Data Protection Impact Assessment

QResearch

Prepared by: Julia Hippisley-Cox (j.hippisley-cox@qmul.ac.uk)

Version: 1.0

QResearch DPIA

Document Owner:	**Julia Hippisley-Cox**
Status:	**Final**
Version:	**1.0**
Current Version Date:	**02.10.2025**
Review date:	

Revision History			
Author	Description	Version	Date
Julia Hippisley-Cox	1st version following review and approval by Paul Smallcombe, records and information compliance manager, Queen Mary University of London.	1.0	02.10.2025

Final Approval		
Name - Position	Version	Date

Introduction

You should fill out this document at the start of any major project involving the use of personal data, or if you are making a significant change to an existing process.

To learn more about the process of completing a DPIA, please refer to the accompanying guidance document provided on the following webpage: https://www.qmul.ac.uk/governance-and-legal-services/governance/information-governance/data-protection/data-protection-impact-assessments/

Step 1: Identify a Need for a DPIA

Explain broadly what the project aims to achieve and what type of processing it involves. You may find it helpful to refer or link to other documents, such as a project proposal. Summarise why you identified the need for a DPIA.

Project Aim

QResearch (www.qresearch.org) is an de-identified medical research database which is used to answer key research questions relevant to the delivery and care of patients in the NHS including the development and validation of widely used clinical prediction tools such as QRISK, QDiabetes, QFracture, QAdmissions which are recommended by NICE for use across the NHS.

QResearch GP data is linked to NHSE secondary care data and QMUL provides secure access to de-identified linked data for academics employed by UK universities who are undertaking research projects of benefit to the NHS. Research is also undertaken to evaluate the uptake, safety and effectiveness of medicines, risks and benefits of interventions, natural history of disease and health inequalities. QResearch is also used to undertake health inequalities research to determine whether there are differences between different groups (e.g. by age, gender, ethnic group, deprivation, geography) in in natural history of disease, survival from different conditions and also to evaluate access to, uptake, safety or effectiveness of health care interventions. QResearch is also used to undertake commissioned research. For example, the NHS or DHSC or its arms lengths bodies can commission a research project to answer policy questions. Research data is only accessed by researchers employed by UK universities.

Type of processing involved

At the point of extraction of personal data from the GP systems via the IM1 platform, the data are immediately de-identified. The NHS number is scrambled using a one-way hashing algorithm to create a de-identified code which cannot be reverse. The NHS number is then removed. The postcode is converted to a deprivation score and then removed. All free text comments, names, addresses, emails, contact phone numbers are removed from the data. Dates of birth are rounded to years of birth. Only coded data with a de-identified code which cannot be reversed, is then stored in the QResearch database. All personal data are removed/deleted at the first opportunity at the point of extraction after de-identification.

Need for DPIA

The DPIA screening tool and associated discussion with Paul Smallcombe at Queen Mary University of London on 26.9.2025, concluded that DPIA is not required. However, this DPIA has been completed for reassurance purposes.

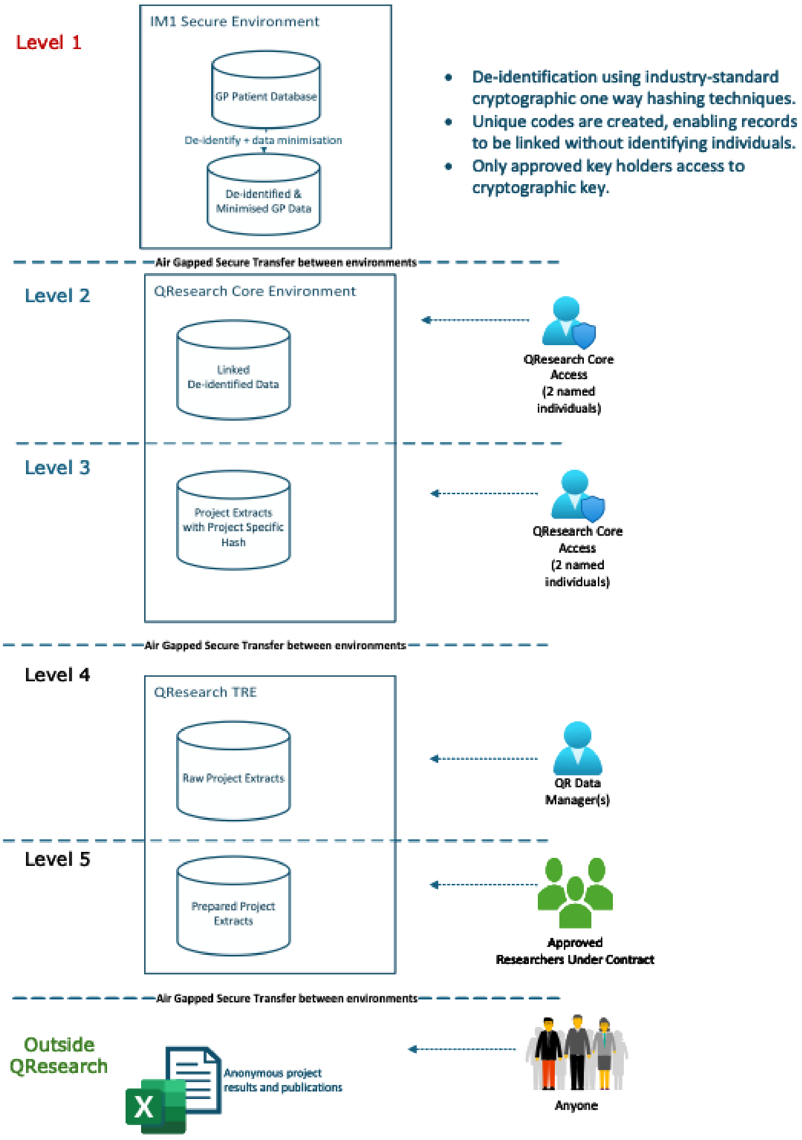
Step 2: Describe the Processing

Describe the nature of the processing: how will you collect, use, store and delete data? What is the source of the data? Will you be sharing data with anyone? You might find it useful to refer to a flow diagram or another way of describing data flows. What types of processing identified as likely high risk are involved?

Data Source

The source of the data is medical health records from general practices in England which contribute to the QResearch database. Figure 1 shows the flow of data and the different levels of access

Figure 1 Data Flow diagram


Level 1
IM1 Secure Environment
GP Patient Database
De-identify + data minimisation
De-identified & Minimised GP Data
De-identification using industry-standard cryptographic one way hashing techniques.
Unique codes are created, enabling records to be linked without identifying individuals.
Only approved key holders access to cryptographic key.
Air Gapped Secure Transfer between environments
Level 2
QResearch Core Environment
Linked De-identified Data
QResearch Core Access (2 named individuals)
Level 3
Project Extracts with Project Specific Hash
QResearch Core Access (2 named individuals)
Air Gapped Secure Transfer between environments
Level 4
QResearch TRE
Raw Project Extracts
QR Data Manager(s)
Level 5
Prepared Project Extracts
Approved Researchers Under Contract
Air Gapped Secure Transfer between environments
Outside QResearch
Anonymous project results and publications
Anyone

Data Collection and processing

Data are collected from the GP practices via the IM1 platform provided by the GP system suppliers (Optum and TPP). The data are de-identified at the point of collection and no personal data is stored.

- Unique codes, known as GUIDS, are generated using one-way industry-standard cryptographic hashing techniques that enable records to be linked without identifying individuals
- Only NHS England approved key holders have access to the cryptographic keys, ensuring that patient identities remain protected. Approved key holders do not have access to the QResearch database.
- Postcodes are removed and replaced with a deprivation score.
- All free text comments, titles, names, addresses, emails, mobile numbers, telephone numbers, IP addresses, booking references, scanned images, medical drawings, letters and other attachments are removed from the data
- Dates of birth are rounded to year of birth.
- Only de-identified coded data with unique de-identified GUIDS are then stored in the QResearch database.

To further enhance security, QResearch applies a second layer of hashing that is project-specific when data are shared with researchers for each research project. This means that even if the same de-identified data is used across multiple research projects, each dataset is uniquely hashed for its intended purpose. This additional cryptographic transformation ensures that data cannot be cross-linked between projects, significantly reducing the risk of re-identification and reinforcing the principle of data minimisation

Data Storage and Use

Once the data has been de-identified, the data are stored in a MS SQL database on Queen Mary University of London owned servers and used for observational medical research.

Data Deletion

Personal data are deleted immediately the datafiles have been anonymised at the point of data extraction.

Sharing of Data

No personal data are shared with other organisations. Only de-identified data from the medical research database are shared with academic researchers that have ethical approval to undertake research projects who are based in the UK.

High risk types of processing

None

Describe the scope of the processing: what is the nature of the data, and does it include special category or criminal offence data? How much data will you be collecting and using? How often? How long will you keep it? How many individuals are affected? What geographical area does it cover?

Nature of Data

The data is health record data from patients registered with GP practices in England that contribute to the QResearch database. Whilst the GP data extract includes special category data, the method of processing means it can’t be linked to an individual person. There are no criminal offence data or genetic data. Whilst there may be genetic diagnostic codes in the GP record, for example, stating that someone has a disease with a genetic component (e.g. Down’s syndrome) but there is no access to an individuals’ complex genome or actual gene sequencing as these are not stored in GP records. Similarly, there is no criminal records except where the GP has used a SNOMED-CT code to record an offence.

Ethnic origin data is stored in the GP records and this is required for analysis and research purposes.

Amount of Data Collected and Used

The minimum data is collected for the purpose stated above. The data are de-identified and anonymised at the point of data extraction i.e. at the very first opportunity. Only the de-identified anonymised data will be stored. Any personal data will be deleted as soon as the anonymisation has been done, again at the first opportunity.

Data Collection Frequency

The data will be collected once or twice each year.

Number of Individuals Affected

The QResearch database includes approximately 40 million patients.

Geographical Area

England.

Describe the context of the processing: what is the nature of your relationship with the individuals? How much control will they have? Would they expect you to use their data in this way? Do they include children or other vulnerable groups? Are there prior concerns over this type of processing or security flaws? Is it novel in any way? What is the current state of technology in this area? Are there any current issues of public concern that you should factor in? Are you signed up to any approved code of conduct or certification scheme (once any have been approved)?

Nature of Relationship with Data Subjects

The GP practices are the original data controller for the personal data. Optum and TPP are data processors for the GP practices to make the data available on the IM1 platform. QMUL is the data controller for anonymisation process at the point of extraction from the IM1 platform. QMUL is also the data controller for the resulting anonymised medical research database (QResearch). QMUL has no relationship with the individual patients. All data in QResearch are de-identified and it is not possible for researchers to re-identify this data.

Individuals’ Control

Individuals are able to opt out of their GP sharing their health records for research and planning by completing a type 1 opt out form (https://www.nhs.uk/using-the-nhs/about-the-nhs/opt-out-of-sharing-your-health-records/).

If a patient has registered a Type 1 Opt-out with their GP practice where they are currently registered, none of their data will be made available or included in the data extraction. This approach to uphold the Type 1-Opt out has been discussed nationally by NHS England with stakeholders and has the support of the British Medical Association and Royal College of General Practitioners.

Individuals’ Expected Use of Data

There is a privacy notice which is available either from the GP practice or via the QResearch website which means people will know about the research uses of their data. https://www.qresearch.org/information/patient-information-and-privacy-notice/

Children and Other Vulnerable Groups

Whilst data will be collected on patients of all ages (including children and vulnerable groups) it is not possible to identify individuals from the data.

Prior Concerns and Security Flaws

There have been no prior issues of concern or security flaws

Novelty of Processing

There is no novel of processing. Other academic institutions and organizations undertake similar activities.

Relevant Processing Technologies

Industry standard cryptographic techniques are used to encrypt the NHS number to create an identifier which is unique to QResearch and cannot be reverse engineered and hence does not identify an individual. Only NHS England approved key holders have access to the cryptographic keys, ensuring that patient identities remain protected. Approved key holders do not have access to the QResearch database.

Potential Current Issues of Public Concern

There are no current issues of public concern. We have patient and public representatives on the QResearch Advisory Board and QResearch Scientific Committee and these have not raised any concerns. Indeed, the public are likely to be reassured about the care we are taking to ensure that users meet the criteria for database access.

Details of any Approved Code of Conduct and Certification

All staff accessing data need to have completed information governance and research training. QResearch is also governed by the NHS England Data Security Protection Tool Kit https://www.dsptoolkit.nhs.uk/OrganisationSearch/8HX86

Describe the purposes of the processing: what do you want to achieve? What is the intended effect on individuals? What are the benefits of the processing – for you, and more broadly?

Desired Outcome of Processing

Benefits of Processing

The purpose of the processing is to anonymise data efficiently, securely and effectively so that only de-identified data are stored in the medical research database. The mechanism for de-identifying the NHS Number is also applied to other NHS data sets by NHS England which are supplied to QResearch such as hospital and mortality data. This mechanism enables the GP data, one de-identified, to be linked to hospital and mortality data.

The benefit is then that the data can be used for medical research to generate new knowledge to improve understanding of disease or the management of patient care in the NHS.

Effects on Participants

Whilst there is no direct effect or benefits on participants (as participants cannot be identified from the resulting database), patients can benefit from the results of medical research using their de-identified data.

Step 3: Consultation Process

Consider how to consult with relevant stakeholders: describe when and how you will seek individuals' views – or justify why it's not appropriate to do so. Who else do you need to involve within Queen Mary? Do you need to ask any processors to assist? Do you plan to consult information security experts, or any other experts?

Consultation with individuals Involved within QMUL

We have consulted

- Gerry Collins, Head of Contracts at JRMO, QMUL
- Maria Frovola, Associate Commercialisation Manager, Queen Mary Innovations
- Mr Paul Smallcombe, Records & Information Compliance Manager, QMUL
- Mr Michael Garvey Eckett, Data and AI Research Governance Lead

Organisations outside of QMUL

We have consulted with

- QResearch Advisory Board which has patient and professional representation.
- Richard Langley - Head of Information Governance, NHS England
- The IM1 team at NHS England
- Dr Tom Nichols, RCGP Joint Chair of the Joint Committee on IT, representing the GP profession.
- Optum (GP system supplier)

Consultation with Information Security Experts

We have consulted with information security experts Dancing Houses Consulting Ltd who provide IT software and database support for QResearch under contract with QMUL.

Step 4: Assess Necessity and Proportionality

Describe compliance and proportionality measures, in particular: what is your lawful basis for processing? Does the processing actually achieve your purpose? Is there another way to achieve the same outcome? How will you prevent function creep? How will you ensure data quality and data minimisation? What information will you give individuals? How will you help to support their rights? What measures do you take to ensure processors comply? How do you safeguard any international transfers?

Lawful Basis for Processing

The lawful bases for processing personal data for this study are **Article 6(1)(e) of UK GDPR**, which states the processing is necessary to perform a task in the public interest and **Article 9(2)(j) of UK GDPR**, which states the 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 the Data Protection Act 2018. This is supported by the University Charter which states research as a purpose of the organisation.

In addition, because the research requires access to identifiable NHS patient information without consent at the point of extraction in order to de-identify it before storing the data in the QResearch database, we will apply for **section 251 support** from the Confidentiality Advisory Group to set aside the common-law duty of confidentiality. Any personal data will be handled in accordance with the UK GDPR and the Data Protection Act 2018

Describe how Purpose Achieved

The purpose is achieved by extracting medical data from GP practice, anonymising at the point of extraction and then storing the data in a research database so that the data can be analysed by researchers.

Function Creep

The QResearch database has defined ethics committee approval, governance processes (including an oversight Advisory Board) and data sharing agreements which limit the purposes for which the data can be used. These measures will prevent function creep.

Data Quality

Data gathered by GP practice as part of routine care. There are some data quality controls at the point of data entry in the GP systems. The IM1 platform, which is used for data extraction, has been extensively tested by NHS England and organisations which use the platform ensuring data integrity. Data quality checks are also run on the resulting anonymised data to ensure data has transferred correctly.

Data Minimisation

Only the minimum data needed for research purpose are stored in the database.

Informing Data Subjects

GP practices display transparency notices in their waiting rooms and on their websites to inform patients their data is being used for research. There is also a transparency notice on the QResearch website.

https://www.qresearch.org/information/patient-information-and-privacy-notice/

Supporting Data Subjects' Rights

This is not applicable as patients cannot be identified.

Ensuring Compliance with GDPR

GDPR compliance is ensured by only collecting minimal, none-sensitive data; ensuring information is available through a privacy notice; ensuring users can access their information to check accuracy and request updates and deletions; ensuring there is a regular review of this policy at least annually. Only approved personnel with training can access data.

Safeguarding International Transfers

There are no international transfers of personal data. All the data reside in the UK at all times.

Risk Key					
Likelihood	**Impact**				
	1 - Negligible	2 - Minor	3 - Moderate	4 - Major	5 - Catastrophic
1 – Rare	**Low** **1**	**Low** **2**	**Low** **3**	**Low** **4**	**Low** **5**
2 – Unlikely	**Low** **2**	**Low** **4**	**Low** **6**	**Medium** **8**	**Medium** **10**
3 – Possible	**Low** **3**	**Low** **6**	**Medium** **9**	**Medium** **12**	**High** **15**
4 – Likely	**Low** **4**	**Medium** **8**	**Medium** **12**	**High** **16**	**High** **20**
5 – Almost Certain	**Low** **5**	**Medium** **10**	**High** **15**	**High** **20**	**High** **25**

Risk Type	Risk Score
Low	1-7
Medium	8-14
High	15-25

Step 5: Identify and Assess Risks				
Describe source of risk and nature of potential impact on individuals. Include associated compliance and corporate risks as necessary.		**Likelihood** 1 - 5	**Impact** 1 - 5	**Risk Score** Likelihood x Impact = Risk Score
1	Personal data is kept for longer than needed. The risk of this is very low as personal data is de-identified at the point of extraction and no personal data is stored in the QResearch database.	1	2	2

Step 6: Identify Measures to Reduce Risk				
Options to reduce or eliminate risk: Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5.		**Effect on Risk** (choose from Eliminated, Reduced or Accepted)	**Residual Risk**	**Measure Approved** Yes/No
1	Not applicable as there are no medium or high risks identified in step 5			

Step 7: Sign Off and Record Outcomes			**Notes**
Approvals	Name - Position	Date	
Measures Approved by:	Julia Hippisley-Cox	02/10/2025	Integrate actions back into project plan, with date and responsibility for completion
Residual Risk Approved by:			If accepting any residual high risk, consult the ICO before going ahead
DPO Advice Provided by:	Paul Smallcombe – Records & Information Compliance Manager	02/10/2025	DPO should advise on compliance, step 6 measures and whether processing can proceed
Summary of DPO Advice: No medium or high risks. Processing can proceed.			
DPO Advice Accepted/Overruled by:			If overruled, you must explain your reasons
Comments:			
Consultation Responses Reviewed by:			If your decision departs from individuals' views, you must explain your reasons
Comments:			
DPIA kept under review by:			The DPO should also review ongoing compliance with DPIA

Integration of Risk Mitigation Measures			
Decide who is responsible for the integration of the risk mitigation measures which were decided and agreed upon in step 6. Record the completion status.		**Responsibility**	**Status** (RAG)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			