

Fair Processing Information for the Education and Diabetes Project

About this document

This fair processing information document (also known as the privacy notice or patient notification) has been developed specifically for a Cardiff University research project. It is important because it concerns the processing of confidential patient information connected with health records which are to be processed for use in research without patients' consent. This document explains how the organisations responsible for looking after the source datasets can share them with Cardiff University for our research. It also includes information on how patients can opt-out of having their data included in the study.

This document was last updated on 29/4/2022 (version 3.1). Please email the project lead, Rob French, at frenchr3@cardiff.ac.uk with any queries relating to different versions of this document, if you think some information is missing, or have suggestions on how we could present this information better.

Essential info repeated on the front page

Our patient panel requested that we include the information on three key topics at the top of this fair processing document:

Project information: Further information on the whole research project (going beyond the data processing) is included on the [Cardiff University project website](#), though a slightly more user-friendly source is the [page on the current funder's website](#) ([Welsh version here](#)) or the [blog at the start of the latest phase of this work](#).

How patients can 'opt-out' their data from the research: Patients can opt-out in three ways: (i) by opting out of all health data research using the national opt-out, (ii) by opting out of specific administrative datasets by contacting the controllers of those datasets, or (iii) opting out of our specific research study by contacting our research project before the data provider extracts the data. Please note that once the processors delete the real-world person IDs (Name, date of birth, NHS number) from the data, it is not possible for us to re-identify and remove people from the study.

Patients' views on this data processing: Researchers must consult people with diabetes on how they plan to use their data for research. Information about how we consulted young people is provided on the [webpage for the first 2021 PPIE workshop](#), the [workshop report](#), and the [blog from that event by Charlotte Austin, the Diabetes UK lead for Type 1 diabetes](#). We have also recorded the presentations given in our patient engagement sessions, including [Alex Bailey from the Medical Research Council Regulatory Support Centre speaking about using health data for research](#), [Lucie Burgess speaking about the Diabetes UK perspectives on data](#), [Rob French speaking about the specific data usage in the research study](#), and also [feedback from three of the participants](#). In the second workshop, the young people produced a video explaining how researchers use data relating to young people with diabetes. Although the video and the report produced from that workshop are not yet released, there is a version on [YouTube](#). One can also view the [video for the paediatric diabetes audit](#) from that session. We are conducting further patient engagement this year; the 'official' website for that work is [here](#). Please look up the project on the [DUK website page](#) or contact the patient engagement lead, Karen Rigby, directly at steadfast@diabetes.org.uk to get involved.

Introduction

This fair processing document aims to inform research participants about how their administrative health and education records will be processed and used for our research study in a fair, legal, and ethical way. This document will be made available to participants via the [project website](#). Cardiff University will use the data for a research project which tracks the education and health of young people with diabetes; we show the data linkage mapped against age in Figure 1 below

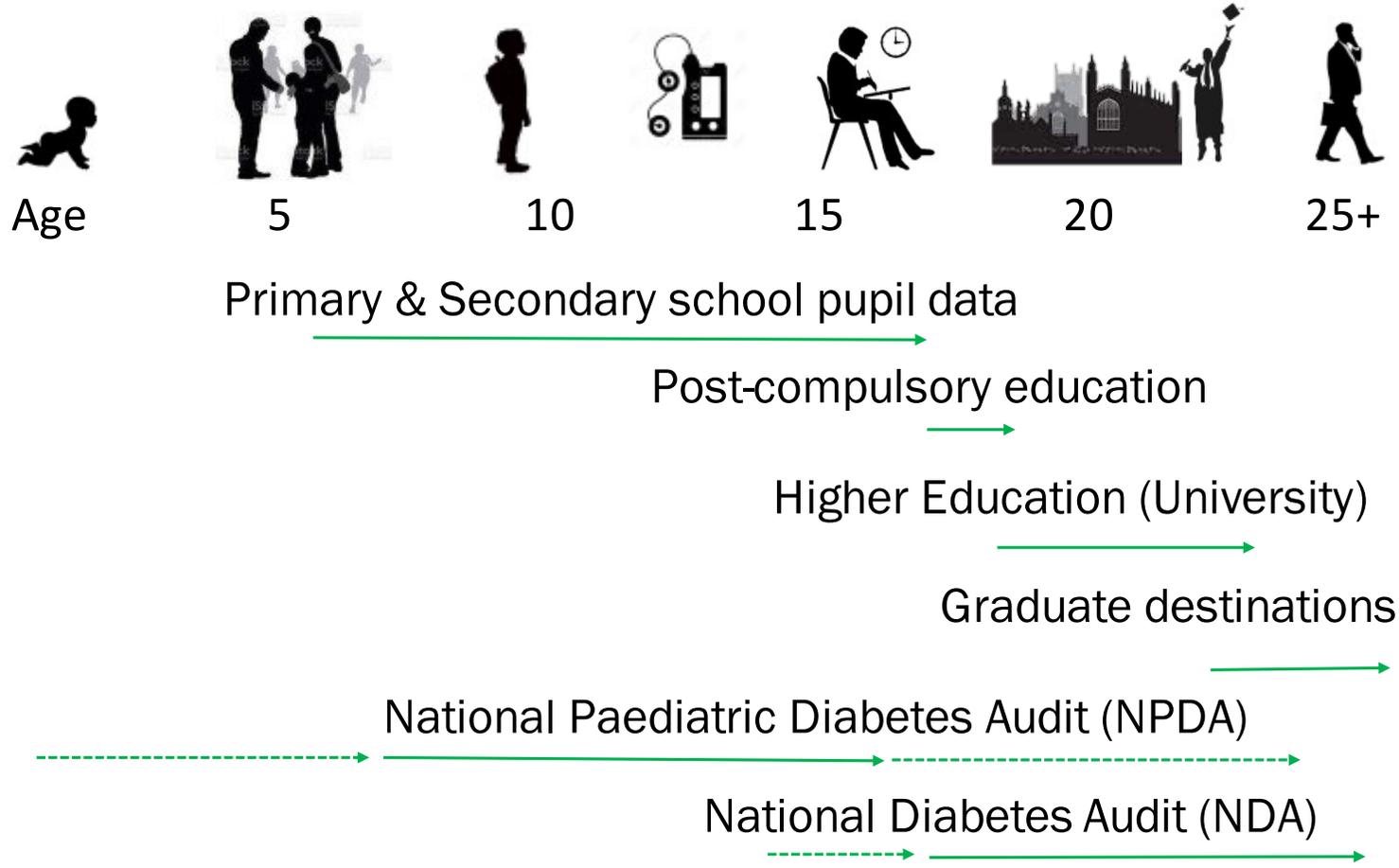
The fair processing for this project relates to the flow of identifying information (NHS number, name, postcode, date of birth, gender) from the health data providers NHSD (NHS Digital) and RCPC (Royal College of Paediatrics and Child Health) to the data processors.

The project has four data processors who use the confidential patient to create a deidentified linkage ID on behalf of the data providers. For the Welsh linkage, the processor is the NHS Wales Informatics Service (NWIS). For the English linkage, the processors include the Department for Education (DfE), NHS Digital Leeds Data Production Team (NHSD-LDP), and the Office for National Statistics Data Access Platform (ONS-DAP). We summarise the flows in Figures 2 to 7 below. These flows are for the data processors to create a unique but deidentified ID to facilitate linking administrative data for health research.

There are three legal justifications for using the data:

1. Because the data shared for linkage are 'personal data', the legal basis for our project to use such data is ['Article 6 1 \(e\)'](#) (Public Task). In other words, our legal basis is for Cardiff University (a public authority) to perform a task (research) in the public interest.
2. Because those parts of the data related to health are 'special category' (more sensitive) personal data, the legal basis for our to process such data is ['Article 9 2 \(j\)'](#) (Processing in the Public Interest). In other words, the processing will support a public benefit, and appropriate safeguards are in place.
3. The legal basis for processing patient identifiable information without consent is the ['Section 251'](#) exemption from the common law duty of confidentiality. The Section 251 approval letters for this are available on the [project webpage](#). More information is given on the specific flows which relate to this in the 'How data will be flowed for linkage' section.

Figure 1: Age Coverage for each of the linked datasets



Aims of the project and data linkage

The overall aim of the research is to quantify the associations between educational outcomes (e.g. school absence) and health outcomes (e.g. blood glucose levels). The benefit to health and social care is the providing of evidence to practitioners to help them to take account of the links between education and diabetes management to help inform differential care, for example, for those with low educational attainment meeting the challenges of independence after leaving compulsory schooling or those struggling to cope with self-management amongst the other pressures of moving to university.

Objectives of the project and data linkage

The Health Research Authority (HRA) ensures good governance for research studies in health and social care research and is responsible for the process under which we seek to share confidential patient information. This approved process facilitates the deidentified linkage across databases. Specifically, we will link the diabetes-related health measures from diabetes audits to educational records of individuals collected by schools and universities to quantify (i) how trajectories of diabetes management (HbA1c measures) are associated with educational outcomes (ii) how educational outcomes affect trajectories of diabetes management (iii) jointly model both education and health trajectories to quantify the effects of common factors.

To link datasets, the data providers will share identifiers (NHS number, name, date of birth, gender, and postcode) with the data processors to generate deidentified linkage IDs to facilitate linkage in data linkage repositories summarised in Figures 1 and 2 below. There are two data linkage repositories in this study where the analysis of the deidentified linked data will take place. The Welsh data linkage repository is the 'Secure Anonymised Information Linkage Databank (SAIL) at Swansea University. The English data linkage repository is the Office for National Statistics – Secure Research Service (ONS-SRS).

The data is shared between the data provider and linkage repository in two parts to separate the health information from the IDs of the person. First, the patient identifiers are not sent directly to the data linkage repository, only to the processor, who uses them to create a privacy-preserving linkage ID. Once the processors have assigned the linkage ID to a dataset, they delete the real-world IDs as soon as possible, typically within three weeks of receiving the data. The data providers share the substantive health information directly to the data repository; there are no real-world person IDs in that dataset. The data repositories check the linked data before it

is shared with the researcher for analysis to ensure that the linkage has not inadvertently made cases re-identifiable.

How data will be flowed for linkage

The aim of moving data (known as data flows) is to bring together schools data, university data, paediatric diabetes audit data, and adult diabetes audit data. Unfortunately, due to technical and information governance constraints, this cannot be done in a single data repository but must be done in two repositories. The first has all the data except English schools data; the second has all the data except Welsh schools data. Researchers use statistical techniques to combine the analysis from the two repositories into a single model.

Figure 2 shows how the different datasets (except English schools data) will flow into the first repository. The dashed arrows coming from the top show the flows of substantive health and education data; these do not include real-world identifiers directly relating the data to the individual concerned. The solid arrows show the flows of identifiers containing information that directly identifies individuals (NHS number, name, date of birth and postcode). These identifier datasets do not contain any substantive data alongside the identifiers. To flow the identifiers relating to health data requires the research team to get a special approval (known as a Section 251 exemption), ensuring sufficient safeguards and patient information are in place to protect the patient's privacy. One of the safeguards is using another organisation to replace the real-world IDs with a project-specific ID which cannot be traced back to the original real-world IDs. These project-specific IDs are reunited with the substantive data in the repository using an additional common ID created by the original data provider.

Figure 3 shows how the different datasets (except Welsh schools data) will flow into the second repository. Again, the dashed arrows show substantive health and education data flows, but no real-world identifiers directly relating the data to the individual concerned. The flows of real-world identifiers shown by the solid arrows present a greater challenge than the equivalent flows for the first repository shown in Figure 2. The paediatric diabetes audit does not contain the patient's name, but the processor used to create the project-specific linkage ID from the real-world IDs cannot process NHS numbers, so we use NHS Digital to add each person's name to the list of real-world IDs. As for the first repository, Cardiff University must obtain support from HRA (Section 251) to ensure the proper safeguards and information sharing are in place. In our application for Section 251 support, we included some alternative processors for the real-world IDs to check if they could do a better job; however, we are not testing those alternative approaches yet.

Figure 2: Data flows (excluding English schools data) into the first data repository

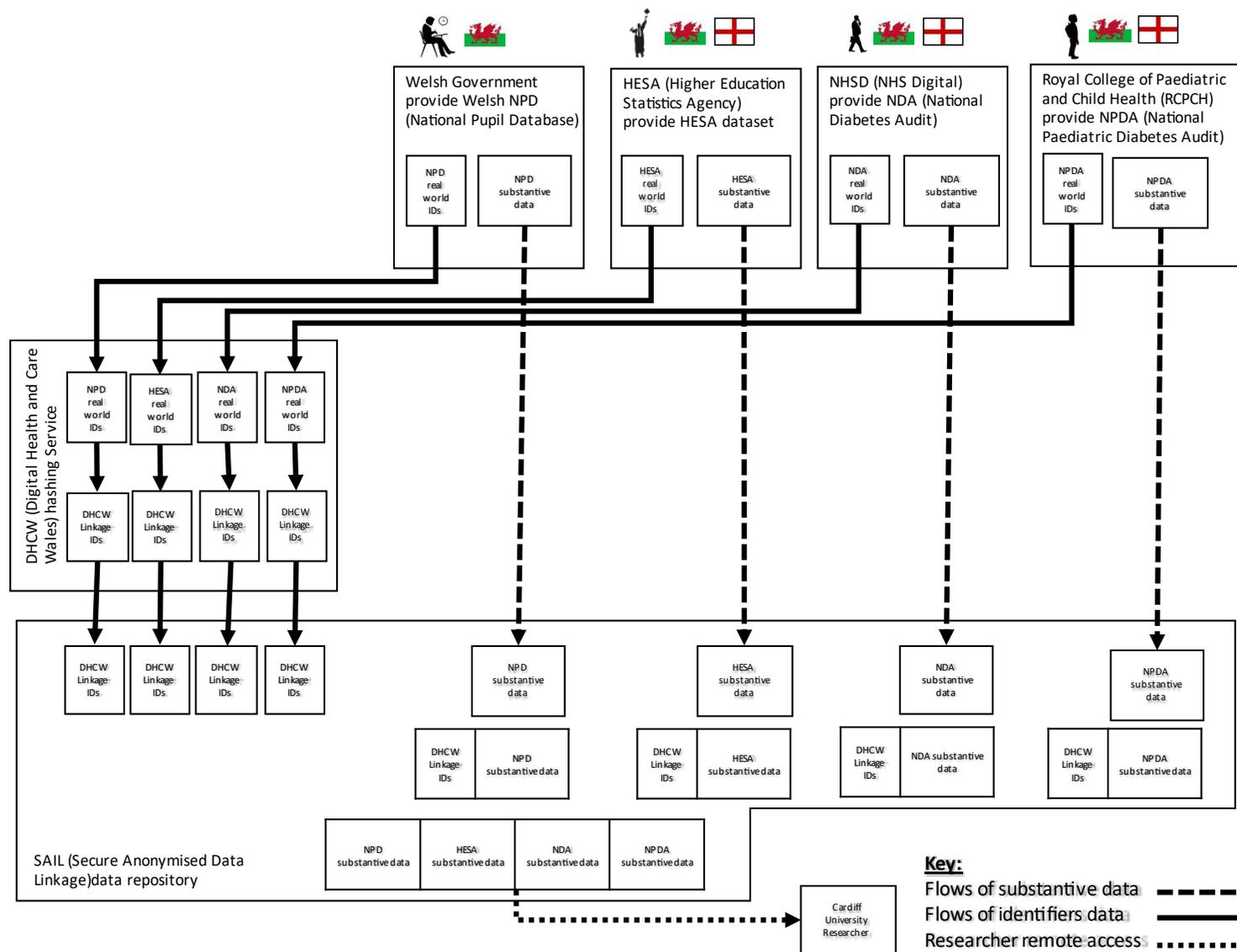
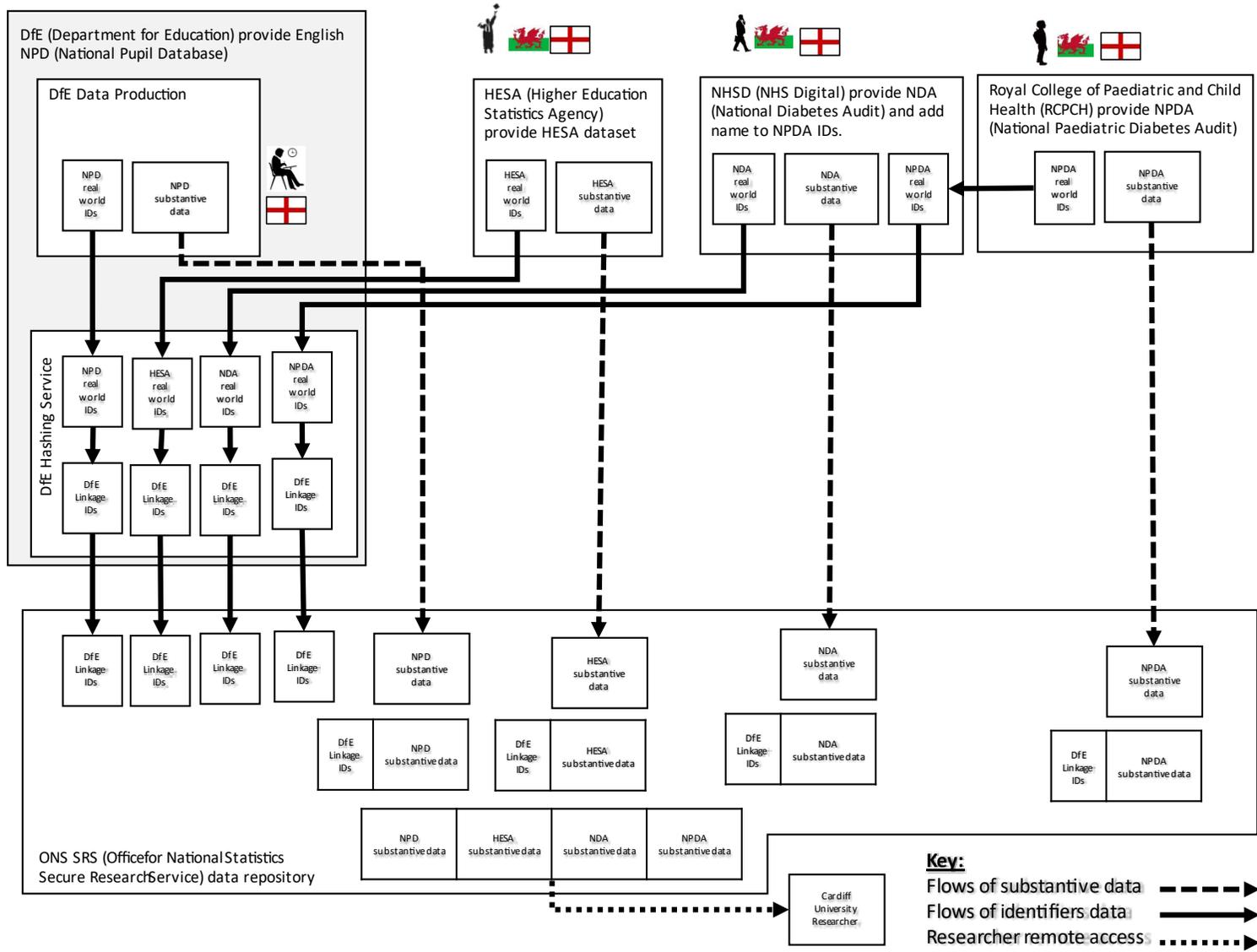


Figure 3: Data flows (excluding Welsh schools data) into the second data repository



The identity of the researcher and their relationship with the University/department

The Data Controller for the project is The University of Cardiff. The principal investigator is [Dr Rob French](#), Senior Research Fellow in the School of Medicine at Cardiff University. The primary statistician working on the data will be the lead researcher Robert French, other researchers within the project team may also work on the data, but this work has not yet been allocated to named individuals. All staff accessing data must meet the stringent requirements set out by the Office for National Statistics. Though person identifiers are shared to enable linkage of the datasets, identifiers are deleted before the data reaches the repository, so it is technically impossible for these identifiers to be made available to any researchers, including the lead researchers.

Opting out of the study

Participants can opt-out of any combination of the datasets for which identifiers will be flowed for linkage (i) the National Paediatric Diabetes Audit (ii) The National Diabetes Audit (iii) Higher Education Statistics Agency Dataset (iv) Welsh Government schools data (v) DfE English schools data. Opting-out can be done directly through the relevant audit processes, with full details on each data provider's website. For the National Paediatric Diabetes Audit, a patient can opt-out by informing the staff at the child's Paediatric Diabetes Unit. To opt out of the National Diabetes Audit, the patient can inform their GP or the secondary care unit responsible for the patient's diabetes management. To opt out of the Higher Education Dataset, a person can inform the Higher Education Statistics Agency (customer.services@hesa.ac.uk). For opting out of the Welsh schools data, a person can directly contact the Welsh Government (customerhelp@gov.wales). To opt-out of the English schools data, a person can directly contact the DfE (data.sharing@education.gov.uk). The information on specific datasets is up to date at the time of writing, but things can change, and we may not be made aware of changes; please contact our project for help if you are struggling with this option.

One can opt-out by contacting the research project; we will forward the request to the relevant data provider. To opt-out in this way, patients should contact the administrative lead for the research project: Mrs Lynne Hope, University Hospital Wales, 02920 742182 hopelm@cardiff.ac.uk or the academic lead for the project: Dr Robert French, University Hospital Wales 02920 743703, frenchr3@cardiff.ac.uk.

If you do not wish to join the study or choose to withdraw from the study, this will have no impact on your medical care. Opting out of this study must happen prior to the data flows from the data providers to the data processors (this has already occurred for Wales and is scheduled for May 2022 for England, there will be annual refreshes each November). After the processing takes place, the identifiers dataset will have flowed to the processors to create a linkage ID. Beyond this point, the processors do not retain any identifiable information, so there is no scope for re-identifying and removing people from the study.'

The Cardiff University Data Protection Officer is Andrew Lane; he can be contacted by email at lanea5@cardiff.ac.uk or by phone at 029208 74164.

If you wish to report any issues or make a complaint regarding this study that cannot be resolved with the project lead or the Cardiff University Data Protection Officer, please contact the [Health Research Authority Confidentiality Advisory Group](#) at cag@hra.nhs.uk or 0207 104 8100.