. However, sporadic cases of catastrophic failure of implants, e.g. 3M Capital and metal-on-metal hip replacements, have reawakened interest in follow up. Certainly some of this has been driven by patients themselves, with a culture of patients knowing that there can be failure that is not associated with symptoms.

With increasing demand and increasing patient expectations coupled to diminishing resources, rational evidence-based changes to practice are essential. Current routine follow-up of arthroplasty costs the NHS in the region of £100m per year. Reduction in follow-up is therefore seen as an easy cost-saving measure. However, there has been no robust cost-effective study of how follow-up should be conducted. The recent Briggs report on Improving the Quality of Orthopaedic Care within the National Health Service in England states that all patients should receive appropriate follow-up to detect complications and disease recurrence early [2]. Disinvestment in follow-up as a cost-saving strategy cannot be justified, unless a robust evidence-base is established. Similarly, benefits of more expensive regular follow-up may be limited. Whilst we are aware that research is currently on-going to evaluate new technologies for monitoring patients at a distance, these technologies are themselves expensive, and before they are employed into routine clinical practice an evidence-base for arthroplasty follow-up must first be established. Follow-up must take into account a variety of factors, including implant type, the joint involved and patient factors and a decision to alter follow-up pathways must consider the long-term impact on patients, health professionals, and the NHS as a whole. This project is built around the hypothesis that a comprehensive, evidence-based, stratification algorithm to determine appropriate follow-up for an individual patient may provide a more cost-effective strategy for orthopaedic follow-up service delivery.

Rationalising the current diversity of follow-up practices will enable substantial savings for the NHS and focus the use of NHS resources on those patients most at need. Implementation of appropriate follow-up will reduce the burden on both patients and the NHS in terms of outpatient visits and clinical tests that do not add benefit, while optimising detection of potential problems and thus ensuring patients are not harmed. We envisage the outputs to be readily applicable to the wider NHS, and internationally, at the end of the grant period.

Why is this research needed now?

Research on follow-up is currently piecemeal. Robust, collaborative research is required to definitively address the question of how total hip and knee arthroplasty follow-up should be conducted. Key to this is ensuring that any decision to disinvest does not result in patient harm. This body of work involves key bodies across the UK and has the potential for huge impact across the NHS and internationally with appropriate dissemination.

A recent report from the Academy of Medical Royal Colleges highlighted the increasing pressure on the NHS to preserve standards of care in an environment of growing demand and increasingly constrained budgets [22]. The report challenges the NHS to consider waste in terms of unnecessary use of clinical resources and low value services. They highlight the widespread overuse of tests and interventions which bring little benefit to patients, which in some cases may do more harm than good (such as tests involving ionising radiation) and which prevent NHS resources from being used to bring the best health outcomes to patients. Disinvestment in unnecessary procedures is a key step

in focusing NHS spend, optimising health outcomes and improving patient care. However, such disinvestment must be based on robust evidence to ensure that the interest of the patient remains at the centre of NHS care.

This project is also particularly timely given the urgent need highlighted by Monitor to identify mechanisms to close the NHS funding gap whilst ensuring that the interests of patients remain protected and that the standard of service provision is not compromised [23]. Key areas for investigation within the Monitor report, and which this project is designed to address, include improving productivity within existing services; ensuring that the right care is delivered in the right settings; developing new, innovative ways of delivering health care and allocating spending more rationally. These areas directly align with the remit of HS&DR.

Aims and Objectives

Research question

What are the consequences of disinvestment in hip and knee arthroplasty follow-up?

Objectives:

- To identify who needs follow up and when this should occur for primary total hip and total and uni-compartmental knee arthroplasty surgery by making use of routine data
- To understand the patient journey (in primary and secondary care) to revision surgery by recruiting patients admitted for elective and emergency revision surgery
- To establish how and when patients are identified for revision and why some patients are missed from regular follow-up and present acutely with fracture around the implant (peri-prosthetic fracture), by using prospective and retrospective data
- To identify the most appropriate and cost-effective follow-up pathway to minimise potential harm to patients by undertaking cost-effectiveness modelling
- To provide evidence- and consensus-based recommendations on how follow-up of primary hip and knee arthroplasty should be conducted

Study Design

Overview

UK SAFE comprises of 3 complementary evidence synthesis work-packages to inform a final consensus process:

- **Work Package 1: Systematic Review**
  A systematic review of the clinical and cost effectiveness literature and evidence synthesis will be performed in line with published guidelines [1-5]. Defined inclusion/exclusion criteria, outcomes and cost parameters will be agreed with our PPI and clinical experts. Scoping searches identified 3-4000 papers. Paper selection and data extraction will be undertaken by 2 experienced reviewers. This will provide a robust evidence base for the cost-effectiveness modelling (WP3) and consensus guideline development (WP4).

- **Work Package 2: Determining when, which and how patients present for revision surgery**
  Routine data from 5 national datasets will be used: Clinical Practice Research Database (CPRD) [6], ResearchOne (RO) [7], Hospital Episode Statistics (HES) [8], National Joint Registry (NJR) [9], and Patient Reported Outcome Measures (PROMs) [10] along with prospective data from patients who have revision surgery to understand when and which patients present for revision surgery, and to understand how they are currently identified for revision surgery.

    - **Work Package 2a: Routine data**
      Three linked datasets will be used: (a) CPRD-HES pre-exists at the University of Oxford, (b) RO-HES will be constructed and analysed at the University of Leeds, and (c) NJR-HES-PROMS will be constructed and analysed at the University of Oxford. Sub-packages (a) and (b) will provide a primary care view (e.g. prior diagnoses, prescribing) and include different, representative patient populations for cross-validation, (c) will provide a secondary care view (e.g. surgeon, procedure details).

    - **Work Package 2b: Prospective Cohort**
      Research nurses (RN) at 25 centres will identify elective and emergency patients coming for revision surgery in a 12-month period. We will aim to recruit 675 patients listed for elective revision, as well as emergency patients. Participants will complete a questionnaire about how they came to have revision surgery at that time, including referral route and symptoms prior to surgery. The RN will supplement participant data with data extracted from the medical notes.

- **Work Package 3: Health economics**
  From Work Package 1 and Work Package 2, we will be able to infer from the literature and patterns in the data, informed by our previous work [13], the probability that individuals were identified and listed for revision surgery by routine recall, GP referral or emergency presentation. WP2 datasets will be used to quantify primary/secondary healthcare resource use for current practice and alternative care models through estimation of NHS costs and health-related quality of life (HRQoL). A Markov model [14] will simulate long-term costs and

quality adjusted life years associated with each care model. Primary care costs will include prescriptions, consultations. Hospital costs will be derived by grouping hospital episodes into Health Resource Groups [15]. Panel data regression analysis [16-18] will be used to estimate hospital costs conditional on patient characteristics and co-morbidities. HRQoL and transition probabilities will be derived from the linked datasets, published literature and inputs from expert elicitation to test the care pathway assumptions.

- **Work Package 4: Consensus process**
  Evidence from WP1-3 will feed into a Delphi-consensus process, using the NICE Guideline development model, and involving 25-30 participants including patients, surgeons, GPs, and commissioners to determine appropriate follow-up care pathways for total hip and knee arthroplasty. The output will be a policy document which includes a stratification algorithm for appropriate follow-up for an individual patient.

This document relates only to Work Package 2a (RO-HES) of UK SAFE. Other work packages will be documented separately.

Work Package 2a (RO-HES)

Work Package 2a (RO-HES) has the following objective:

- To determine when, which and how patients present for revision surgery

To fulfil this objective, Work Package 2a (RO-HES) will undertake a retrospective cohort study using the following sources of routine NHS data:

1. **Hospital Episode Statistics (HES)[1]:** a database controlled by NHS Digital containing patient data relating to A&E, inpatient and outpatient episodes
2. **ResearchOne[2]:** a database controlled by The Phoenix Partnership containing de-identified patient data from primary care settings that use the SystmOne clinical information system

Population

Patients will be initially identified for the study from a record of an inpatient admission to a hospital in England in a specific index period (1st April 2000 – 31st March 2015) for one of the following procedures: i) hip replacement, ii) hip revision, iii) knee replacement, or iv) knee revision.

The scope of all hospitals in England has been chosen to enable *variation in follow-up for hip and knee replacements and revisions between hospitals* to be analysed. The time period of 15 years over which patients will be identified has been chosen to enable *variation in follow-up for hip and knee replacements and revisions for patients over time* to be analysed.

NHS Digital will undertake identification of patients from their Hospital Episode Statistics (HES) Inpatient data in accordance with the following inclusion and exclusion criteria:

Inclusion Criteria

- Patient has an inpatient admission to a hospital in England where:

[1] See http://www.digital.nhs.uk/hes
[2] See http://www.researchone.org

- Arrival Date is between 1st April 2000 and 31st March 2015 (Index Period), **AND**
- Patient is aged 18 or over[3], **AND**
- Patient has a coded procedure[4] for:

    - Hip Replacement, **OR**
    - Hip Revision, **OR**
    - Knee Replacement, **OR**
    - Knee Revision

Exclusion Criteria

- Patient has registered a Type 2 objection with NHS Digital to prevent their identifiable data from any health and social care setting being released[5].

Sample Size

Based on data provided within the National Joint Registry (NJR) Annual Report[6], 800,683 primary hip replacements, 89,023 hip revisions, 875,585 primary knee replacements and 54,278 knee revisions were reported (subject to specific inclusion and exclusion criteria) in period from 1 April 2003 to 31 December 2015. Therefore, for a comparable period within the study, we anticipate that around 900,000 hip replacements and 1,000,000 knee replacements (1,900,000 joint replacements in total) will be identified.

Data Items

For included patients, the following data items will be obtained from NHS Digital for the period from 1st April 2000 to 31st March 2015:

- Details of all A&E episodes where "*A&E Diagnosis"* has a value starting with "05" (Dislocation/Fracture/Joint Injury/Amputation) and "A&E diagnosis - anatomical area" has a value of "28" (Hip), "29" (Groin), "30" (Thigh), "31" (Knee), or "32" (Lower Leg). For each episode, the following items will be included:

Data Item	Description
Record Identifier	This is a record identifier that is created by the HES system.
Arrival date	The arrival date of a patient in the A&E department.
Duration to Conclusion	The time (expressed as a whole number of minutes) between the patient's arrival and conclusion of their attendance or treatment (whichever is later).

[3] Patients will only be included if they are aged 18 or over at the index episode. However, it is important to note that data items may be obtained for those patients from NHS Digital and ResearchOne for a period in which the patient was aged under 18.
[4] Definitions based on those provided within the "OPCS Codes relevant to procedures recorded on the NJR" document published by the National Joint Registry (NJR) – see http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Data%20collection%20forms/OPCS%20Procedure%20codes%20relevant%20to%20NJRv4.pdf
[5] See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469290/Data-Provision-Notice_Patient_Objections_Management_19.10.15.pdf
[6] See http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Reports/13th%20Annual%20Report/07950%20NJR%20Annual%20Report%202016%20ONLINE%20REPORT.pdf

Attendance category	An indication of whether a patient is making an initial or follow-up attendance within a particular A&E Department.
Department type	A classification of A&E department type according to the activity carried out.
A&E Diagnosis	The A&E diagnosis code recorded for an A&E attendance.
A&E diagnosis - anatomical area	The A&E diagnosis anatomical area (a classification of parts of the human body)
A&E diagnosis - anatomical side	The A&E diagnosis anatomical side (an indication of the side of the human body).
Diagnosis scheme in use	The Coding Scheme basis of the Diagnosis.
A&E Investigation	The A&E investigation recorded for an A&E attendance.
Number of investigations	Number of investigations
Number of treatments	Number of treatments
A&E treatment	The A&E treatment recorded for an A&E attendance. The CDS allows an unlimited number of treatments to be submitted, however, only the first 12 treatments are available within HES.
IMD Decile group	This field uses the IMD Overall Ranking to identify which one of ten groups a Super Output Area belongs to, from most deprived through to least deprived.
IMD Overall Rank	The Index of Multiple Deprivation (IMD) overall ranking is made by combining the seven IMD Domain scores.
3-digit Provider Code	A provider code is a unique code that identifies an organisation acting as a health care provider.
5-digit Provider Code	A provider code is a unique code that identifies an organisation acting as a health care provider.
Provider Type	Healthcare provider type.
Age on arrival	This field contains the age in whole years on arrival, calculated from arrival date and DOB.
Carer support indicator	This field contains a code which states whether carer support is available to the patient at home or other normal residence. This does not include any paid support or support from a voluntary organisation unless the patient is normally resident in a nursing home, group home or residential care home.
Ethnic Category	This field contains a code which specifies some ethnic groups and some nationalities.
Date of Birth - month and year	Month and year of date of birth only. Day is not made available
Postcode district	Contains the outward portion of the patient's postcode (ie all characters to the left of the space).
Sex of patient	This field contains a code which defines the sex of the patient.
NHS Number Valid Flag	This field indicates whether the NHS Number supplied is valid or not.
Postcode Found	Field confirms if postcode is valid.
Encrypted HES ID	This field contains a unique identifier for each individual patient.

- Details of all Inpatient episodes where *Treatment Specialty* has a value of “110” (Trauma and orthopaedics) or “410” (Rheumatology). For each episode, the following items will be included:

Data Item	Description
Record Identifier	This is a record identifier that is created by the HES system.
Date of admission	This field contains the date the patient was admitted to hospital at the start of a hospital spell.
Method of admission	This field contains a code which identifies how the patient was admitted to hospital.
Date of decision to admit	This field contains the date on which a consultant, or another member of the clinical staff, decided to admit the patient to a hospital.
Waiting time	This field contains the difference in days between the date on which it was decided to admit the patient (elecdate) and the actual admission date (admidate).
Calculation of Elecdur	This field returns the elecdur but excludes admissions from emergency, so only includes eledur where the method of admission (admimeth) is 11 or 12
First regular day or night admission	This field indicates whether the episode falls within a sequence of regular day and night admissions and, if so, whether it is the first or subsequent episode within the sequence.
Source of admission	This field contains a code which identifies where the patient was immediately prior to admission.
Cause code	External cause of injury or poisoning.
All Diagnosis codes	There are twenty fields (fourteen before April 2007 and seven before April 2002), diag_01 to diag_20, which contain information about a patient's illness or condition.
Operation status code	Status of operation.
All Operative procedure codes	There are twenty-four fields (twelve before April 2007 and four prior to April 2002), oper_01 to oper_24, which contain information about a patient's operations. The field oper_01 contains the main (ie most resource intensive) procedure. The other fields contain secondary procedures.
Date of operation	This field contains the dates for operations recorded in the operation codes (opertn_nn) field.
Pre-operative duration	This derived field contains the difference in days between the date the episode started (epistart) and the date of the main operation (opdte_01).
Post-operative duration	This derived field contains the difference in days between the date of the main operation (opdte_01) and the date the episode ended (epiend).
Main specialty	This field contains a code that defines the specialty under which the consultant is contracted. It can be compared with tretspef, the specialty under which the consultant worked.
Treatment specialty	This field contains a code that defines the specialty in which the consultant was working during the period of care. It can be compared with mainspef, the specialty under which the consultant is contracted.
Destination on discharge	This field contains a code which identifies where the patient was due to go on leaving hospital.
Method of discharge	This field contains a code which defines the circumstances under which a patient left hospital.
Date of discharge	This field contains the date on which the patient was discharged from hospital. ; It is only present in the record for the last episode of a spell.
Discharge ready date	The date that a patient was medically ready for discharge from a hospital bed, but couldn't be discharged, therefore qualifying for delayed discharge payments

Episode duration	This field contains the difference in days between the episode start date (epistart) and the episode end date (epiend).
Date episode ended	This field contains the date on which a patient left the care of a particular consultant, for one of the following reasons: Discharged from hospital (includes transfers) or moved to the care of another consultant.
Episode order	This field contains the number of the episode within the current spell. All spells start with an episode where epiorder is 01.
Date episode started	This field contains the date on which a patient was under the care of a particular Consultant.
Episode status	This field tells you whether the episode had finished before the end of the HES datayear (ie whether the episode was still 'live' at midnight on 31 March).
Beginning of spell	This derived field contains a code that defines whether the episode is the first of a spell and whether the spell started in the current or previous year. Other maternity events are excluded.
Duration of spell	This derived field contains the difference in days between the admission date (admidate) and the discharge date (epiend) provided the discharge method (dismeth) confirms that the spell has finished.
End of spell	This field contains a code which defines whether the episode is the last of a spell. It is set for finished episodes (episode status - epistat - is 3) for general, delivery or birth episodes (episode type - epitype - is 1, 2 or 3) provided the discharge method (dismeth) confirms that the spell has finished.
Ward type at start of episode	This field contains a code that defines the characteristics of a ward. The code has six parts: AABCDEF.
Provider code - 3 character	A provider code is a unique code that identifies an organisation acting as a health care provider. The code is managed by the National Administrative Codes Service (NACS) and supports the identification of organisations exchanging information within the NHS.
Provider code - 5 character	A provider code is a unique code that identifies an organisation acting as a health care provider. The code is managed by the National Administrative Codes Service (NACS) and supports the identification of organisations exchanging information within the NHS.
Provider type	Healthcare provider type
Duration of elective wait	The number of days that a patient waited from the date when a decision was taken for treatment to when they received the treatment.
Method of Admission - Waiting List	Calculation determining patients whose method of admission was from the waiting list
Age on admission	A patient's age, in years, at the date of admission.
Ethnic category	This field contains a code that specifies some ethnic groups and some nationalities. It was introduced from the 1995-96 data year
Date of Birth - month and year	Month and year of date of birth only. Day is not made available.
Postcode district of patient's residence	Contains the outward portion of the patient's postcode (ie all characters to the left of the space).
Sex of patient	This field contains a code which defines the sex of the patient.
NHS Number valid flag	This field indicates whether the NHS Number supplied is valid or not.
Postcode Found	Field confirms if postcode is valid.
IMD Decile group	This field uses the IMD Overall Ranking to identify which one of ten groups a Super Output Area belongs to, from most deprived through to least deprived.

IMD Overall Rank	The Index of Multiple Deprivation (IMD) overall ranking is made by combining the seven IMD Domain scores.
Encrypted HES ID	This field contains a unique identifier for each individual patient.

- Details of all Outpatient episodes where *Treatment Specialty* has a value of "110" (Trauma and orthopaedics) or "410" (Rheumatology). For each episode, the following items will be included:

Data Item	**Description**
Record Identifier	This is a record identifier that is created by the system.
Appointment date	The date when an appointment was scheduled.
Last DNA or patient cancelled date	This is recorded when patients who have been offered an appointment date have missed this date with or without advance notice.
Days waiting	'Waiting' gives the period in days between the date of the appointment date and either the referral request received date (reqdate) or the DNA (did not attend) date, if given.
Referral request received date	This field records the date the referral request was received by the healthcare provider.
Waiting calculation indicator	Waitind indicates how and whether waiting time has been calculated.
Attendance type	A field derived from 'first appointment' (firstatt) and 'attended or did not attend' (attended), used to identify if the attendance occurred and whether it was the first or subsequent.
Attended or did not attend	This indicates whether or not a patient attended for an appointment. If the patient did not attend it also indicates whether or not advanced warning was given.
First attendance	Gives an indication of whether a patient is making a first attendance or follow-up attendance, and whether the consultation was face-to-face or via telephone/telemedicine consultation.
Medical staff type seeing patient	Gives information about the type of care professional staff dealing with the patient during a consultant outpatient attendance, or nurse or midwife contact.
Outcome of attendance	This records the outcome of an outpatient attendance.
Priority type	This is the priority of a request for services in the case of services to be provided by a consultant, it is as assessed by or on behalf of the consultant.
Service type Requested	Describes the terms of reference for the referral request.
Source of referral	A classification which is used to identify the source of referral of each consultant outpatient episode.
Diagnosis	There are twelve fields (two before April 2007), diag_01 to diag_12, which contain information about a patient's illness or condition.
Operative Procedure	There are twenty-four fields (twelve before April 2007), oper_01 to oper_24, which contain information about a patient's operations.
Main specialty	A code that defines the specialty under which the consultant is contracted. Compare with 'treatment specialty' (tretspef), the specialty under which the consultant worked
Operation status code	Status of operation.

Treatment specialty	This field contains a code that defines the specialty in which the consultant was working during the period of care. It can be compared with mainspef, the specialty under which the consultant is contracted.
Provider code (3 character)	A provider code is a unique code that identifies an organisation acting as a health care provider.
Provider code (5 character)	A provider code is a unique code that identifies an organisation acting as a health care provider.
Provider type	Healthcare provider type.
Age on day of appointment	This derived field, calculated from appointment date (apptdate) and date of birth (dob), contains the patient's age in whole years.
Ethnic category	This field contains a code that specifies some ethnic groups and some nationalities.
Date of birth – month and year	Month and year of date of birth only. Day is not made available.
Postcode district of patient's residence	Contains the outward portion of the patient's postcode (ie all characters to the left of the space).
Sex of patient	This field contains a code which defines the sex of the patient.
NHS Number valid flag	This field indicates whether the NHS Number supplied is valid or not.
Postcode Found	Field confirms if postcode is valid.
IMD Decile Group	This field uses the IMD Overall Ranking to identify which one of ten groups a Super Output Area belongs to, from most deprived through to least deprived.
IMD Overall Ranking	The IMD overall ranking is made by combining the seven IMD Domain scores.
Encrypted HES ID	This field contains a unique identifier for each individual patient.

In addition, we define a subset of patients who are identified in accordance with the criteria above and who also fulfil the following criteria:

- Patient is registered to a General Practice that has opted into ResearchOne and has not individually opted out of ResearchOne

For this subset of patients, the following additional data items will be obtained from ResearchOne for the period prior to the 1st April 2015:

- Details of demographic attributes, which will include:

Data Item	Description
Patient Identifier	Unique ResearchOne Identifier for patient
Date Of Birth (mm/yyyy)	Month and year of patient's birth.
Date Of Death (mm/yyyy)	Month and year of patient's death.
Gender	Gender of patient

- Details of socio-economic attributes, which will include:

Data Item	Description
Patient Identifier	Unique ResearchOne Identifier for patient

Start Date	Month and year at which patient started residence at address.
End Date	Month and year at which patient ended residence at address.
Type Of Address	Specific type of address, e.g. private residential or community establishment.
IMD Rank	Index of multiple deprivation (IMD) rank calculated from patient address.
IMD Score	Index of multiple deprivation (IMD) score calculated from patient address.
Sector Level Postcode	5 digit sector level postcode for patient address.

- Details of all coded events where *CTV3 Concept Id* matches a code included in: i) QOF-related definitions[7] for Asthma, Atrial Fibrillation, Blood Pressure, Cancer, CHD, CVD, CKD, COPD, Dementia, Depression, Diabetes, Epilepsy, HF, Hypertension, Learning Disability, Obesity, Osteoporosis, Psychosis, Schizophrenia or Bipolar-Affective Disorder, PAD, Palliative Care, Rheumatoid Arthritis, Smoking, Stroke, Stroke (TIA), ii) expert clinician definitions for Joint Pain, Hip Replacement, Hip Revision, Knee Replacement, Knee Revision, or iii) expert clinician definitions for relevant referrals to Trauma/Orthopaedics or Rheumatology. For these events, the following items will be included:

Data Item	**Description**
Patient Identifier	Unique ResearchOne Identifier for patient
EventId	Reference to event in which referral occurred.
EventDate	Date of event
CTV3 Concept Id	CTV3 code for diagnosis/observation.
NumberValue	Number value associated with CTV3 code
LowerBound	Lower bound associated with CTV3 code
UpperBound	Upper bound associated with CTV3 code
NumericRangeComparisonMethod	Operator to use for numeric range comparisons

- Details of all prescriptions for drugs where *BNFChapter* matches a chapter included in expert clinician definitions for Opioid Analgesics, Non Steroidal Anti-Inflammatory Drugs, Non Opioid Analgesics and Compound Analgesics, Rubefacients, Topical NSAIDS, Capsaicin and Poultices, Drugs That Suppress The Rheumatic Disease Process, or Corticosteroids. For these events, the following items will be included:

Data Item	**Description**
Patient Identifier	Unique ResearchOne Identifier for patient
MedicationId	Unique reference to specific prescription
EventId	Reference to event in which prescription occurred.

[7] See http://www.hscic.gov.uk/qof

EventDate	Date of event in which prescription occurred.
StartDate	Start date of prescription
EndDate	End date of prescription
DrugNumber	Unique reference for prescribed drug.
Dose	Dosage of drug
Quantity	Quantity of drug
DrugStatus	
RepeatMedicationId	Reference to repeat medication (if applicable)
MedicationEnded	Indicator of whether medication has ended.
DateMedicationEnded	Date on which medication was ended.
ReasonMedicationEnded	Reason medication was ended.

- Details of all repeat prescriptions for drugs where *BNFChapter* matches a chapter included in expert clinician definitions for Opioid Analgesics, Non Steroidal Anti-Inflammatory Drugs, Non Opioid Analgesics and Compound Analgesics, Rubefacients, Topical NSAIDS, Capsaicin and Poultices, Drugs That Suppress The Rheumatic Disease Process, or Corticosteroids. For these events, the following items will be included:

Data Item	**Description**
Patient Identifier	Unique ResearchOne Identifier for patient
Repeat Medication Id	Unique reference for repeat prescription
Event Id	Reference to event in which referral occurred.
Event Date	Date on which event occurred
Start Date	Start date of repeat prescription.
End Date	End date of repeat prescription.
Drug Number	Reference to drug prescribed.
Dose	Dosage of drug prescribed
Quantity	Quantity of drug prescribed
Drug Status	
Maximum Issues Authorised	Maximum number of issues authorized.
Medication Review Date	Date of review for medication
Medication Ended	Indicator for whether medication ended
Date Medication Ended	Date on which medication was ended
Reason Medication Ended	Reason for which medication was ended

- Details of all non-coded referrals, which will include the following items:

Data Item	**Description**
Patient Identifier	Unique ResearchOne Identifier for patient

Event Id	Reference to event in which referral occurred.
Date Of Referral	Date on which referral occurred.
Type Of Referral	Type of referral
Urgency Of Referral	Urgency of referral
Reason For Referral	Clinical reason for referral
Referral To	Destination for referral

- Details of all practice registrations, which will include the following items:

Data Item	Description
Patient Identifier	Unique ResearchOne Identifier for patient
Practice Number	Pseudonym for the GP practice.
Start Date	Date at which patient registration at practice started.
End Date	Date at which patient registration at practice ended.
Applied GMS	Whether registration was Applied GMS.
Fully GMS	Whether registration was Fully GMS.
Temporary Resident Short	Whether registration was Temporary Resident Short.
Temporary Resident Long	Whether registration was Temporary Resident Long.
Temporary Resident Telephone	Whether registration was Temporary Resident Telephone.
Immediate Necessary Treatment	Whether registration was Immediate Necessary Treatment.
Emergency	Whether registration was Emergency.
Child Health	Whether registration was Child Health.
Contraception	Whether registration was Contraception.
Maternity	Whether registration was Maternity.
Minor Surgery	Whether registration was Minor Surgery.
Private	Whether registration was Private.
Other	Whether registration was Other.
WalkIn	Whether registration was Walk-In.

- Additional detail relating to any coded event, prescriptions, repeat prescription and referrals, which will include the following items:

Data Item	Description
Event Id	Reference to event in which referral occurred.
EventDate	Date of event
Method	Method of interaction associated with event, e.g. ‘Face to Face’ appointment.

For any drug referenced in the prescriptions or repeat prescriptions, the following items will be included:

Data Item	Description
Drug Number	Unique drug number
FullName	Full name of drug
Pack	Type of pack for drug
Form	Delivery form of drug
Strength	Strength of drug
BNFChapter	BNF chapter for drug.

For any GP practice referenced in the practice registrations, the following items will be included:

Data Item	Description
Practice Number	Unique ResearchOne Identifier for GP practice
List Size	List size at 1st April 2015
No of Partners	Number of partners at 1st April 2015
Description	Type of organization, e.g. General Practice.
LiveOnSystmOneSince	Date from which the GP practice was on SystmOne.

Further details on the patient selection criteria and the data items that will be obtained for patients from both NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne) can be found in Appendix 1.

Methods

NHS Digital will provide the data items relating to A&E, Inpatient and Outpatient episodes described above for the identified patient population to the University of Leeds. Within this data, each patient will be uniquely referenced with a (non-personal) identifier (HES ID). Additionally, NHS Digital will generate a pseudonym for each one of these patients by applying a defined cryptographic hash function to a concatenation of NHS number and a project-specific "salt" file supplied by the University of Leeds. NHS Digital will provide this set of pseudonyms to The Phoenix Partnership.

The Phoenix Partnership will generate a pseudonyms for each of the patients in the ResearchOne database in an analogous manner to NHS Digital and use these pseudonyms to determine those patients that are present in the list provided by NHS Digital. Data items relating to General Practice for this subset of patients will be provided to the University of Leeds by The Phoenix Partnership. Within this data, each patient will be uniquely referenced with a (non-personal) identifier (RO ID). Additionally, The Phoenix Partnership will provide a file to NHS Digital that maps each unique RO ID to the associated pseudonym.

NHS Digital will use the file provided by The Phoenix Partnership and the pseudonyms previously generated for the identified patient population to generate a file that maps each HES ID to an RO

ID. NHS Digital will provide this mapping file to the University of Leeds to enable the data items obtained from NHS Digital and The Phoenix Partnership for each patient to be linked for analysis.

Further details on the data linkage methodology can be found in Appendix 2.

Analysis

Survival analysis will be used to model time to revision. To determine the follow up time of revision, a smoothed Nelson-Aalen cumulative hazard rate will be examined to identify any peak in the mid-long term risk of revision. To identify patients most likely to require revision, proportional hazards regression modelling will be used to identify pre, peri- and post-operative predictors of mid-late term revision. The date of the first incidence of a subject's hip or knee replacement will be used as the start time. The event of interest in all time-to-event models will be the first recorded revision operation.

Cox proportional hazards regression modelling will be used to identify pre, peri- and post-operative predictors of mid-late term revision (defined as more than 5y post primary surgery). Should testing reveal that the proportional hazards assumption is not valid, then parametric modelling will be used instead. Shared frailty will be modelled with a random effect for hospitals/providers and another for general practice, since it is anticipated that both hospital practice regarding follow up and GP behaviour regarding referral with influence the survival time of the primary joint replacement. Competing risks, including mortality and comorbidities following the primary surgery will be considered. Linearity of continuous predictors will be assessed using fractional polynomial regression modelling or splines. Missing data will be handled by using multiple imputation methods using the ICE (Imputation by Chained Equations) procedure.

Ethical Issues

Benefits

Use of routinely collected data from EHRs has significant advantages over conventional sources of research data. Such data enables a wide range of patients, health and healthcare dimensions, and timeframes to be analysed in a cost-effective and timely manner. Studies can be undertaken at a scale and granularity that would not be viable (due to time and resource constraints) using alternative (study-specific) sources of research data.

Within Work Package 2a (RO-HES), use of such data provides an opportunity to analyse when, which and how patients present for revision surgery at a national scale and with patient-level granularity. For a subset of patients, data from both primary and secondary care settings will be available. Such data will provide a comprehensive set of patient factors for consideration during analysis. Variation can be studied between hospitals and patients over time to understand the factors that have a significant effect on when, which and how patients present for revision surgery.

For certain descriptive studies, patient-level data may not be required to undertake the necessary analysis. Aggregate data obtained from data sources may be sufficient to answer the research question. However, the use of aggregate data in this study would significantly limit the granularity of analysis that could be undertaken and the insights that could be derived in relation to revision surgery. Aggregation criteria across all dimensions of analysis (e.g. patients and time) would have to be specified *a priori* by the research team and provided to the data sources. Data sources would also require the capability and capacity to process their data in accordance with this aggregation criteria. Aggregate data would also preclude any joint consideration of data from primary and secondary care for each patient, which would further limit the granularity of analysis and insights that could be derived.

In summary, use of routinely collected data from EHRs in Work Package 2a (RO-HES) has the following benefits:

- Cost-effective research methodology
- Detailed understanding of when, how and which patients present for revision surgery.
- Detailed understanding of how follow-up models vary between hospitals (over time) in England.

These benefits feed into the wider promised benefits of UK SAFE, which include:

- Identification of more appropriate and cost-effective follow-up pathways (that minimise potential harm to patients) to be identified within subsequent work packages.
- Reduced service costs for NHS through the identification of the most cost-effective follow-up model for hip and knee replacement.

Risks

Data held by NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne) is derived from data collected that is routinely collected in Electronic Health Records (EHRs) by healthcare professionals (under a duty of confidentiality) for the purpose of care provision. Use of such data for secondary purposes is subject to different ethical and legal considerations. Any use of patient identifiable data for secondary purposes (e.g. research) requires patient consent or another relevant legal basis (e.g. Section 251 support). Work Package 2a (RO-HES) does not

require any patient identifiable data to be obtained by the research team, and explicit consent will not be sought from each individual patient whose data is to be obtained for the study. An application will be made to the Confidentiality Advisory Group (CAG) to confirm whether Section 251 support is required by the study[8].

Whilst no patient identifiable data is obtained by the research team, risks remain for those patients whose data are included. The most significant risk for patients is their (re-)identification from the data obtained for the study from NHS Digital and The Phoenix Partnership. Re-identification of a patient from the data may enable particular (potentially sensitive) data items relating to health and healthcare to be associated with that participant by a member of the research team, a member of staff at the data provider, or a third party who gains access to the data. Such association by these parties, and any subsequent use of the association by these parties, would likely be considered a breach of confidentiality and a violation of patient privacy. This could have negative social, financial and health/wellbeing consequences for the patient.

Further details regarding the confidentiality and privacy risks associated with Work Package 2a (RO-HES) are provided in Appendix 3.

Measures

In order to mitigate the risk relating to confidentiality and privacy whilst retaining the benefits of the work, a number of different measures will be taken within the study:

Patient Opt-Out

Use of routinely collected data from NHS Digital (Hospital Episode Statistics) will draw upon the Patient Objection Management process[9] to enable patients to opt-out of the use of their data for purposes other than care provision. Patients can express Type 1 and Type 2 objections[10] through the Patient Objection Management process in order to prevent the use of their identifiable data for purposes other than direct care. Such objections relate specifically to identifiable data and pseudonymised data requests that go through the DARS process at NHS Digital are considered to be compliant with the ICO code of practice on anonymisation[11]. Type II objections would not typically be honoured on requests that are not considered to contain identifiable data. Work Package 2a (RO-HES) does not require identifiable data and therefore Type II objections would not typically be honoured on such a request. However, to acknowledge the importance of honouring patient objections to the use of their data (whether or not the data is considered to be identifiable), we will request that NHS Digital honour Type II objections on data requests relating to Work Package 2a (RO-HES).

For clarity, patients who have issued a Type I or II objection will not be considered in: i) the clinical data received from NHS Digital by the University of Leeds, or ii) the list of pseudonyms provided by NHS Digital to The Phoenix Partnership for linkage purposes.

---

[8] Following submission of an application to the Confidentiality Advisory Group at the Health Research Authority (22nd June 2017), and following subsequent correspondence with the Health Research Authority, NHS Digital and The Phoenix Partnership, it was determined that Section 251 support was not required for the study and the application to the Confidentiality Advisory Group with withdrawn (4th October 2017).

[9] See http://www.hscic.gov.uk/media/12822/Guide-to-confidentiality-in-health-and-social-care/pdf/HSCIC-guide-to-confidentiality.pdf

[10] See http://content.digital.nhs.uk/yourinfo

[11] See http://content.digital.nhs.uk/article/7092/Information-on-type-2-opt-outs

Additionally, use of routinely collected data from The Phoenix Partnership (ResearchOne) will draw upon ResearchOne's consent model[12]. GP practices that use the SystmOne clinical information system can opt-in to ResearchOne on an organisational basis. As a consequence, data for all patients registered to that GP practice are included on ResearchOne. Any patient registered to a GP practice that has opted-in to ResearchOne can opt-out of ResearchOne on an individual basis. Only patients whose GP practice has opted-in to ResearchOne and who have not individually opted-out of ResearchOne are included on ResearchOne.

For clarity, patients whose GP practice have not opted-in to ResearchOne, or whose GP practice has opted-in to ResearchOne but who have individually opted-out of ResearchOne will not be considered in: i) the clinical data received from ResearchOne by the University of Leeds, or ii) the mapping file provided by The Phoenix Partnership to NHS Digital for linkage purposes.

Anonymisation

In adherence to national guidance on information governance in the health and social care system[13], Work Package 2a (RO-HES) has been designed to minimise the need for access to identifiable information. No patient identifiable data[14] is included in the data items that are supplied for each patient by either NHS Digital or The Phoenix Partnership. Data provided by NHS Digital as part of pseudonymised data requests is considered to be compliant with the ICO code of practice on anonymisation[15]. Additionally, data contained within ResearchOne has been subject to a process of anonymisation[16], which involves the omission and transformation of specific data items to reduce the risk that a patient can be (re-)identified from the data. ResearchOne has received a favourable ethical opinion as a research database from an NHS Research Ethics Committee (REF: 11/NE/0184) and has received a decision from the National Information Governance Board (NIGB) that Section 251 support is not required[17] as there is no disclosure of identifiable data.

Data Minimisation

Use of anonymised data (as described above) significantly reduces the risk that individual patients can be associated with any (potentially sensitive) data items relating to health and healthcare released by these organisations for use in the study. However, there remains a risk to patients that they can be (re-)identified using unique combinations of data items obtained for the study in conjunction with data items from externally available datasets. It is not feasible for the research team to determine *a priori* every potentially identifying combination of values for the different data items, and every potential knowledge set (i.e. externally available datasets) of an attacker, and to amend the data requested accordingly. Critically, the determination of potentially identifying combination of values for the different data items would require access to the data itself before the request is made.

[12] See http://www.researchone.org/documentation/
[13] See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf
[14] See http://content.digital.nhs.uk/media/12053/Obtaining-consent-for-patient-identifiable-and-or-sensitive-data-V2-050613/pdf/obtaining_consent_for_patient_identifiable_and_or_sensitive_data_V2_050613.pdf
[15] See http://content.digital.nhs.uk/article/7092/Information-on-type-2-opt-outs
[16] See http://www.researchone.org/wp-content/uploads/2013/09/ResearchOne-Database-System-Summary_V2.pdf
[17] See http://www.researchone.org/wp-content/uploads/2013/09/ResearchOne-NIGB-ECC.pdf

To mitigate the risk that patients can be (re-)identified using unique combinations of data items obtained for the study, Work Package 2a (RO-HES) has been designed to minimise the number of data items that are obtained for each patient from both data sources to those that have been deemed by members of study management group as *necessary* and *sufficient* to robustly answer the research question.

Further details on the inclusion/exclusion criteria for patients and the data items that will be obtained for these patients can be found in Appendix 1.

Linkage

Work Package 2a (RO-HES) requires data from NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne) that relates to the same patient to be linked. As previously stated, no patient identifiable data[18] is included in the data items that are supplied for each patient by either NHS Digital or The Phoenix Partnership. To enable linkage to be undertaken without the requirement for NHS Digital or The Phoenix Partnership to release patient identifiable data, pseudonyms will be used as the basis for linkage. In accordance with national guidance on information governance in the health and social care system[19], such linkage will be undertaken by NHS Digital (previously Health and Social Care Information Centre).

University of Leeds will inform NHS Digital and The Phoenix Partnership of the specific process by which pseudonyms must be generated. This process will stipulate the application of a specific one-way cryptographic hash function (SHA-512) to a concatenation of NHS number and a project-specific "salt" file supplied by the University of Leeds. The "salt" file represents a random string of defined length that is used as an additional input to the pseudonym generation process to reduce the risk that: i) NHS numbers could be recovered from pseudonyms, and ii) pseudonyms generated from NHS numbers for different projects (and any associated data) could be matched.

NHS Digital and The Phoenix Partnership will generate pseudonyms for patients using this methodology. NHS Digital will communicate a set of pseudonyms to The Phoenix Partnership to indicate those patients for whom primary care data is required. The Phoenix Partnership will communicate a mapping from pseudonym to RO-specific identifier to NHS Digital for each patient for whom primary care data is available. NHS Digital will generate a mapping from HES-specific identifier to RO-specific identifier using the mapping provided by The Phoenix Partnership. NHS Digital will provide this mapping file to University of Leeds to enable linkage of the data from NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne) for analysis.

Importantly, University of Leeds do not receive any pseudonyms from NHS Digital or The Phoenix Partnership. University of Leeds receives: i) data from NHS Digital (Hospital Episode Statistics) where patients are referenced by a HES-specific identifier, ii) data from The Phoenix Partnership (ResearchOne) where patients are referenced by a RO-specific identifier, and iii) a file from NHS Digital which maps HES-specific identifiers to RO-specific identifiers. Linkage of the HES-specific identifier to RO-specific identifier is undertaken at NHS Digital.

Further details on the linkage methodology can be found in Appendix 2.

---

[18] See http://content.digital.nhs.uk/media/12053/Obtaining-consent-for-patient-identifiable-and-or-sensitive-data-V2-050613/pdf/obtaining_consent_for_patient_identifiable_and_or_sensitive_data_V2_050613.pdf

[19] See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf

Internal Approvals

Use of data from NHS Digital is subject to an approval process[20] that "verifies that there is an appropriate legal basis for accessing the requested data and that appropriate safeguards are in place to ensure that you will store and handle data safely and securely". Data requests are evaluated by the HES Information Asset Owner (IAO) and Independent Group Advising on the Release of Data (IGARD)[21] [22] (formerly Data Access Advisory Group (DAAG)). Data is not released by NHS Digital without relevant approvals.

Use of data from ResearchOne is also subject to an approval process[23] in which data requests are evaluated by the ResearchOne Project Committee (RPC). The RPC assesses data requests based on criteria defined by the ResearchOne Database Committee (RDC) and by the research aims of ResearchOne. The RPC assesses the risks posed by each proposal and refer any disputes to the RDC for advice and guidance.

Agreements

Use of data from NHS Digital by University of Leeds is governed by a Data Sharing Framework Contract (DSFC)[24]. Specific Data Sharing Agreement(s) that will be established for the specific data required by this study will be underpinned by this DSFC[25].

Use of data from ResearchOne is subject to terms of use specified in a data request[26] which relates to the project, governance arrangements, and the scope, format and intended use of data. The format of the data request form is defined by The Phoenix Partnership, as Data Controller for ResearchOne, and requires approval from the ResearchOne Project Committee[27]. Data is only released for specified research purposes. Individual researchers who will have access to the data must also complete a Confidentiality Agreement[28] that specifies the conditions under which access to the data is permitted.

Additionally, University of Leeds will establish Data Processing Agreements with NHS Digital (if applicable)[29] and The Phoenix Partnership[30] to govern specific data processing tasks that will be undertaken by these parties on behalf of the University of Leeds.

---

[20] See http://content.digital.nhs.uk/article/6888/DARS-Process
[21] See http://content.digital.nhs.uk/media/23491/IGARD-Terms-of-Reference-v15/pdf/IGARD_Terms_of_Reference_v1.5.pdf
[22] Approval for this work has now been received from the Information Asset Owner (September 2020) and the Independent Group Advising on the Release of Data (September 2020).
[23] See http://www.researchone.org/wp-content/uploads/2013/08/TPP-Research-Database_Protocol_V1.4.pdf.
[24] See http://content.digital.nhs.uk/media/15731/Data-Sharing-Framework-Contract-Guidance/pdf/Data_Sharing_Framework_Contract_Guidance_v1_-_FINAL.pdf
[25] A Data Sharing Agreement between the University of Leeds and NHS Digital has been established for this work (October 2020) (REF: NIC-147997).
[26] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc
[27] Data request for this work was submitted by the University of Leeds to The Phoenix Partnership in January 2018 and approval from the ResearchOne Project Committee has been most recently received for this work in February 2019.
[28] See http://www.researchone.org/wp-content/uploads/2013/09/ResearchOne-Database-System-Summary_V2.pdf
[29] Following correspondence with NHS Digital in relation to another research project based at the University of Leeds (LP-MAESTRO – IRAS ID: 178391), which uses the same linkage methodology between Hospital Episode Statistics and ResearchOne, it was determined that a Data Processing Agreement between the University of Leeds and NHS Digital was not required (August 2017). UK-SAFE WP2a (RO-HES) proceeded on the same basis using LP-MAESTRO as the precedent.
[30] A Data Processing Agreement between the University of Leeds and The Phoenix Partnership has been established for this work (December 2018).

Information Security

To minimise any risk that data could be accessed by an unauthorised party, Work Package 2a (RO-HES) will adhere to strict information security practices.

To minimise any risk that data could be accessed by an unauthorised part in transit, secure transfer mechanisms will be used. For transfer of data between NHS Digital and The Phoenix Partnership, or NHS Digital and University of Leeds, the secure transfer mechanism provided by NHS Digital (SEFT) will be used. For transfer of data between The Phoenix Partnership and University of Leeds, the secure transfer mechanism provided by Leeds Institute for Clinical Trials Research (SFT) will be used.

To minimise any risk that data could be accessed by an unauthorised party once received by the University of Leeds, data will be managed in an infrastructure provided by the Leeds Institute for Clinical Trials Research (LICTR) that is compliant with NHS Data Security and Protection Toolkit. Access to the data will be restricted to specific members of the research team who have undertaken appropriate information security training, and who have been approved by the Principal Investigator. A log will be maintained of those members of the research team with access to the data.

To minimise any risk that data published by the research team discloses the identity or attributes of specific patients, all published outputs will be assessed against the NHS Anonymisation Standard[31] and ICO's Anonymisation: managing data protection risk - code of practice[32] and approved by the Principal Investigator prior to publication.

[31] See http://www.isb.nhs.uk/documents/isb-1523/amd-20-2010/1523202010spec.pdf
[32] See https://ico.org.uk/media/1061/anonymisation-code.pdf

Approvals

Work Package 2a (RO-HES) will not proceed before it has obtained (minimally) the following approvals/decisions:

- **Favourable opinion from a National Health Service (NHS) Research Ethics Committee.[33]**
    - If required, any additional approvals/decisions on which this opinion is contingent will be obtained.

- **Decision from Health Research Authority's Confidentiality Advisory Group (CAG) as to whether Section 251 support is required.[34]**
    - If required, Section 251 support will be obtained.

[33] Favourable opinion was received from an NHS Research Ethics Committee on 8th August 2017 (REF: 17/YH/0250), subject to Management Permission, which was received on 9th September 2017.

[34] Following submission of an application to the Confidentiality Advisory Group at the Health Research Authority (22nd June 2017) (REF: 17/CAG/0122), and following subsequent correspondence with the Health Research Authority, NHS Digital and The Phoenix Partnership, it was determined that Section 251 support was not required for the study and the application to the Confidentiality Advisory Group with withdrawn (4th October 2017).

Data Management

Project Infrastructure

Work Package 2a (RO-HES) will use the infrastructure provided by the Leeds Institute for Clinical Trials Research (LICTR)[35] at the University of Leeds to securely manage the data received from NHS Digital and The Phoenix Partnership, and to underpin the terms agreed between the University of Leeds and the data controllers (NHS Digital and The Phoenix Partnership) of the databases (Hospital Episode Statistics and ResearchOne) used in this work. The LICTR infrastructure has NHS Data Security and Protection Toolkit[36] approval (REF: ECC0010) and is named the Data Sharing Framework Agreement that currently exists between the University of Leeds and NHS Digital.

Secure data transfer between the organizations involved in the methodology will be performed using one of the following services:

- SFT Service[37] provided by LICTR
- SEFT Service[38] provided by NHS Digital

We envisage that the use of the SEFT service will be mandated by NHS Digital for any transfer to or from NHS Digital. For all other transfers, we envisage that the SFT service will be used. Accordingly, the methodology description that follows is based on the premise that:

- A named individual at UoL and The Phoenix Partnership have an active account on the SEFT service (NHS Digital)
- A named individual at The Phoenix Partnership has an active account on the SFT service (UoL).

University of Leeds is registered on the Data Protection Public Register held by the Information Commissioner's Office (Data Protection Registration Number: Z553814X) and has a set of has a set of defined Information Security Policies[39].

Long Term Storage

In accordance with the terms that will be established with the data controllers (NHS Digital and The Phoenix Partnership) and the databases used in this work (Hospital Episode Statistics and ResearchOne), data that is provided for Work Package 2a (RO-HES) will be retained by the University of Leeds for a defined time period. Following this retention period, the data will be securely deleted.

To ensure provenance and repeatability of the research, data relating to the following will be retained within the University of Leeds for 15 years:

- Criteria from which all data was extracted from Hospital Episode Statistics and ResearchOne
- Methodology used by NHS Digital to link data between sources
- Criteria used to filter and process the data for analysis

[35] See http://medhealth.leeds.ac.uk/info/400/leeds_institute_of_clinical_trials_research/
[36] See https://www.dsptoolkit.nhs.uk/
[37] See Appendix 2a
[38] See Appendix 2b
[39] See https://it.leeds.ac.uk/info/116/policies

Dissemination and Outputs

Plans for Dissemination

This project will deliver the first research-supported, best-for-patient, joint-specific, cost-effective recommendations for follow-up care, providing a gold standard for clinical excellence, and follow-up advice for patients, surgeons, purchasers and health services. Value is not limited to the UK, but has massive global potential.

Nationally, the outputs, in the form of an executive summary statement of the agreed pathway/s will be disseminated through appropriate NHS Networks, the NHS England Elective Orthopaedics Subcommittee, the NHS Institute for Innovation and Improvement and professional societies. Dissemination will be key to developing a culture of 'finding the best way of doing something and doing it everywhere' to significantly reduce wastage of clinical resources and optimise NHS spend.

We have support of the British Hip Society (BHS), British Orthopaedic Association (BOA), British Association for Surgery of the Knee (BASK), NHS England, Arthroplasty Care Practitioners Association (ACPA), the National Joint Registry and three Leeds-based CCGs for this research and for dissemination activities. We will put forward the consensus statement to each society's AGM for adoption as a resolution.

Internationally, dissemination platforms are already in place through the International Society of Arthroplasty Registers (ISAR) and the European Federation of National Associations of Orthopaedics and Traumatology (EFORT).

Overall and individual work-packages findings will be disseminated through a variety of media. Abstracts will be submitted to major British and international orthopaedic conferences, and separate relevant meetings, including the Health Economics Study Group and Exploiting Existing Data for Health Research conference. We will look to present at the NIHR Methodology Conference and NHS Management conferences and events. Manuscripts will be submitted to appropriate peer-reviewed journals, including general medical, orthopaedic and management journals.

Patient dissemination will be supported through the LMBRU PPI forum and website and our strong ties with Arthritis Care. We will hold a PPI conference at the end of the study. We will encourage our PPI representatives to be involved in presentations, with support from research staff, and they will help to ensure conference material is appropriate. With our lay representatives we will write a lay summary for publication in a patient publication such as The Patient or Inspire.

Expected Output of Research/Impact

Upon completion, this research will have major immediate effect on national NHS planning and budgeting and patient well-being. The outputs will be evidence-based support for timing of follow up and identification of the most cost-effective follow-up model. This fits directly within the NHS framework for improving outcomes from elective procedures. Rationalising current diversity of follow up practices should enable substantial savings for the NHS. Novel follow-up strategies, such as creating a rapid access pathway after joint replacement for symptomatic patients will be examined. We envisage outputs to be readily applicable to the wider NHS, not only hip and knee but also other joint replacements.

The impact will be to reduce the burden on patients and the NHS in terms of outpatient visits and clinical tests that do not add benefit, while optimising detection of potential problems. From an NHS

perspective, this work will provide NHS managers with economic and clinical information on arthroplasty follow-up to inform service planning and delivery, and the role of arthroplasty practitioners in this service; provide orthopaedic surgeons with guidance on follow-up, including patient and economic considerations of factors involved; produce arthroplasty follow-up guidelines for adoption by the relevant specialist societies and inclusion with information for their members. From a patient perspective, this work will help to inform patients about follow-up practice and empower them to make choices about future healthcare relating to their joint arthroplasty.

At the end of the project, a policy document will be created with support of the relevant societies, NHS England, CCGs and patient representation. It is anticipated that this will include a stratification algorithm to determine appropriate follow-up for an individual patient, taking into account, for example, implant type and patient factors, and that recommended follow-up pathways for hip may differ to those for knee. This advisory document will be disseminated to all stakeholders, including orthopaedic surgeons, arthroplasty surveillance professionals and NHS managers. With the committed support of these key organisations, we anticipate that these guidelines will be positively received and that implementation will be widespread. It is our ambition for the recommended follow-up pathway/s defined by this programme of work to be adopted for all hip and knee replacement patients in the UK and internationally.

References

1. Lefebvre C, Manheimer E, et al. Chapter 6: Searching for studies. In: Higgins JPT, Green S (Eds.). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration. 2001.
2. Noyes J, Popay J, et al. Chapter 20: Qualitative research and Cochrane reviews. In: Higgins JPT, Green S (Eds.). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration. 2011.
3. Centre for Reviews and Dissemination (2009) Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care. CRD: York.
4. Higgins JPT & Green S (Eds.) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration (Available from www.cochrane-handbook.org)
5. http://www.prisma-statement.org.
6. http://www.cprd.com
7. http://www.researchone.org
8. http://www.hscic.gov.uk/hes
9. http://www.njrcentre.org.uk.
10. http://www.hscic.gov.uk/proms
11. Hunt LP, Ben-Shlomo Y, 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. 2013 Sep 28; 382(9898):1097-1104.
12. Liddle AD, Judge A, et al. 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. 2014 Oct 18; 384(9952):1437-1445.
13. Smith LK. A Survey of the Current State of Hip Arthroplasty Surveillance in the United Kingdom. Musculoskeletal care. 2014 Jul 28.
14. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics. 1998 Apr; 13(4):397-409.
15. Department of Health. NHS reference costs 2009-2010, 2011.
16. Basu A, Rathouz PJ. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. Biostatistics. 2005 Jan; 6(1):93-109.
17. Clarke P, Leal J, et al. Estimating the cost of complications of diabetes in Australia using administrative health-care data. Value Health. 2008 Mar-Apr; 11(2):199-206.
18. Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. J Health Econ. 1998 Jun; 17(3):247-281.
19. National Joint Registry for England and Wales 11th Annual report 2014 http://www.njrreports.org.uk/. 2014.
20. Briggs T. Getting it right first time. Improving the quality of orthopaedic care within the National Health Service in England. . 2012.
21. The management of hip fracture in adults. NICE Clinical Guidance CG124. http://www.nice.org.uk/CG124. 2011.
22. Academy of Medical Royal Colleges, Protecting resources, promoting value: a doctor's guide to cutting clinical waste in clinical care. November 2014.
23. Monitor. Closing the NHS funding gap: how to get better value health care for patients. IRREP 22/13. October 2013.

24. Pinedo Villanueva RA, Turner D, et al. Cost-effectiveness of Total Hip Replacements: a comprehensive model. *World Conference on Osteoporosis, Osteoarthritis and Musculoskeletal Diseases*. Seville, Spain 2014.
25. Pinedo-Villanueva R, Turner D, et al. Primary care costs attributable to osteoarthritis before total hip replacement. Osteoarthritis and Cartilage. 2014 April 2014; 22, Supplement(0):S210-S211.
26. Pinedo-Villanueva R, Turner D, et al. Outcomes after total hip replacement. Osteoarthritis and Cartilage. 2014 April 2014; 22, Supplement(0):S214.
27. Pinedo-Villanueva R. Expert elicitation to estimate transition probabilities for a model assessing the cost-effectiveness of an outcome prediction tool for total hip replacements in the UK. *Health Economists Study Group Summer conference*. Warwick, UK 2013.

Appendix One: Data Specification

Aim and Scope

This document aims to provide a description of the data that is proposed for use within Work Package 2a (RO-HES) of Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE). The document provides details of the inclusion/exclusion criteria for patients and the data items required for each patient, along with a brief justification for why these data items are required to address the research question.

Intended Audience

This document is intended for individuals or organizations that are required to take an ethical, legal or technical view in relation to the use of (linked) routine NHS data within the research project.

Data Specification

Work Package 2a (RO-HES) of Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE) will use the following sources of routine NHS data:

1. **Hospital Episode Statistics (HES)[40]:** a database controlled by NHS Digital containing patient data relating to A&E, inpatient and outpatient episodes
2. **ResearchOne[41]:** a database controlled by The Phoenix Partnership containing de-identified patient data from primary care settings that use the SystmOne clinical information system

Patients and data items have been selected that are *necessary* and *sufficient* to robustly answer the research question. No patient identifiable data[42] is included in the data items that are required for each patient from either NHS Digital or The Phoenix Partnership. We describe the criteria for patient selection and data items included for each patient from each of the Hospital Episode Statistics (HES) datasets and ResearchOne below.

Hospital Episode Statistics (A&E)

The data items contained within HES A&E data extracts can be found at: http://content.digital.nhs.uk/media/18619/HES-AE-Data-Dictionary/pdf/DD_AE_v2.pdf. We use the data items described in this document to define the criteria for patient selection and the data items included for each patient.

Patient Selection

A&E episodes will be included where:

- Patient has AT LEAST ONE HES Inpatient Episode where:

    - Arrival Date (ARRIVALDATE) is BETWEEN 1st April 2000 AND 31st March 2015 (Index Period)

    AND

[40] See http://www.digital.nhs.uk/hes
[41] See http://www.researchone.org
[42] See http://content.digital.nhs.uk/media/12053/Obtaining-consent-for-patient-identifiable-and-or-sensitive-data-V2-050613/pdf/obtaining_consent_for_patient_identifiable_and_or_sensitive_data_V2_050613.pdf

- Age on arrival (ARRIVALAGE) >= 18

AND

- Procedure code (PROCODE4) is ANY of:

  - W371,W378,W379,W381,W388,W389,W391,W398,W399, W521 AND (Z843 OR Z761 OR Z756), W531 AND (Z843 OR Z761 OR Z756), W541 AND (Z843 OR Z761 OR Z756), W581 AND (Z843 OR Z761 OR Z756), W931, W938, W939, W941, W948, W949, W951, W958, W959, ***Z843, Z761, Z756*** (Hip Primary[43])

    OR

  - W370,W372,W373, W374, W380, W382, W383, W384, W392, W393, W394 AND (Y032 OR Y037), W395, W462, W472, W482, W522 AND (Z843 OR Z761 OR Z756), W523 AND (Z843 OR Z761 OR Z756), W532 AND (Z843 OR Z761 OR Z756), W533 AND (Z843 OR Z761 OR Z756), W542 AND (Z843 OR Z761 OR Z756), W543 AND (Z843 OR Z761 OR Z756), W544 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W572 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W574 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W582 AND (Z843 OR Z761 OR Z756), W932, W933, W940, W942, W943, W952, W953, W954, ***Y032, Y037, Z843, Z761, Z756*** (Hip Revision[43])

    OR

  - O181, O188, O189, W401, W408, W409, W411, W418, W419, W421, W428, W429, W521 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W531 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W541 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W581 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), ***Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Primary[43])

    OR

  - 0180, O182, 0183, O184, W400, W402, W403, W404, W410, W412, W413, W414, W420, W422, W423, W424 AND (Y032 OR Y037), W425, W522 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W523, W533, W542, W543, W544, W553, W564, W574, W582, W603, W613, W641, W642, ***Y032, Y037, Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Revision[43])

AND

- Patient has not registered a Type 2 objection with NHS Digital to prevent their identifiable data from any health and social care setting being released by NHS Digital[44].

[43] Definitions based on those provided within the "OPCS Codes relevant to procedures recorded on the NJR" document published by the National Joint Registry (NJR) – See http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Data%20collection%20forms/OPCS%20Procedure%20codes%20relevant%20to%20NJRv4.pdf

[44] See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469290/Data-Provision-Notice_Patient_Objections_Management_19.10.15.pdf

Data Items

For each episode identified using the criteria above, a number of data items will be obtained. We describe each of these data items below and provide a justification for use of the data items in the analysis.

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Attendance	Arrival date	The arrival date of a patient in the A&E department.	Any	Required to characterise the timing within and between episodes/spells and determine *when* patients present for present for hip/knee replacement/revision and related episodes.
	Duration to Conclusion	The time (expressed as a whole number of minutes) between the patient's arrival and conclusion of their attendance or treatment (whichever is later).	Any	
	Attendance category	An indication of whether a patient is making an initial or follow-up attendance within a particular A&E Department.	Any	
	Department type	A classification of A&E department type according to the activity carried out.	Any	Required to characterise the service in which the episode took place and determine *how* patients present for hip/knee replacement/revision and related episodes.
Clinical Diagnoses	A&E Diagnosis	The A&E diagnosis code recorded for an A&E attendance.	05 (Dislocation/fracture/joint injury/amputation)	Required to characterise the episode by clinical factors such that episodes relating to hip/knee replacement/revision can be included. Episodes will be included that match the inclusion criteria on BOTH *A&E Diagnosis* AND *A&E diagnosis - anatomical area*
	A&E diagnosis - anatomical area	The A&E diagnosis anatomical area (a classification of parts of the human body)	28 (Hip), 29 (Groin), 30 (Thigh), 31 (Knee), 32 (Lower Leg)	
	A&E diagnosis - anatomical side	The A&E diagnosis anatomical side (an indication of the side of the human body).	Any	
	Diagnosis scheme in use	The Coding Scheme basis of the Diagnosis.	Any	Required to enable potential consideration of variation in recording based on the specific diagnosis scheme used by a provider.

Clinical Investigations	A&E Investigation	The A&E investigation recorded for an A&E attendance.	Any	Required to characterise the episode by clinical factors and determine *how* patients present for revision surgery.
	Number of investigations	Number of investigations	Any	
Clinical Treatments	Number of treatments	Number of treatments	Any	Required to characterise the episode by clinical factors and determine *how* patients present for revision surgery.
	A&E treatment	The A&E treatment recorded for an A&E attendance. The CDS allows an unlimited number of treatments to be submitted, however, only the first 12 treatments are available within HES.	Any	
Geographical	IMD Decile group	This field uses the IMD Overall Ranking to identify which one of ten groups a Super Output Area belongs to, from most deprived through to least deprived.	Any	Required to characterise patients by socio-economic factors and determine *which* patients present for hip/knee replacement/revision and related episodes.
	IMD Overall Rank	The Index of Multiple Deprivation (IMD) overall ranking is made by combining the seven IMD Domain scores.	Any	
Organisation Data	3-digit Provider Code	A provider code is a unique code that identifies an organisation acting as a health care provider.	Any	Required to enable analysis of follow-up for hip/knee revision/replacement within and between specific providers.
	5-digit Provider Code	A provider code is a unique code that identifies an organisation acting as a health care provider.	Any	
	Provider Type	Healthcare provider type.	Any	
Patient Data	Age on arrival	This field contains the age in whole years on arrival, calculated from arrival date and DOB.	Any	Required to characterise patients by demographic factors and determine *which* patients present for hip/knee replacement/revision and related episodes.
	Carer support indicator	This field contains a code which states whether carer support is available to the patient at home or other normal residence. This does not include any paid support or support from a voluntary organisation unless the patient is normally resident in a nursing home, group home or residential care home.	Any	

	Ethnic Category	This field contains a code which specifies some ethnic groups and some nationalities.	Any	
	Date of Birth - month and year	Month and year of date of birth only. Day is not made available	Any	
	Postcode district	Contains the outward portion of the patient's postcode (ie all characters to the left of the space).	Any	
	Sex of patient	This field contains a code which defines the sex of the patient.	Any	
	NHS Number Valid Flag	This field indicates whether the NHS Number supplied is valid or not.	Any	Required to determine whether an invalid NHS number may have affected ability to link to data in ResearchOne.
	Postcode Found	Field confirms if postcode is valid.	Any	Required to determine whether the postcode was validated and therefore whether the postcode district can be considered valid.
	Encrypted HES ID	This field contains a unique identifier for each individual patient.	Any	Required to link to other episodes for the same patient within HES (A&E), HES (Inpatient), and HES (Outpatient), and to data in ResearchOne (via mapping file form NHSD).
System Data	Record Identifier	This is a record identifier that is created by the HES system.	Any	Required by NHS Digital to be included in the supplied data.

Hospital Episode Statistics (Inpatient)

The data items contained within HES Inpatient data extracts can be found at: http://content.digital.nhs.uk/media/18618/HES-Admitted-Patient-Data-Dictionary/pdf/DD_APC_v2.pdf. We use the data items described in this document to define the criteria for patient selection and the data items included for each patient.

Patient Selection

Inpatient episodes will be included where:

- Patient has AT LEAST ONE HES Inpatient Episode where:

    - Arrival Date (ARRIVALDATE) is BETWEEN 1st April 2000 AND 31st March 2015 (Index Period)

    AND

    - Age on arrival (ARRIVALAGE) >= 18

    AND

    - Procedure code (PROCODE4) is ANY of:

        - W371,W378,W379,W381,W388,W389,W391,W398,W399, W521 AND (Z843 OR Z761 OR Z756), W531 AND (Z843 OR Z761 OR Z756), W541 AND (Z843 OR Z761 OR Z756), W581 AND (Z843 OR Z761 OR Z756), W931, W938, W939, W941, W948, W949, W951, W958, W959, ***Z843, Z761, Z756*** (Hip Primary[45])

          OR

        - W370,W372,W373, W374, W380, W382, W383, W384, W392, W393, W394 AND (Y032 OR Y037), W395, W462, W472, W482, W522 AND (Z843 OR Z761 OR Z756), W523 AND (Z843 OR Z761 OR Z756), W532 AND (Z843 OR Z761 OR Z756), W533 AND (Z843 OR Z761 OR Z756), W542 AND (Z843 OR Z761 OR Z756), W543 AND (Z843 OR Z761 OR Z756), W544 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W572 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W574 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W582 AND (Z843 OR Z761 OR Z756), W932, W933, W940, W942, W943, W952, W953, W954, ***Y032, Y037, Z843, Z761, Z756*** (Hip Revision[43])

          OR

        - O181, O188, O189, W401, W408, W409, W411, W418, W419, W421, W428, W429, W521 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W531 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W541 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W581 (Z846 OR Z765 OR

[45] Definitions based on those provided within the "OPCS Codes relevant to procedures recorded on the NJR" document published by the National Joint Registry (NJR) – see http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Data%20collection%20forms/OPCS%20Procedure%20codes%20relevant%20to%20NJRv4.pdf

Z845 OR Z844 OR Z774 OR Z787), ***Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Primary[43])

OR

- 0180, O182, 0183, O184, W400, W402, W403, W404, W410, W412, W413, W414, W420, W422, W423, W424 AND (Y032 OR Y037), W425, W522 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W523, W533, W542, W543, W544, W553, W564, W574, W582, W603, W613, W641, W642, ***Y032, Y037, Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Revision[43])

AND

- Patient has not registered a Type 2 objection with NHS Digital to prevent their identifiable data from any health and social care setting being released by NHS Digital[46].

[46] See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469290/Data-Provision-Notice_Patient_Objections_Management_19.10.15.pdf

Data Items

For each episode identified using the criteria above, a number of data items will be obtained. We describe each of these data items below and provide a justification for use of the data items in the analysis.

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Admissions; Period of Care	Date of admission	This field contains the date the patient was admitted to hospital at the start of a hospital spell.	Any	Required to characterise the timing within and between episodes/spells and determine *when* patients present for present for hip/knee replacement/revision and related episodes.
	Method of admission	This field contains a code which identifies how the patient was admitted to hospital.	Any	
	Date of decision to admit	This field contains the date on which a consultant, or another member of the clinical staff, decided to admit the patient to a hospital.	Any	
	Waiting time	This field contains the difference in days between the date on which it was decided to admit the patient (elecdate) and the actual admission date (admidate).	Any	
	Calculation of Elecdur	This field returns the elecdur but excludes admissions from emergency, so only includes eledur where the method of admission (admimeth) is 11 or 12	Any	
	First regular day or night admission	This field indicates whether the episode falls within a sequence of regular day and night admissions and, if so, whether it is the first or subsequent episode within the sequence.	Any	
	Source of admission	This field contains a code which identifies where the patient was immediately prior to admission.	Any	Required to characterise from where the patient was admitted and to determine *how* and *who* presents for revision surgery.

Clinical	Cause code	External cause of injury or poisoning.	Any	Required to characterise the episode by clinical factors and determine *how* patients present for revision surgery.
	All Diagnosis codes	There are twenty fields (fourteen before April 2007 and seven before April 2002), diag_01 to diag_20, which contain information about a patient's illness or condition.	Any	
	Operation status code	Status of operation.	Any	
	All Operative procedure codes	There are twenty-four fields (twelve before April 2007 and four prior to April 2002), oper_01 to oper_24, which contain information about a patient's operations. The field oper_01 contains the main (ie most resource intensive) procedure. The other fields contain secondary procedures.	Any	
	Date of operation	This field contains the dates for operations recorded in the operation codes (opertn_nn) field.	Any	Required to characterise the timing within and between episodes/spells and determine *when* patients present for present for hip/knee replacement/revision and related episodes.
	Pre-operative duration	This derived field contains the difference in days between the date the episode started (epistart) and the date of the main operation (opdte_01).	Any	
	Post-operative duration	This derived field contains the difference in days between the date of the main operation (opdte_01) and the date the episode ended (epiend).	Any	
Clinical; Period of Care	Main specialty	This field contains a code that defines the specialty under which the consultant is contracted. It can be compared with tretspef, the specialty under which the consultant worked.	Any	Required to characterise the episode by clinical factors such that episodes that are likely to relate to hip/knee replacement/revision can be included.
	Treatment specialty	This field contains a code that defines the specialty in which the consultant was working during the period of care. It can be compared with mainspef, the	110 (Trauma and orthopaedics) OR 410 (Rheumatology)	

		specialty under which the consultant is contracted.	**NOTE:** It is assumed that any episode that fulfils the patient selection criteria would be included using this inclusion criteria, and therefore the index episode(s) would be included for each patient. For example, a hip replacement procedure would always be classified under either '*Trauma and orthopaedics'* OR '*Rheumatology'*.	
Discharges; Period of Care	Destination on discharge	This field contains a code which identifies where the patient was due to go on leaving hospital.	Any	Required to characterise the episode by the discharge and determine *which* and *how* patients present for revision surgery.
	Method of discharge	This field contains a code which defines the circumstances under which a patient left hospital.	Any	
	Date of discharge	This field contains the date on which the patient was discharged from hospital. ; It is only present in the record for the last episode of a spell.	Any	Required to characterise the timing within and between episodes/spells and determine *when* patients present for present for hip/knee replacement/revision and related episodes.
	Discharge ready date	The date that a patient was medically ready for discharge from a hospital bed, but couldn't be discharged, therefore qualifying for delayed discharge payments	Any	
Episodes and Spells; Period of Care	Episode duration	This field contains the difference in days between the episode start date (epistart) and the episode end date (epiend).	Any	Required to characterise the timing within and between episodes/spells and determine *when* patients present for present for hip/knee replacement/revision and related episodes.
	Date episode ended	This field contains the date on which a patient left the care of a particular consultant, for one of the following reasons: Discharged from hospital (includes transfers) or moved to the care of another consultant.	Any	
	Episode order	This field contains the number of the episode within the current spell. All	Any	

		spells start with an episode where epiorder is 01.		
	Date episode started	This field contains the date on which a patient was under the care of a particular Consultant.	Any	
	Episode status	This field tells you whether the episode had finished before the end of the HES datayear (ie whether the episode was still 'live' at midnight on 31 March).	Any	
	Beginning of spell	This derived field contains a code that defines whether the episode is the first of a spell and whether the spell started in the current or previous year. Other maternity events are excluded.	Any	
	Duration of spell	This derived field contains the difference in days between the admission date (admidate) and the discharge date (epiend) provided the discharge method (dismeth) confirms that the spell has finished.	Any	
	End of spell	This field contains a code which defines whether the episode is the last of a spell. It is set for finished episodes (episode status - epistat - is 3) for general, delivery or birth episodes (episode type - epitype - is 1, 2 or 3) provided the discharge method (dismeth) confirms that the spell has finished.	Any	
	Ward type at start of episode	This field contains a code that defines the characteristics of a ward. The code has six parts: AABCDEF.	Any	Required to characterise services and enable analysis of follow-up for hip/knee revision/replacement within and between specific providers.
Organisation	Provider code - 3 character	A provider code is a unique code that identifies an organisation acting as a health care provider. The code is managed by the National Administrative Codes Service	Any	Required to enable analysis of follow-up for hip/knee revision/replacement within and between specific providers.

		(NACS) and supports the identification of organisations exchanging information within the NHS.		
	Provider code - 5 character	A provider code is a unique code that identifies an organisation acting as a health care provider. The code is managed by the National Administrative Codes Service (NACS) and supports the identification of organisations exchanging information within the NHS.	Any	
	Provider type	Healthcare provider type	Any	
Patient Pathway	Duration of elective wait	The number of days that a patient waited from the date when a decision was taken for treatment to when they received the treatment.	Any	Required to characterise the timing within and between episodes/spells and determine *when* patients present for present for hip/knee replacement/revision and related episodes.
	Method of Admission - Waiting List	Calculation determining patients whose method of admission was from the waiting list	Any	Required to determine whether admission was from a waiting list, to determine *which* and *how* patients present for revision surgery, and to analysis variation within and between providers.
Patient Data	Age on admission	A patient's age, in years, at the date of admission.	Any	Required to characterise patients by demographic factors and determine *which* patients present for hip/knee replacement/revision and related episodes.
	Ethnic category	This field contains a code that specifies some ethnic groups and some nationalities. It was introduced from the 1995-96 data year	Any	
	Date of Birth - month and year	Month and year of date of birth only. Day is not made available.	Any	

	Postcode district of patient's residence	Contains the outward portion of the patient's postcode (ie all characters to the left of the space).	Any	
	Sex of patient	This field contains a code which defines the sex of the patient.	Any	
	NHS Number valid flag	This field indicates whether the NHS Number supplied is valid or not.	Any	Required to determine whether NHS number can be considered to be valid.
	Postcode Found	Field confirms if postcode is valid.	Any	Required to determine whether post code can be considered to be valid.
	Encrypted HES ID	This field uniquely identifies a patient across all data years.	Any	Required to link to other episodes for the same patient within HES (A&E), HES (Inpatient), and HES (Outpatient), and to data in ResearchOne (via mapping file form NHSD).
Socio-Economic	IMD Decile group	This field uses the IMD Overall Ranking to identify which one of ten groups a Super Output Area belongs to, from most deprived through to least deprived.	Any	Required to characterise patients by socio-economic factors and determine *which* patients present for hip/knee replacement/revision and related episodes.
	IMD Overall Rank	The Index of Multiple Deprivation (IMD) overall ranking is made by combining the seven IMD Domain scores.	Any	
System Data	Record Identifier	This is a record identifier that is created by the HES system.	Any	Required by NHS Digital to be included in the supplied data.

Hospital Episode Statistics (Outpatient)

The data items contained within HES Outpatient data extracts can be found at: http://content.digital.nhs.uk/media/1359/HES-Hospital-Episode-Statistics-HES-Outpatient-Data-Dictionary/pdf/HES_Outpatient_DD_Dec10.pdf. We use the data items described in this document to define the criteria for patient selection and the data items included for each patient.

Patient Selection

Outpatient episodes will be included where:

- Patient has AT LEAST ONE HES Inpatient Episode where:

    - Arrival Date (ARRIVALDATE) is BETWEEN 1st April 2000 AND 31st March 2015 (Index Period)

    AND

    - Age on arrival (ARRIVALAGE) >= 18

    AND

    - Procedure code (PROCODE4) is ANY of:

        - W371,W378,W379,W381,W388,W389,W391,W398,W399, W521 AND (Z843 OR Z761 OR Z756), W531 AND (Z843 OR Z761 OR Z756), W541 AND (Z843 OR Z761 OR Z756), W581 AND (Z843 OR Z761 OR Z756), W931, W938, W939, W941, W948, W949, W951, W958, W959, ***Z843, Z761, Z756*** (Hip Primary[47])

          OR

        - W370,W372,W373, W374, W380, W382, W383, W384, W392, W393, W394 AND (Y032 OR Y037), W395, W462, W472, W482, W522 AND (Z843 OR Z761 OR Z756), W523 AND (Z843 OR Z761 OR Z756), W532 AND (Z843 OR Z761 OR Z756), W533 AND (Z843 OR Z761 OR Z756), W542 AND (Z843 OR Z761 OR Z756), W543 AND (Z843 OR Z761 OR Z756), W544 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W572 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W574 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W582 AND (Z843 OR Z761 OR Z756), W932, W933, W940, W942, W943, W952, W953, W954, ***Y032, Y037, Z843, Z761, Z756*** (Hip Revision[43])

          OR

        - O181, O188, O189, W401, W408, W409, W411, W418, W419, W421, W428, W429, W521 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W531 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W541 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W581 (Z846 OR Z765 OR

[47] Definitions based on those provided within the "OPCS Codes relevant to procedures recorded on the NJR" document published by the National Joint Registry (NJR) – see http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Data%20collection%20forms/OPCS%20Procedure%20codes%20relevant%20to%20NJRv4.pdf

Z845 OR Z844 OR Z774 OR Z787), ***Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Primary[43])

OR

- 0180, O182, 0183, O184, W400, W402, W403, W404, W410, W412, W413, W414, W420, W422, W423, W424 AND (Y032 OR Y037), W425, W522 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W523, W533, W542, W543, W544, W553, W564, W574, W582, W603, W613, W641, W642, ***Y032, Y037, Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Revision[43])

AND

- Patient has not registered a Type 2 objection with NHS Digital to prevent their identifiable data from any health and social care setting being released by NHS Digital[48].

[48] See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469290/Data-Provision-Notice_Patient_Objections_Management_19.10.15.pdf

Data Items

For each episode identified using the criteria above, a number of data items will be obtained. We describe each of these data items below and provide a justification for use of the data items in the analysis.

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Appointments	Appointment date	The date when an appointment was scheduled.	Any	Required to characterise the timing within and between episodes/spells and determine *when* patients present for present for hip/knee replacement/revision and related episodes.
	Last DNA or patient cancelled date	This is recorded when patients who have been offered an appointment date have missed this date with or without advance notice.	Any	
	Days waiting	'Waiting' gives the period in days between the date of the appointment date and either the referral request received date (reqdate) or the DNA (did not attend) date, if given.	Any	
	Referral request received date	This field records the date the referral request was received by the healthcare provider.	Any	
	Waiting calculation indicator	Waitind indicates how and whether waiting time has been calculated.	Any	
	Attendance type	A field derived from 'first appointment' (firstatt) and 'attended or did not attend' (attended), used to identify if the attendance occurred and whether it was the first or subsequent.	Any	Required to characterise the attendance and to determine *how* and *who* presents for revision surgery.
	Attended or did not attend	This indicates whether or not a patient attended for an appointment. If the patient did not attend it also indicates whether or not advanced warning was given.	Any	
	First attendance	Gives an indication of whether a patient is making a first attendance or follow-up attendance, and whether the	Any	

		consultation was face-to-face or via telephone/telemedicine consultation.		
	Medical staff type seeing patient	Gives information about the type of care professional staff dealing with the patient during a consultant outpatient attendance, or nurse or midwife contact.	Any	
	Outcome of attendance	This records the outcome of an outpatient attendance.	Any	
	Priority type	This is the priority of a request for services in the case of services to be provided by a consultant, it is as assessed by or on behalf of the consultant.	Any	
	Service type Requested	Describes the terms of reference for the referral request.	Any	
	Source of referral	A classification which is used to identify the source of referral of each consultant outpatient episode.	Any	
Clinical	Diagnosis	There are twelve fields (two before April 2007), diag_01 to diag_12, which contain information about a patient's illness or condition.	Any	Required to characterise the episode by clinical factors such that episodes that are likely to relate to hip/knee replacement/revision can be included and to determine *how* patients present for revision surgery.
	Operative Procedure	There are twenty-four fields (twelve before April 2007), oper_01 to oper_24, which contain information about a patient's operations.	Any	
	Main specialty	A code that defines the specialty under which the consultant is contracted. Compare with 'treatment specialty' (tretspef), the specialty under which the consultant worked	Any	
	Operation status code	Status of operation.	Any	
	Treatment specialty	This field contains a code that defines the specialty in which the consultant was working during the period of care. It can be compared with mainspef, the	110 (Trauma and orthopaedics) OR 410 (Rheumatology)	

		specialty under which the consultant is contracted.		
Organisation	Provider code (3 character)	A provider code is a unique code that identifies an organisation acting as a health care provider.	Any	Required to enable analysis of follow-up for hip/knee revision/replacement within and between specific providers.
	Provider code (5 character)	A provider code is a unique code that identifies an organisation acting as a health care provider.	Any	
	Provider type	Healthcare provider type.	Any	
Patient	Age on day of appointment	This derived field, calculated from appointment date (apptdate) and date of birth (dob), contains the patient's age in whole years.	Any	Required to characterise patients by demographic factors and determine *which* patients present for hip/knee replacement/revision and related episodes.
	Ethnic category	This field contains a code that specifies some ethnic groups and some nationalities.	Any	
	Date of birth – month and year	Month and year of date of birth only. Day is not made available.	Any	
	Postcode district of patient's residence	Contains the outward portion of the patient's postcode (ie all characters to the left of the space).	Any	
	Sex of patient	This field contains a code which defines the sex of the patient.	Any	
	NHS Number valid flag	This field indicates whether the NHS Number supplied is valid or not.	Any	Required to determine whether NHS number can be considered to be valid.
	Postcode Found	Field confirms if postcode is valid.	Any	Required to determine whether post code can be considered to be valid.
	Encrypted HES ID	This field uniquely identifies a patient across all data years. It is generated by matching records for the same patient using a combination of NHS Number, local patient identifier, postcode, sex and date of birth.	Any	Required to link to other episodes for the same patient within HES (A&E), HES (Inpatient), and HES (Outpatient), and to data in ResearchOne (via mapping file form NHSD).
Socio-Economic	IMD Decile Group	This field uses the IMD Overall Ranking to identify which one of ten groups a Super Output Area belongs	Any	Required to characterise patients by socio-economic factors and determine *which* patients present for hip/knee

		to, from most deprived through to least deprived.		replacement/revision and related episodes.
	IMD Overall Ranking	The IMD overall ranking is made by combining the seven IMD Domain scores.	Any	
System Data	Record Identifier	This is a record identifier that is created by the HES system.	Any	Required by NHS Digital to be included in the supplied data.

ResearchOne

ResearchOne contains data items relating to primary care, including diagnoses, prescriptions and referrals. We define the criteria for patient selection based upon inclusion in the HES data extracts, and define the criteria for the data items included for each patient based on specific diagnoses and prescriptions that are required for analysis.

Patient Selection

Data will be included for all patients for whom:

- Patient has AT LEAST ONE HES Inpatient Episode where:

    - Arrival Date (ARRIVALDATE) is BETWEEN 1st April 2000 AND 31st March 2015 (Index Period)

    AND

    - Age on arrival (ARRIVALAGE) >= 18

    AND

    - Procedure code (PROCODE4) is ANY of:

        - W371,W378,W379,W381,W388,W389,W391,W398,W399, W521 AND (Z843 OR Z761 OR Z756), W531 AND (Z843 OR Z761 OR Z756), W541 AND (Z843 OR Z761 OR Z756), W581 AND (Z843 OR Z761 OR Z756), W931, W938, W939, W941, W948, W949, W951, W958, W959, ***Z843, Z761, Z756*** (Hip Primary[49])

          OR

        - W370,W372,W373, W374, W380, W382, W383, W384, W392, W393, W394 AND (Y032 OR Y037), W395, W462, W472, W482, W522 AND (Z843 OR Z761 OR Z756), W523 AND (Z843 OR Z761 OR Z756), W532 AND (Z843 OR Z761 OR Z756), W533 AND (Z843 OR Z761 OR Z756), W542 AND (Z843 OR Z761 OR Z756), W543 AND (Z843 OR Z761 OR Z756), W544 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W572 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W574 AND ((Y032 OR Y037) AND (Z843 OR Z761 OR Z756)), W582 AND (Z843 OR Z761 OR Z756), W932, W933, W940, W942, W943, W952, W953, W954, ***Y032, Y037, Z843, Z761, Z756*** (Hip Revision[43])

          OR

        - O181, O188, O189, W401, W408, W409, W411, W418, W419, W421, W428, W429, W521 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W531 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W541 (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W581 (Z846 OR Z765 OR

[49] Definitions based on those provided within the "OPCS Codes relevant to procedures recorded on the NJR" document published by the National Joint Registry (NJR) – see http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Data%20collection%20forms/OPCS%20Procedure%20codes%20relevant%20to%20NJRv4.pdf

Z845 OR Z844 OR Z774 OR Z787), ***Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Primary[43])

OR

- 0180, O182, 0183, O184, W400, W402, W403, W404, W410, W412, W413, W414, W420, W422, W423, W424 AND (Y032 OR Y037), W425, W522 AND (Z846 OR Z765 OR Z845 OR Z844 OR Z774 OR Z787), W523, W533, W542, W543, W544, W553, W564, W574, W582, W603, W613, W641, W642, ***Y032, Y037, Z846, Z765, Z845, Z844, Z774, Z787*** (Knee Revision[43])

AND

- Patient has not registered a Type 2 objection with NHS Digital to prevent their identifiable data from any health and social care setting being released by NHS Digital[50].

AND

- Patient is registered to a General Practice that has opted into ResearchOne and has not individually opted out of ResearchOne.

[50] See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469290/Data-Provision-Notice_Patient_Objections_Management_19.10.15.pdf

Data Items

For each patient identified using the criteria above, a number of data items will be obtained. These data items will be obtained for the period prior to the 1st April 2015.

We describe each of these data items below and provide a justification for use of the data items in the analysis.

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Patient	Patient Identifier	Unique ResearchOne Identifier for Patient	Associated with a pseudonym that is contained in a list provided by NHSD.	Required to link to other data for the same patient within ResearchOne and to data in HES (A&E), HES (Inpatient), and HES (Outpatient) (via mapping file form NHSD).
	Date Of Birth (mm/yyyy)	Month and year of patient's birth.	Any	Required to characterise patients by demographic factors over time and determine *which* patients present for hip/knee replacement/revisions. Potential for cross-validation with IMD contained in related HES episodes.
	Date Of Death (mm/yyyy)	Month and year of patient's death.	Any	
	Gender	Gender of patient	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Address	Patient Identifier	Unique ResearchOne Identifier for Patient	Associated with a pseudonym that is contained in a list provided by NHSD.	Required to link to other data for the same patient within ResearchOne and to data in HES (A&E), HES (Inpatient), and HES (Outpatient) (via mapping file form NHSD).
	Start Date	Month and year at which patient started residence at address.	Any	Required to characterise patients by socio-economic factors over time and determine *which* patients present for hip/knee replacement/revisions. Potential for cross-validation with IMD contained in related HES episodes.
	End Date	Month and year at which patient ended residence at address.	Any	
	Type Of Address	Specific type of address, e.g. private residential or community establishment.	Any	
	IMD Rank	Index of multiple deprivation (IMD) rank calculated from patient address.	Any	

	IMD Score	Index of multiple deprivation (IMD) score calculated from patient address.	Any	
	Sector Level Postcode	5 digit sector level postcode for patient address.	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Coded Events	Patient Identifier	Unique ResearchOne Identifier for Patient	Associated with a pseudonym that is contained in a list provided by NHSD.	Required to link to other data for the same patient within ResearchOne and to data in HES (A&E), HES (Inpatient), and HES (Outpatient) (via mapping file form NHSD).
	EventId	Reference to event in which referral occurred.	Any	Required to characterise patients by clinical factors (diagnoses, symptoms and procedures) over time and determine *which* patients present for hip/knee replacement/revisions. Potential for cross-validation of hip/knee revision/replacement codes with procedure codes contains in related HES episodes.
	EventDate	Date of event	Any	
	CTV3 Concept Id	CTV3 code for diagnosis/observation.	Matches a code included in the QOF-related definitions[51] for: • Asthma • Atrial Fibrillation • Blood Pressure • Cancer • CHD • CVD • CKD • COPD • Dementia • Depression • Diabetes • Epilepsy • HF • Hypertension • Learning Disability • Obesity	

[51] See http://www.hscic.gov.uk/qof

			• Osteoporosis • Psychosis, Schizophrenia or Bipolar-Affective Disorder • PAD • Palliative Care • Rheumatoid Arthritis • Smoking • Stroke • Stroke (TIA) OR Matches a code included in expert clinician definitions for: • Joint Pain • Hip Replacement • Hip Revision • Knee Replacement • Knee Revision OR Matches a code included in expert clinician definitions for a relevant referral to: • Trauma/Orthopaedics • Rheumatology	
	NumberValue	Number value associated with CTV3 code	Any	
	LowerBound	Lower bound associated with CTV3 code	Any	
	UpperBound	Upper bound associated with CTV3 code	Any	
	NumericRangeComparisonMethod	Operator to use for numeric range comparisons	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification

Drug	Patient Identifier	Unique ResearchOne Identifier for Patient	Associated with a pseudonym that is contained in a list provided by NHSD.	Required to link to other data for the same patient within ResearchOne and to data in HES (A&E), HES (Inpatient), and HES (Outpatient) (via mapping file form NHSD).
	MedicationId	Unique reference to specific prescription	Any	Required to characterise patients by clinical factors (prescriptions) over time and determine *which* patients present for hip/knee replacement/revisions.
	EventId	Reference to event in which prescription occurred.	Any	
	EventDate	Date of event in which prescription occurred.	Any	
	StartDate	Start date of prescription	Any	
	EndDate	End date of prescription	Any	
	DrugNumber	Unique reference for prescribed drug.	Any	
	Dose	Dosage of drug	Any	
	Quantity	Quantity of drug	Any	
	DrugStatus		Any	
	RepeatMedicationId	Reference to repeat medication (if applicable)	Any	
	MedicationEnded	Indicator of whether medication has ended.	Any	
	DateMedicationEnded	Date on which medication was ended.	Any	
	ReasonMedicationEnded	Reason medication was ended.	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Events	Event Id	Reference to event in which referral occurred.	Any	Required to characterise service interaction of patients with GP and the method by which details of care were recorded, and determine *which* and *how*
	EventDate	Date of event	Any	

	Method	Method of interaction associated with event, e.g. 'Face to Face' appointment.	Any	patients present for hip/knee replacement/revisions.

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Organisation	Practice Number	Unique ResearchOne Identifier for GP practice		Required to assign patients to GP practices at a point in time and to consider practice-level variations in *when*, *which* and *how* patients present for hip/knee replacement/revisions.
	List Size	List size at 1st April 2015	Any	Required to characterise GP practices for consideration of practice-level variation.
	No of Partners	Number of partners at 1st April 2015	Any	
	Description	Type of organization, e.g. General Practice.	Any	
	LiveOnSystmOne Since	Date from which the GP practice was on SystmOne.	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Referral	Patient Identifier	Unique ResearchOne Identifier for Patient	Associated with a pseudonym that is contained in a list provided by NHSD.	Required to link to other data for the same patient within ResearchOne and to data in HES (A&E), HES (Inpatient), and HES (Outpatient) (via mapping file form NHSD).
	Event Id	Reference to event in which referral occurred.	Any	Required to characterise patients by clinical factors (referrals) over time and determine *when*, *which and how* patients present for hip/knee replacement/revisions.
	Date Of Referral	Date on which referral occurred.	Any	
	Type Of Referral	Type of referral	Any	
	Urgency Of Referral	Urgency of referral	Any	
	Reason For Referral	Clinical reason for referral	Any	
	Referral To	Destination for referral	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Registration	Patient Identifier	Unique ResearchOne Identifier for Patient	Associated with a pseudonym that is contained in a list provided by NHSD.	Required to link to other data for the same patient within ResearchOne and to data in HES (A&E), HES (Inpatient), and HES (Outpatient) (via mapping file form NHSD).
	Practice Number	Pseudonym for the GP practice.	Any	Required to characterise patients by their GP practice over time and consider practice-level effects in *when*, *which* and *how* patients present for hip/knee replacement/revisions.
	Start Date	Date at which patient registration at practice started.	Any	
	End Date	Date at which patient registration at practice ended.	Any	
	Applied GMS	Whether registration was Applied GMS.	Any	
	Fully GMS	Whether registration was Fully GMS.	Any	
	Temporary Resident Short	Whether registration was Temporary Resident Short.	Any	
	Temporary Resident Long	Whether registration was Temporary Resident Long.	Any	
	Temporary Resident Telephone	Whether registration was Temporary Resident Telephone.	Any	
	Immediate Necessary Treatment	Whether registration was Immediate Necessary Treatment.	Any	
	Emergency	Whether registration was Emergency.	Any	
	Child Health	Whether registration was Child Health.	Any	
	Contraception	Whether registration was Contraception.	Any	
	Maternity	Whether registration was Maternity.	Any	
	Minor Surgery	Whether registration was Minor Surgery.	Any	
	Private	Whether registration was Private.	Any	
	Other	Whether registration was Other.	Any	
	WalkIn	Whether registration was Walk-In.	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Repeat Drug	Patient Identifier	Unique ResearchOne Identifier for Patient	Associated with a pseudonym that is contained in a list provided by NHSD.	Required to link to other data for the same patient within ResearchOne and to data in HES (A&E), HES (Inpatient), and HES (Outpatient) (via mapping file form NHSD).
	Repeat Medication Id	Unique reference for repeat prescription	Any	Required to characterise patients by clinical factors (repeat prescriptions) over time and determine *which* patients present for hip/knee replacement/revisions.
	Event Id	Reference to event in which referral occurred.	Any	
	Event Date	Date on which event occurred	Any	
	Start Date	Start date of repeat prescription.	Any	
	End Date	End date of repeat prescription.	Any	
	Drug Number	Reference to drug prescribed.	Any	
	Dose	Dosage of drug prescribed	Any	
	Quantity	Quantity of drug prescribed	Any	
	Drug Status		Any	
	Maximum Issues Authorised	Maximum number of issues authorized.	Any	
	Medication Review Date	Date of review for medication	Any	
	Medication Ended	Indicator for whether medication ended	Any	
	Date Medication Ended	Date on which medication was ended	Any	
	Reason Medication Ended	Reason for which medication was ended	Any	

Data Category	Data Item	Description	Inclusion/Exclusion Criteria	Justification
Drug Reference	Drug Number	Unique drug number	Any	Required to distinguish between different drugs and forms of drug that are prescribed to patients.
	FullName	Full name of drug	Any	

	Pack	Type of pack for drug	Any	
	Form	Delivery form of drug	Any	
	Strength	Strength of drug	Any	
	BNFChapter	BNF chapter for drug.	Matching expert clinician defined BNF Chapters: • Opioid Analgesics (4.7.2) • Non Steroidal Anti-Inflammatory Drugs (10.1.1) • Non Opioid Analgesics and Compound Analgesic Preparations (4.7.1, 4.7.1.1, 4.7.1.2, 4.7.1.3, 4.7.1.4) • Rubefacients, Topical NSAIDS, Capsaicin and Poultices (10.3.2, 10.3.2.1, 10.3.2.2, 10.3.2.3, 10.3.2.4) • Drugs That Suppress The Rheumatic Disease Process (10.1.3, 10.1.3.1, 10.1.3.2, 10.1.3.3, 10.1.3.4, 10.1.3.5, 10.1.3.6) • Corticosteroids (10.1.2.1, 10.1.2.2)	

Appendix Two: Data Linkage Methodology

Aim and Scope

This document aims to provide a description of the data linkage methodology that is proposed for use within Work Package 2a (RO-HES) of Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE). The document provides a background to aspects of the methodology proposed, details of the infrastructure on which the methodology will be enacted, and details the specific data flows between organisations required by the methodology.

Intended Audience

This document is intended for individuals or organizations that are required to take an ethical, legal or technical view in relation to the use of (linked) routine NHS data within the research project.

Background

Pseudonymisation

The proposed linkage methodology will use patient pseudonyms to enable the clinical data from NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne) to be linked for individual patients without requiring patient identifiable data to be transmitted outside of these organizations. The process of pseudonymization is performed by NHS Digital and The Phoenix Partnership prior to transmission of clinical data outside their organization.

The use of patient pseudonyms for linkage is motivated by two requirements:

1. Individual patients must be uniquely and consistently identifiable across the clinical data provided by the different organizations.
2. No patient identifiable data[52] must flow out of the different organizations that are providing clinical data.

Patient identifiable is held for each patient by the organizations that will provide clinical data. The specific data items (e.g. date of birth and surname) and the quality of these data items (e.g. accuracy and completeness) will vary across organizations. In order for a patient to be uniquely and consistently identifiable across the clinical data of different organizations, the data items used must be captured by each organization with high quality.

NHS number aims to provide a unique and consistent reference for patients across the health system. National initiatives have emphasized the importance of the accurate and complete recording of NHS number within clinical data (in 2007-2008, 97% of inpatient activity was associated with a valid NHS number[53]). Consequently, the organizations involved in the methodology are likely to hold a valid NHS number for the majority of patients that are referenced in their clinical data. However, the explicit use of data items such as NHS number, surname and date of birth to uniquely

[52] See http://content.digital.nhs.uk/media/12053/Obtaining-consent-for-patient-identifiable-and-or-sensitive-data-V2-050613/pdf/obtaining_consent_for_patient_identifiable_and_or_sensitive_data_V2_050613.pdf
[53] See http://content.digital.nhs.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HES-patient-ID/pdf/HESID_Methodology.pdf

and consistently identify patients across clinical data would require such data to flow out of the different organizations. This would therefore violate the second requirement.

We propose the use of pseudonyms to enable individual patients to be uniquely and consistently identifiable across the clinical data provided by the different organizations, whilst ensuring that no patient identifiable data[54] is required to flow out of these organizations. Whilst NHS number is considered to be patient identifiable data and therefore cannot be transmitted outside of the organizations without violating our second requirement, NHS number can be used as an input to a pseudonymization process. This pseudonymization process would generate a pseudonym from the NHS number which is itself unique and consistent (given an identical NHS number in identical format), but from which the recovery of the NHS number would be typically infeasible. Under the assumption that the pseudonymization process is undertaken in a uniform manner across each of the different organizations, the pseudonym that is generated for each patient from their NHS number will be identical across the different organizations. These pseudonyms would then provide a basis for matching the clinical data of patients across the different organizations.

Pseudonym Generation

In order to generate a pseudonym from an NHS number, we require a function that exhibits the following properties:

- One-way: computationally infeasible to recover the NHS number from the pseudonym
- Collision-resistant: each unique NHS number produces a unique pseudonym

Additionally, the function must exhibit the following practical properties:

- Accessibility: different organizations involved in the project have the technical capability to execute the function.
- Tractability: computational resources required to execute the function are not prohibitive

There are two potential candidates for the function:

- SHA-512[55]: a strong cryptographic hash function that generates a 512-bit hash value (i.e. pseudonym) from a given input (e.g. NHS Number).
- PBKDF2: a key derivation function that generates a hash value (i.e. pseudonym) of specified length by applying a chosen cryptographic function (e.g. SHA-512) a stated number of times to a given input.

Both functions would require the integration of a cryptographic salt into the pseudonym generation process. The cryptographic salt is a random value that is appended to the input of the function in order to prevent dictionary attacks and rainbow table attacks. These attacks use a set of hash values that are generated using the function for a set of known inputs (dictionary attack) or all possible inputs (rainbow table attack). These pre-computed hash values can be compared against those included within a dataset to discover the input from which the hash value was generated. A cryptographic salt defends against such attacks by including an additional random value in the input to the function, such that both the salt and the input value would need to be considered in any attack on the hash values.

[54] See http://content.digital.nhs.uk/media/12053/Obtaining-consent-for-patient-identifiable-and-or-sensitive-data-V2-050613/pdf/obtaining_consent_for_patient_identifiable_and_or_sensitive_data_V2_050613.pdf

[55] Other variants in the SHA2 family of cryptographic functions, e.g. SHA-256, provide reduced security.

We consider both SHA-512 and PKBDF2 (using SHA-512) to exhibit the properties of one-way and collision-resistant in a sufficiently robust manner. Additionally, as both functions are included within a variety of programming and scripting languages, we consider both functions to exhibit the accessibility property. Therefore, the choice of function is based upon the tractability property.

In contrast to SHA-512, PBKDF2 has been purposefully designed to ensure that the generation of hash values is computationally intensive. The extent of this computational intensity is determined by the number of times it is specified that the chosen cryptographic hash function must be applied to generate the hash. The higher the number of applications, the greater the computational intensity of the hash function, and the more time required to generate a hash value. The prevailing opinion regarding the numbers of applications given the computational power of present day machines is now over 100,000. Increasing the computational intensity required to generate a hash value provides some defence against the attacks previously described, as the attacker must have knowledge of the iterations and undertake that number of iterations to generate each hash value.

Whilst increasing the computation required to generate hash values can provide some defence against attacks, it also introduces a significant computational burden to the legitimate generation of hash values by organizations involved in the project. The different organizations would be required to generate hash values (pseudonyms) for large numbers of inputs (NHS numbers). For both ResearchOne and NHS Digital, hash values will need to be generated for millions of inputs to enable matching of patients from one dataset with any patients in their dataset.

Using a simple timing simulation[56], we have determined that the computation of 1000 hash values using PDKDF2 (with SHA-512) with 100,000 iterations took 135 seconds, and the computation of 10,000 hash values took 1345 seconds. If we consider the time to increase linearly with the number of hash values to be generated, then we estimate that the generation of 1,000,000 hash values would require around 135000 seconds (~1.5 days), and 10,000,000 would require around 1350000 (~15.5 days).

To ensure tractability for the organizations involved in the project, we propose the use of the SHA-512 hash function (in conjunction with a 512-bit cryptographic salt). We acknowledge the PBKDF2 may provide some additional defence against cryptographic attacks on the pseudonyms generated. However, we consider the use of SHA-512 to be acceptable for the project when associated context is considered, e.g. the specific organizations involved and their associated governance arrangements.

Alternatives to Pseudonymization

Homomorphic encryption[57] has been proposed as an alternative to the use of pseudonyms for secure linkage of healthcare data[58]. Whilst we acknowledge the significant value in such an approach, there are two significant challenges to the feasibility of such an approach for research projects such as UK-SAFE, which use data for large numbers of patients from a number of different organizations.

Firstly, each organization would be required to execute a cryptographic protocol that has significant technical complexity. All organizations involved in the project would be required to have this technical capability in order for the protocol to be feasible. However, the organizations will vary

[56] Simulation was performed on an Intel(R) Core (TM) i7-2600 CPU @ 3.40GHz with 16GB of memory (RAM).
[57] See P.Pallier. *Composite-Residuosity Based Cryptography: An Overview*, Cryptobytes 5:1 (2002): 20-26.
[58] See Khaled El Emam and Luc Arbuckle. *Anonymizing Health Data*. 2013.

significantly in their technical capability, and some organizations are likely to not have the requisite capability to support the protocol. Some form of technical assessment would have to be undertaken for each organization prior to the project to determine whether the protocol could be supported.

Secondly, under the assumption that each organization is technical capable of supporting the protocol, the issue of the computational resources required to undertake the protocol must be addressed. El Emam and Arbuckle[58] highlight that performance is an issue for such protocols, as the number of encrypted computations required is high. There is currently a lack of evidence to demonstrate that the performance of such a protocol would be acceptable for large numbers of patients in a project such as UK SAFE.

Therefore, we have determined that use of a linkage methodology based on homomorphic encryption is not feasible for this project

Data Linkage Methodology

Project Infrastructure

Work Package 2a (RO-HES) of Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE) will use the infrastructure provided by the Leeds Institute for Clinical Trials Research (LICTR)[59] at the University of Leeds to securely manage the data received from NHS Digital and The Phoenix Partnership, and to underpin the terms agreed between the University of Leeds and the data controllers (NHS Digital and The Phoenix Partnership) of the databases (Hospital Episode Statistics and ResearchOne) used in this work. The LICTR infrastructure has NHS Data Security and Protection Toolkit[60] approval and underpins the Data Sharing Framework Agreement that currently exists between the University of Leeds and NHS Digital.

Secure data transfer between the organizations involved in the methodology will be performed using one of the following services:

- SFT Service[61] provided by LICTR
- SEFT Service[62] provided by NHS Digital

We envisage that the use of the SEFT service will be mandated by NHS Digital for any transfer to or from NHS Digital. For all other transfers, we envisage that the SFT service will be used. Accordingly, the methodology description that follows is based on the premise that:

- A named individual at UoL and The Phoenix Partnership have an active account on the SEFT service (NHS Digital)
- A named individual at The Phoenix Partnership has an active account on the SFT service (UoL).

Data Sources

Work Package 2a (RO-HES) of Towards UK post Arthroplasty Follow-up rEcommendations (UK SAFE) will use the following sources of routine NHS data:

[59] See http://medhealth.leeds.ac.uk/info/400/leeds_institute_of_clinical_trials_research/
[60] See https://www.dsptoolkit.nhs.uk/
[61] See Appendix 2a
[62] See Appendix 2b

1. **Hospital Episode Statistics (HES)[63]:** a database controlled by NHS Digital containing patient data relating to A&E, inpatient and outpatient episodes

2. **ResearchOne[64]:** a database controlled by The Phoenix Partnership containing de-identified patient data from primary care settings that use the SystmOne clinical information system

Specific data items will be obtained from Hospital Episode Statistics for a specified population of patients. Specific data items will then be obtained from ResearchOne for a subset of these patients whose primary care data is present on ResearchOne. Data relating to individual patients from these two sources will be linked together for analysis.

Principles

The data linkage methodology has been developed to minimize the risk that any clinical data could be attributed to an individual, whilst ensuring that technical and economic burden on the organizations involved is not prohibitive.

The methodology adheres to the following principles:

1. No patient identifiable data[65] flows out of NHS Digital or The Phoenix Partnership.
2. Minimal clinical data to answer the research question flows out of NHS Digital and The Phoenix Partnership.
3. NHS Digital and The Phoenix Partnership provide clinical data directly to University of Leeds.
4. The Phoenix Partnership provide a set of unique (non-personal) identifiers and associated pseudonyms to NHS Digital.
5. No pseudonyms flow into the University of Leeds.
6. Unique (non-personal) identifiers from different organizations are linked at NHS Digital by matching the associated pseudonyms.
7. NHS Digital supplies mapping files to University of Leeds that map unique (non-personal) identifiers provided by the The Phoenix Partnership to unique (non-personal) identifiers provided by NHS Digital
8. University of Leeds aggregates clinical data from different organizations using the mapping files provided by NHS Digital.

Prerequisites

The description of the data linkage methodology is based on the following prerequisites:

1. NHS Digital and The Phoenix Partnership have prior knowledge of the clinical data that they are to provide[66].
2. NHS Digital and The Phoenix Partnership have prior knowledge of the pseudonymization process and the technical capability to undertake this process.

We now describe each data flow and process that comprises the methodology.

---

[63] See http://www.digital.nhs.uk/hes
[64] See http://www.researchone.org
[65] See http://content.digital.nhs.uk/media/12053/Obtaining-consent-for-patient-identifiable-and-or-sensitive-data-V2-050613/pdf/obtaining_consent_for_patient_identifiable_and_or_sensitive_data_V2_050613.pdf
[66] See Appendix 1

Data Flows

Figure 1 provides an overview of the data flows associated with the proposed data linkage methodology.

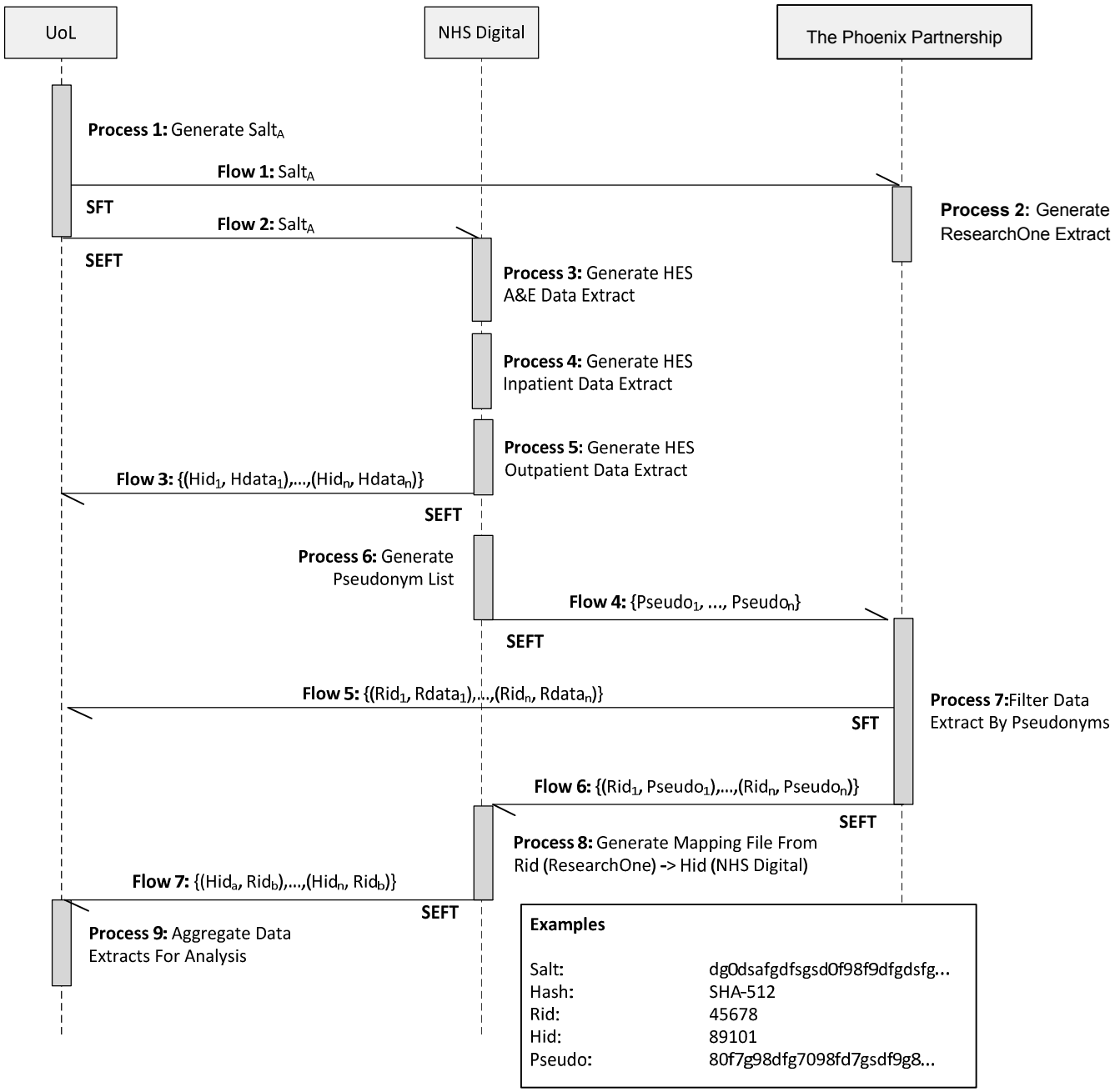


Figure 1: Overview of the Data Linkage Methodology

Further details on the specific processes and flows are provided below:

Process 1: Generate Salt (UoL)

This process is undertaken at the UoL and involves the generation of a cryptographic salt that will be used in the generation of patient pseudonyms by NHS Digital and The Phoenix Partnership. The generated salt file will only be used by these organisations for this specific linkage.

The generated salt will be saved to a file that is stored securely at the UoL.

Flow 1: Salt (UoL to The Phoenix Partnership)

This flow will be initiated by the UoL and involves the communication of the cryptographic salt generated by the UoL for this specific execution of the data linkage methodology to The Phoenix Partnership.

The file will be securely transmitted to The Phoenix Partnership using the SFT service.

Once received by the The Phoenix Partnership, the salt file will be stored securely and access will be restricted to members of staff involved in the data extraction and pseudonymization.

Following the supply of the clinical data to UoL (Flow 5) and mapping file to NHS Digital (Flow 6), The Phoenix Partnership will securely delete the salt file.

Flow 2: Salt (UoL to NHS Digital)

This flow will be initiated by the UoL and involves the communication of the cryptographic salt generated by the UoL for this specific execution of the data linkage methodology to NHS Digital.

The file will be securely transmitted to NHS Digital using the SEFT service.

Once received by the NHS Digital, the salt file will be stored securely, with access restricted to only those members of staff involved in the data extraction and pseudonymization.

Following the supply of the mapping files to UoL (Flow 7), NHS Digital will securely delete the salt file.

Process 2: Generate ResearchOne Extract (The Phoenix Partnership)

The Phoenix Partnership will generate a ResearchOne extract from SystmOne based on a supplied specification[67]. For each patient, The Phoenix Partnership will generate a unique identifier that is independent of any directly identifiable patient information.

To facilitate linkage of the data provided by The Phoenix Partnership (ResearchOne) to the data provided by NHS Digital (Hospital Episode Statistics), The Phoenix Partnership will also generate a pseudonym for each patient in the data extract using the method described in the Background section and the cryptographic salt provided by UoL in Flow 1.

The generated identifiers will then be appended to the extract to produce the following:

Pseudonym	Rid	RData		
		Date of Birth	Date of Death	
7cd3735db9426c9fa9ca60256c3bcc0a32f325a17d300fd5ce7f637ce703844d	950001	01/05/1956		…

Figure 2: Data Extract (ResearchOne)

Further details regarding the generation of pseudonyms by The Phoenix Partnership can be found in the ResearchOne Database Protocol document (see http://www.researchone.org/wp-content/uploads/2013/08/TPP-Research-Database_Protocol_V1.4.pdf).

[67] See Appendix 1

Process 3: Generate HES A&E Data Extract (NHS Digital)

NHS Digital will generate an extract from HES A&E based on a supplied data specification[68]. For each patient, NHS Digital will generate a unique identifier that is independent of any directly identifiable patient information data items.

To facilitate linkage of the data provided by NHS Digital (Hospital Episode Statistics) to the data provided by The Phoenix Partnership (ResearchOne), NHS Digital will also generate a pseudonym for each patient in the data extract using the method described in the Background section and the cryptographic salt provided by UoL in Flow 2.

The generated identifiers will then be appended to the extract to produce the following:

Pseudonym	TId	$Hdata_{A\&E}$		
		Arrival Mode	Attendance Category	…
7cd3735db9426c9fa9ca60256c3bcc0a32f325a17d300fd5ce7f637ce703844d	10001	2	1	…
…	…	…	…	…

Figure 3: Data Extract (NHS Digital – A&E)

Process 4: Generate HES Inpatient Data Extract (NHS Digital)

NHS Digital will generate an extract from HES Inpatient based on a supplied data specification[68]. For each patient, NHS Digital will generate a unique identifier that is independent of any directly identifiable patient information.

To facilitate linkage of the data provided by NHS Digital (Hospital Episode Statistics) to the data provided by The Phoenix Partnership (ResearchOne), NHS Digital will also generate a pseudonym for each patient in the data extract using the method described in the Background section and the cryptographic salt provided by UoL in Flow 2.

The generated identifiers will then be appended to the extract to produce the following:

Pseudonym	TId	$HData_{inpatient}$		
		Admission Date Check Flag	Date of Admission	…
7cd3735db9426c9fa9ca60256c3bcc0a32f325a17d300fd5ce7f637ce703844d	10001	0	01/04/2013	
…	…	…	…	…

Figure 4: Data Extract (NHS Digital – Inpatient)

Process 5: Generate HES Outpatient Data Extract (NHS Digital)

NHS Digital will generate an extract from HES Outpatient based on a supplied data specification[68]. For each patient, NHS Digital will generate a unique identifier that is independent of any directly identifiable patient information.

[68] See Appendix 1

To facilitate linkage of the data provided by NHS Digital to the data provided by The Phoenix Partnership, NHS Digital will also generate a pseudonym for each patient in the data extract using the method described in the Background section and the cryptographic salt provided by UoL in Flow 2.

The generated identifiers will then be appended to the extract to produce the following:

Pseudonym	TId	$HData_{outpatient}$		
		Appointment Date	Attendance Type	…
7cd3735db9426c9fa9ca60256c3bcc0a32f325a17d300fd5ce7f637ce703844d	10001	01/04/2013	1	
…	…	…	…	…

Figure 5: Data Extract (NHS Digital – Outpatient)

Flow 3: HES A&E, Inpatient and Outpatient Data Extract (NHS Digital to UoL)

This flow will be initiated by NHS Digital and involves communication of the HES A&E, Inpatient and Outpatient data extracts to the UoL. Prior to communication, the data extract structures will be amended to remove the pseudonym field to produce the following:

HId	$Hdata_{A\&E}$		
	Admission Method	Admission Type	…
10001	2	1	
…	…	…	…

Figure 6: Data Extract (NHS Digital – A&E)

HId	$HData_{inpatient}$		
	Admission Date Check Flag	Date of Admission	…
10001	0	01/04/2013	
…	…	…	…

Figure 7: Data Extract (NHS Digital - Inpatient)

Hid	$HData_{outpatient}$		
	Appointment Date	Attendance Type	…
10001	01/04/2013	1	
…	…	…	…

Figure 8: Data Extract (NHS Digital – Outpatient)

The data extracts will be transmitted by NHS Digital to UoL using SEFT service.

Once received by the UoL, the data extract will be stored securely and access will be restricted to relevant members of the research team.

Process 6: Generate Patient Pseudonyms (NHS Digital)

To facilitate linkage of the data provided by the NHS Digital (Hospital Episode Statistics) to the data provided by The Phoenix Partnership (ResearchOne), NHS Digital will generate a list containing pseudonyms for each unique patient contained in the HES Inpatient data extracts described in Process 4 whose admission occurred within the defined index period:

Pseudonym
7cd3735db9426c9fa9ca60256c3bcc0a32f325a17d300fd5ce7f637ce703844d
…

Figure 9: List of Patient Pseudonyms (NHS Digital)

Flow 4: Pseudonyms (NHS Digital to The Phoenix Partnership)

This flow will be initiated by NHS Digital and involves communication of a list of pseudonyms for whom primary care data is required from ResearchOne to The Phoenix Partnership.

The pseudonyms will be transmitted by NHS Digital to The Phoenix Partnership using the SEFT service.

Once received by The Phoenix Partnership, the pseudonyms will be stored securely and access will be restricted to members of staff involved in the data extraction and pseudonymization.

Process 7: Filter Data Extract By Pseudonyms (The Phoenix Partnership)

Based on the pseudonyms provided by NHS Digital in Flow 4, The Phoenix Partnership will filter the data extract produced in Process 2 to include only those patients whose pseudonym is contained in the list provided by NHS Digital.

This will produce a data extract of identical form to that produced in Process 2.

Flow 5: ResearchOne Extract (The Phoenix Partnership to UoL)

This flow will be initiated by The Phoenix Partnership and involves communication of the ResearchOne data extract to the UoL. Prior to communication, the data extract structure will be amended to remove the pseudonym field to produce the following:

Rid	RData		
	Date of Birth	Date of Death	…
950001	01/05/1956		
…	…	…	…

Figure 10: Data Extract (ResearchOne)

The data extract will be transmitted by The Phoenix Partnership to UoL using the SFT service.

Once received by the UoL, the data extract will be stored securely and access will be restricted to relevant members of the research team.

Flow 6 (The Phoenix Partnership to NHS Digital)

This flow will be initiated by The Phoenix Partnership and involves communication of the pseudonyms and corresponding patient identifiers to NHS Digital for the purpose of linkage. The file will have the following form:

Pseudonym	RId
7cd3735db9426c9fa9ca60256c3bcc0a32f325a17d300fd5ce7f637ce703844d	950001
…	…

Figure 11: Pseudonym Data File

This data file will be securely transmitted to NHS Digital using the SEFT service.

Following the supply of the mapping file to UoL (Flow 7), NHS Digital will securely delete the pseudonym data file.

Process 8: Generate Mapping File (NHS Digital)

NHS Digital will generate a file that maps the unique identifiers used for patients in Hospital Episode Statistics and the unique identifiers used for patients in ResearchOne using the pseudonyms with which these identifiers are associated. By simply matching the pseudonyms provided by the The Phoenix Partnership with the pseudonyms generated for the HES data by NHS Digital, the corresponding patient identifiers can be found for both the HES data and the ResearchOne data.

A file can then be produced that contains all those IDs that refer to the same patient (as determined by identical pseudonyms) in the HES and ResearchOne data:

HId	RId
10001	950001
…	…

Figure 12: Mapping File (HES to ResearchOne)

Flow 7: Mapping File (NHS Digital to UoL)

This flow will be initiated by NHS Digital and involves communication of the mapping file (HES to ResearchOne) to the UoL.

The mapping file will be transmitted by NHS Digital to UoL using the SEFT service.

Once received by the UoL, the mapping file will be stored securely, with access restricted to selected members of the research team.

Process 9: Aggregate Data Extracts for Analysis (UoL)

UoL will aggregate the clinical data received from the NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne) using the mapping files provided by NHS Digital. The patient identifier in each extract will be matched to the patient identifiers in the other extracts on the

basis of the mapping file. This will enable construction of a longitudinal record for each patient that includes data from the three different sources.

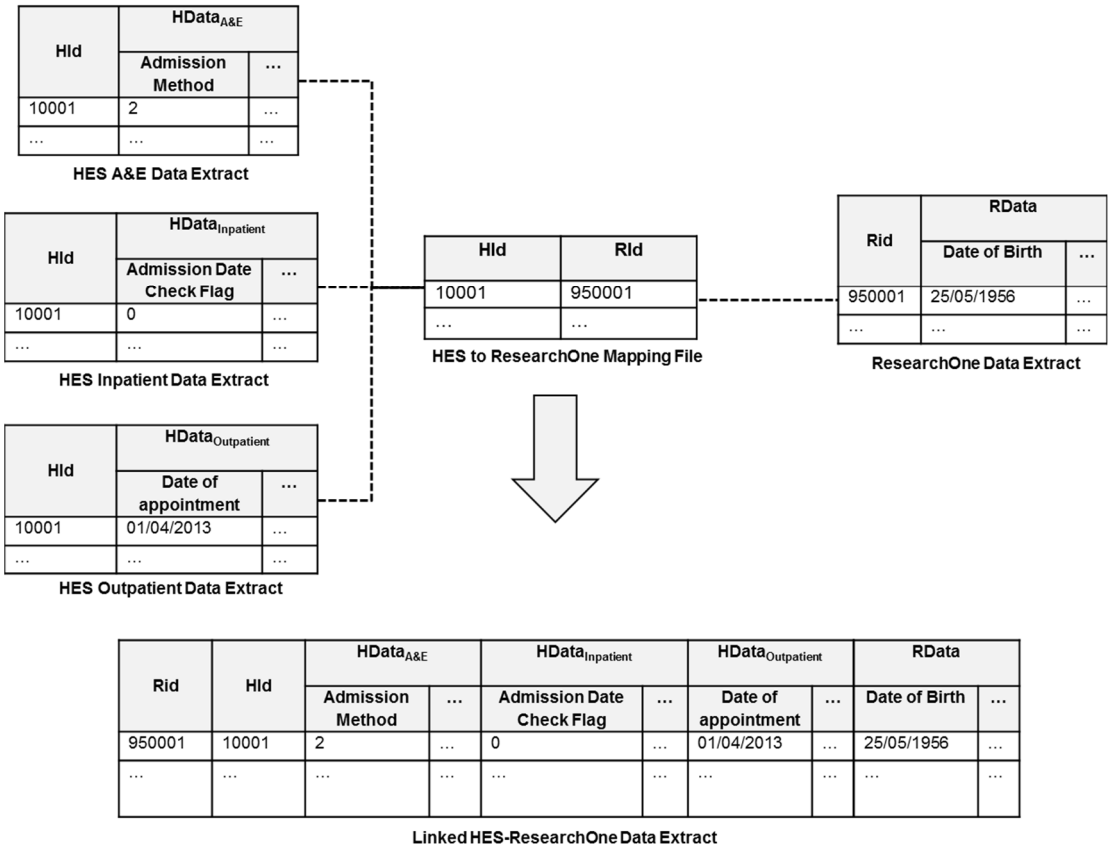


Figure 13: Aggregation of Data Extracts from NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne)

Related Usage and Approvals

The linkage methodology described was developed for another funded NIHR (HS&DR) project at the Leeds Institute for Health Sciences: *Liaison Psychiatry: Measurement and Evaluation of Service Types, Referral Patterns and Outcomes* (REF: 13/58/08), which requires data from NHS Digital and Research to be linked for analysis.

For the specific work stream in LP-MAESTRO that uses this methodology (WS2P1), LP-MAESTRO has received: i) a favourable opinion from an NHS Research Ethics Committee (REF: 16/NS/0025), and ii) a decision from the Health Research Authority's Confidentiality Advisory Group (CAG) that Section 251 support is not required (REF: 16/CAG/0037).

Appendix 2a: SFT Service

ctru System Documentation V1 16/11/2010	Document Title	*Biscom Secure File Transfer – Security Statement*				
	Author	*AB*	Version	1	Date	19/06/2014

Overview

Biscom Secure File Transfer (SFT) is implemented within the LICTR as a secure service for transferring messages and files with partners in a secure manner.

Security

Tiered architecture means a single public facing proxy server is the only part of the system exposed to the internet. This server stores no data and routes all requests to internal servers. All files are stored in an encrypted format (AES 256) on the SFT application server. During transit to and from an end user all data is encrypted using SSL.

The product is FIPS 140-2 certified. FIPS is a US Government standard that describes the encryption and security requirements that IT products should meet in order to be used for transferring sensitive documents. It also covers authentication requirements for end users to make sure the product is accessed securely. Details of Biscoms FIPS certification can be obtained from here:
http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/140sp/140sp1906.pdf

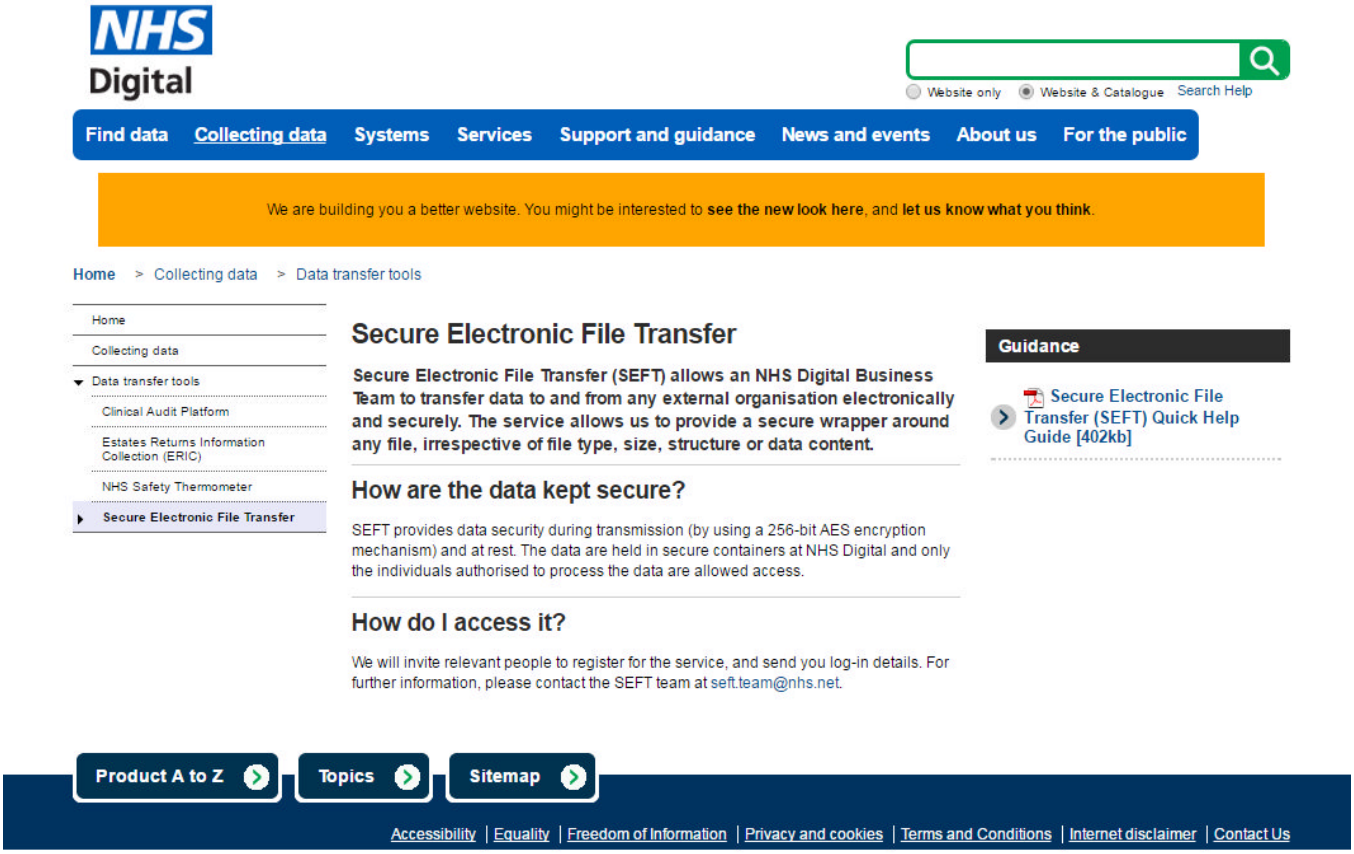
A local copy of this document is also kept in the same folder as this statement. Although a US standard the certification is recognised world-wide as a standard for IT systems using encryption.

All servers involved in the SFT service are isolated by the use of firewalls. The servers operate off domain and do not share credentials with other systems. The diagram below shows a high level overview of how data is transferred between an end user and the Secure File Transfer service.

Compliance

The LICTR IS System Administrator has the ability to enable a compliance role linked to the SFT service. The role allows all transfers by the system to be audited, even if the transfer has been deleted. The compliance data shows the transfer details, its recipients, who accessed/ downloaded files and when it occurred as well as details about each file.

Appendix 2b: SEFT Service



Further details can be found at: http://content.digital.nhs.uk/seft.

Appendix Three: Privacy Impact Assessment

Aim and Scope

This document provides a Privacy Impact Assessment for Work Package 2a (RO-HES) of the NIHR-funded research project: Towards UK post Arthroplasty Follow-up rEcommendations (UK-SAFE). The document follows the "Privacy Impact Assessment Template" provided in Annex Two of "Conducting Privacy Impact Assessments Code of Practice" by the Information Commissioner's Office[69].

Please note: *This document may be subject to amendment over the lifetime of the project in response to consideration by (and feedback from) external advisory bodies, including NHS Research Ethics Committee (REC) and the Health Research Authority's Confidentiality Advisory Group (CAG), and ongoing consideration by (and feedback from) the Project Management Group (PMG) and Independent Advisory Group (IAG).*

Intended Audience

This document is intended for individuals or organisations that are required to take an ethical, legal or technical view in relation to the use of (linked) routine NHS data within the research project. Feedback on this document is welcomed.

Background

Aims

Towards UK post Arthroplasty Follow-up rEcommendations (UK-SAFE) is a research project that is based at Leeds Institute for Rheumatic and Musculoskeletal Medicine (LIRMM) and sponsored by the University of Leeds. The project is funded by the National Institute for Health Research (NIHR) under its Health Services and Delivery Research (HS&DR) funding stream.

The overall aim of UK-SAFE is to determine the consequences of disinvestment in hip and knee arthroplasty follow-up.

The objectives are:

- To identify who needs follow up and when this should occur for primary total hip and total and uni-compartmental knee arthroplasty surgery by making use of routine data
- To understand the patient journey (in primary and secondary care) to revision surgery by recruiting patients admitted for elective and emergency revision surgery
- To establish how and when patients are identified for revision and why some patients are missed from regular follow-up and present acutely with fracture around the implant (peri-prosthetic fracture), by using prospective and retrospective data
- To identify the most appropriate and cost-effective follow-up pathway to minimise potential harm to patients by undertaking cost-effectiveness modelling
- To provide evidence- and consensus-based recommendations on how follow-up of primary hip and knee arthroplasty should be conducted

[69] See https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf

Work Package 2 of UK-SAFE will use routine data from 5 national datasets will be used: Clinical Practice Research Database (CPRD)[70], ResearchOne (RO)[71], Hospital Episode Statistics (HES)[72], National Joint Registry (NJR)[73], and Patient Reported Outcome Measures (PROMs)[74] along with prospective data from patients who have revision surgery to understand when and which patients present for revision surgery, and to understand how they are currently identified for revision surgery

Work Package 2a (RO-HES) is a specific sub-package of work that will use the following sources of routine NHS data:

1. **Hospital Episode Statistics (HES)[75]:** a database controlled by NHS Digital containing patient data relating to A&E, inpatient and outpatient episodes
2. **ResearchOne (RO)[76]:** a database controlled by The Phoenix Partnership containing de-identified patient data from primary care settings that use the SystmOne clinical information system

Data obtained from these sources for a defined patient population will be linked to construct longitudinal records covering primary and secondary care. These records will be analysed determine when, which and how patients present for revision surgery.

Benefits

With increasing demand and increasing patient expectations coupled to diminishing resources, rational evidence-based changes to practice are essential. Current routine follow-up of arthroplasty costs the NHS in the region of £100m per year. Reduction in follow-up is therefore seen as an easy cost-saving measure. However, there has been no robust cost-effective study of how follow-up should be conducted. The recent Briggs report on Improving the Quality of Orthopaedic Care within the National Health Service in England states that all patients should receive appropriate follow-up to detect complications and disease recurrence early.

Disinvestment in follow-up as a cost-saving strategy cannot be justified, unless a robust evidence-base is established. Similarly, benefits of more expensive regular follow-up may be limited. Whilst we are aware that research is currently on-going to evaluate new technologies for monitoring patients at a distance, these technologies are themselves expensive, and before they are employed into routine clinical practice an evidence-base for arthroplasty follow-up must first be established. Follow-up must take into account a variety of factors, including implant type, the joint involved and patient factors and a decision to alter follow-up pathways must consider the long-term impact on patients, health professionals, and the NHS as a whole. This project is built around the hypothesis that a comprehensive, evidence-based, stratification algorithm to determine appropriate follow-up for an individual patient may provide a more cost-effective strategy for orthopaedic follow-up service delivery.

Rationalising the current diversity of follow-up practices will enable substantial savings for the NHS and focus the use of NHS resources on those patients most at need. Implementation of appropriate follow-up will reduce the burden on both patients and the NHS in terms of outpatient visits and

[70] See https://www.cprd.com
[71] See http://www.researchone.org
[72] See http://digital.nhs.uk/hes
[73] See http://www.njrcentre.org.uk
[74] See http://content.digital.nhs.uk/proms
[75] See http://digital.nhs.uk/hes
[76] See http://www.researchone.org

clinical tests that do not add benefit, while optimising detection of potential problems and thus ensuring patients are not harmed. We envisage the outputs to be readily applicable to the wider NHS, and internationally, at the end of the grant period.

Justification

Work Package 2a (RO-HES) requires a Privacy Impact Assessment (PIA) based on the following rationale:

- Use of data that is routinely captured for individual patients for the primary purpose of *direct care*.
- Use of data that is captured by healthcare professionals under a duty of *confidentiality* to patients.
- Use of data that is *personal* and contains data items that are likely to be considered *sensitive*.
- Use of data without explicit consent of each individual patient.
- Use of data captured in different care settings: primary and secondary care.
- Use of data from two different sources (NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne)).
- Linkage of data for individual patients from the two data sources.

Privacy Impact Assessment

The aim of this PIA is to identify the risks relating to privacy that arise within Work Package 2a (RO-HES) and to demonstrate the measures that have been taken to mitigate these risks, whilst ensuring that the aim and objectives of the projects can be fulfilled.

Information Flows

Work Package 2a (RO-HES) will use data obtained from two data sources: NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne). The data required from these sources and the methodology by which data from these sources will be linked together have been designed to ensure that: i) the research question can be robustly answered, ii) any risks to confidentiality and privacy are minimised, and iii) the technical and economic burden on the organizations involved is not prohibitive.

Figure 1 provides an overview of data flows between University of Leeds (UoL), NHS Digital and The Phoenix Partnership. Further detail regarding the data obtained from each source and the methodology by which the data will be linked together can be found in Appendix 1 and Appendix 2.

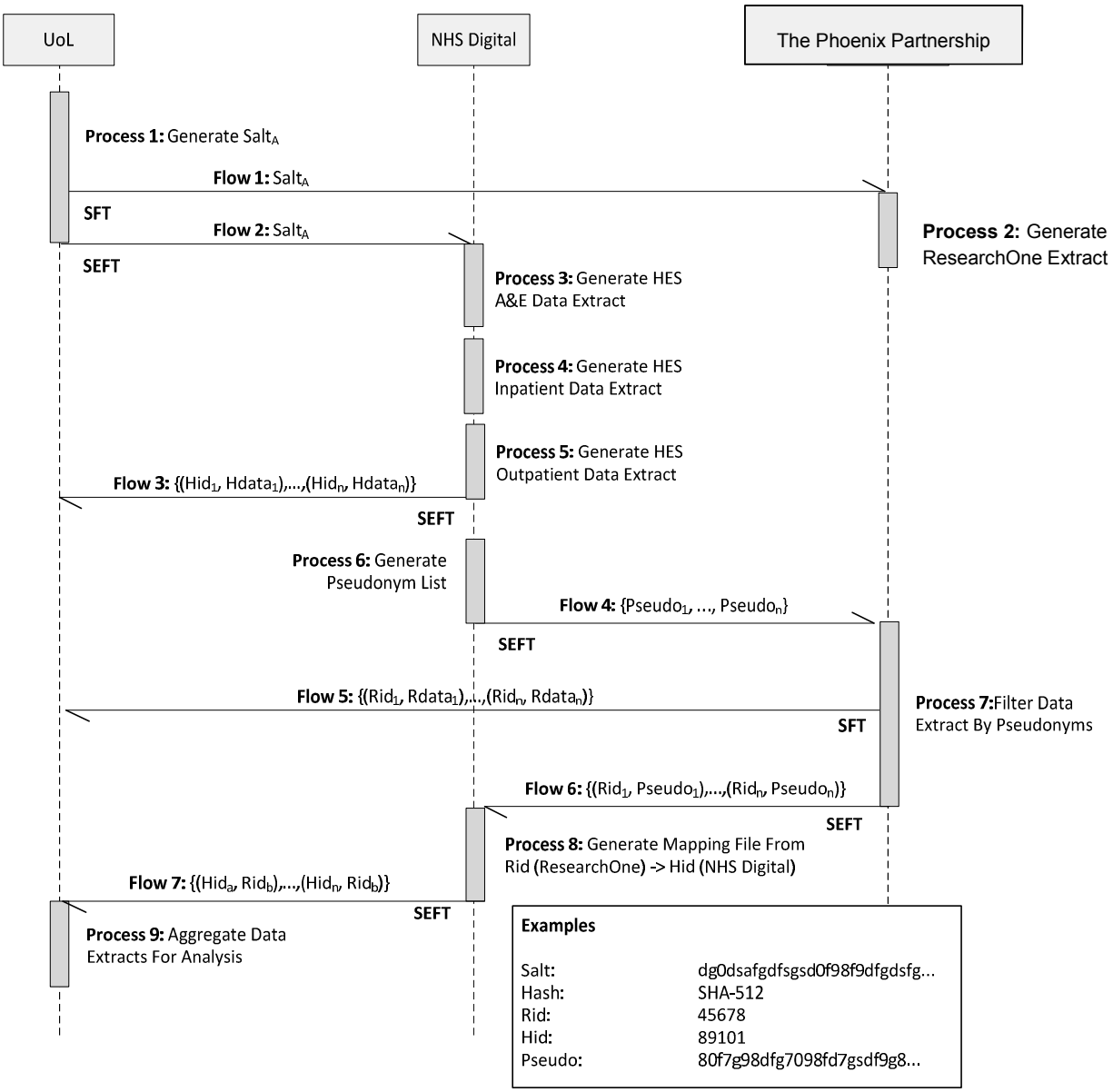


Figure 14: Overview of the Information Flows

Scope

Patients will be initially identified for the study from a record of an inpatient admission to a hospital in England in a specific index period (1st April 2000 – 31st March 2015) for one of the following procedures: i) hip replacement, ii) hip revision, iii) knee replacement, or iv) knee revision.

The scope of all hospitals in England has been chosen to enable *variation in follow-up for hip and knee replacements and revisions between hospitals* to be analysed. The time period of 15 years over which patients will be identified has been chosen to enable *variation in follow-up for hip and knee replacements and revisions for patients over time* to be analysed.

Based on data provided within the National Joint Registry (NJR) Annual Report[77], 800,683 primary hip replacements, 89,023 hip revisions, 875,585 primary knee replacements and 54,278 knee

revisions were reported (subject to specific inclusion and exclusion criteria) in period from 1 April 2003 to 31 December 2015. Therefore, for a comparable period within the study, we anticipate that around 900,000 hip replacements and 1,000,000 knee replacements (1,900,000 joint replacements in total) will be identified.

Consultation Details

In order to identify and address privacy risks associated with Work Package 2a (RO-HES), a number of different parties have been, or will be, consulted:

- **Project Management Group**
  UK-SAFE's Project Management Group have been consulted in relation to the Ethics Protocol, including the Data Specification (Appendix 1), Data Linkage Methodology (Appendix 2), Privacy Impact Assessment (Appendix 3) and Data Processing (Appendix 4). Members of the PMG have provided feedback on the contents of these documents. In relation to the Data Specification, members of the PMG have been involved in the refinement of data items to ensure that those included are both *necessary* and *sufficient* to robustly answer the research question.

- **NHS Digital**
  NHS Digital will be consulted once an application is submitted for the data required by the project[78]. This application will be evaluated by the Information Asset Owner (IAO)[79] for Hospital Episode Statistics (HES) and the Independent Group Advising on the Release of Data (IGARD)[80] to verify that there is an appropriate legal basis for accessing the requested data and that appropriate safeguards are in place to ensure that data is stored and handled safely and securely. The application must be approved by both the IAO and IGARD prior to release of the data by NHS Digital[81].

- **The Phoenix Partnership**
  The Phoenix Partnership will be consulted once an application is submitted for the data required by the project. This application will be evaluated by ResearchOne Project Committee (RPC)[82] against the aims and objectives of the ResearchOne database. The application must be approved by the RPC prior to release by ResearchOne[83].

NHS Digital and The Phoenix Partnership have been previously consulted regarding the linkage methodology to be used in Work Package 2a (RO-HES) (see Annexe 2). This linkage methodology was developed for use in Work Stream 2 (Phase 1) of another funded NIHR

[78] Following application to NHS Digital, the specification of data to be supplied from Hospital Episode Statistics has been amended in accordance with requirements specified by the production team at NHS Digital. We expect the planned analysis to remain feasible on the basis of the data items supplied by NHS Digital.
[79] See http://content.digital.nhs.uk/article/6888/DARS-Process
[80] See http://content.digital.nhs.uk/media/23491/IGARD-Terms-of-Reference-v15/pdf/IGARD_Terms_of_Reference_v1.5.pdf
[81] Data request for this work was submitted by the University of Leeds to NHS Digital in January 2018 and approval has now been received for this work from the Information Asset Owner (September 2020) and from the Independent Group Advising on the Release of Data (September 2020).
[82] See http://www.researchone.org/wp-content/uploads/2013/08/TPP-Research-Database_Protocol_V1.4.pdf
[83] Data request for this work was submitted by the University of Leeds to The Phoenix Partnership in January 2018 and approval from the ResearchOne Project Committee has been most recently received for this work in February 2019.

(HS&DR) project at the Leeds Institute for Health Sciences: Liaison Psychiatry: Measurement and Evaluation of Service Types, Referral Patterns and Outcomes – LP-MAESTRO (REF: 13/58/08). NHS Digital and The Phoenix Partnership were consulted in the development of the methodology for LP-MAESTRO to ensure that relevant technical and organisational constraints were considered.

- **NHS Research Ethics Committee**
  Approval will be sought from an NHS Research Ethics Committee[84] for Work Package 2a (RO-HES) whose role is to safeguard the rights, safety, dignity and well-being of research participants. The work will only proceed on receipt of a favourable ethical opinion from the committee, and on fulfilment of any conditions on which this opinion is predicated.[85]

- **Confidentiality Advisory Group**
  Advice will be sought from the Health Research Authority's Confidentiality Advisory Group (CAG)[86] as to whether it is considered that Work Package 2a (RO-HES) involves disclosure of patient identifiable data without consent, and therefore that Section 251 support is required. If it is determined that Section 251 support is required, a recommendation for such support will be obtained. The work will only proceed: i) on receipt of a decision that Section 251 support is not required, or ii) on receipt of a decision that Section 251 support is required and is then successfully obtained.[87]

North of Scotland NHS Research Ethics Committee and the Confidentiality Advisory Group have been previously consulted in relation to Work Stream 2 (Phase 1) of another funded NIHR (HS&DR) project at the Leeds Institute for Health Sciences: Liaison Psychiatry: Measurement and Evaluation of Service Types, Referral Patterns and Outcomes – LP-MAESTRO (REF: 13/58/08). This work employs the same linkage methodology between the same data sources: NHS Digital (Hospital Episode Statistics) and The Phoenix Partnership (ResearchOne). LP-MAESTRO Work Stream 2 (Phase 1) received a favourable opinion from North of Scotland NHS Research Ethics Committee (REF: 16/NS/0025) and a decision from the Health Research Authority's Confidentiality Advisory Group (CAG) that Section 251 support was not required (REF: 16/CAG/0037).

- **University of Leeds**
  Approval and signature will be sought from the Principal Investigator, Sponsor's Representative and Information Guardian at the University of Leeds as part of the submission process to an NHS Research Ethics Committee and the Confidentiality Advisory Group to

[84] See http://www.hra.nhs.uk/about-the-hra/our-committees/research-ethics-committees-recs/
[85] Favourable opinion was received from an NHS Research Ethics Committee on 8th August 2017, subject to Management Permission, which was received on 9th September 2017.
[86] See http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/
[87] Following submission of an application to the Confidentiality Advisory Group at the Health Research Authority (22nd June 2017), and following subsequent correspondence with the Health Research Authority, NHS Digital and The Phoenix Partnership, it was determined that Section 251 support was not required for the study and the application to the Confidentiality Advisory Group with withdrawn (4th October 2017).

ensure the sufficiency of technical, ethical and legal measures from an organisational perspective.[88]

- **Leeds Institute of Clinical Trials Research (LICTR)**
  LICTR have been consulted to confirm their willingness and ability to provide an infrastructure for secure management of data for Work Package 2a (RO-HES) within the University of Leeds that exhibits the requisite information security standards. The infrastructure at LICTR is compliant with the NHS Data Security and Protection Toolkit[89] and has the necessary policies and standards in place to achieve this compliance.

LICTR have been previously consulted and provided an infrastructure for secure management of data for another funded NIHR (HS&DR) project at the Leeds Institute for Health Sciences: Liaison Psychiatry: Measurement and Evaluation of Service Types, Referral Patterns and Outcomes – LP-MAESTRO (REF: 13/58/08).

[88] Approval for the project was received from the University of Leeds on 14th June 2017, prior to submission of the project to the Health Research Authority for consideration by an NHS Research Ethics Committee and the Confidentiality Advisory Group.
[89] See https://www.dsptoolkit.nhs.uk/

Privacy and Related Risks

Privacy Issue	Risk To Individuals	Compliance Risk	Associated Organisation/Corporate Risk
Data intercepted in transit between organisations by an external party	Data available to external parties who may have motivation and ability to (re-) identify individuals or groups of individuals and to use for undefined purposes with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data intentionally or unintentionally transmitted to an external party by a data provider	Data available to external parties who may have motivation and ability to (re-) identify individuals or groups of individuals and to use for undefined purposes with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data intentionally or unintentionally transmitted to an external party by the research team	Data available to external parties who may have motivation and ability to (re-) identify individuals or groups of individuals and to use for undefined purposes with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data is accessed by an external party at a data provider	Data available to external parties who may have motivation and ability to (re-) identify individuals or groups of individuals and to use for undefined purposes with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data is accessed by an external party at the University of Leeds	Data available to external parties who may have motivation and ability to (re-) identify individuals or groups of individuals and to use for undefined purposes with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust

[90] At the time of authoring Version 1 of this document, the applicable legislation was Data Protection Act 1998 (see http://www.legislation.gov.uk/ukpga/1998/29/contents). Since 25th May 2018, the applicable legislation is the Data Protection Act 2018 (see https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted) and the General Data Protection Regulation (see https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN).

Data used for purposes outside of the research project by a data provider	Data used for purposes that have not been ethically approved.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data used for purposes outside of the research project by research team	Data used for purposes that have not been ethically approved.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data that is unnecessary for research is requested by the research team	Increased risk that an individual or group of individuals can be (re-) identified in the data by the research team or an external party with negative financial, social and health/wellbeing consequences.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data that is unnecessary for research is supplied to the research team by data providers	Increased risk that an individual or group of individuals can be (re-) identified in the data by the research team or an external party with negative financial, social and health/wellbeing consequences.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Research team intentionally or unintentionally (re-) identify an individual or groups of individuals from the data	Identification of individuals or group of individuals in the data and associated with sensitive data items by research team with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Research team externally publish data that enables (re-) identification of an individual or group of individuals from the data	Identification of individuals or group of individuals in the data and associated with sensitive data items by external parties and use for purposes outside of the research project with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data providers retain data beyond the specified end date of the project	Increased risk that data is subject to intentional or unintentional transmission to or access by an external party, or that data is used for purposes outside of the research project with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust
Data providers retain data beyond the specified end date of the project	Increased risk that data is subject to intentional or unintentional transmission to or access by an external party, or that data is used for purposes outside of the research project with negative financial, social and health/wellbeing consequences for individuals.	Data Protection Act[90]	- Regulatory action - Reputational damage - Loss of public trust

Privacy Solutions

Risk	Solution(s)	Result (Risk Eliminated, reduced or accepted?)	Evaluation (is impact of individuals after implementation of solutions a justified, compliant and proportionate response to aims of project)
Data intercepted in transit between organisations by an external party	• Data transferred between parties using secure transfer mechanisms where data is encrypted and transmitted between named individuals in each organisation[91]. • No patient identifiable data are transmitted between organisations[91]. • Patients linked across data providers using pseudonyms generated from NHS numbers[91]. • Cryptographic salt file[91] sent separately from other data to each organisation by University of Leeds and used for this specific linkage scenario only[91]. • Pseudonyms only transmitted with ResearchOne-specific identifiers between The Phoenix Partnership and NHS Digital to enable linkage by NHS Digital[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91]. • Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92].	Risk reduced	Yes
Data intentionally or unintentionally transmitted to an external party by a data provider	• Data transferred between parties using secure transfer mechanisms where data is encrypted and transmitted between named individuals in each organisation[91]. • No patient identifiable data are transmitted between organisations[91]. • Patients linked across data providers using pseudonyms generated from NHS numbers[91].	Risk reduced	Yes

[91] See Appendix 2
[92] See Appendix 1

	• Cryptographic salt file sent separately from other data to each organisation by University of Leeds and used for this specific linkage scenario only[91]. • Pseudonyms only transmitted with ResearchOne-specific identifiers between The Phoenix Partnership and NHS Digital to enable linkage by NHS Digital[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91]. • Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92]. • Data Processing Agreement to be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. ~~• Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~		
Data intentionally or unintentionally transmitted to an external party by the research team	• Data to be securely managed in an infrastructure at LICTR that is compliant with NHS Data Security and Protection Toolkit[93]. • No patient identifiable data are transmitted between organisations[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91]. • Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92]. • Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[94] and Medical Research Council[95]. • Any members of research team with access to the data will have an employment contract with the University of Leeds	Risk reduced	Yes

[93] See https://www.dsptoolkit.nhs.uk/
[94] See http://it.leeds.ac.uk/info/273/information_security_training/1061/information_security_awareness_training
[95] See http://byglearning.co.uk/mrcrsc-lms/course/category.php?id=1 - "Research, GDPR and Confidentiality"

	• University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[97] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available.		
Data is accessed at a data provider by an external party	• Data Processing Agreement to be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. • ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	Risk reduced	Yes
Data is accessed at the University of Leeds by an external party	• Data to be securely managed in an infrastructure at LICTR that is compliant with NHS Data Security and Protection Toolkit[93]. • No patient identifiable data are transmitted between organisations[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91]. • Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92]. • University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which	Risk reduced	Yes

[96] See http://content.digital.nhs.uk/media/14847/Draft----HSCIC-Data-Sharing-Framework-Contract/pdf/HSCIC_Data_Sharing_Framework_Contract.pdf for sample agreement.
[97] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

	data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[98] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available..		
Data used for purposes outside of the research project by a data provider	• Data Processing Agreement to be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. • ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	Risk reduced	Yes
Data used for purposes outside of the research project by research team	• University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[99] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available. • Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will have an employment contract with the University of Leeds	Risk reduced	Yes

[98] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc
[99] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

Data that is unnecessary for research is requested by the research team	• Data specification will be approved by the Principal Investigator as part of the Ethics Protocol. • Data specification will be approved by an NHS Research Ethics Committee as part of the Ethics Protocol. • Application to NHS Digital for data items included in the data specification must be approved by Information Asset Owner (IAO) for Hospital Episode Statistics (HES) and Independent Group Advising on the Release of Data (IGARD). • Application to ResearchOne for data items included in the data specification must be approved by the ResearchOne Project Committee (RPC).	Risk reduced	Yes
Data that is unnecessary for research is supplied to the research team by data providers	• Data specification will be formally specified in application for data from NHS Digital (Hospital Episode Statistics). • Data specification will be formally specified in application for data from The Phoenix Partnership (ResearchOne). • Data Processing Agreement will be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. • ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	Risk reduced	Yes
Research team intentionally or unintentionally (re-) identify an individual or groups of individuals from the data	• University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[100] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require	Risk reduced	Yes

[100] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

	approval from the ResearchOne Project Committee prior to any data being made available. • Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will have an employment contract with the University of Leeds • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[94] and Medical Research Council[95].		
Research team externally publish data that enables (re-) identification of an individual or group of individuals from the data	• Approval of all outputs derived from data by Principal Investigator. • University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[101] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available. • Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will have an employment contract with the University of Leeds • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[94] and Medical Research Council[95].	Risk reduced	Yes
Data providers retain data beyond the specified end date of the project	• Data Processing Agreement will be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security.	Risk reduced	Yes

[101] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

	• ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~		
Research team retain data beyond the specified end date of the project	• University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. [102] • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[103] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available. • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[94] and Medical Research Council[95]. • Approval from an NHS Research Ethics Committee will included defined project end date.	Risk reduced	Yes

PIA Outcomes

Risk	Approved Solution	To Be Approved By
Data intercepted in transit between organisations by an external party	• Data transferred between parties using secure transfer mechanisms where data is encrypted and transmitted between named individuals in each organisation[91]. • No patient identifiable data are transmitted between organisations[91]. • Patients linked across data providers using pseudonyms generated from NHS numbers[91].	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group

[102] A Data Sharing Agreement between the University of Leeds and NHS Digital has been established for this work in October 2020 (REF: NIC-147997).
[103] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

	• Cryptographic salt file sent separately from other data to each organisation by University of Leeds and used for this specific linkage scenario only[91]. • Pseudonyms only transmitted with ResearchOne-specific identifiers between The Phoenix Partnership and NHS Digital to enable linkage by NHS Digital[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91]. • Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92].	
Data intentionally or unintentionally transmitted to an external party by a data provider	• Data transferred between parties using secure transfer mechanisms where data is encrypted and transmitted between named individuals in each organisation[91]. • No patient identifiable data are transmitted between organisations[91]. • Patients linked across data providers using pseudonyms generated from NHS numbers[91]. • Cryptographic salt file sent separately from other data to each organisation by University of Leeds and used for this specific linkage scenario only[91]. • Pseudonyms only transmitted with ResearchOne-specific identifiers between ResearchOne and NHS Digital to enable linkage by NHS Digital[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91]. • Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92]. • Data Processing Agreement to be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. ~~• Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group
Data intentionally or unintentionally transmitted to an external party by the research team	• Data to be securely managed in an infrastructure at LICTR that is compliant with NHS Data Security and Protection Toolkit[104]. • No patient identifiable data are transmitted between organisations[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91].	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group

[104] See https://www.dsptoolkit.nhs.uk/

	• Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92]. • Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[105] and Medical Research Council[106]. • Any members of research team with access to the data will have an employment contract with the University of Leeds • University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[107] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[108] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available..	
Data is accessed at a data provider by an external party	• Data Processing Agreement to be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. • ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group

[105] See http://it.leeds.ac.uk/info/273/information_security_training/1061/information_security_awareness_training
[106] See http://byglearning.co.uk/mrcrsc-lms/course/category.php?id=1 - "Research, GDPR and Confidentiality"
[107] See http://content.digital.nhs.uk/media/14847/Draft----HSCIC-Data-Sharing-Framework-Contract/pdf/HSCIC_Data_Sharing_Framework_Contract.pdf for sample agreement.
[108] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

Data is accessed at the University of Leeds by an external party	• Data to be securely managed in an infrastructure at LICTR that is compliant with NHS Data Security and Protection Toolkit[93]. • No patient identifiable data are transmitted between organisations[91]. • No pseudonyms generated from patient identifiable data (NHS number) included in the data received by the University of Leeds[91]. • Data specification includes only data items that are necessary and sufficient to robustly answer the research question[92]. • University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[109] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available..	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group
Data used for purposes outside of the research project by a data provider	• Data Processing Agreement to be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. • ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	• Principal Investigator
Data used for purposes outside of the research project by research team	• University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security.	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group

[109] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

	• Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[110] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available. • Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will have an employment contract with the University of Leeds	
Data that is unnecessary for research is requested by the research team	• Data specification will be approved by the Principal Investigator as part of the Ethics Protocol. • Data specification will be approved by an NHS Research Ethics Committee as part of the Ethics Protocol. • Application to NHS Digital for data items included in the data specification must be approved by Information Asset Owner (IAO) for Hospital Episode Statistics (HES) and Independent Group Advising on the Release of Data (IGARD). • Application to ResearchOne for data items included in the data specification must be approved by the ResearchOne Project Committee (RPC).	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group
Data that is unnecessary for research is supplied to the research team by data providers	• Data specification will be formally specified in application for data from NHS Digital (Hospital Episode Statistics). • Data specification will be formally specified in application for data from The Phoenix Partnership (ResearchOne.) • Data Processing Agreement will be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security • ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	• Principal Investigator
Research team intentionally or unintentionally (re-) identify an individual or	• University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security.	• Principal Investigator • University of Leeds (Sponsor)

[110] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

groups of individuals from the data	• Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[111] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available. • Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will have an employment contract with the University of Leeds • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[94] and Medical Research Council[95].	• NHS Research Ethics Committee • Confidentiality Advisory Group
Research team externally publish data that enables (re-) identification of an individual or group of individuals from the data	• Approval of all outputs derived from data by Principal Investigator. • University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[112] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available..	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group

[111] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc
[112] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

	• Any members of research team with access to the data will be approved by Principal Investigator and recorded. • Any members of research team with access to the data will have an employment contract with the University of Leeds • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[94] and Medical Research Council[95].	
Data providers retain data beyond the specified end date of the project	• Data Processing Agreement will be established between University of Leeds and The Phoenix Partnership formalising processing activity required and detailing obligations with respect to information security. • ~~Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.~~	• Principal Investigator
Research team retain data beyond the specified end date of the project	• University of Leeds has a Data Sharing Framework Contract between NHS Digital and University of Leeds[96] which details obligations with respect to information security. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[113] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available. • Any members of research team with access to the data will undertake Information Governance training provided by University of Leeds[94] and Medical Research Council[95]. • Approval from an NHS Research Ethics Committee will included defined project end date.	• Principal Investigator • University of Leeds (Sponsor) • NHS Research Ethics Committee • Confidentiality Advisory Group

[113] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

Updates to PIA Solutions

PIA solutions defined above were included in Version 1 of the Ethics Protocol, which received approval from the Principal Investigator, University of Leeds (Sponsor) and NHS Research Ethics Committee. As detailed previously, following correspondence with the Health Research Authority, NHS Digital and The Phoenix Partnership, it was determined that Section 251 support was not required for the study and the application to the Confidentiality Advisory Group was withdrawn. Therefore, the PIA solutions were not subject to consideration by the Confidentiality Advisory Group.

Within Version 2 of the Ethics Protocol, we include updates to the status of the solutions proposed above. The table below summarises these changes:

Solution	Update
Data Processing Agreement to be established between University of Leeds and NHS Digital formalising processing activity required and detailing obligations with respect to information security.	Following correspondence with NHS Digital in relation to another research project based at the University of Leeds (LP-MAESTRO – IRAS ID: 178391), which uses the same linkage methodology between Hospital Episode Statistics and ResearchOne, it was determined that a Data Processing Agreement between the University of Leeds and NHS Digital was not required (August 2017). UK-SAFE WP2a (RO-HES) proceeded on the same basis using LP-MAESTRO as the precedent. Therefore, this solution has been removed and struck out in the PIA Outcomes section above.
Data Sharing Agreements will be established with NHS Digital and ResearchOne on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security.	NHS Digital explicitly refer to a Data Sharing Agreement. However, the Phoenix Partnership do not explicitly refer to a Data Sharing Agreement. Therefore, for clarity, we now represent the solution using two separate points. • Data Sharing Agreement will be established with NHS Digital on behalf of the University of Leeds formalising the specific terms under which data is made available for the project and detailing obligations with respect to information security. • Project-specific terms under which data will be made available from ResearchOne by The Phoenix Partnership, including terms relating to information security, will be specified by the University of Leeds in the ResearchOne data request form[114] (format defined by The Phoenix Partnership as data controller for ResearchOne) and which will require approval from the ResearchOne Project Committee prior to any data being made available.

[114] See http://www.researchone.org/wp-content/uploads/2014/05/EoI_and_DataRequest_V6.doc

Responsibilities

The table below captures current key actions relating to the project, along with the dates and person(s) responsible for their completion. As stated previously, this document may be subject to amendment over the lifetime of the project in response to consideration by (and feedback from) external advisory bodies, including NHS Research Ethics Committee (REC) and the Health Research Authority's Confidentiality Advisory Group (CAG), and ongoing consideration by (and feedback from) the Project Management Group (PMG) and Independent Advisory Group (IAG). Subsequent actions relating to the project will be detailed in response to this consideration.

Action To Be Taken	Date For Completion Of Actions	Responsibility For Action
Preparation of ethics protocol (and appendices) and completion of relevant entries in IRAS form	13/04/2017	Chris Smith
Approval of IRAS and CAG forms and associated documents (Chief Investigator)	01/05/2017	Philip Conaghan
Approval of IRAS and CAG forms and associated documents (Sponsor's Representative)	01/05/2017	University of Leeds – Faculty of Medicine and Health Research Governance
Approval of IRAS and CAG forms and associated documents (Information Guardian)	01/05/2017	Roger Gair
Submission of application and supporting documents to NHS Research Ethics Committee	22/06/2017	Lema Vernon
Submission of application and supporting documents to Confidentiality Advisory Group	22/06/2017	Lema Vernon
Overall project oversight	Ongoing	Philip Conaghan

Actions above are detailed up to the point at which Version 1 of the protocol was submitted to the Health Research Authority for review by an NHS Research Ethics Committee and the Confidentiality Advisory Group.

Any concerns regarding privacy can be directed to: Professor Philip Conaghan (Principal Investigator) or Dr Chris Smith (Senior Research Fellow/Data Scientist).

Appendix 3a: Re-Identification Attacks

The PIA has identified risks relating to privacy that arise within Work Package 2a (RO-HES) and demonstrated the measures that have been taken to mitigate these risks, whilst ensuring that the aim and objectives of the projects can be fulfilled. However, we acknowledge that there remains a residual risk that data required within the project could be used by an attacker who gains access to the data (either in transit or at rest), and who has the requisite motivation and ability, to re-identify individuals or groups of individuals and to associate sensitive data items within these individuals.

Attack Classes

We describe below two specific classes of attack that may be undertaken and how the measures taken within the project have mitigated the risk of such attacks.

Cryptographic Attack

This attack involves an attacker deriving the NHS numbers used by NHS Digital and The Phoenix Partnership to generate the patient pseudonyms. The attacker could use externally available datasets to resolve one or more of these NHS numbers to other patient identifiable data (e.g. name and address) to re-identify one or more individuals. In isolation, NHS number would enable an attacker to infer membership of an individual or group of individuals in a particular cohort of interest to a study. In conjunction with a ResearchOne ID, and subject to an additional ability to access the data items provided for each patient by The Phoenix Partnership from ResearchOne, the attacker has the potential to these data items with one or more NHS numbers and to therefore associate one or more individuals with these data items.

The attack would require an attacker to gain access to the pseudonyms either:

(1) *in transit* between NHS Digital and The Phoenix Partnership, or

(2) *at rest* within NHS Digital or The Phoenix Partnership

To gain access to a set of pseudonyms *in transit*, the attacker would be required to intercept data on a secure channel provided by the SEFT service (NHS Digital) between The Phoenix Partnership and NHS Digital. Access to the channel is subject to authentication and authorization processes, and data is strongly encrypted (AES-256) for transmission along this channel. Therefore, the risk that an attacker could gain access to this channel and intercept data is considered to be very low.

To gain access to a set of pseudonyms *at rest*, the attacker would be required to gain access to the storage at NHS Digital and The Phoenix Partnership. Such risks reside solely within the control of NHS Digital and The Phoenix Partnership, and should be managed by appropriate systems, standard operating procedures and policies within these organizations. Data processing agreements will be established between NHS Digital and the University of Leeds[115], and The Phoenix Partnership and the University of Leeds[116] to specify obligations of these parties in relation to information security within the project.

---

[115] Following correspondence with NHS Digital in relation to another research project based at the University of Leeds (LP-MAESTRO – IRAS ID: 178391), which uses the same linkage methodology between Hospital Episode Statistics and ResearchOne, it was determined that a Data Processing Agreement between the University of Leeds and NHS Digital was not required (August 2017). UK-SAFE WP2a (RO-HES) proceeded on the same basis using LP-MAESTRO as the precedent.

[116] A Data Processing Agreement between the University of Leeds and The Phoenix Partnership has been established for this work (December 2018).

Let us consider a scenario in which an attacker has gained access to a set of pseudonyms. The attacker must derive the NHS number from which each pseudonym was generated. However, the pseudonym will be generated by applying a strong cryptographic hash function (SHA-512) to a concatenation of a 512 bit cryptographic salt file and NHS number for each patient. This would require either:

(1) access to the salt file, or

(2) ability to derive or consider the salt file

To access the salt file, the attacker would be required to either intercept the salt file in transit, or gain access to the salt file at rest from University of Leeds, NHS Digital or The Phoenix Partnership. The salt file is transferred (in isolation) between the University of Leeds and the other organizations on a secure channel provided by either the SEFT (NHS Digital) service or the SFT (University of Leeds) service. As previously discussed, access to the secure channel is considered to be very low risk. Access to the salt file at rest would require access to the storage at University of Leeds, NHS Digital or The Phoenix Partnership. Data will be securely managed in an infrastructure that is compliant with NHS Data Security and Protection Toolkit at the University of Leeds. As previously described, data processing agreements will be established between NHS Digital and the University of Leeds[117], and The Phoenix Partnership and the University of Leeds[118] to specify obligations of these parties in relation to information security within the project.

To derive or consider a salt file, the attacker would be required to determine the salt from one or more digests. However, the salt file has been chosen to be of a sufficient length (512 bits) to make such derivation via a cryptographic attack sufficiently costly and resource-intensive to be infeasible for an attacker.

Let us consider a scenario in which the attacker has managed to derive the salt file, or has gained access to the salt file, the attacker must then iterate over all possible NHS numbers (or those NHS numbers of interest), to generate a pseudonym by applying the hash function to the salt and each NHS number. This would produce a lookup table of pseudonyms to NHS numbers, by which the corresponding NHS number could be found for any given pseudonym. Generation of such a lookup table is considered to be computationally feasible for most attackers.

However, the data linkage methodology separates transmission of pseudonyms (and their associated ResearchOne IDs) from the data items received from The Phoenix Partnership for these patients. In the event that the pseudonyms were compromised through a cryptographic attack, and corresponding NHS numbers could be derived, these NHS numbers could not be associated with any clinical data without a further attack to gain access to the related data items (either from ResearchOne or elsewhere). Moreover, the methodology uses a cryptographic salt file that is generated and applied for this specific usage scenario. Therefore, the pseudonym generated for a given NHS number in this usage scenario would be different from the pseudonym generated for that NHS number in a different usage scenario (using a different cryptographic salt). This limits the

[117] Following correspondence with NHS Digital in relation to another research project based at the University of Leeds (LP-MAESTRO – IRAS ID: 178391), which uses the same linkage methodology between Hospital Episode Statistics and ResearchOne, it was determined that a Data Processing Agreement between the University of Leeds and NHS Digital was not required (August 2017). UK-SAFE WP2a (RO-HES) proceeded on the same basis using LP-MAESTRO as the precedent.

[118] A Data Processing Agreement between the University of Leeds and The Phoenix Partnership has been established for this work (December 2018).

extent to which pseudonyms obtained in an attack can be combined with those generated in other usage scenarios to assist re-identification.

Jigsaw Attack

This attack involves an attacker gaining access to the data items provided for each patient by NHS Digital and The Phoenix Partnership and using these data items in conjunction with data items from externally available datasets to re-identify an individual or group of individuals and associate these individuals with clinical data.

The attack would require an attacker to gain access to the data items either:

(1) *in transit* between NHS Digital, The Phoenix Partnership and University of Leeds, or

(2) *at rest* within NHS Digital, The Phoenix Partnership or University of Leeds

To gain access to a set of pseudonyms *in transit*, the attacker would be required to intercept data on a secure channel provided by the SEFT service (NHS Digital) between The Phoenix Partnership or University of Leeds and NHS Digital, or on a secure channel provided by the SFT service (University of Leeds) between The Phoenix Partnership and University of Leeds. Access to these channels is subject to authentication and authorization processes, and data is strongly encrypted (AES-256) for transmission along the channels. Therefore, the risk that an attacker could gain access to this channel and intercept data is considered to be very low.

To gain access to clinical data at rest, the attacker would be required to gain access to the storage of The Phoenix Partnership, NHS Digital, or to UoL. Data will be securely managed in an infrastructure that is compliant with NHS Data Security and Protection Toolkit at the University of Leeds. As previously described, data processing agreements will be established between NHS Digital and the University of Leeds[119], and The Phoenix Partnership and the University of Leeds[120] to specify obligations of these parties in relation to information security within the project.

Let us consider a scenario in which an attacker gained access to data items supplied by NHS Digital and/or The Phoenix Partnership. These data items contain no patient identifiable data. Each patient is referenced with a unique (non-personal) identifier in the data from NHS Digital (HES ID) and The Phoenix Partnership (ResearchOne ID). Accordingly, the data items cannot be associated with an individual without further background knowledge. However, within the clinical data that is obtained from the different organizations there is likely to be combinations of specific data items that are unique to a single patient. Using combinations of these data items, in conjunction with externally available datasets, it may be feasible for certain patients to be re-identified.

The set of potential jigsaw attacks are based upon the classes of data item that are included in those supplied by NHS Digital from Hospital Episode Statistics and by The Phoenix Partnership from ResearchOne[121]. These data items contain a number of different classes of data, including:

- *Patient* references (e.g. Unique (non-personal) patient identifier: ABC123)

[119] Following correspondence with NHS Digital in relation to another research project based at the University of Leeds (LP-MAESTRO – IRAS ID: 178391), which uses the same linkage methodology between Hospital Episode Statistics and ResearchOne, it was determined that a Data Processing Agreement between the University of Leeds and NHS Digital was not required (August 2017). UK-SAFE WP2a (RO-HES) proceeded on the same basis using LP-MAESTRO as the precedent.

[120] A Data Processing Agreement between the University of Leeds and The Phoenix Partnership has been established for this work (December 2018).

[121] See Appendix 1

- *Temporal* references (e.g. Arrival Date: 2016-01-01).
- *Geographical* references (e.g. Sector Level Postcode: LS1 7).
- *Organisational* references (e.g. Provider Code: XYZ12).
- *Clinical* references (e.g. Primary Diagnosis: J01).
- *Demographic* references (e.g. Ethnicity: White British).

Data items from one or more classes can be combined to characterise an "event" that is unique (both within and outside the dataset). For example, an event with the following (fictional) values for a set of data items may be unique in the Hospital Episode Statistics (A&E) dataset:

- Arrival Date: 2016-01-01
- Sector Level Postcode: LS1 7
- Primary Diagnosis: J01
- Ethnicity: White British.

Externally available datasets that contain one or more of the same data items as this event may be used to determine additional data items about this event. The values of these additional data items may increase the probability that the event can be attributed to a specific individual up to a point where the event can be uniquely attributed to a specific individual with certainty. Consequently, sensitive data items (e.g. Primary Diagnosis) could be attributed to that individual. Events can also be grouped together using the value of a specific data item to increase the probability with which the events can be attributed to a specific individual. For example, all events with a unique (non-personal) patient identifier of ABC123 could be grouped. Unique patterns can be identified between events in this groups. For example, unique temporal patterns in attendances to A&E. Externally available datasets can be used to associate these unique patterns with a specific individual.

In addition to the classes of data item that are included in the supplied datasets, the set of potential jigsaw attacks are also based on the availability to the attacker of external datasets that contain relevant references. Minimally, it can be assumed that an attacker will have access to any publicly available datasets. For example, data relating to certain organisations or geographical areas that may be published as open data. Additionally, certain attackers may have access to non-publicly available datasets that may be relevant. For example, data acquired through personal experience or available through specific personal circumstances.

Construction of potential attacker models based on knowledge of different externally available datasets, and application of specific generalisation and suppression techniques by NHS Digital and The Phoenix Partnership based on these attacker models and a defined measure of (re-)identifiability would not be feasible within the scope of this project. This would require consideration of a vast number of different permutations and may result in significant degradation of the supplied data to the extent that it becomes insufficient to robustly answer the research question. Accordingly, we acknowledge the risk associated with jigsaw attacks and seek to address this risk with a number of different measures, including use of only those data items that are necessary and sufficient to answer the research question, adherence to high standards of information security through use of a secure infrastructure at the University of Leeds, and establishment of data sharing and data

processing agreements with NHS Digital[122] and The Phoenix Partnership to specify obligations of these parties in relation to information security within the project [123].

[122] Following correspondence with NHS Digital in relation to another research project based at the University of Leeds (LP-MAESTRO – IRAS ID: 178391), which uses the same linkage methodology between Hospital Episode Statistics and ResearchOne, it was determined that a Data Processing Agreement between the University of Leeds and NHS Digital was not required (August 2017). University of Leeds confirmed that LP-MAESTRO could proceed on this basis. This work also proceeded on this basis using LP-MAESTRO as the precedent.
[123] A Data Processing Agreement between the University of Leeds and The Phoenix Partnership has been established for this work.

Appendix Four: Data Processing

Who is undertaking this research project?

UK-SAFE Work Package 2a (RO-HES) is being undertaken by a research team based in the Leeds Institute of Health Sciences:

Leeds Institute of Health Sciences
Level 10, Worsley Building
Clarendon Way
Leeds
LS2 9NL
United Kingdom
Email: lihs@leeds.ac.uk
Web: https://medicinehealth.leeds.ac.uk/homepage/146/leeds_institute_of_health_science

Leeds Institute of Health Sciences is situated within the Faculty of Medicine and Health at the University of Leeds:

Faculty of Medicine and Health
Leeds
LS2 9JT
United Kingdom
Web: https://medicinehealth.leeds.ac.uk

The University of Leeds is registered as a data controller with the Information Commissioner's Office (Registration Number: Z553814X).

The Data Protection Officer for the University of Leeds is David Wardle who can be contacted by email on d.wardle@adm.leeds.ac.uk or by post to University of Leeds, 11.72 EC Stoner Building, Leeds, LS2 9JT, United Kingdom.

What are the purpose(s) for which data are being processed within this research project?

UK-SAFE Work Package 2a (RO-HES) is a discrete sub-project within the UK-SAFE project. UK-SAFE will examine the requirements for arthroplasty follow-up and produce evidence and consensus-based recommendations as to how, when and on whom follow-up should be conducted.

UK-SAFE is funded by the National Institute for Health Research (NIHR) – Health Services and Delivery Research (HS&DR) Programme (REF: 14/70/146).

For further information, see: https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/1470146.

What is the legal basis under which data is processed within this research project?

Under the General Data Protection Regulation (GDPR), a legal basis is required for the processing of personal data. The legal basis under which data is processed within UK-SAFE Work Package 2a (RO-HES) is:

- Article 6(1)(e) – “processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller”

Processing of data concerning health (and other ‘special categories’ of data) requires the fulfilment of an additional condition under GDPR. The specific condition fulfilled by the processing of such data within UK-SAFE Work Package 2a (RO-HES) is:

- Article 9 (2) (j) – “processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and interests of the data subject.”

For further information, see: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/.

Which sources of NHS data are being used?

Work Package 2a (RO-HES) of UK-SAFE will use NHS data from two different sources:

- **NHS Digital (NHSD)**
  NHS Digital is the national provider of information, data and IT systems for commissioners, analysts and clinicians in health and social care. It is responsible for collecting, analysing and presenting national health and social care data. Hospital Episode Statistics (HES) is a dataset that is held by NHSD which includes data on all hospital episodes within all hospitals in England.

  For further information, see: http://www.digital.nhs.uk

- **The Phoenix Partnership (TPP)**
  TPP is a healthcare technology company. ResearchOne is a health and care research database developed by TPP in partnership with the University of Leeds and the UK Government’s Technology Strategy Board. The database consists of clinical and administrative data that have been drawn from the electronic patient records currently held on the TPP SystmOne clinical system.

  For further information, see: http://www.researchone.org/

How is the data being collected?

Work Package 2a (RO-HES) will use NHS data that is collected routinely by clinicians and healthcare professionals as part of direct care. Such data is typically collected within the Electronic Health Records maintained by NHS organizations.

No additional or amended data is being collected for the purpose of Work Package 2a (RO-HES).

Data required for the project will be provided from the data sources detailed above.

What categories of data will be obtained from these sources?

Work Package 2a (RO-HES) will obtain different data from the two sources. The data that will be obtained has been identified by the research team as being necessary and sufficient to robustly answer the research question.

The following data will be obtained:

(1) NHS Digital (NHSD)
Data relating to Accident and Emergency, Inpatient and Outpatient hospital episodes will be obtained for patients with an attendance at a hospital in England that matches specific criteria. This data will include: diagnoses, procedures, patient demographics and provider organizations.

For further information, see: http://content.digital.nhs.uk/hes

(2) The Phoenix Partnership (TPP)
Data relating to primary care will be obtained from ResearchOne for patients with a hospital episode included in the data obtained from NHS Digital. This data will include: diagnoses, prescriptions, referrals, and patient demographics.

For further information, see: http://www.researchone.org/research-faqs/

Will my data be used by the research project?

Data relating to **hospital admissions** will be included in the data obtained for use in Work Package 2a (RO-HES) of UK-SAFE if:

1. You have been previously admitted to a hospital in England

   AND

2. Your admission was in the specific index period chosen for the study between (1st April 2000 and 31st March 2015)

   AND

3. You were aged 18 or over at the time of this admission

   AND

4. Your admission involved one of the following procedures: i) hip replacement, ii) hip revision, iii) knee replacement, or iv) knee revision.

   AND

5. You have not registered an opt-out through the national opt-out programme

Additionally, data relating to **primary care** will be included if:

6. Your hospital admissions data are included (see 1-5).

AND

7. Your General Practice has opted-in to ResearchOne, and you have not opted out of ResearchOne.

How will patient privacy be protected?

Work Package 2a (RO-HES) of UK-SAFE will use *anonymised* data for analysis purposes. This means is that no patient identifiable data, such as name and address, will be included in the data obtained by the research team at the University of Leeds. Instead, patients will be uniquely referenced using non-personal identifiers that can't be decoded by the researchers to re-identify individuals. Additionally, Work Package 2a (RO-HES) will use only that data which is necessary and sufficient to robustly answer the research question.

The data sources (NHSD and TPP) will remove all patient identifiable data and use unique (non-personal) identifiers to reference patients in the data supplied to the University of Leeds. Additionally, pseudonyms will be generated for patients by these sources in such a manner that it enables patients to be consistently identified across different sources of data, without enabling the specific individual to whom the data relates to be determined. NHSD will use these pseudonyms to generate a file that enables the research team to match unique (non-personal) identifiers generated by NHSD to unique (non-personal) identifiers generated by TPP. NHS Digital will provide this file to the research team at the University of Leeds to enable data relating to each patient from the two sources to be linked for analysis.

All data received at the University of Leeds will be securely stored within a secure computer environment that demonstrates compliance with relevant security assurances (see https://www.dsptoolkit.nhs.uk, Reference: ECC0010), and will be managed in accordance with terms agreed with the data controllers of the Hospital Episode Statistics (NHS Digital) and ResearchOne (The Phoenix Partnership). Access to the data will be restricted to a small number of named and approved members of the research team.

What are my rights in relation to the data?

The General Data Protection Regulation provides the following rights for individuals in relation to their personal data:

- The right to be informed
- The right of access
- The right of rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights in relation to automated decision making and profiling

For further information, see: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/.

A summary of these rights and their applicability in respect of the UK-SAFE WP2a (RO-HES) project is provided in the table below.

Right	Applies to UK—SAFE WP2a (RO-HES)	Rationale
The right to be informed	Yes	Details regarding the project and its use of data is provided within this privacy notice. Additional information regarding the project can be found on the website of the research funder: https://www.journalslibrary.nihr.ac.uk/hsdr/1470146.
The right of access	No	The research team at the University of Leeds are not able to identify specific data subjects in the data provided by NHS Digital and The Phoenix Partnership. The right of access does not apply to UK-SAFE WP2a (RO-HES) as the research team are unable to determine the data that relates to a specific data subject.
The right of rectification	No	The research team at the University of Leeds are not able to identify specific data subjects in the data provided by the NHS Digital and The Phoenix Partnership. The right of rectification does not apply to UK-SAFE WP2a (RO-HES) as the research team are unable to determine the data that relates to a specific data subject.
The right to erasure	No	The research team at the University of Leeds are not able to identify specific data subjects in the data provided by the NHS Digital and The Phoenix Partnership. The right of erasure does not apply to UK-SAFE WP2a (RO-HES) as the research team are unable to determine the data that relates to a specific data subject.
The right to restrict processing	No	The research team at the University of Leeds are not able to identify specific data subjects in the data provided by the NHS Digital and The Phoenix Partnership. The right to restrict processing does not apply to UK-SAFE WP2a (RO-HES) as the research team are unable to determine the data that relates to a specific data subject.
The right to data portability	No	The research team at the University of Leeds are not able to identify specific data subjects in the data provided by the NHS Digital and The Phoenix Partnership. The right to data portability does not apply to UK-SAFE WP2a (RO-HES) as the research team are unable to determine the data that relates to a specific data subject.

The right to object	No	Your data will not be included in the study if you have registered an opt-out through the national opt-out programme prior to the point at which the study population is determined by NHS Digital. For further information, see: https://digital.nhs.uk/services/national-data-opt-out-programme. If you have not registered an opt-out through the national opt-out programme and you fulfil the criteria of the study population (see "Will my data be used by the research project") then your data relating to hospital admissions will be included in the study. If your data relating to hospital admissions is included in the study, your primary care data will not be included in the study if you have registered an opt-out from the ResearchOne database with your General Practice prior to the data being extracted from the ResearchOne database for the study. For further information, see: http://www.researchone.org/documentation/. The research team at the University of Leeds are not able to identify specific data subjects in the data provided by NHS Digital and The Phoenix Partnership. Once the data has been received by the research team from NHS Digital and The Phoenix Partnership, the right to object does not apply to UK-SAFE WP2a (RO-HES) as the research team are unable to determine the data that relates to a specific data subject.
Rights in relation to automated decision making and profiling	No	The analysis performed by the research team will attempt to determine how, when and on whom follow-up for joint replacement should be conducted. Patients referenced in the data will be analysed in respect of their health by automated means to understand variation in follow-up for hip and knee replacement over time across different hospitals. The analysis performed by the research team will not produce automated decisions about the specific patients who are referenced in the data. Analysis will informed recommendations regarding the timing of follow-up and the identification of the most cost-effective follow-up model. It is anticipated that a policy document will be created by the project that will include a stratification algorithm to determine appropriate follow-up for a patient, taking into account a variety of factors, such as age.

What approvals will be obtained for the research?

Data will be retained by the research project for two years after the end of the project to facilitate publication of reports.

What approvals will be obtained for the research?

UK-SAFE WP2a (RO-HES) has successfully obtained:

1. Favourable opinion from the Leeds East NHS Research Ethics Committee (REF: 17/YH/0250).
2. Decision based on correspondence with the Health Research Authority (REF: 17/CAG/0122), NHS Digital (as data controller for Hospital Episode Statistics) and The Phoenix Partnership (as data controller for ResearchOne) that Section 251 support is not required.
3. Approval from the Independent Group Advising on the Release of Data (IGARD) at NHS Digital (REF: NIC-147997).
4. Approval from the ResearchOne Project Committee (REF: R1_2020_E_02_UK_SAFE_Extended).

What if I have further questions?

Any questions regarding UK-SAFE WP2a (RO-HES) can be directed to the research team at the Leeds Institute of Health Sciences, who can be contacted using the details provided above. You have the right to lodge a complaint with the Information Commissioner's Office (https://ico.org.uk) who is the supervisory authority for information rights in the United Kingdom.