

**HDRUK**  
Health Data Research UK



# COVID-19 Data and Connectivity National Core Study

PROTECT NCS Conference 5<sup>th</sup> May 2022



## Our mission

HDR UK's mission is to unite the UK's health data to enable discoveries that improve people's lives

Our 20-year vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health.

### Core funders




### Programme funders





# Data and Connectivity National Core Study:

## Supporting COVID-19 response with research data infrastructure

Surveillance & Epidemiology	Longitudinal Health and Wellbeing	Clinical Trials Infrastructure & Support	Transmission & Environment	Immunity	
<p><b>Professor Ian Diamond (ONS)</b></p> <p>Collecting and analysing data to understand incidence and prevalence broadly and in different settings in order to inform response measures.</p>	<p><b>Professor Nishi Chaturvedi (UCL)</b> <b>Professor Jonathan Sterne (University of Bristol)</b></p> <p>Understanding the impact of Covid-19 on long term health (including long covid) to inform the design of mitigating policies.</p>	<p><b>Professor Patrick Chinnery (MRC)</b></p> <p><i>Establishing infrastructure to run large scale trials for Covid-19 drugs and vaccines without disrupting trials for other diseases.</i></p>	<p><b>Professor Andrew Curran (HSE)</b></p> <p>Taking samples to aid understanding of transmission of the disease in workplace, transport and public places.</p>	<p><b>Professor Paul Moss (University of Birmingham)</b></p> <p>Understanding serology as a useable predictor of immunity against Covid.</p>	

**Data and Connectivity**  
**Professor Andrew Morris (HDR UK in partnership ONS)**  
 Making UK-wide health and administrative data available for linkage and accessible to catalyse COVID-19 research.

# Data and Connectivity National Core Study

## Core aims: “Data as Infrastructure”

1

**Map the Data** for national data science driven research efforts related to COVID-19 and National Core Studies

2

**Accelerate access** to UK-wide priority data relevant to COVID-19 for research

3

**Address important questions** by leveraging the UK’s health data science capability

- Support the response across the four nations
- Involve patients and the public to demonstrate trustworthiness



# A model of collaboration across the UK health data ecosystem

## Trusted Research Environment Delivery partners

Public Health  
Scotland

epcc

HSC

Office for  
National Statistics

SAIL DATABANK

NHS  
Digital

British Heart Foundation  
Data Science Centre  
Led by Health Data Research UK

OpenSAFELY

## Stakeholders

UKRI  
UK Brain Banks Network  
Medical  
Research  
Council

NHS  
Barts Health  
NHS Trust

ISARIC 4C

CO-CONNECT

NIHR | National Institute  
for Health Research

Public Health  
England

CHIEF  
SCIENTIST  
OFFICE

BREATHE  
Health Data Research Hub

CIPHA  
Combined Intelligence for  
Population Health Action

COVID-19 LONGITUDINAL HEALTH AND  
WELLNESS  
NATIONAL CORE  
STUDY

Genomics  
england  
ADRUK

icnarc  
25  
years

ONE  
LONDON

NHS  
England

HSE  
Health & Safety  
Executive

GenOMICC

COVID-19  
GENOMICS  
UK CONSORTIUM

PIONEER  
Health Data Research Hub

Q RESEARCH

Imperial College  
London



# HDR Innovation Gateway – ‘one stop shop’ for data access

Building the “Gateway” to Findable Accessible Interoperable and Reusable (FAIR) datasets, tools and resources

## Gateway to health data and tools for research

Search, discover and request access to hundreds of datasets, tools and resources for your research. Join the thousands of researchers and scientists worldwide who are already using the Gateway for research and scientific discovery.

🔍 I'm looking for...

★ [Advanced Search](#)

datasets 754	tools 182	data uses 734	courses 199	papers 1,961	dataset requests 489	registered users 2,086	searches last month 9,338
-----------------	--------------	------------------	----------------	-----------------	-------------------------	---------------------------	------------------------------



For Researchers



For Data Custodians



For Patients and Public

*“Really impressed with this resource. I think as a gateway to search by data type and indication, it’s a really powerful tool.”*

**David Leather, GSK**

<https://www.healthdatagateway.org/>

**95 NCS datasets now discoverable and accessible**

# The Gateway provides a common, transparent approach to data access management based around the Five Safes framework

The screenshot shows the 'Data Access Request' interface. At the top, it says 'Health Data Research Innovation Gateway' with a search icon and a notification bell. Below this is a dark blue header with 'Data Access Request' and 'ALLIANCE > SAIL', and a 'Save now' button with a close icon. On the left, a sidebar menu is circled in orange, listing: 'About this application', 'Safe people', 'Safe project', 'Safe data', 'Safe settings', 'Safe outputs', and 'Files'. The main content area is titled 'About this application' and contains the text: 'Preparation is key to a successful data access request. You need to be able to demonstrate how you will ensure safe use of patient data and the potential for public benefit. The steps below are intended to help you get off to a good start.' Below this is a step indicator '1' in a green circle, followed by the heading 'Select the datasets you need' and the text 'The datasets you select may impact the questions being asked in this application form. You cannot change this'. At the bottom, there is a progress indicator '4/73 questions answered | 5FC0-31C3-D1E8-1B7C-1E02-BAF7' and three buttons: 'Contributors', 'Submit application' (highlighted with a green border), and 'Next'.

# Enabling secure access via national network of Trusted Research Environments

- Trusted Research Environments (TREs), also known as ‘Data Safe Havens’, are **highly secure spaces for researchers to access sensitive data**.
- They are based on the idea that researchers **should access and use data within a single secure environment**.
- TREs have **multiple layers of security and safeguards in place**, designed to minimise the risk of anyone’s data being misused.
- **Multiple examples of TREs** operating successfully in this way, both for healthcare data and other potentially sensitive data

**What is a TRE?**  
A TRE is a **Trusted Research Environment**. Also known as ‘Data Safe Havens’, TREs are highly secure computing environments that provide remote access to health data for approved researchers to use in research that can save and improve lives.

**Why are they important?**

- TREs make research safer. Making data available through a TRE means that people can be confident that their personal health data is accessed securely and their privacy protected.
- TREs help make research efficient, collaborative and cost effective, providing rich data that enables deep insights which will go on to improve healthcare and save lives.
- TREs provide approved researchers with a single location to access valuable datasets. The data and analytical tools are all in one place, a bit like a secure reference library.

**How is my data safeguarded?**  
Health data should always be kept safe and secure, and used responsibly to ensure privacy. Health Data Research UK ensures these high standards are met by promoting the use of the ‘Five Safes’ model across all TREs.

- Safe People**  
Only trained and specifically accredited researchers can access the data
- Safe Projects**  
Data is only used for ethical, approved research with the potential for clear public benefit
- Safe Settings**  
Access to data is only possible using secure technology systems – the data never leaves the TRE
- Safe Data**  
Researchers only use data that have been de-identified to protect privacy
- Safe Outputs**  
All research outputs are checked to ensure they cannot be used to identify subjects

Learn more about TREs and discover examples of how TREs are being used to enable life-saving health research. [Learn more about TREs](#)



## Working with patients and the public...

Working together, we want to ensure the patient and public voice influences and shapes our work in “Data and Connectivity” through



## Data & Connectivity NCS: progress so far..

### NCS Data Assets

- **95 datasets** now made FAIR including
  - Viral sequencing
  - Vaccination
  - Census data linked to GP and hospital data
  - Geospatial, socioeconomic data

### Developing cross-TRE innovation

- **Implementing a Cohort Discovery tool for all NCS datasets** (in partnership with CO CONNECT)

### Enhancing and improving user experience

- Streamlined **Data Access Request** process by
  - Implementing a new standardised form across providers
  - Testing new **platform-approval models** of access
  - Piloting a **unified 4-nations governance** process

# Research using NCS data assets and infrastructure informing UK and international policy: >600 projects and 246 papers (and preprints) using NCS data assets and infrastructure\*

## Vaccine Safety and Effectiveness

THE LANCET

Two-dose ChAdOx1 nCoV-19 vaccine protection against COVID-19 hospital admissions and deaths over time: a retrospective, population-based cohort study in Scotland and Brazil

Prof Srinivasa Vittal Katikireddi, PhD<sup>1</sup> · Thiago Cerqueira-Silva<sup>2</sup> · Eleftheria Vasileiou, PhD<sup>3</sup> · Prof Chris Robertson, PhD<sup>4</sup> · Sarah Amele, PhD<sup>5</sup> · Jiafeng Pan, PhD<sup>6</sup> · et al. [Show all authors](#) · [Show footnotes](#)

### Risk of thrombocytopenia and thromboembolism after covid-19 vaccination and SARS-CoV-2 positive testing: self-controlled case series study

BMJ 2021;374 doi: <https://doi.org/10.1136/bmj.n1931> (Published 27 August 2021)  
Cite this as: BMJ 2021;374:n1931

### Association of COVID-19 vaccines ChAdOx1 and BNT162b2 with major venous, arterial, or thrombocytopenic events: A population-based cohort study of 46 million adults in England

William N. Whiteley<sup>1</sup> · Samantha Ip<sup>2</sup> · Jennifer A. Cooper<sup>3</sup> · Thomas Bolton<sup>4</sup> · Spencer Keene<sup>5</sup> · Venexia Walker<sup>6</sup> · Rachel Denholm<sup>7</sup> · Ashley Akbari<sup>8</sup> · Elosa Omigie<sup>9</sup> · Sam Hollings<sup>10</sup> · Emanuele Di Angelantonio<sup>11</sup> · Spiros Denaxas<sup>12</sup> · Angela Wood<sup>13</sup> · et al.

### COVID-19 hospital admissions and deaths after BNT162b2 and ChAdOx1 nCoV-19 vaccinations in 2.57 million people in Scotland (EAVE II): a prospective cohort study

Utkarsh Agrawal, PhD<sup>1</sup> · Prof Srinivasa Vittal Katikireddi, PhD<sup>2</sup> · Prof Colin McCowan, PhD<sup>3</sup> · Rachel H Mulholland, MSc<sup>4</sup> · Amaya Azcoaga-Lorenzo, PhD<sup>5</sup> · Sarah Amele, MSc<sup>6</sup> · et al. [Show all authors](#)

## Viral and host genomics

THE UNIVERSITY of EDINBURGH

Home **Research output** Profiles Research Units Projects Datasets Prizes Activities

Severity of Omicron variant of concern and vaccine effectiveness against symptomatic disease: national cohort with nested test negative design study in Scotland

The NEW ENGLAND JOURNAL of MEDICINE

BNT162b2 and ChAdOx1 nCoV-19 Vaccine Effectiveness against Death from the Delta Variant

nature <https://doi.org/10.1038/s41586-022-04576-6>

Accelerated Article Preview

Whole genome sequencing reveals host factors underlying critical Covid-19

## Impact of COVID-19 across different groups

<https://doi.org/10.1038/s41467-022-28248-1> OPEN

A population-based cohort study of obesity, ethnicity and COVID-19 mortality in 12.6 million adults in England

Thomas Yates<sup>1,2,3\*</sup>, Annabel Summerfield<sup>3</sup>, Cameron Razieh<sup>1,2,4</sup>, Amitava Banerjee<sup>5,6</sup>

### Trends and clinical characteristics of COVID-19 vaccine recipients:

a federated analysis of 57.9 million patients' primary care records *in situ* using OpenSAFELY

ARTICLES **nature medicine**

<https://doi.org/10.1038/s41591-021-01666-2>

OPEN

SARS-CoV-2 infection and COVID-19 vaccination rates in pregnant women in Scotland

### BMJ Open Sociodemographic inequality in COVID-19 vaccination coverage among elderly adults in England: a national linked data study

Vahe Nafilyan<sup>1,2</sup>, Ted Dolby<sup>1</sup>, Cameron Razieh<sup>3,4,5</sup>



## Priorities to end of March 2023 (and beyond!)

### Addressing high priority areas of unmet data need:

- ✓ **Outbreak Data Analysis Platform (ODAP)**– progress linking outbreak data from ISARIC4C, COG-UK, GenOMMIC, PHOSP, ICNARC, clinical records, research studies. Finalising single panel, multi custodian, proportionate and robust data access Governance process. **Vision is to create single UK-wide capability for pandemic data science in partnership with PHAs**
- ✓ **Enabling, rapid, regional linked acute admissions data flows for research:** currently available national data feeds are dated (up to a 6-week time delay), do not include people still in hospital, and lack diagnostic granularity. This is a gap in the UK health data research ecosystem – collaborating with 9 UK wide regions to implement regional linked feeds and scale up across the UK
- ✓ **Further streamline data access** at both researcher and custodian to custodian level
- ✓ **Secure legacy** of the national network of TRE data infrastructures and NCS curated data assets