HDRUK
Health Data Research UK

Making Health Data more Accessible for Research

Susheel Varma PhD MBA FBCS, Director of Engineering

07 June 2021

National institute to unite, improve and use health data for research

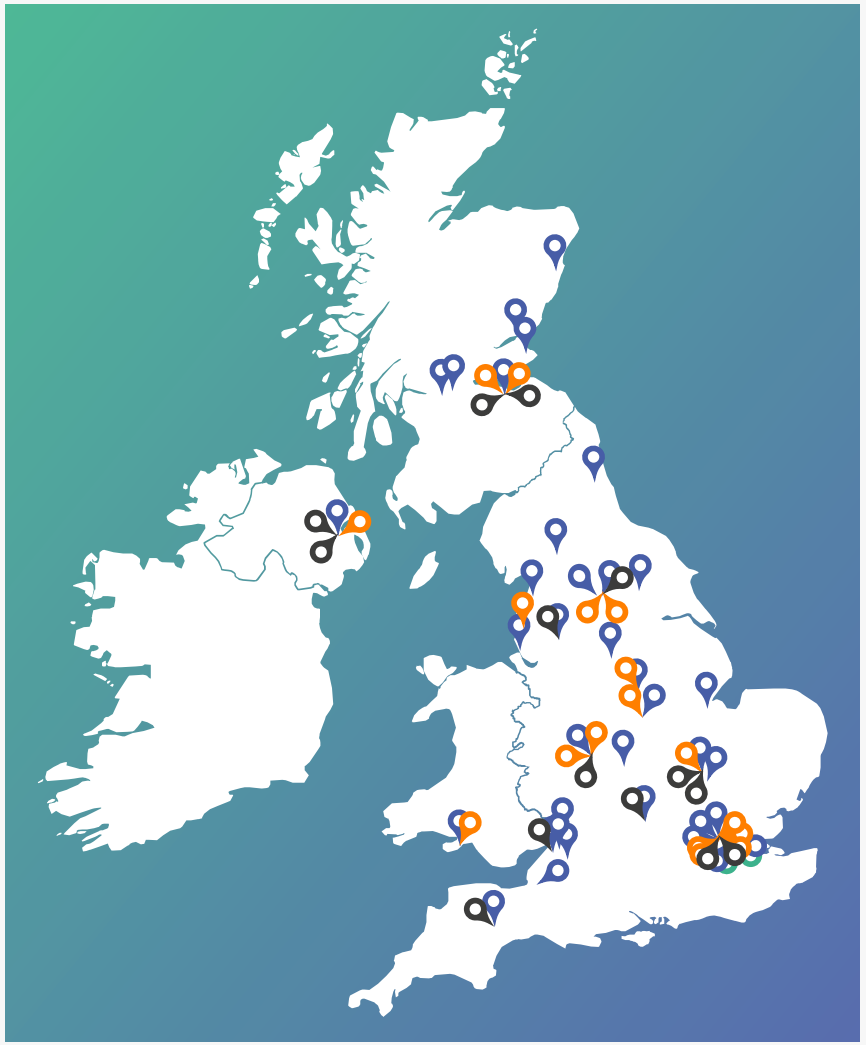
HDRUK
Health Data Research UK

CENTRAL TEAM OFFICES

Wellcome Trust
Great Ormond Street
DRIVE Unit

RESEARCH LOCATIONS

HDR UK Cambridge
HDR UK London
HDR UK Midlands
HDR UK North
HDR UK Oxford
HDR UK Scotland
HDR UK South-West
HDR UK Wales and
Northern Ireland



HEALTH DATA RESEARCH HUBS

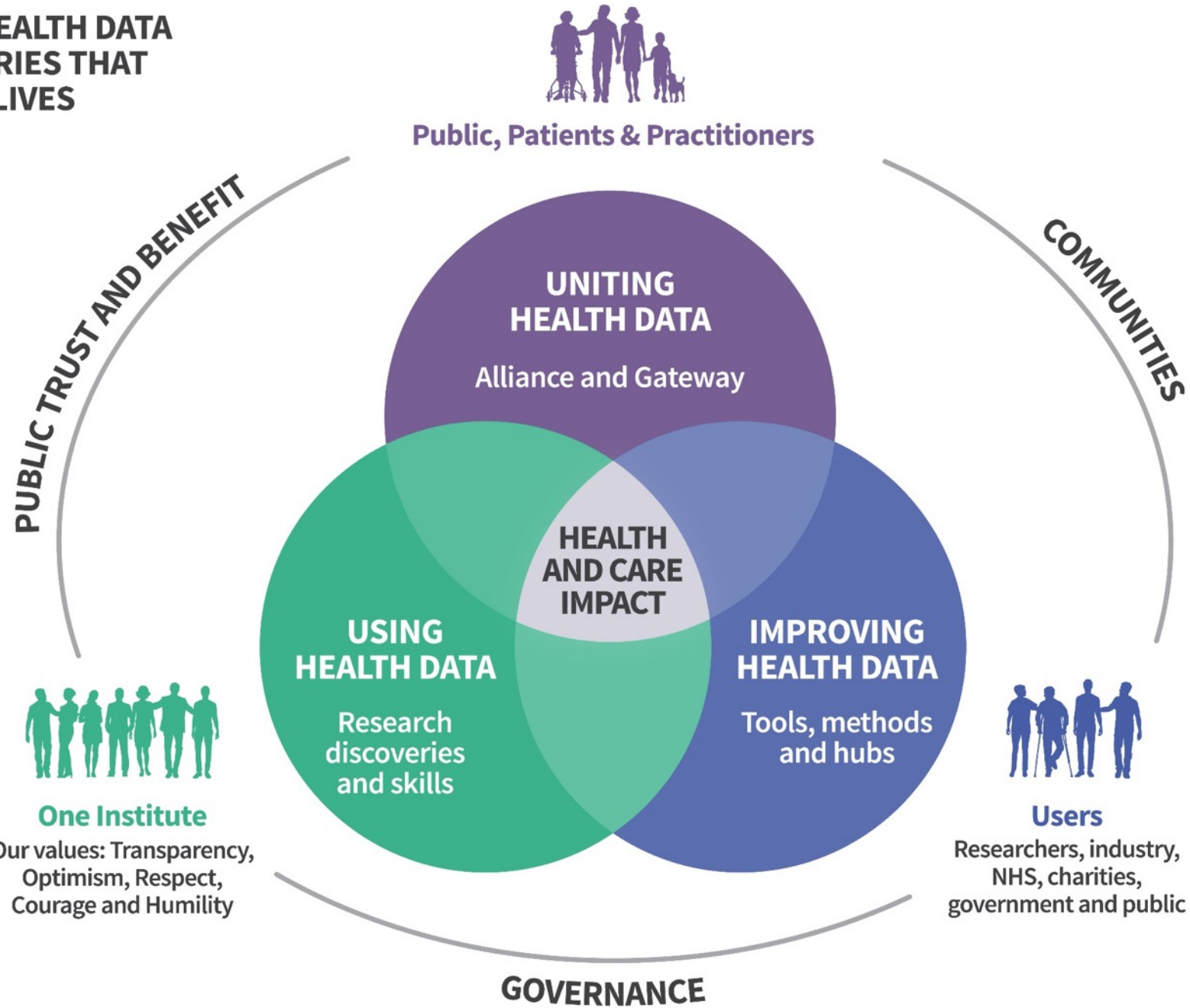
BREATHE
DATA-CAN
Discover-NOW
Gut Reaction
INSIGHT
PIONEER
NHS DigiTrials
BHF Data Science Centre

TRAINING LOCATIONS (Masters and PhD)

Belfast
Birmingham
Bristol
Cambridge
Edinburgh
Exeter
Leeds
London
Manchester
Oxford

UNITING THE UK'S HEALTH DATA TO ENABLE DISCOVERIES THAT IMPROVE PEOPLE'S LIVES

HDRUK
Health Data Research UK



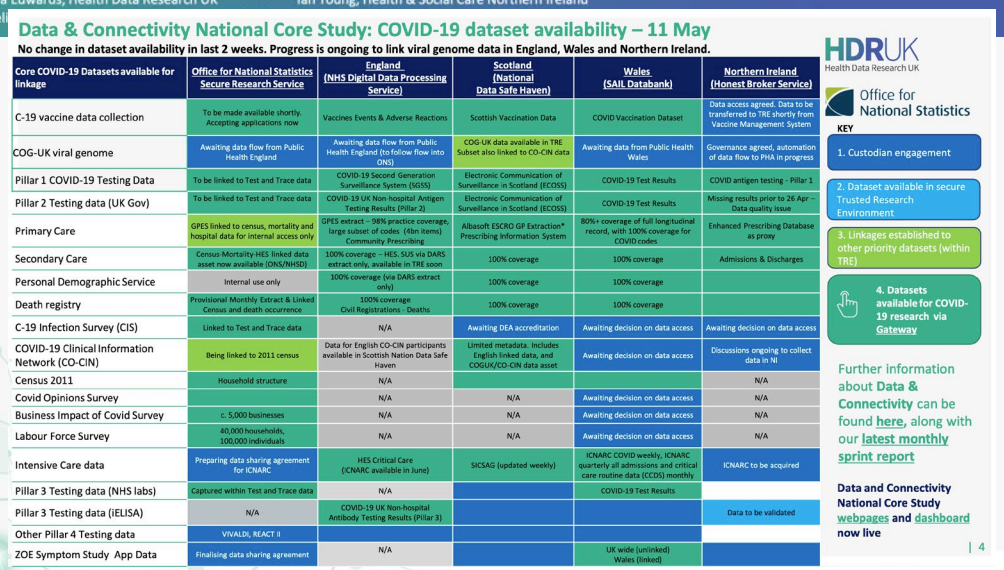
Our three key priorities at the onset of the COVID-19 pandemic

HDRUK
Health Data Research UK

1. Co-ordinate and connect national data science driven research efforts related to COVID-19
2. Accelerate access to UK-wide priority data relevant to COVID-19 for research
3. Leverage the best of the UK's health data science capability to address the wider impact of the COVID-19 pandemic, supporting vulnerable groups that will be hardest hit

Support the UK government response through regular reporting to SAGE

Regular reports to SAGE



Data & Connectivity National Core Study: COVID-19 dataset availability – 11 May

No change in dataset availability in last 2 weeks. Progress is ongoing to link viral genome data in England, Wales and Northern Ireland.

Core COVID-19 Datasets available for linkage	Office for National Statistics Secure Research Service	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)
C-19 vaccine data collection	To be made available shortly. Accepting applications now	Vaccines Events & Adverse Reactions	Scottish Vaccination Data	COVID Vaccination Dataset	Data access agreed. Data to be transferred to TRE shortly from Vaccine Management System
COG-UK viral genome	Awaiting data flow from Public Health England	Awaiting data flow from Public Health England (to follow flow into ONS)	COG-UK data available in TRE Subset also linked to CO-CIN data	Awaiting data from Public Health Wales	Governance agreed, automation of data flow to PHA in progress
Pillar 1 COVID-19 Testing Data	To be linked to Test and Trace data	COVID-19 Second Generation Surveillance System (SGSS)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID-19 Test Results	COVID antigen testing - Pillar 1
Pillar 2 Testing data (UK Gov)	To be linked to Test and Trace data	COVID-19 UK Non-hospital Antigen Testing Results (Pillar 2)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID-19 Test Results	Missing results prior to 26 Apr – Data quality issue
Primary Care	GPES linked to census, mortality and hospital data for internal access only	GPES extract – 98% practice coverage, large subset of codes (4bn items) Community Prescribing	Albasoft ESCRO GP Extraction* Prescribing Information System	80%+ coverage of full longitudinal record, with 100% coverage for COVID codes	Enhanced Prescribing Database as proxy
Secondary Care	Census-Mortality-HES linked data asset now available (ONS/NHSD)	100% coverage – HES. SUS via DARS extract only, available in TRE soon	100% coverage	100% coverage	Admissions & Discharges
Personal Demographic Service	Internal use only	100% coverage (via DARS extract only)	100% coverage	100% coverage	
Death registry	Provisional Monthly Extract & Linked Census and death occurrence	100% coverage Civil Registrations - Deaths	100% coverage	100% coverage	
C-19 Infection Survey (CIS)	Linked to Test and Trace data	N/A	Awaiting DEA accreditation	Awaiting decision on data access	Awaiting decision on data access
COVID-19 Clinical Information Network (CO-CIN)	Being linked to 2011 census	Data for English CO-CIN participants available in Scottish Nation Data Safe Haven	Limited metadata. Includes English linked data, and COGUK/CO-CIN data asset	Awaiting decision on data access	Discussions ongoing to collect data in NI
Census 2011	Household structure	N/A			N/A
Covid Opinions Survey		N/A	N/A	Awaiting decision on data access	N/A
Business Impact of Covid Survey	c. 5,000 businesses	N/A	N/A	Awaiting decision on data access	N/A
Labour Force Survey	40,000 households, 100,000 individuals	N/A	N/A	Awaiting decision on data access	N/A
Intensive Care data	Preparing data sharing agreement for ICNARC	HES Critical Care (ICNARC available in June)	SICSAG (updated weekly)	ICNARC COVID weekly, ICNARC quarterly all admissions and critical care routine data (CCDS) monthly	ICNARC to be acquired
Pillar 3 Testing data (NHS labs)	Captured within Test and Trace data	N/A		COVID-19 Test Results	
Pillar 3 Testing data (iELISA)	N/A	COVID-19 UK Non-hospital Antibody Testing Results (Pillar 3)			Data to be validated
Other Pillar 4 Testing data	VIVALDI, REACT II				
ZOE Symptom Study App Data	Finalising data sharing agreement	N/A		UK wide (unlinked) Wales (linked)	

HDRUK Health Data Research UK

Office for National Statistics

KEY

1. Custodian engagement
2. Dataset available in secure Trusted Research Environment
3. Linkages established to other priority datasets (within TRE)
4. Datasets available for COVID-19 research via Gateway

Further information about **Data & Connectivity** can be found here, along with our latest monthly sprint report

Data and Connectivity National Core Study webpages and dashboard now live

4

Health data research community highlights this period

HDRUK Health Data Research UK

Office for National Statistics

35 COVID-19 taskforce calls with **183** clinical and health data research leaders engaged

1,222 COVID-19 pre-print publications, and **147** papers published

771 academic, industry and NHS participants in COVID-19 Slack channel with 10 sub-channels

111 health data research questions identified

Patient and Public Voice Feedback - Lot of positive progress has been made but as we continue to move out of a national lockdown, we must sustain momentum:

- Those who are immunosuppressed have no certainty about vaccine effectiveness as there is no/limited evidence. Understanding and publicising the resulting data must be a priority to ensure immunosuppressed people do not put themselves at risk unnecessarily.
- Whilst more and more of the population become vaccinated and evidence has shown effectiveness against the UK variant, it is imperative to better understand and communicate the effects of the vaccine on the different and emerging variants as well as the effect of the variants on transmission.
- Urgent research is needed in the broader long-term follow up of COVID-19 patients, looking at all age groups and all of those ranging from clinically vulnerable to "healthy" adults.
- Urgently, with restrictions becoming much lighter on 17 May 2021 onwards, research needs to explore the effects on cases, hospitalisations and deaths and should include the impact of removing the requirement for pupils to wear masks in schools.

Click here to read more feedback

Click here for a list of regularly updated COVID-19 research questions from the

Research topics with new insights generated in last 2 weeks

HDRUK Health Data Research UK

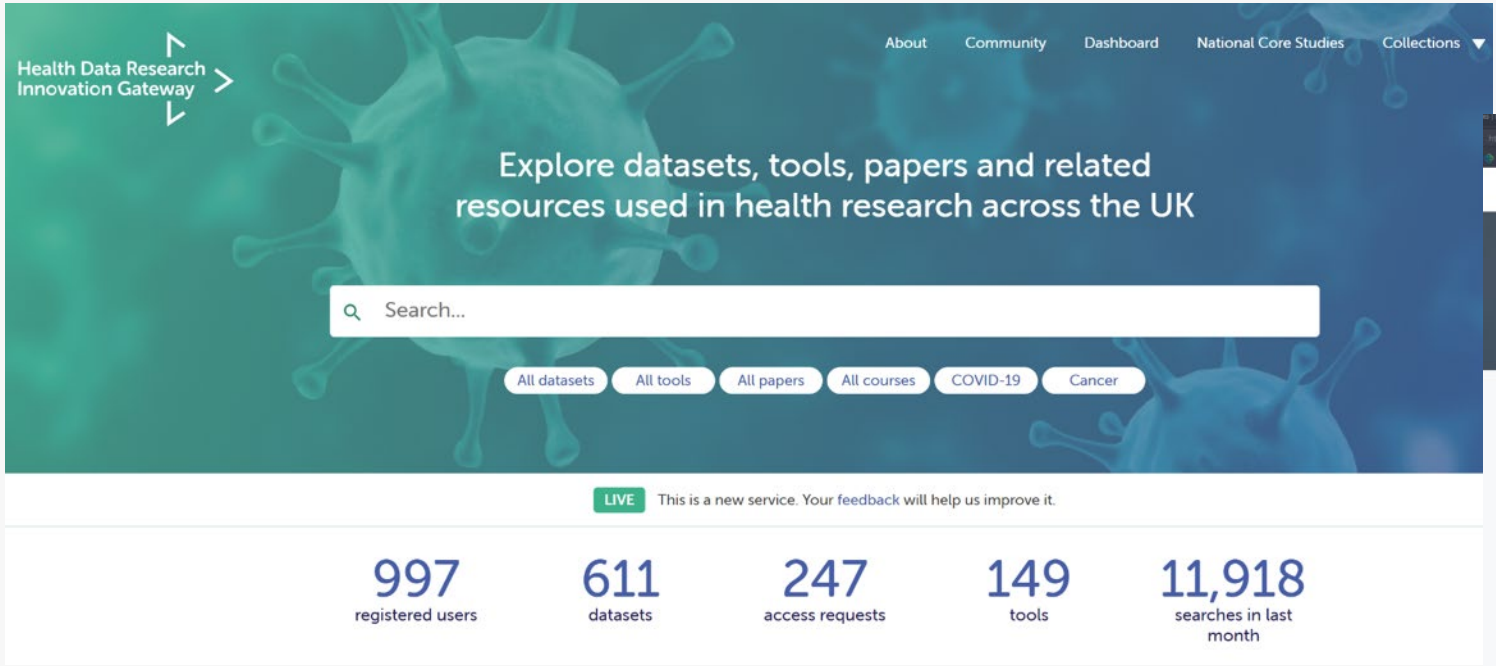
Health data research outputs on COVID-19 continues to grow, now reaching 1,222 (+8) non-peer-reviewed pre-prints & 147 (+9) published papers.

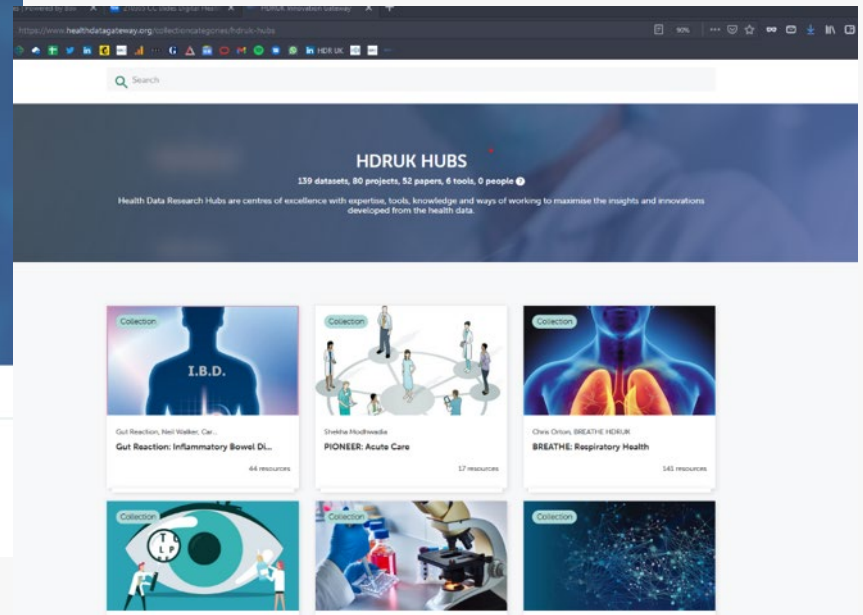
Topic	Insights from ongoing studies (links provide further details):
Surveillance & Epidemiology	• A multi-ethnic population study using linked primary and secondary care electronic health records of >15,000 patients with hepato–pancreato–biliary conditions such as liver cancer, pancreatic cancer, and gallstones found that the risk of COVID-19 was heightened for patients with a history of substance misuse.
Immunity & Vaccines	• Real-world data from the vaccine rollout in Scotland suggest that the 1st dose of the Pfizer-BioNTech and the Oxford-AstraZeneca vaccines are 91% and 88% effective at reducing COVID-19 hospitalisation, respectively. The study used the Early Pandemic Evaluation and Enhanced Surveillance of COVID-19—EAVE II—database comprising linked vaccination, primary care, real-time reverse transcription-PCR testing, and hospital admission patient records for the entire Scottish population. • Preliminary analyses combining vaccination and COVID-19 testing data for University Hospitals Birmingham healthcare workers revealed that a high proportion of workers vaccinated with a 1st dose acquired COVID-19 – calling for healthcare workers to remain vigilant after a 1st vaccine dose. • A real-world study using PCR-test results from a representative cohort of >300,000 participants as part of the Office for National Statistics (ONS) COVID-19 Infection Survey (CIS) has found that a single dose of Oxford-AstraZeneca or Pfizer-BioNTech vaccines, or two doses of Pfizer-BioNTech, reduced new infections. Both vaccines appear to be effective against the UK variant. • As part of the same survey, a separate study of post-vaccine antibody levels in 45,965 UK adults found that vaccine response differs with demographics (including age, gender and previous infection status) and – if vaccine supplies become limited – prioritising vaccines for people not previously infected, and 2nd doses to those over the age of 60 may be warranted.
Longitudinal health & wellbeing	• Analyses of COVID-19 Symptom Study survey responses from >2.7 million people in the US, UK,, and Sweden found no association between use of non-steroidal anti-inflammatory drugs (including Aspirin) and COVID-19 infection – suggesting patients and healthcare providers should continue to use these drugs to prevent cardiovascular disease, treat chronic secondary pain, etc. • Analyses by the OpenSAFELY Collaborative of linked patient-level data revealed that people using routinely prescribed oral anti-coagulants (n=70,464) had a lower risk of severe COVID-19 outcomes and showed no evidence that warafin users (n=372,746) have an increased risk of COVID-19 outcomes (versus other anti-coagulants). These results provide reassurance that these treatments can continue to be used safely during the pandemic.
Transmission & Environment	• Responses from >20,000 participants of the VirusWatch household study indicate that people living in deprived areas are more likely to use public transport, work/attend school outside the home, and visit essential shops –suggesting interventions to prevent exposure during these activities may reduce risk inequalities. • A matched cohort study of >200,000 care home residents using primary care electronic health from the Clinical Practice Research Datalink Aurum Database in England confirmed the disproportionate impact of the 1st wave on the care home population and highlights the need to protect these individuals in future outbreaks. • A study exploring feasibility of point-of-care COVID-19 testing from ~200 care home staff and residents across 4 sites suggests that both standard operating procedures and training materials require adjustments specific to the care home environment to minimise sources of contamination.
Clinical Trials	• Recent results from ATOMIC2 an open-label, randomised superiority clinical trial at 19 centres in the UK found that azithromycin (an anti-inflammatory antibiotic) does not reduce hospitalisation in patients (n=298) with mild-moderate COVID-19.

Building a Gateway for researchers and innovators to access health data

www.healthdatagateway.org

HDRUK
Health Data Research UK





- **Access to request over 600+ datasets**, 149 tools, 195 educational courses and 1,242 publications
- A **dedicated collection of datasets for each Hub – 140 in total**
- **Facilitated 247 requests to access health datasets**, in particular actively supporting the government's National Core Studies into COVID-19 ("Data and Connectivity")
- **Open Source by default** – https://github.com/HDRUK

"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

Convening the sector through the UK Health Data Research Alliance (50 members and growing…)

Open, Federated and Interoperable Ecosystem for Health Data Research in the UK

HDRUK
Health Data Research UK

ACADEMIA, NHS, INDUSTRY

Public, Patient and Practitioner Involvement and Engagement

National Core Studies

Basic Research | Thematic Research | Priority Research | Cross-Sectoral Research

CLINICAL PRACTICE & PUBLIC HEALTH

Health Data Research Innovation Gateway

DISCOVER → ACCESS → ANALYSE → COMMUNICATE

Search metadata | Signup / Login | Provision Analysis | Publish Results

Discover Cohorts | Data Access Req. | Airlock In/Egress

Health Data Research Innovation Gateway

APPLIED ANALYTICS
TEXT ANALYTICS
HUMAN PHENOME
BHF DSC, HUBS, PRUK

... | FED ML | MODEL LIBRARY | DISEASE ATLAS | SYNTH DATA | ...

UNDERSTANDING CAUSE OF DISEASE
IMPROVING PUBLIC HEALTH
ENHANCED CLINICAL TRIALS
BETTER PERSONALISEDCARE

INTEROPERABLE IFRASTRUCTURE

Federation Core Services

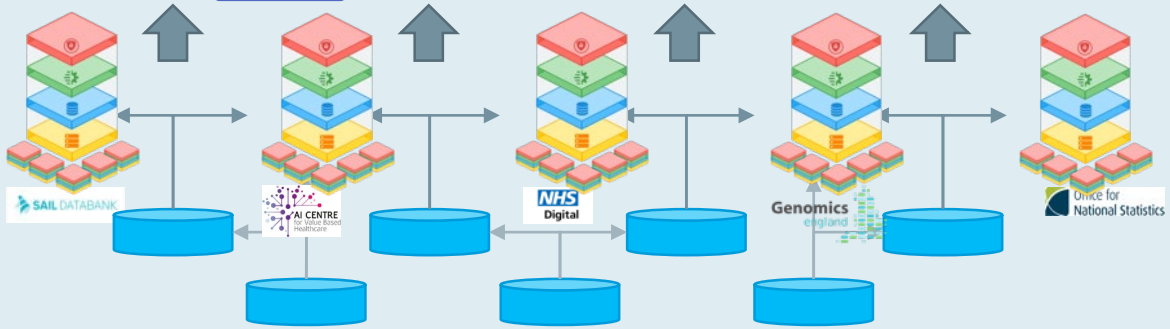
Metadata catalogues | Cohort Query Engine | Identity & Accreditation | Data Access Requests | Workflow Execution | Code/Data Transfer | Results Aggregation

Federation Core Services

INTEROPERABLE INFRASTRUCTURE

TRUSTED RESEARCH ENVIRONMENTS

Metadata catalogues | Query Execution | Access & Authorisation | DAR Processing | Task Execution Service | Airlock / Escrow



TRUSTED RESEARCH ENVIRONMENTS

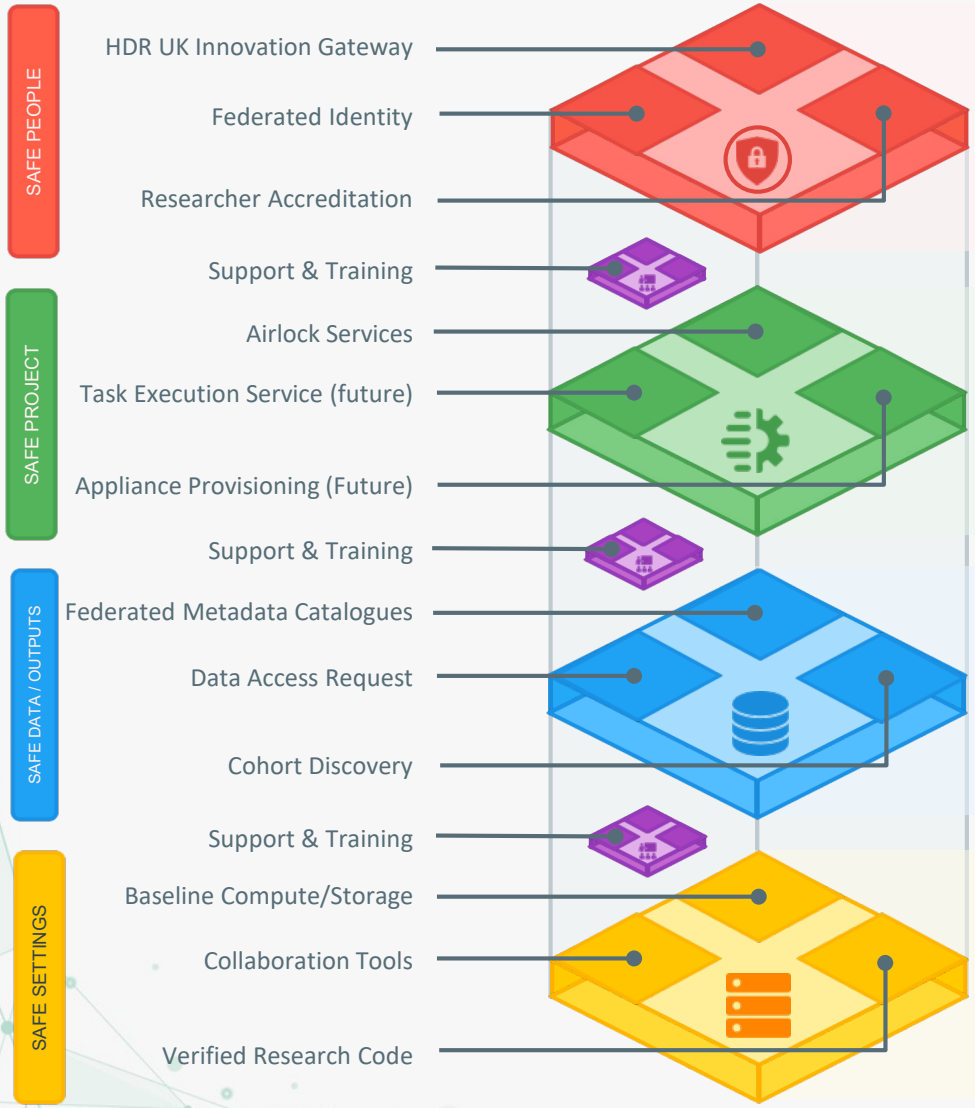
DATA SOURCES

DATA SOURCES

Public Health, Primary Care, Secondary Care, Test & Trace, Administrative Datasets with increased adoption of data standards and interoperability

Open, Federated and Interoperable Technology Stack for Trusted Research Environments

HDR UK
Health Data Research UK



Identity Federation

Provides authenticated, authorized and auditable access to federated resources using standardized single sign-on and identity federation

Analytics Federation

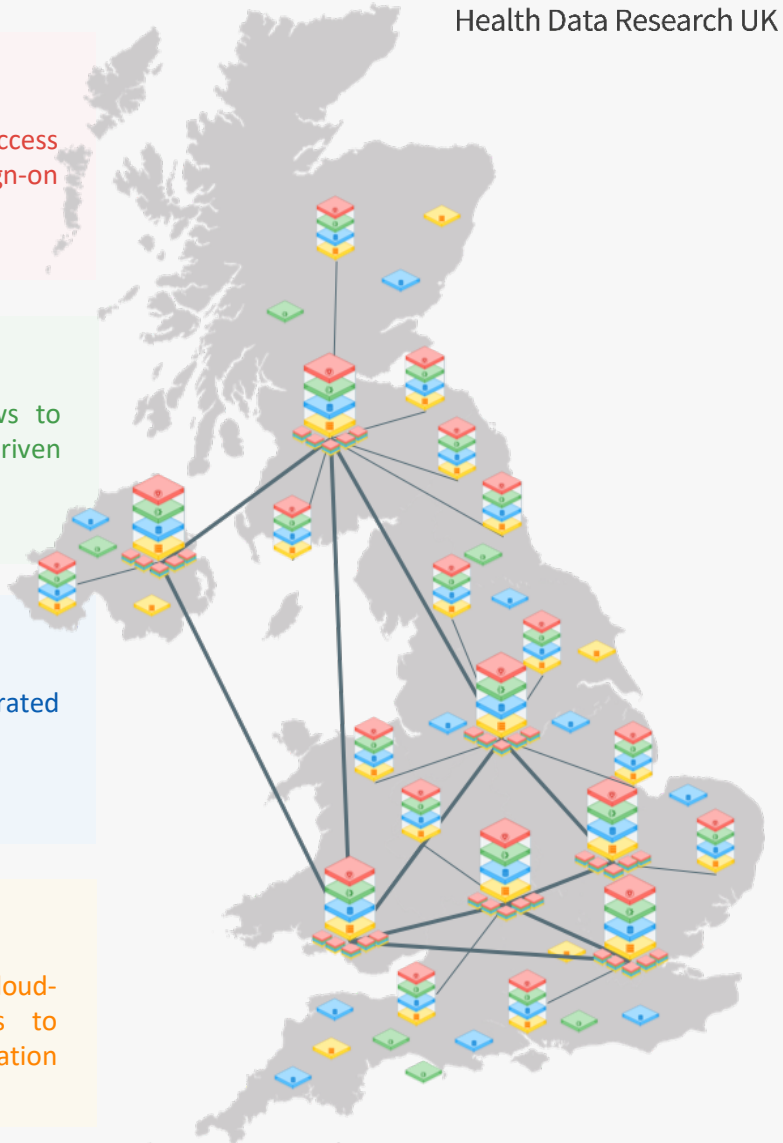
Reuse and combine portable tools and workflows to enhance healthcare delivery with advanced data-driven translational insights.

{Meta}Data Federation

Discover, explore, organize and securely access federated data for accelerating the translation into healthcare.

Infrastructure Federation

Immediate access to advanced & flexible hybrid cloud-based computational resource including access to specialized accelerators and container orchestration services.



Layout for illustration purpose only

An exemplar of working with public and patients

828 submissions to public survey on COVID-19 vaccine

Strong **Public Advisory Board** providing strategic guidance on all our work

Consultation on COVID work across **7 UK-wide patient and public networks with 168 responses.**

16,500 contacts with patient & public contributors across the institute in 2020

> I am glad you're involving me from what seems to be the beginning so that you can actually take my concerns and address them whilst helping the greater good
>
> **Patient / Public Voice Rep**

92 people consulted to inform decisions on methods and process for clinical trial recruitment

> It is essential that the public is included in this ground-breaking work.
>
> **Margaret Rogers**
> Member of the HDR UK Public Advisory Board

Health Data Research UK Ecosystem

UK Health Data Research Alliance

HDRUK Health Data Research UK

1,242 COVID-19 pre-prints **158** published papers.
Better Care, Understanding Causes of Disease, Clinical Trials & Public Health

PRINCIPLE | PGS Catalog | RECOVERY | COVID-19 GENOMICS UK CONSORTIUM

293 COVID-19 research projects using national data custodian data
300 Health Data Research Hub contracts with industry & academia

Gut Reaction | DATA-CAN | Discover-NOW | INSIGHT | PIONEER | BREATHE | DIGITRIALS | APPLIED ANALYTICS | HUMAN PHENOME | BHF DATA SCIENCE CENTRE

>22,000 patient and public participants in health data research
HDR Standards: Trusted Research Environments, Data Utility, Federation

Streamlined data access request process, harmonised across TREs. Time from application to active research: **3-14 days** [1]

[1] Wales and Scotland data only

C-19

Example: Zoe COVID-19 symptom tracker – most frequently accessed dataset

Example: Hubs + CVD-COVID-19 For the first time, linked health data resource covering **54.4 million people**

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

112 datasets set up in **5** national trusted research environments – by National Core Studies Data & Connectivity

HSC Public Health Agency Research and Development | NHS Digital | Office for National Statistics | SAIL DATABANK | epcc | Public Health Scotland | The Scottish Government

Health Data Research Innovation Gateway

646 discoverable datasets
>16,000 monthly searches

Hubs impact and insights

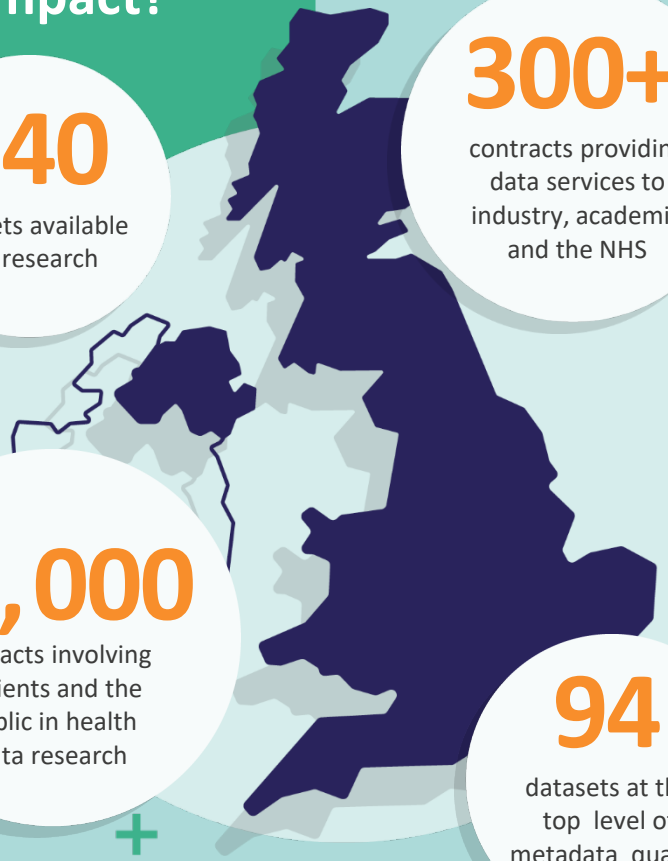
Making datasets more useful, so researchers can better understand data

HDRUK
Health Data Research UK

What is our impact?

140
datasets available for research

300+
contracts providing data services to industry, academia and the NHS



16,000
contacts involving patients and the public in health data research

94
datasets at the top level of metadata quality

Providing crucial insights on COVID-19

NHS DigiTrials
Provided access and linked data for the RECOVERY Trial, saving thousands of lives.

DATA-CAN
Estimated excess cancer patient deaths.

Gut Reaction
Enabled 34,000 people with inflammatory bowel disease to assess their risk.

BHF Data Science Centre
Enabled safe, secure access to health data for over 55 million people in the UK, to investigate the impact of COVID-19 on cardiovascular disease.

PIONEER
Analysis of real-time data discovered an increase in venous thromboembolic events (VTE); developed guidance which is now being used in over 60 countries.

Discover-NOW
Supported a range of projects across London.

BREATHE
Provided policymakers maps and hotspot data from 4.4.m people using the ZOE/ Kings College London Symptom Study.

INSIGHT
Supported real world data analysis to identify impact on hospital eye services.

Data enabled clinical trials - RECOVERY

Challenge

To rapidly identify treatments that may be beneficial for people hospitalised with suspected or confirmed COVID-19, to save lives and reduce the burden on the NHS

Solution

The RECOVERY trial was set up quickly to enable access and use of large-scale data so that effective treatments could be identified quickly and made available.

The NHS DigiTrials Hub and HDR UK's clinical trials team enabled access to linked data from hospital admissions, discharge, treatments, deaths, and COVID test results.

Impact

- The trial protocol set up in less than a fortnight and the first patient recruited within 9 days
- **38,000 patients have been recruited through 180 hospitals, the largest randomised trial for COVID-19 treatments**
- Within 4 months, the trials had discovered the **first successful treatment for COVID-19 - dexamethasone** (estimated to have saved 650,000 lives across the world)
- Within a year the trial had discovered tocilizumab as an effective treatment
- RECOVERY has ruled out 4 drugs - including hydroxychloroquine - as viable treatments (an equally valuable insight from clinical trials)

"The widespread collection of healthcare data and trustworthy tools for bringing them together mean that we can reignite and extend that opportunity – using existing NHS data from GPs and hospitals to assess the impact of new treatments for a range of diseases, and bringing benefits to patients faster and more efficiently than ever before."
Martin Landray, HDR UK Research Director, Clinical Trials

https://www.recoverytrial.net/results

Real-world validation of COVID-19 vaccine effectiveness through rapid use of linked, national data

BREATHE Health Data Research Hub

Challenge

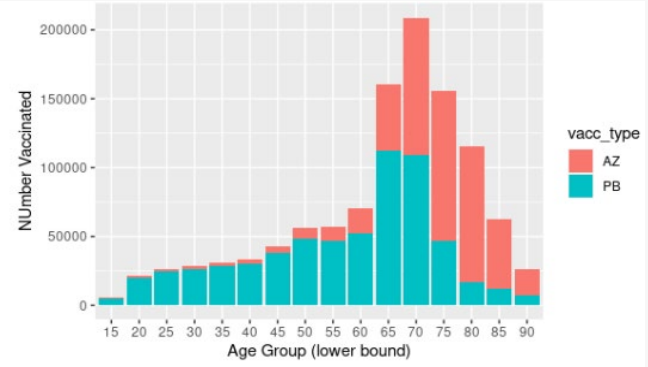
- To **confirm the real-world effectiveness of COVID-19 vaccines** to establish the success of the roll out is urgent

Solution

- Enabling access to the EAVE II dataset, which links vaccination, primary care, COVID-19 testing, hospitalisation and mortality records for 99% Scottish population (5.4M people), through the Scottish National Safe Haven
- This setting provided access for approved researchers in a **safe and secure way** and allowed **rapid analysis to data**, with publication of findings within weeks of data collection

Impact

- Analysis showed that the Pfizer and AstraZeneca vaccines reduced hospital admissions by 85% and 94% respectively (and this result was maintained for adults >80 years)
- The results provided early insights that the **COVID-19 vaccination is working on a population-wide level and provides confidence in the continuing roll out**



"This study is the first to describe the country-wide effect of the approved COVID-19 vaccines on preventing severe illness and hospital admissions… the brilliant news is that this suggests the vaccine delivery programme is working."

Dr Josie Murray,
Public Health Scotland

Effectiveness of First Dose of COVID-19 Vaccines Against Hospital Admissions in Scotland: National Prospective Cohort Study of 5.4 Million People

21 Pages • Posted: 19 Feb 2021

New large scale data assets - CVD-COVID-UK

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

HDRUK
Health Data Research UK

Challenge

At the start of the COVID-19 pandemic, approved researchers were unable to access national, linked health data across the whole UK population to carry out analysis that would support healthcare and public health policy.

Solution

- BHF Data Science Centre has developed a new Trusted Research Environment (TRE), in partnership with NHS Digital, **providing researchers with secure access to linked health data from primary and secondary care, registered deaths, COVID-19 laboratory and vaccination data and cardiovascular specialist audits**
- This data set covers the population of England of 57m people), with similar linked data made available for Scotland and Wales (>8 million people)

Impact

- Through the CVD-COVID-UK consortium (130 members from 40 UK research organisations), **>50 analysts are – for the first time ever - analysing linked health data on >65 million people to address COVID-19 related research questions**
- Linking data from different health settings is crucial to define cardiovascular events, COVID-19 diagnoses and key patient characteristics - **this resource enables statistically powerful research and includes all age groups, ethnicities , geographic locations, and socioeconomic, health and personal characteristics**
- **130 Collaborators, 50 analysts working in the TRE**

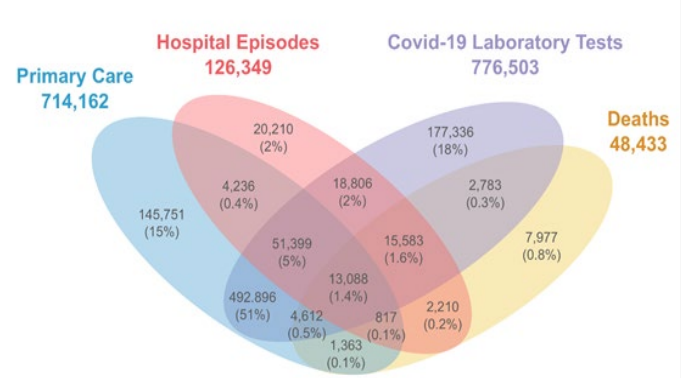


Figure - the data sources for 960,000 COVID-19 diagnoses in England from Jan-Oct 2020

"Because of our partnership with NHS Digital, researchers can now access health data at a scale that a year ago was hardly even imaginable."

Cathie Sudlow

Director, BHF Data Science Centre

Reference:

https://www.medrxiv.org/content/10.1101/2021.02.22.21252185v2

The extraordinary role of health data science in the response to COVID-19

HDRUK
Health Data Research UK

Lesson 1: Interoperable data are a key resource to help save lives

Lesson 3: The public and patients are enablers, and important advocates

Lesson 2: Getting value from data requires people with a wide range of skills working together

Further lessons from the front lines of health data research in the UK

HDRUK
Health Data Research UK

- **Start with well-defined research questions** and work backwards to identify key assets, weak links (e.g. social care) and streamline data flows – spatial, temporal and multi-modal
- **Being transparent and open with public and patients** - PPIE groups often accelerate research when engaged from the outset, instead of it being an after-thought
- **Establishing an open, interoperable and federated infrastructure** is a key enabler for resilience and synergy between national and international activities
- **Be cognizant not to build Yet Another Data Silo** – with urgency and pressure it is human nature to centralise without strengthening the underlying roots
- **Invest in data quality, interoperability and digital maturity** across all stakeholders of the health ecosystem with a particular **emphasis on diversity**
- **Build mechanisms for SAFE access to FAIR data** - SAFE People, SAFE Project, SAFE Data, SAFE Settings and SAFE Outputs.
- **Develop, train and sustain a international community of interdisciplinary Health Data Scientists** to work with diverse health data safely and securely
- **Need for a "Deep Retrospective"** to discover good, bad and alternate paths to success – there will never be a panacea but we can share what we learnt

…to enable discoveries that improve people's lives


HDRUK
Health Data Research UK
“We can only see a short distance ahead, but
we can see plenty there that needs to be done.”
~ Alan Turing (d 07 June 1954)
Find out more:
www.hdruk.ac.uk
www.healthdatagateway.org
@HDR_UK