

Health Data Research UK

Applicant information

Job description and person specification

- Post:** Health Data Scientist (based within HDR UK)
- Location:** Flexible. Primary base could be anywhere in the UK, but may be required to travel to London, Birmingham or Liverpool in connection with the role
- Duration:** 6 months (secondments welcome)
- Reporting to:** Professor Cathie Sudlow, BHF Data Science Centre/Dr Sophie Williams, Barts Health NHS Trust

About Health Data Research UK

Health Data Research UK (HDR UK) is the UK's national institute for health data science. Our mission is to unite the UK's health data to enable discoveries that improve people's lives. Our vision is that every health and care interaction and research endeavour will be enhanced by access to large scale data and advanced analytics.

We have established national research programmes that use data at scale, and we are building an infrastructure to enable the responsible access and analysis of this data. Our work is structured around three key themes:

- Uniting health data – including our work with data custodians through the [UK Health Data Research Alliance](#) and on making health data discoverable and accessible through the [Health Data Research Innovation Gateway](#).
- Improving health data – incorporating contributions from HDR UK's [Human Phenome](#) and [Applied Analytics](#) priority areas, the [Health Data Research Hubs](#) and the [BHF Data Science Centre](#), including our work on improving data quality and standards, on data curation, and on developing tools and methods for sharing, linking and analysing data.
- Using health data –including our work on research discoveries and skills development across four national priority areas: [Understanding the Causes of Disease](#), [Clinical Trials](#), [Public Health](#) and [Better Care](#).

We are delivering this strategy through our inclusive, team-oriented One Institute ethos - bringing together NHS, universities, research institutes and charities.

About the Data and Connectivity National Core Study

The [National Core Studies](#) (NCS) programme is an important initiative, enabling the UK to use health data and research to inform our long-term response to COVID-19 and accelerating progress in establishing a world-leading health data and research infrastructure for the future.

The programme comprises six National Core Studies (Epidemiology and Surveillance, Clinical Trials Infrastructure (Vaccines and Therapeutics), Transmission and Environment, Immunity, Longitudinal Health & Wellbeing and [Data and Connectivity](#) led by experts from across the UK. Data and connectivity is a cross cutting study led by HDRUK in partnership with the Office for National Statistics, which connects the other five studies and accelerates the approach to answering key research questions the by **enabling streamlined data access and analysis**.

The Data and Connectivity NCS study aims are:

- To build a national health data research capability to support COVID-19 research questions.
- To ensure datasets are discoverable and accessible.
- To establish linkages to answer the research questions prioritised in the other National Core Studies.
- To harmonise methods and standards in NCS research.

About the BHF Data Science Centre

The [British Heart Foundation \(BHF\) Data Science Centre](#), is building on a £10m initial investment from the BHF to deliver the data and data science needed to address some of the most pressing challenges in heart and circulatory health research. For more information please see our [brochure](#).

The centre works in partnership with patients, the public, NHS, researchers and clinicians to promote the safe, ethical and scientifically robust use of data for research into the causes, prevention and treatment of all diseases of the heart and circulation (including, for example, heart attacks, heart failure, heart rhythm disorders, stroke, peripheral vascular disease and vascular dementia).

The BHF Data Science Centre does not hold data itself. Instead, it works with relevant data custodians, including through the UK Health Data Research Alliance and Health Data Research Innovation Gateway, to provide knowledge and expertise to help researchers from the NHS, academia and industry find, access, understand, connect and analyse the UK's unique cardiovascular 'big data' that is distributed across national registries, NHS electronic medical records and other relevant datasets.

The BHF Data Science Centre is coordinating the [CVD-COVID-UK/COVID-IMPACT](#) programme of work which, in collaboration with NHS Digital, has for the first time linked de-identified data across an individual's healthcare journey for 96% of the English population. This equates to **>10 billion rows of data spanning birth to death and covering across primary care, hospitalisation, medication, COVID-19 test and vaccination data and specialist cardiovascular audit and registry data**. This data is brought together in a secure environment that protects patient privacy but allows safe researcher collaboration to answer urgent health questions. CVD-COVID-UK/COVID-IMPACT consortium is a collaboration of >200 researchers in over 40 UK academic and NHS organisations to enable vital research into the relationship between cardiovascular disease and COVID-19. COVID-IMPACT is an expansion of this approach looking at the impact of COVID-19 on other health conditions and health related risk factors in the English population.

Publication: [Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource](#)

About Barts Health

Our group of hospitals provide a huge range of clinical services to people in east London and beyond.

We operate from four major hospital sites (The Royal London, St Bartholomew's, Whipps Cross and Newham) and a number of community locations, including Mile End hospital. Around 2.5 million people living in east London look to our services to provide them with the healthcare they need.

The Royal London in Whitechapel is a major teaching hospital providing local and specialist services in state-of-the-art facilities. Whipps Cross in Leytonstone is a large general hospital with a range of local services. Newham in Plaistow is a busy district hospital with innovative facilities such as its orthopaedic centre. Mile End hospital is a shared facility in Mile End for a range of inpatient, rehabilitation, mental health and community services. And St Bartholomew's in the City, London's oldest hospital, is a regional and national centre of excellence for cardiac and cancer care.

As well as district general hospital facilities for three London boroughs, Tower Hamlets, Waltham Forest and Newham, we have the largest cardiovascular centre in the UK, the second largest cancer centre in London, an internationally-renowned trauma team, and the home of the London Air Ambulance. The Royal London also houses one of the largest children's hospitals in the UK, a major dental hospital, and leading stroke and renal units.

We're also proud to be part of UCLPartners, Europe's largest and strongest academic health science partnership. The objective of UCLPartners is to translate cutting edge research and innovation into measurable health gain for patients and populations through partnership across settings and sectors, and through excellence in education.

Purpose of the post

The post holder will support cross regional data analysis for a project which is enabling regional, acute hospital admissions data to support vaccine safety research. To enable surveillance and research on the COVID-19 vaccination programme and other rare events there is a critical need for rapid, near real time and more granular acute care admissions data.

This data is needed to evaluate the safety and effectiveness of the vaccination programme and answer key questions of clinical, regulatory and policy importance, including the incidence of rare adverse events such as vaccine-induced immune thrombocytopenia and thrombosis (VITT). Current national data feeds are compromised by being dated, incomplete or lacking sufficient diagnostic coding granularity.

Recent analyses of national data¹²³ to determine association of COVID-19 vaccines with VITT events have confirmed limitations in these analyses due to time lag in acute admissions data, and lack of linked laboratory (and potentially radiology) data flows. The critical requirements for a data feed therefore are as follows:

- Multisource linked data from hospital admissions, primary care, Covid tests and vaccinations
- Haematology lab data feeds (for platelet counts +/- other haematological indices)
- Defined and enumerated population denominator
- Diagnostic coding (as close to real time as possible)

The [Data and Connectivity National Core Study](#) therefore commissioned a cross regional collaborative group which includes the BHF Data Science Centre, [OneLondon](#) (Discovery), [PIONEER Data Hub](#) (Birmingham and West Midlands) and [CIPHA](#) (Cheshire and Mersey). This group are exploring the feasibility of curating a minimum dataset to demonstrate thrombotic complications from Covid vaccination. If successful, the approach may have value in addressing other important policy-related research questions relevant to vaccine safety and effectiveness. Initial work has resulted in a curated data set across three regions (population ~7.7million), with the aim that an analytical protocol will be implemented across all regions.

This project will a) establish proof of concept, b) support immediate data and connectivity needs of vaccine safety research, and c) explore the added value of enabling such rapid regional data flows with increased granularity (e.g. enhanced ICD coding to distinguish between different types of venous thrombosis or availability of laboratory data for sensitive identification of thrombocytopenia).

The post holder will conduct the analysis of data across two regions, supervised by Dr Will Whitely and Dr Sophie Williams and the lead analysts at each site. This post would suit a data/computer scientist with significant experience in the management and curation of health data and advanced skills in statistical/epidemiological analysis.

Main responsibilities

- Work with existing health data scientists within the BHF Data Science Centre, PIONEER and OneLondon (Discovery) sites to carry out analyses on the curated linked regional data sets as per the cross regional analytical protocol
- Access regional, linked health record data across regions, develop and execute re-usable analytic approaches, and undertake statistical analysis
- Support the data management and curation for the project across each of the regional partners, ensuring co-ordination across teams.
- Manage processes and deliver across limited timescale, communicating progress, challenges and escalating issues where necessary.
- Anticipate, communicate and solve any potential problems that may arise with data management/curation/analysis for the project.

¹ <https://www.medrxiv.org/content/10.1101/2021.08.18.21262222v1>

² <https://www.nature.com/articles/s41591-021-01408-4.pdf>

³ <https://www.bmj.com/content/374/bmj.n1931>

- In collaboration with the wider team, write, organise and curate support documentation for the linked data resources (e.g. data dictionaries, variable mapping tables, data access process documentation, Git repositories).
- Carry out technical validation checks on the linked data sources (e.g. duplicates, linkage errors)
- Prepare and present results in oral and written reports and publications
- Be an active participant and attend the regular wider project management meetings, reporting on progress and presenting analytical results, and relevant Data and Connectivity NCS, BHF Data Science Centre communications and meetings
- We are committed to open source, transparent and reproducible research and the post will be releasing tools, algorithms and approaches under an open-source licence.

Planning and organising

The postholder will join a small team that is embedded in HDR UK but will also work closely with the regions involved in this project to plan and conduct analyses. The post holder will be responsible for planning and day-to-day management of their own workload across this project. This will require good planning and organisational skills, at the same time, the post holder will require a flexible approach to work to changing demands, particularly external changes.

Problem solving

The postholder will need to develop analytic approaches to complex health-related data to underpin complex algorithm development in the face of missing, incomplete or inconsistent data. They will also need to be able to resolve complex data analysis challenges, discussing as required with epidemiologists, statisticians and clinicians in the project team.

Decision making

The postholder will determine the most appropriate tools and approaches for querying, analysing, maintaining and documenting complex health-related data.

Continuous improvement

HDR UK is dedicated to continuous improvement through our quality management system and has achieved ISO 9001 accreditation. The post-holder will review, analyse, identify and implement opportunities for quality improvement within their role and as part of the wider team through our strategy development and internal audit processes.

Key contacts/relationships

The post holder will work in close conjunction with the core cross regional team (led by Luke Readman, CIO, OneLondon) who are leading this work, primarily with Professor Cathie Sudlow (BHF DSC), Dr Will Whitely, (Uni of Edinburgh) and Dr Sophie Williams (Barts Health NHS Trust) and the lead analysts in CIPHA. In addition, they will work with the wider Data and Connectivity NCS and BHF Data Science Centre teams.

As appropriate to project, they will build and maintain effective working relationships across multiple HDR UK teams, partners in the British Heart Foundation Data Science Centre and other key stakeholders.

Knowledge, skills and experience

Experience

- Good first degree and/or higher degree or equivalent experience in one of the following subjects: bioinformatics, biostatistics, computer science, mathematics, statistics, data science, informatics, epidemiology.
- Scripting skills and experience in writing code for re-use in at least one programming language (e.g. SQL, Python).
- Advanced skills in at least one statistical software package (e.g. R).
- Writing, presenting and explaining technical and/ or scientific reports to a wide range of scientific and lay audiences.
- Understanding of information governance, privacy and security issues with using NHS health records.
- Understanding of epidemiological study design and analysis methods.
- Knowledge of commonly used terminologies in health data (e.g. ICD-10, SNOMED CT) and existing phenotyping algorithms, such as those developed by [CALIBER](#).
- Ability and track record of working independently and co-operatively as part of a team
- Experience of using Git and fundamental concepts in source code revision.
- Prior experience working with large-scale health-related longitudinal data, deriving variables from electronic health records and preparing data extracts for analysis.

Skills

- Committed to open source and reproducible research.
- Ability to work accurately, with attention-to-detail
- Excellent communication skills; experience of working in multidisciplinary teams is advantageous.
- Excellent written and verbal communication skills with the ability to communicate effectively and confidently
- Able to clearly communicate technical concepts to a non-technical audience
- Excellent report writing and presentation skills
- Excellent organisational and time management skills, with the ability to work independently as well as manage competing priorities and issues under time pressures.

Dimensions

- This is a full-time role. Travel in the UK may be required to partner organisations
- HDR UK is a national institute, and our activities take place across the UK.

Application Process

Interested applicants are encouraged to contact the project supervisors to discuss further: william.whiteley@ed.ac.uk and Sophielouise.williams@nhs.net

Please apply with a CV and covering letter of no more than 500 words explaining what you can bring to this role to: hdruk@gravitatehr.co.uk

Secondment Information:

Before applying for this secondment post, please ensure that you have:

- Written support of your employer (an institutional supporting letter will be required for the application)
- Have a contract of employment that extends up to 2024 or beyond.

The closing date for this vacancy is 1st December 2021 – 5.00pm

Interview date TBC. Interviews will take place via zoom.

Equal Opportunities Policy Statement

Health Data Research UK is an equal opportunities employer, and as such aims to treat all employees, consultants and applicants fairly. It is our policy to provide employment equality to all, irrespective of:

- Gender, including gender reassignment
- Marital or civil partnership status
- Having or not having dependants
- Religion or belief
- Race (including colour, nationality, ethnic or national origins)
- Disability
- Sexual orientation
- Age

We are opposed to all forms of unlawful and unfair discrimination. All job applicants and employees who work for us will be treated fairly and will not be unfairly discriminated against on any of the above grounds. Decisions about recruitment and selection, promotion, training or any other benefit will be made objectively and without unlawful discrimination.