



DATA INSIGHTS IN A PANDEMIC



Annual Review
2020/2021



Introduction

- About us 02
- Highlights 03
- Director and CEO's review 05
- Strategy 08
- Our approach 09

- Data insights in a pandemic 11
- Uniting health data 34
- Improving health data 42
- One Institute 45



Who we are

Health Data Research UK is the national institute for health data science.

What we do

Our mission is to unite the UK's health and care data to enable discoveries that improve people's lives. We do this by uniting, improving and using health and care data as one national institute.

Our vision

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.

Highlights of the year



A team from DATA-CAN, the Health Data Research Hub for Cancer, led by Charlie Davie, HDR UK Hub Director, and Mark Lawler, working with Alvina Lai and UCL researchers to highlight risk of extra deaths in people with cancer due to COVID-19 crisis

RECOVERY Trial, co-led by Martin Landray, a HDR UK Research Director finds dexamethasone reduces death by up to one third in hospitalised patients with severe respiratory complications of COVID-19 with data via the NHS DigiTrials Health Data Research Hub



954 people participated in HDR UK's first virtual 'One Institute' conference

Members of the HDR UK research community provide new insights into risk of contracting COVID-19 based on age and ethnicity

International COVID-19 Data Research Alliance (ICODA) launched

Establishment Review report for funders published

Creation of COVID-19 false positives tests interactive tool



April 2020

May 2020

June 2020

July 2020

August 2020

UK Health Data Research Alliance initiates work to promote the use of Trusted Research Environments as a key part of the ecosystem for secure use of health data



UK Health Data Research Alliance welcomes 3 new members: Generation Scotland, NIHR Clinical Research Network and the National Consortium of Intelligent Medical Imaging

HDR UK research community, including Laura Shallcross, publish insights into level of COVID-19 and excess deaths in care homes





The first students join the Wellcome PhD programme in health data science, in partnership with the Alan Turing Institute

COVID-19 National Core Studies, including the Data & Connectivity Study led by HDR UK and the Office for National Statistics, established by the UK government

HDR UK and 20 of its partners across the UK, announce the launch of an internship programme for Black health data scientists

500 people attend the UK Health Data Research Alliance Symposium online

New linked health data resource covering 54.4 million people in England made available for research through the BHF Data Science Centre led by Cathie Sudlow, HDR UK Director

The world's first real world data on single dose vaccine effectiveness published by the EAVE II study and BREATHE Health Data Research Hub led by Aziz Sheikh, HDR UK Hub Director



September 2020

October 2020

November 2020

December 2020

January 2021

February 2021

March 2021

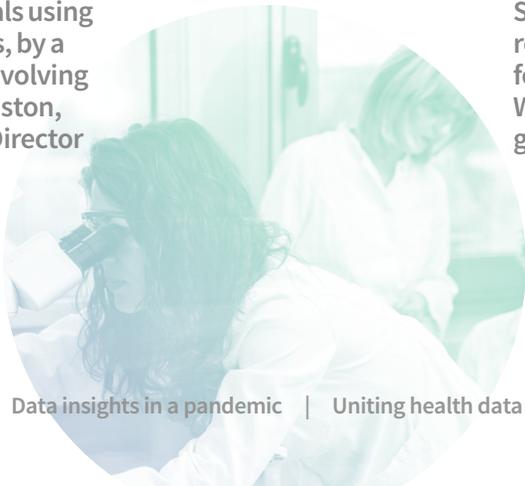
First international standards developed for clinical trials using AI innovations, by a consortium involving Alastair Denniston, HDR UK Hub Director

Using large scale anonymised data from the Zoe COVID-19 Symptom Tracker App, researchers map hotspots for the pandemic across Wales, helping to inform government policy

800 members of public contribute to vaccine research prioritisation

12 new urgent COVID-19 research "driver projects" announced

First nationwide COVID-19 vaccine datasets into made available to researchers on the Health Data Research Innovation Gateway





Andrew Morris Director

Caroline Cake Chief Executive Officer

Data saves lives

Health data research has provided enormous value to people across the UK during the COVID-19 pandemic. By uniting, improving and using data, the research community has enabled clinical trials including RECOVERY and PRINCIPLE which discovered the first effective treatments; learnt how the virus variants are evolving and the impact of new variants; identified vulnerable groups and the direct and indirect impacts of the disease on them; and evaluated the effectiveness of vaccines. This has required partnership working and collaboration on a scale never seen before.

This review of 2020/21 acknowledges and celebrates the hard work and commitment of health data scientists across the UK. It also looks ahead to the potential benefits that data science can make across all areas of health and care, building on the insights from the pandemic.

These benefits will be realised through demonstrating trustworthiness, and partnership working across the NHS, academia, charities, industry and government.

Patient and public trust

Personal health data is sensitive and requires special protection. The response to COVID-19 has reinforced not only how much patients and members of the public are interested in health data science but how willing they are to participate and support. We are grateful

for the sustained involvement from our advisory groups and members of the public who have contributed their time and experience to enable better health data research. In December 2020, over 800 members of the public contributed to our work on vaccine research prioritisation. This involvement makes research better and also demonstrates trustworthiness in how data is being used for research.

HDR UK and the UK Health Data Research Alliance have played a leading role in making the case for the use of Trusted Research Environments (TREs) as a key component of the UK's health data research ecosystem, with our green paper published in July 2020. TREs are a key part of enabling datasets to be accessed securely and protecting patient privacy.

The far-reaching impact of health data research is evidenced in over 2,000 published research papers and over 150 resources openly available on Github involving members of the HDR UK research community, many of which provide crucial insights into COVID-19. We are accelerating reproducible science by bringing together repositories of open standards, data and source code, tackling some of the most important challenges in wrangling multi-model data and generating replicable insights. The power of this "Team Science" approach is celebrated through our open events, virtually attended by 1000s of people, which showcase this research and catalyse further collaboration.



“The work of Health Data Research UK in the areas of data linkage, governance and reach across the four nations is crucial and has really been vital over the last 12 months.”

Sir Patrick Vallance, UK Government Chief Scientific Adviser

Collaboration

The UK Health Data Research Alliance now constitutes over 50 organisations that are working together to align standards and improve access to health data in a secure, transparent way. The HDR Innovation Gateway (the “Gateway”) is maturing to be a unified platform to discover and request access to over 650 health and care datasets. There are over 1,000 researchers registered on the Gateway, and with over 11,500 searches each month from around the world, the Gateway is becoming the “go to” place for researchers to discover

and request access to UK health datasets with the added benefit of giving much-needed transparency to the UK public on what data is available, how they are used and why.

Great science relies on teams of people uniting, improving and using the data. We are training people across the country and across the entire career path from internships, apprenticeships and Masters programmes, to PhD programmes and fellows. Our new virtual learning environment, HDR UK Futures, provides accessible continued

professional development. To date we have provided training for 6,704 health data scientists.

The Health Data Research Hubs combine expertise from the NHS, charities, academia and industry. In their first 18 months they have successfully brought together, improved and enabled access to 157 health datasets and delivered 300 multi-sector projects with over 20,000 patient and public interactions, and 2,300 training activities. The report [Improving UK Health Data](#) shows how the Hubs have informed UK policy decisions on

the effectiveness of COVID-19 vaccines, created tools to improve clinical decision-making in the management of patients with vascular disease, and supported research in cancer, heart disease and hospital care pathways by linking routinely-collected data. The impacts that they have had during the pandemic, include:

- The NHS DigiTrials Health Data Research Hub and HDR UK led work with the PRINCIPLE trial to accelerate recruitment of participants within 24 hours of a positive SARS CoV-2 result in the community, which enabled the addition of the budesonide treatment arm. This has since become the first widely available, inexpensive drug found to shorten recovery times in the community.

- The world’s first real world data on single dose vaccine effectiveness. The EAVE II project and BREATHE Health Data Research Hub and produced a national (5.6m people) data infrastructure, which enabled the first whole country estimates of Astra-Zeneca and Pfizer-BioNTech vaccine effectiveness by vaccine type in different age groups. This was announced by the Prime Minister in February 2021 and influenced national regulatory strategy in the UK, Canada, Denmark, France and Germany (Lancet 2021).

The hub network is growing, with MRC funded hubs in pain and mental health launched in 2021.

Science relies on partnerships. The HDR UK community has engaged in programmes that exemplify this national and international collaboration, including the HDR UK Implementation Projects, Data and Connectivity National Core Studies Programme, the BHF Data Science Centre and the International COVID-19 Data Alliance (ICODA). A common purpose of all is to nurture unrivalled coordination across researchers, infrastructure providers and data custodians.

Impact for all of us

Datasets must be representative and inclusive if the benefits from research using health data is to be relevant and applicable to all. The national data infrastructure, enabled by the COVID-19 Data and Connectivity National Core Study is supporting 885 researchers, working on 323 research projects across five national trusted research environments (ONS and NHS Digital in England, and Scotland, Wales, and Northern Ireland). Working in partnership with NHS Digital and the BHF Data Science Centre, we have created a new linked health data resource covering 54.4 million people – over 96% of the English population (BMJ 2021).

This is led by the CVD-COVID-UK consortium and is available to UK researchers to collaborate in NHS Digital's new secure trusted research environment.

We are creating tools to realise impact from data; for example in partnership with the University of Cambridge, the European Bioinformatics Institute (EMBL-EBI) and the Baker Institute the Polygenic Score Catalog improves the use of polygenic scores to enhance the prediction and understanding of 200 diseases and traits.

We are also ensuring the sector is representative of our society and providing role models for the next generation of health data scientists. The Black internship programme, in partnership with the UK Health Data Research Alliance,

supported 49 interns (60% women) in 23 organisations. Not only will this bring diverse perspectives and skills to our health and science community, it provides an opportunity for young Black people to flourish in STEM careers and contributes to our ambition to train over 10,000 health data scientists.

Looking ahead

A highlight of the year was the resoundingly positive feedback received from our core funders on the submission of our "Establishment Review" in September 2020 – summarising the impact of our work since the institute was formed.

"Alone we can do so little; together we can do so much."

Helen Keller

As we look forward to the coming year, there will be a renewed focus on our purpose and strategy as we prepare for our next quinquennium that starts in April 2023. We are committed to working with our partners and community, strengthening our connections and promoting nationwide and international collaboration for public benefit.

We thank each and every person that has worked with us during the year. The members of our research community who have worked tirelessly to use data safely to enable discoveries that are improving lives. Our patient and public advisors who have given up their own time to advise, guide and challenge our work.

Our partners who have placed their trust in us to convene and enable collaborations across UK-wide programmes. And our colleagues in the central team whose commitment has achieved so much this year.

Andrew Morris

Director

Caroline Cake

Chief Executive Officer

Strategy

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 and biomedical discovery and to enhance public health.

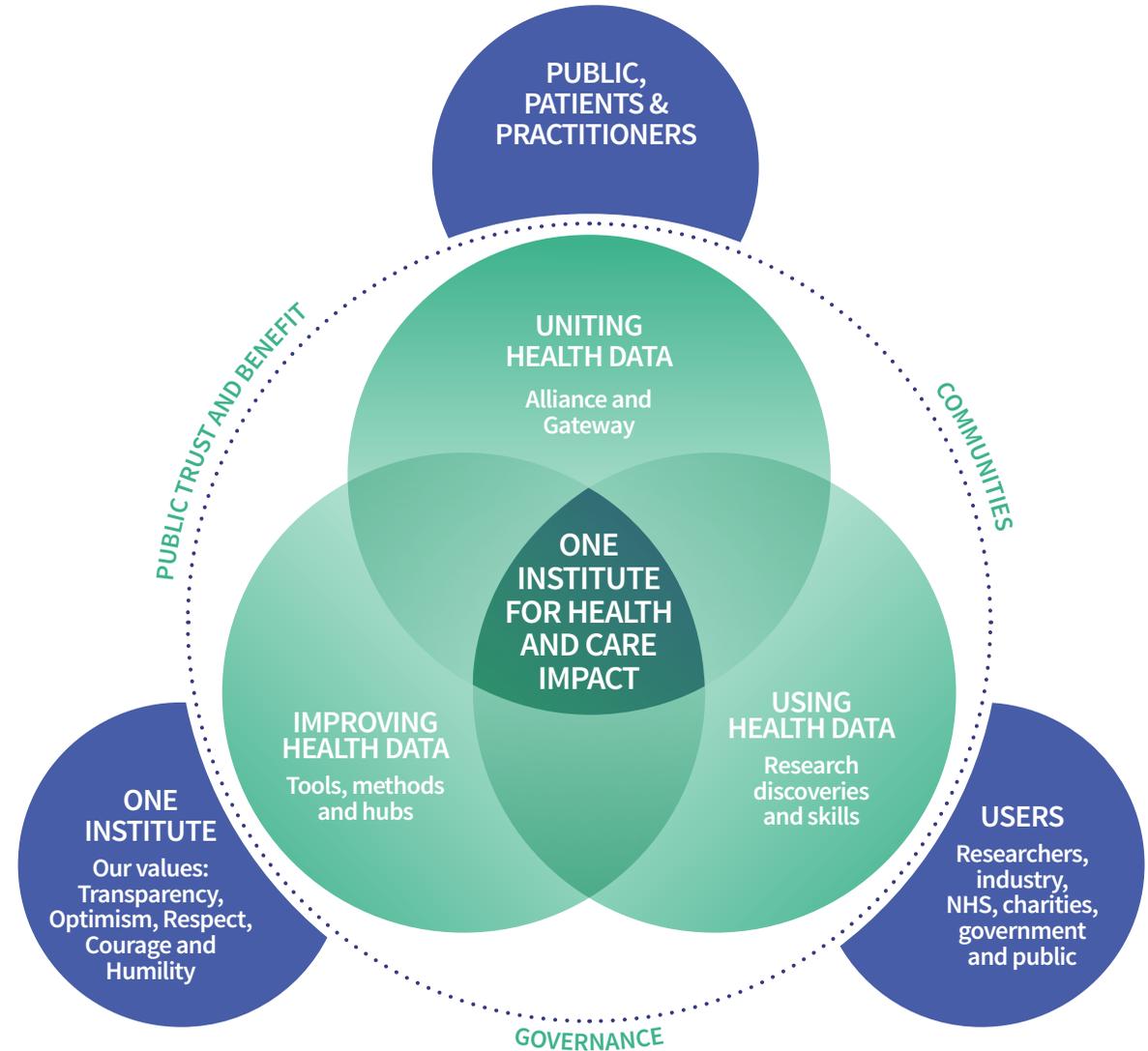
To achieve this, our strategic delivery plan, published in April 2019, focuses on our unique strengths and expertise of building a health data research infrastructure for the UK through:

Uniting health data through the UK Health Data Research Alliance and Innovation Gateway with open standards, and in a way that earns the trust of patients and the public.

Improving health data by providing tools, methods, hubs, and national expertise in health data quality improvement for researchers and innovators.

Using health data by enabling research and innovation, demonstrating novel approaches to health data use, impact at scale, and establishing the next generation of research leaders in health data science.

One institute – we are delivering this strategy through our inclusive, team-oriented ethos – bringing together NHS, universities, research institutes, industry and charities – built on our values of transparency, optimism, respect, courage and humility.



Our approach

Our commitment to Open Science in Health Data Research



We believe that open source is part of a wider open ecosystem approach that encourages open practice in multiple ways – increasing reuse and collaboration.

We aim to accelerate reproducible science and maximise limited resources through open practices, such as open standards, data and access. This means we carry out our research in a transparent way, to encourage collaboration and further use of our data for new insights.

Since HDR UK was established as the national institute for health data science in 2018, we have brought together over 150 repositories of open standards, data and source code, tackling some of the most important challenges in wrangling multi-modal data and generating replicable insights. In addition, almost 90% of HDR UK publications are open access.

We believe the HDR UK community has established a strong open practice on which to foster a community of open innovation. Now we are curating the whole into an exciting, living library for the public good.

In the coming year, we will build on our work with our research software engineers, digital research technologists, academics and others to recognise and reward software development and progress our strategy to train 10,000 data scientists by 2023. We will also contribute back to existing open source platforms and investments to improve reuse and foster a community. We will make our work even more discoverable, not just in individual repositories but across cognate groups of standards, papers, code and other research outputs.

Case study

SPOTLIGHT ON RADAR-BASE

An award-winning open source mHealth platform

Much patient data is gathered during the brief periods when people interact with health services rather than during daily life. More data, from more people, and over longer periods, would enable better research and healthcare.

Mobile and wireless technologies offer a way forward through the remote monitoring of large numbers of volunteers, but many mHealth solutions are less than ideal because they are closed source or lack flexibility.

This year, RADAR-base (Remote Assessment of Disease And Relapses) was further developed with HDR UK support by a team from PhiDataLab including Richard Dobson – an HDR UK associated researcher and Professor of Health Informatics at KCL and UCL. It enables the seamless integration of data streams from various wearables to collect sensor data in real-time for retrospective analysis.

Passive data collection allows monitoring of movement, location, audio, calls and texts and app usage. Active data collection gives the opportunity to ask questions. Information on factors like sleep patterns, heart rate and even social activity can potentially be used as biomarkers for anything from stress to depression. One example is that tracking symptoms related to fatigue may be important in understanding conditions such as Long COVID. If someone is ill, data can be gathered at every moment of the disease continuum. Healthy people can be monitored for risk.

RADAR-base unlocks the potential of wearable and mobile devices for health data research and individual care, and can be linked to a range of popular consumer products. The effectiveness and flexibility of the system has made it ideal for major research projects.

RADAR-base has already been used in national and international research projects, including remotely monitoring early Alzheimer's disease, seizure monitoring in epilepsy patients and remote monitoring in Long COVID.

All components are available as open source software under an Apache 2 license: see the RADARbase Github repository.



[View HDR UK's GitHub Repository here](#)

Our approach

Patient and public involvement and engagement (“PPIE”) – How involving patients and the public improves our work

As a result of the COVID-19 pandemic, the use of data for health research was more visible to the public than ever before - there has rarely been a day when academic and scientific topics have not been headline news.

Involving the public in our work improves the research. It provides a focus on what matter most to patients. It also helps us to demonstrate trustworthiness and build confidence in the use of health data for research and innovation. To really benefit from working with patients, carers and the public, we encourage all teams and projects to be set up in a way that allows for meaningful involvement.

This year we have had over 22,000 patients and public contributors involved in health data research. The work of our Public Advisory Board, lay members in our Governance structure, our COVID-19 Patient and Public Involvement and Engagement Group and groups and panels set up to support the Health Data Research Hubs and research priorities, have all

“In my experience of working with different PPIE panels for research over the last 10 years this ranks among the most rapid, extensive and useful!”

Harry Hemingway, HDR UK Research Director for Human Phenome Strategic Priority

made a significant impact and ensured the work we do has public benefit at its core.

The HDR UK Public Advisory Board (PAB) is our ‘critical friend’ and has a strategic focus on work related to data access, where we know there are challenging and complex questions to be answered and a potential risk to public trust. Over the last year members have shaped approaches to involving the public in our programmes on both a national and international scale, helped develop accessible information whilst ensuring transparency on the Innovation Gateway, and influenced the direction of COVID-19 research. Every

fortnight, members have provided direct and detailed feedback on our COVID-19 reporting to the government’s Scientific Group for Emergencies (SAGE) to demonstrate what is important to patients.

Public involvement and engagement has been embedded in the Health Data Research Hubs since the start, from representation on executive and scientific committees, to reviewing access requests, and involvement in the shaping of research projects. This has helped to make research more relevant, accessible, and transparent. Involving patients and the public in this way

improves the validity of the outcomes and conclusions of data science and ensures public benefit.

For our research around COVID-19, the primary focus for patients and the public was to help shape our approach to data access in addition to their role in our research question prioritisation process, developing research projects and our communication. A clear example of their impact was with the development of online risk calculator – called OurRisk.CoV - to estimate the possibility of death based on previous health factors. The initial feedback and input

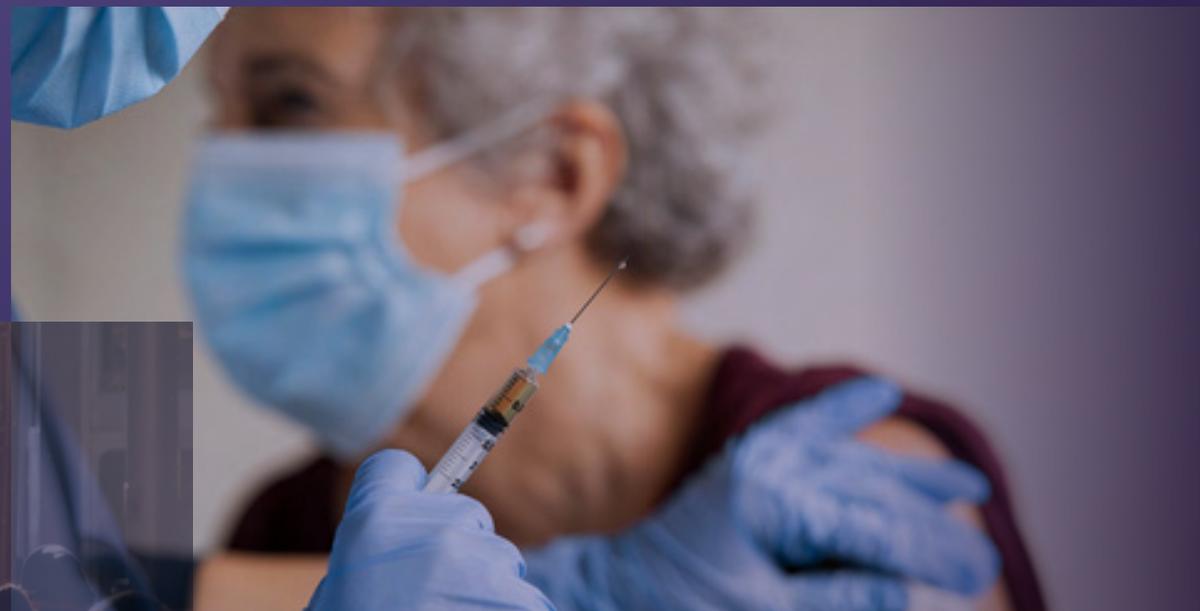
from the group altered the future focus for this work, truly showing the value of involving patients and the public from the beginning of research projects.

The public expect a consistent and higher level of transparency over how health and care data are used and how decisions are made, at a systemic level. Our consultation with the public on vaccine research prioritisation elicited over 800 responses that demonstrated vaccine safety was the primary public concern, and this feedback has been used to inform the work of the National Core Studies.



Read our PPIE Approach, Strategy and Guiding Principles

DATA INSIGHTS IN A PANDEMIC



As the national institute for health data science, HDR UK has been a key part of the UK's research response to the COVID-19 pandemic.

Our strategy has been to mobilise talent and partnerships across the UK to coordinate research efforts, accelerate safe and trustworthy access to priority data and leverage the best of the UK's health data science capability to address the wider impact of the COVID-19 pandemic.

These efforts have helped with understanding the virus, clinical trials for treatments (including proving the benefits of dexamethasone for critical COVID-19 patients), symptom trackers, risk calculators and impacts on vulnerable groups, including cancer patients.

We were also able to provide regular updates and recommendations to the government's Strategic Advisory Group for Emergencies (SAGE) on prioritised health data research related to COVID-19.

Our own portfolio of research quickly pivoted to meet the COVID-19 challenge and generated actionable insights. Members of the HDR UK research community had integral roles in:

- The COVID-19 Genomics UK (COG-UK) consortium, which has delivered large-scale and rapid whole-genome virus sequencing to local NHS centres and the UK government of over 170,000 strains; including identifying the new, more virulent strain first identified in SE England

- Mapping of COVID-19 cases in high-resolution and near real-time, which directly informed public health policy – most notably in Wales.

- Better care of patients: from debunking myths around medicines for pre-existing conditions exacerbating COVID-19 through robust health data research at scale; to understanding the impact of COVID-19 on vulnerable people with health conditions, including those with cancer and heart disease.

“Thank you, without HDR UK's work, a lot of what we have achieved with data this year during COVID-19 would not have been possible.”

Chris Whitty, Chief Medical Officer for England

As the year progressed, our activities rapidly developed into a series of organised programmes to provide services supporting the research community:

ICODA – the International COVID-19 Data Research Alliance (ICODA) and Workbench provide a co-ordinated international platform to enable researchers to access global data to drive rapid insights and speed up the development of treatments.

Research through the Health Data Research Hubs – the original seven Health Data Research Hubs have used their resources to support the UK's response to the COVID-19 pandemic, supporting areas such as infection rates, vaccine rollout, and understanding the impact of COVID-19 on specific disease areas.

HDR Innovation Gateway – COVID-19 datasets, including Vaccination Status and COVID-19 Adverse Reaction datasets, have been listed on Gateway and made available for researchers to request access to.



Discover more about ICODA

“Data saves lives, as the pandemic has starkly illustrated. By continuing to bring the health research community together to use data in a trustworthy way – and engaging with members of the public – HDR UK will help shape the post-pandemic research landscape and support the future development of health data science, both in the UK and globally.”

Caroline Cake, Chief Executive, Health Data Research UK

Data Insights In a Pandemic

Our work to support SAGE



HDR UK has produced 30 concise, regular reports for the Government's Scientific Advisory Group for Emergencies (SAGE) to provide recommendations on research projects to prioritise and an overview of data research insights. Each report was reviewed by patient and public advisors and included their comments.

30

concise, regular reports produced for SAGE since April 2020

All the reports were published on our website providing information on:

- Activity to enable priority datasets for research
- Engagement with data custodians
- Datasets available across Trusted Research Environments; and data linkages
- Datasets available to request access via the HDR Innovation Gateway
- New research projects started
- The level of participation in key UK-wide studies, such as the RECOVERY Trial

The reports were presented to a fortnightly COVID-19 “taskforce” call which, in total, engaged with 183 clinical and research leaders.

HDRUK Health Data Research UK

British Heart Foundation

HSC Public Health Agency Research and Development

NICOR

CANCER RESEARCH UK

Public Health Scotland

NHS Digital

SAIL DATABANK

UK Health Data Research Alliance

COVID-19 Health Data Research

21 July 2020 - Weekly update for SAGE & UKRI/DHSC

Authors:
Andrew Morris, Health Data Research UK
Ben Gordon, Health Data Research UK
Carole Morris, Public Health Scotland
Caroline Cake, Health Data Research UK (lead)
Cathie Sudlow, BHF Data Science Centre
Charlie Davie, DATA-CAN
Clara Fennessy, Health Data Research UK
David Seymour, UK Health Data Research Alliance
John Aston, Home Office (SAGE sponsor)
John Deanfield, NICOR
Mark Parsons, Scotland National Safe Haven

Members of the HDR UK Public Advisory Board
Melissa Lewis-Brown, Health Data Research UK
Ming Tang, NHS England and Improvement
Nilesh Samani, British Heart Foundation
Rhoswyn Walker, Health Data Research UK
Ronan Lyons, SAIL Databank (UKRI/DHSC sponsor)
Sara Hiom, Cancer Research UK
Tom Denwood, NHS Digital
Alice Turnbull, Health Data Research UK
Ian Young, Health & Social Care Northern Ireland

The world's first real world data on single dose vaccine effectiveness

The EAVE II project and BREATHE Health Data Research Hub for Respiratory Health, led by Aziz Sheikh, produced a national data infrastructure which enabled the first whole country estimates of Oxford-AstraZeneca and Pfizer-BioNTech vaccine effectiveness by vaccine type in different age groups. The EAVE II dataset links vaccination, primary care, COVID-19 testing, hospitalisation and mortality records for 99% of the Scottish population - 5.6m people – through the Scottish National Safe Haven.

The infrastructure was developed to allow approved researchers rapid access to and analysis of data in a safe and secure way, with publication of findings within weeks of data collection.

As early as February 2021, analysis from this work showed that the Pfizer-BioNTech and Oxford-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 provided confidence in the continuing roll out.

This was announced by the Prime Minister in February 2021 and influenced national regulatory strategy in the UK, Canada, Denmark, France and Germany (Lancet 2021).



Aziz Sheikh
Professor of Primary Care Research & Development and Director of the Usher Institute at The University of Edinburgh

“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



Find out more about the insights health data research has supported and provided to SAGE as part of our response to COVID-19.

Data Insights In a Pandemic

Data and Connectivity – supporting the research response to COVID-19



In October 2020, the Data and Connectivity National Core Study was established, by the government, led by HDR UK in partnership with the Office for National Statistics to support and accelerate research on COVID-19. Data and Connectivity is a “cross cutting” workstream, supporting the other five National Core Studies into key research questions by enabling streamlined, safe and responsible access to relevant datasets relevant for the pandemic.

The study is the formal continuation of the work we had already started as part of our response to the pandemic; bringing together key assets of the UK data infrastructure by working in partnership with stakeholders from across the four nations to organise health and other data on an unprecedented scale.

72

priority datasets available on the Gateway

Working with partners across the four nations, the programme addresses the existing challenge of bringing together previously fragmented datasets and making them available for researchers to request access via the HDR Innovation Gateway.

Urgent Research Projects

In December 2020, we sponsored 12 urgent research projects following our rapid call for COVID-19 data research initiatives. Among the projects were vital research questions about the effectiveness of COVID-19 vaccines, a study of long-COVID and linked datasets to understand transmission of COVID-19 in schools.

A national linked study to understand socioeconomic inequality in COVID-19 vaccinations among elderly adults in England discovered that those groups most affected by the virus were also most hesitant to be vaccinated. The study also highlights differences in vaccination rates by deprivation, household composition, and disability status, factors disproportionately associated with infection, providing valuable insights for public health policy and community engagement to promote vaccine uptake.

The projects will complete in Summer 2021 and are expected to leave a legacy for future research studies by enhancing the value of data by, for example, creating additional data linkages, improving the quality of data and following best practice in open science, sharing code and tools.

Impact

The Data and Connectivity National Core Study has supported 885 researchers, working on 323 research projects across a network of five national trusted research environments.

By March 2021, there were 72 COVID-19 priority datasets available on the Gateway, and vaccine data flowing to national Trusted Research Environments for researcher use.

855

researchers supported...

...working on

323

research projects

Public and patient involvement

Patient and Public Involvement and Engagement has been invaluable for the programme, particularly in enabling research into the COVID-19 vaccines. Over 800 patients, carers, members of the public and practitioners participated in a two week-long consultation to prioritise health data research questions on COVID-19 vaccines. The findings were shared with our COVID-19 research delivery partners and funders.

HDR UK also worked with AstraZeneca to improve and speed up recruitment into clinical trials, by designing a survey to discover how acceptable the public would find it to be contacted to participate following a positive test for COVID-19, how and who would be best to do so.

A rapid survey yielded 106 responses providing important insights:

- People would rather be contacted by researchers or their GP (over representatives from NHS Test and Trace)
- Reassurance that patient data is only used for research
- A desire for increased understanding of the role of AstraZeneca, that the organisation itself would not be able to access or profit from patient data
- More detail on the trial and potential side effects

The insights led to increased provision of information throughout the recruitment process to improve transparency, including the development of a dedicated website for potential recruits to find out more and provide feedback.

Case study

PUBLIC AND PATIENT INVOLVEMENT

Patient and Public Involvement and Engagement has been invaluable for the programme, particularly in enabling research into the COVID-19 vaccines.

Over 800 patients, carers, members of the public and practitioners participated in a two week-long consultation to prioritise health data research questions on COVID-19 vaccines. The findings were shared with our COVID-19 research delivery partners and funders.

HDR UK also supported AstraZeneca to improve and speed up recruitment into clinical trials, by designing a survey to discover how acceptable the public would find it to be contacted to participate following a positive test for COVID-19, how and who would be best to do so.

A rapid survey yielded 106 responses providing important insights:

- People would rather be contacted by researchers or their GP (over representatives from Test and Trace)
- Reassurance that patient data is only used for research
- A desire for increased understanding of the role of AstraZeneca, that the organisation itself would not be able to access or profit from patient data
- More detail on the trial and potential side effects

The insights led to increased providing in transparency and information through the recruitment process, including the development of a dedicated website for potential recruits to find out more and provide feedback.

Data Insights In a Pandemic

ICODA: A global health data response to COVID-19



In June 2020, the COVID-19 Therapeutics Accelerator announced a grant to HDR UK and partners to establish a new International Data Research Alliance and data analysis Workbench to support the rapid development of insights and treatments to combat the global effects of COVID-19.

Working collaboratively to unite and harness the power of health data from around the world, ICODA (the International COVID-19 Data Alliance) aims to enable discoveries to treat and prevent COVID-19, as well as co-ordinate an effective global health data response to health challenges of the future.

We offer data contributors a streamlined process to facilitate safe and responsible access to data, with rigorous data governance and a secure workbench environment with a wide range of analytical tools. We have prioritised Driver Projects – specific research questions, bringing together a range of datasets for researchers to analyse to deliver health impact.



Driver Project 1 is a holistic evaluation of the efficacy and safety of treatments for COVID-19. Several thousand clinical COVID-19 trials are in progress globally. As these trials are evaluating the benefit/risk of potential COVID-19 treatment options, it is vital that the scientific community can interrogate this data as it emerges.

The summary level data from the most rigorous of these trials across industry, academia and government is being included in the ICODA Workbench and made available to the community of researchers. With their support, we are strengthening our collective efforts and enabling discoveries to help address this global challenge.

Driver project 2 focuses on the International Perinatal Outcomes in the Pandemic (iPOP) Study, exploring the impact of the pandemic lockdown on preterm births worldwide, including variances across countries.

Whether the pandemic is worsening or unexpectedly improving new-born health, our research is providing critical new information to shape prenatal care strategies throughout (and well beyond) the pandemic.

The iPOP Study and ICODA, recently featured in Nature, currently involves over 100 researchers in more than 40 countries, including 22 lower- and middle- income countries (LMICs), with access to data on 2.4m births. Together, they are leveraging the most disruptive and widespread changes of our lifetime to make rapid discoveries about preterm births.

ICODA is funded by the COVID-19 Therapeutics Accelerator, a large-scale initiative started by the Bill & Melinda Gates Foundation, Wellcome, Mastercard with additional support from Minderoo Foundation, and other donors.

Case study

GRAND CHALLENGES ICODA DATA SCIENCE PILOT INITIATIVE

In December 2020, the Bill & Melinda Gates Foundation and the Minderoo Foundation supported ICODA through a pilot data science initiative to address major questions about how to reduce the harm of COVID-19 and future pandemics.

The initiative aims to unite data and develop processes, analytical tools and infrastructure to achieve rapid scientific progress and impact; and act as a pilot for the ongoing Grand Challenges that seek to develop global collaboration and innovative approaches to major health challenges.

Ten projects are ongoing, representing 23 organisations in 19 countries.



Data Insights In a Pandemic

Data-enabled clinical trials for public benefit



Routinely collected healthcare data holds the key to transforming clinical trials. However, it is estimated that fewer than 5% of UK trials between 2013 and 2018 were accessing data from healthcase systems.

Work undertaken by HDR UK in 2019 to improve clinical trials has supported two of the most significant scientific breakthroughs of the response to COVID-19 in 2020.

The RECOVERY Trial

In early March 2020, little was known about COVID-19 and there were no effective treatments, with around one in four patients admitted to hospital dying. The RECOVERY Trial, co-led by Martin Landray, HDR UK Research Director, and Peter Horby was set up in just nine days to test drugs against COVID-19, with over 38,000 patients now recruited through 180 hospitals in the largest randomised trial for COVID-19 treatments.

The NHS DigiTrials Hub and HDR UK's clinical trials team, led by Marion Mafham, enabled safe access to linked data from hospital admissions, discharge, treatments, deaths, and COVID-19 test results.

By June 2020, RECOVERY had identified that dexamethasone, a low-cost and widely available drug, reduced death rates. Within hours of the RECOVERY Trial team announcing its findings, England's Chief Medical Officer had recommended its use across the NHS. Its use was soon worldwide. The trial has saved hundreds of thousands of lives by identifying which treatments are effective for hospitalised COVID-19 patients.

Within a year the trial had also discovered tocilizumab as an effective treatment for COVID-19.

RECOVERY has also ruled out 4 drugs – including hydroxychloroquine - as viable treatments (an equally valuable insight).

Many findings from RECOVERY have contradicted the expectations of substantial sections of the medical and scientific community – underlining the value of large-scale randomised tests in providing a clear evidence base.

“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.”

*Professor Sir Martin Landray,
HDR UK Research Director, Clinical Trials*



Professor Sir Martin Landray



Case study

The PRINCIPLE Trial

The PRINCIPLE trial was created to discover COVID-19 treatments for people aged over 50 who are at high risk of complications, which could be taken at home quickly and therefore prevent them from needing to attend hospital.

By early October 2020, the trial had registered 1,630 eligible participants; with several barriers impacting the ability to recruit more people due to the nature of the pandemic.

HDR UK enabled the use of linked datasets to identify people with a positive COVID-19 test within 24 hours from the UK's NHS Test and Trace system and Summary Care Record data, massively accelerated recruitment to 4,671 patients by 31 March 2021.

HDR UK also led the crucial public engagement to ensure patients were informed and comfortable with the recruitment process; an exercise which also helped modify and improve the process.

The work undertaken to support the trial demonstrated a new way of enhancing early recruitment to a clinical trial in a community setting and discovered that budesonide, a cheap drug, commonly used to treat asthma, can help people at home recover more quickly from COVID-19.



[Read about PRINCIPLE in more detail](#)



[Visit the PRINCIPLE Trial website](#)

“The creative energy that has gone into making the progress you outline could well be saving lives soon. You are making history.”

Chris Butler, Principle PRINCIPLE Trial Lead

Data Insights In a Pandemic

Using health data to provide better care



Our Better Care programme aims to improve people's lives by equipping clinicians and patients in the UK with the best possible data-based information to make decisions about their care.

HDR UK's Better Care National Priority, led by Simon Ball, encompasses a number of partnership programmes across the HDR UK research community, demonstrating the real-world application of the institute's work in clinical settings.

In the last twelve months the programme has expanded geographically to include two new partnership regions in the South-West, led by Jonathan Sterne and North of England led by Munir Pirmohammed.

Case study

DEVELOPING A NEW MODEL FOR TRIAGING COVID-19 PATIENTS

A collaboration of NHS doctors, service planners, and academics working across the HDR UK South-West Better Care Partnership, used computer modelling to assess the value of potential strategies for triaging COVID-19 patients for intensive care.

This study compared no triage, with a triage approach prioritising patients with the greatest chance of surviving admission based on age. It found that the triage approach had little impact on the number of lives saved; but can result in an 8% reduction in number of life-years lost, because the lives of younger patients are saved.

Improved results could be achieved by admitting all patients but considering early discharge for those with worse survival chances given arrival of others with better prognoses and lack of spare capacity. Under this "reverse triage" strategy, total life-years lost can be reduced by 12%, based on a conservative assessment of the likelihood of death following early discharge.

"Using data, modelling and analytics, this study allows us to give more effective guidance to clinicians to support decision making and deliver real-world benefits. It is fantastic example showing how health data and collaborations between doctors, planners and academics can inform NHS decision making in the crucial ICU setting."

*Jonathan Sterne, Research Director,
HDR UK South-West Partnership*



An innovative partnership
with the Health Foundation



In addition to ongoing programmes, a ground-breaking partnership with the Health Foundation has delivered three rapid research “catalyst projects”, completed in 12 months.

RADAR (Risk Algorithms for
Decision Support and Adverse
Outcomes Reduction)

This project in North-West London, which also involved HDR UK’s Discover-NOW data hub, resulted in significantly improved prediction accuracy of artificial intelligence (AI) models to provide calculated risk information for patients with diabetes.

This enabled better-informed decisions around treatment and self-management for patients so that hospital admissions are reduced. Feedback also showed a real appetite for the use of these type of tools and the desire for patients to engage more with their own data.

Using diabetes as an exemplar, this project will provide the opportunity to use real-world data for the improvement of outcomes in patients with other long-term conditions.

“The Health Foundation is delighted to see the outputs of these projects, demonstrating real-world impact of using data-driven approaches in three crucial areas of healthcare. Working with HDR UK has validated the importance of a partnership approach to understand how we can best use data for health research.”

Adam Steventon, Director of Data Analytics at the Health Foundation

BREATHE – AI-driven improved
clinical decision-making tools
to manage a life-long chronic
condition (Cystic Fibrosis)

Coordinated by Royal Papworth NHS Foundation Trust across five organisations, this project demonstrated that patients with cystic fibrosis can use a Bluetooth device in the home to frequently monitor their condition and more accurately predict future flare ups.

It also showed the enormous value of empowering patients by providing the tools they need to understand and manage their conditions themselves, creating a tailored approach to care and a reduction in the costly burden of requiring all patients to attend every clinic.

The projects validate the use of artificial intelligence and patient datasets to provide clinicians with new insights that enable improved care; and have the potential to be developed and scaled across other clinical areas.

Data Insights In a Pandemic

Understanding the causes of diseases



We use health data in its multiple forms to understand the causes of disease and discover new targeted treatments rather than just addressing symptoms. Led by John Danesh, our community of experts work to advance understanding of disease prediction, causation and progression.

The overarching hypothesis is that insights into biology and the causes of disease (aetiology) can be revealed by integration of information, at scale, on genomics, other biomolecular traits, and high-resolution electronic health records. Our ultimate vision is to create new informatics infrastructures and data science methods that help achieve a deep integration of biology, biomedicine, and population health science.

Among a number of projects this year, using health data at scale, highlight the impact that our team of researchers are having in this field.

SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe

Drawing on cohort and routine record data on 12.5 million individuals from dozens of countries, HDR UK researchers alongside others developed and validated a new cardiovascular risk score tool (“SCORE2”) adopted by the European Society of Cardiology (ESC) 2021 European Guidelines on CVD Prevention in Clinical Practice, and its associated ESC software app.

This work has generated major media coverage, including several national UK papers (e.g., Times, Daily Telegraph, Daily Mirror, Sun, Daily Express), in addition to ~50 regional newspapers.

Case study

THE POLYGENIC SCORE CATALOG – AN OPEN DATABASE FOR REPRODUCIBILITY AND SYSTEMATIC EVALUATION

Polygenic risk scores (PGSs) have immense potential for helping predict people’s likelihood of developing many serious illnesses. They work by aggregating the effects of a multitude of small genetic variants into a single number. However, progress has been hindered by a lack of common standards or best practice.

The development of Polygenic Risk Score Reporting Standards (PRS-RS) and a new Polygenic Score (PGS) Catalog to establish the ‘minimal information’ that should be reported by authors of PGS research to ensure analytic validity, transparency and reproducibility is a major step forward.

The catalogue has transformative potential to help a young area of research become of real clinical value. It is already being extensively used and holds more than 800 standardised polygenic scores and 140 papers, demonstrating its capacity to harmonise and unify them in one database.

The PGScatalog (www.PGSCatalog.org) is an open access community platform for PGS studies, where researchers can deposit their PGS and each PGS is consistently annotated with relevant and high-quality metadata – ensuring PGS data can be reproduced and reused to enable new uses and evaluation of PGS predictive ability.

Case study

UNCOVERING EXISTING DRUGS THAT CAN TREAT COVID-19

Developing new treatments for any disease can be costly and time-consuming. When there is an urgent global health crisis, the need to act swiftly by finding out if existing treatments might be effective can therefore have huge benefit. Large datasets – including some combined by HDR UK from several studies – make it possible to identify the genetic variations most common in COVID-19 patients, and then to identify existing drugs which are already known to target those variations.

The international team, which was co-led by HDR UK researchers Liam Gaziano and Adam Butterworth at the University of Cambridge also looked at gene expression variations – in other words, differences in how cells ‘read’ the instructions from DNA about what they should do and when.

Having identified which genetic variations were most common in COVID-19 patients, the researchers then cross-referenced this with information on existing drugs. This enabled them to find out which drugs are already proven to work against the genetic variations seen in COVID-19 patients.

The team identified two proteins that are common in patients with COVID-19 and particularly associated with hospitalisation: ACE2 and IFNAR2. Drugs already exist which target both of these proteins, with ACE2 inhibitors commonly prescribed to treat heart problems and high blood pressure.

The team therefore suggests that, in trying to find drugs to treat early-stage COVID-19, we should prioritise trials of medicines that target these two proteins. And by prescribing

medications to people shortly after they have been diagnosed with COVID-19, it may be possible to reduce the number of people becoming seriously ill with the virus. This would mean fewer people having to go to hospital or into critical care, and hopefully fewer deaths.

More broadly, the study has shown the value of collaboration and sharing in health data research. The team analysed data from more than 7,500 patients who had been hospitalised with COVID-19, much of it made available through the COVID-19 Host Genetics Initiative. This is an international consortium to which researchers are contributing information they already have from other studies.

The team at the University of Cambridge, for instance, has conventionally focused its research on heart disease. But it was quickly possible to link the team’s existing data to COVID-19 diagnoses, identifying which patients in an existing heart disease study had been diagnosed with the virus.

This sort of uniting and reuse of data is helping researchers make quick, robust and swift discoveries – and will have an impact on future studies, whether on COVID-19 or other conditions.

In many cases, the data we need to identify what contributes to poor public health already exists in GP surgeries and hospitals – the challenge is to create the infrastructure to link it all together.

Our public health theme, led by Ronan Lyons, enables data science to transform public health research through linking to data beyond health care, for example, to other government sectors, organisations, and data on environments that influence health. By linking data beyond the individual, we can better evaluate risk factors, outcomes and potential for interventions that target related individuals or groups. By enabling whole country comparisons, we inform national policy development for the UK, improving health of all.

Case study

WALES MORBIDITY COHORT (WMC)

Wales Morbidity Cohort is the most detailed total population multimorbidity routine data cohort available with detailed GP, hospital, laboratory results, mortality and Census data linked on the Welsh population of 3.1m from 2000-2019.

When COVID-19 struck, the HDR UK Wales team was seconded to work for Chief Medical Officer for Wales, using WMC to support the urgent public health response.

A multi-level set of privacy-protecting linkages at individual, household, school, care home and health facilities were developed, which enabled rapid analysis to be conducted on the spread, determinants and population impacts of the disease. The effectiveness of counter-measures to be evaluated and the findings fed directly to the Welsh Government COVID19 Technical Advisory Group (TAG) and onwards to SAGE prior to publication and more widely through pre-print and rapid publication routes.

This group further contributed substantially to leading UK and international initiatives, including COVID19 TAG, NERV TAG and SAGE groups using epidemiological modelling and data science to directly inform national policy in England, Wales, Scotland, and Northern Ireland. The group produced influential thought pieces on data science opportunities in ageing, dementia, and surgical research.

This research has also been leveraged to support efficient collaboration with disease specific hubs in HDR UK, including BREATHE, DATA-CAN, BHF Data Science Centre and others.

Case study

MAPPING COVID-19 ACROSS NEIGHBOURHOODS IN REAL-TIME

To help understand the localised spread of COVID-19, this study adopted sophisticated geo-spatial modelling methods to estimate near-real-time prevalence of infections at small-area resolution using data from the “Zoe” COVID-19 Symptom Study app. The maps generated provide the first fine-scale, UK-wide assessment of the geographical distribution of probable COVID-19 infections.

The project demonstrated the value of a real-time spatio-temporal inferential mapping platform for public health efforts during the emergence and spread of infectious diseases. This study is an exemplar of how a combination of skills (health informatics, statistics and geography) is needed to provide insights to inform local and national government policy at a UK level throughout the COVID-19 pandemic. The system combines real-time data sources and rapid analytical tools and is believed to be the first work that can give predictions at Lower-layer Super Output Area resolution in near-real time.

Building on the HDR UK ‘One Institute’ principles, the paper has prototyped and delivered data infrastructures and analysis pipelines capable of delivering timely and insightful analytics to all levels of government. This research has been used by the devolved administrations for pandemic planning, for example in identifying local hotspots that previous regional-level mapping may have masked.

The COVID Symptom Study app holds data on over 4 million individuals and is located securely in and extracted from the Secure Anonymised Information Linkage (SAIL) Databank.

The study was conducted by a strong research team consisting of mainly HDR UK members from multiple medical schools across the UK, representing a collaboration led by BREATHE – The Health Data Research Hub for Respiratory Health. The work showcases key attributes of HDR UK research values – team science, scale, open science, patient and public involvement, and equality, diversity and inclusion.



[Read the paper in full](#)

Phenomics is the study of how the environment and a person's lifestyle interact with the expression of their genes to influence their health and risk of disease. Led by Harry Hemingway, our phenomics theme aims to connect data, phenomics ontologies and analytics seamlessly across the infrastructure.

£1.25m

NHS Trust revenue gained in a year

The Phenotype Library

In order to study the Phenome, HDR UK researchers tap into the enormous amount of health data contained within Electronic Health Records (EHRs). They develop tools and methods to unlock this data and make it useable and useful, for healthcare professionals and the patients they care for.

These tools include the Phenotype Library, the world's largest open-access library of reproducible phenotyping algorithms for defining disease, behavioural risk factors and biomarkers in EHRs - which now has over 2000 monthly users. We pioneered NLP approaches to convert clinical notes into analysable data and have developed an open-source healthcare analytics platform (COGSTACK MEDCAT) that has been implemented across multiple hospitals internationally.

This work has led to impacts on patient care, service improvement and audit, recruitment to trials, and population research. Using this engineering, we have delivered prognostic research which has informed policy and practice, including a model and online tool for understanding excess mortality over one year during the COVID-19 pandemic.

Other highlights this year include HDR UK researchers being among the first in the world to demonstrate the near shutdown of cancer services in response to COVID-19 and its potential impact on excess mortality, which helped inform the re-start of cancer services by NHS England's National Cancer Director.

An HDR UK-led team of clinicians, epidemiologists, statisticians, data scientists and informaticians also used

large-scale samples of NHS patient data to develop a simple risk calculation tool for policymakers, researchers, and the public to estimate excess deaths from COVID-19, ("OurRisk.CoV"). To date, the tool has had over 1.3 million views.

Risk calculator developed to show how underlying health conditions can affect mortality rates in COVID-19 pandemic Researchers at University College London, University College London NHS Trust, University of Cambridge and HDR UK collaborated to find out how underlying health conditions could affect mortality rates during the COVID-19 pandemic. The study revealed that having an underlying health condition, such as heart disease or diabetes, increases a person's risk of death fivefold over the next year. We developed a prototype online risk

calculator – called OurRisk.CoV – showing how age, sex and underlying health conditions can affect mortality rates depending on varying rates of infection within the population and varying overall impact of the COVID-19 pandemic.

CogStack

CogStack is a platform to support open-source healthcare analytics within the NHS and represents a major advance in the capacity to extract and analyse unstructured data from electronic health records (EHRs). It operates in near real-time and is user-friendly, allowing data to be accessed by asking straightforward questions. It can be used for everything from large-scale research and business intelligence to planning patient personalised care.

It has been used for successful projects including: large-scale research investigating the effectiveness of the NEWS2 hospital early warning score system to predict 14-day outcomes for the most seriously ill COVID-19 patients; improving the safety of prescribing Methotrexate for rheumatology patients; improving the use of the antipsychotic medication Clozapine; and identifying outpatient orthopaedic procedures missed by manual coding – leading to annual NHS Trust revenue gain of over £1.25m a year.

Data Insights In a Pandemic

Applied Analytics



Through our Applied Analytics theme, led by Dave Robertson, we match knowledge of analytical tools, such as machine learning and artificial intelligence, with large-scale health datasets to show their potential to inform health and care delivery. We help the right people to gain access to the data they need, in a way that protects the data, maintains patient anonymity, and allows innovation.

173

publications produced
by Applied Analytics
researchers this year

To date, our Applied Analytics researchers have produced 173 publications, seven open-source software tools and two databases. They have also delivered fundamental national data engineering to support prognostic research, which in turn has informed policy and practice, including through participation the FDA's Clinical Trials Transformation Initiative guidelines on mobile technologies in clinical trials.

An example of how Applied Analytics successfully pivoted to provide insights into COVID-19 is the analysis of text written by doctors being used to find and extract patterns and hidden nuances within medical notes of those who have tested positive for COVID-19.

It found that people taking ACE inhibitors (medicines used to treat high blood pressure or diabetes) were no more likely to experience a severe form of COVID-19 than those who did not take them. This work was also featured in a Statement of Opportunities from UKRI – “Transforming our world with AI”.

Data Insights In a Pandemic

National Implementation Projects



Sharing resources such as code and algorithms increases reuse and lends to quicker, more efficient science and collaboration. Our National Implementation Projects are “jewels in the crown”, bringing together leading experts from across the UK to create the tools and technologies needed to generate new knowledge from health datasets.

National Reproducible Machine Learning Project

The project, led by Aiden Doherty, Chris Holmes and Martin Landray, is bringing together data science, machine learning, health data from wearables, and reproducibility with the aim of harnessing their power to provide trustworthy clinical insights.

To improve standards, robustness and reporting of machine learning methods, as well as promote a culture of reproducibility to emerging health data scientists, a draft paper proposing reporting guideline protocols was submitted to BMJ Open.

We are investigating the potential of synthetic datasets as a way of providing research data while protecting patient privacy. In addition to progressing this work on synthetic datasets, the team is also exploring the development of tools to support reproducible machine learning as a way

of supporting UK health data scientists working restrictive safe haven environments.

National Phenomics Resource

The project, led by Spiros Denaxas, is developing new tools to help analyse electronic health records (EHRs), maximising their benefits by advancing the creation of phenotype algorithms to identify disease status, onset and progression.

The scoping and prototyping of a phenotype library (a comprehensive, open-access resource providing researchers with information, tools and phenotyping algorithms for UK electronic health records) has made substantial progress. By the end of the year, the HDR UK Caliber phenotype library had 353 phenotype algorithms and another 569 been processed; with more to be added in the coming year.

All of the executable forms of Caliber (and other) phenotypes are now online, can be downloaded and executed against a user’s own datasets. This represents major progress towards creating the default phenotype reference point for academia, clinicians, industry and guidelines for community that are trustworthy and transparent.

National Text Analytics Resource

This project, led by Richard Dobson and Angus Roberts, is pioneering the greater use of information stored in electronic health records (EHRs) to help tailor treatment for individual patients and for better, safer healthcare.

Text analytics is already improving research, hospital service planning and clinical decision making in areas such as severe mental illness, multimorbidity and cancer, as well as being used for COVID-19.

One exemplar is a clinician-led project which created dashboards for care teams to use text (and other) analytics for patients on a day-to-day basis. In a mental health trust this would, for example, allow a doctor to gather valuable information ahead of a patient’s psychosis medication review, or for a team to liaise with local GPs on patients they share or to look at wider population groups.

To promote sharing and open access, 95 tools have been made available through the HDR UK Innovation Gateway; and a set of GitHub resources, providing more detail on implementation has been made available. Conversations are underway with NHSX for improving data collaboration and on developing the sustainable models for the new technologies.

National Multimorbidity Resource

Led by Colin McCowan, the team is bringing together six datasets with anonymised information on over 10 million people to improve care by discovering more about what illnesses are found together, how they develop as people age and which cause the most problems. This is already yielding results, for example in work to identify clusters of multimorbid conditions that seem to be associated with hospital or care home admission.

Led from Swansea and working closely with the Chief Medical Officer in Wales analysis has been provided for the Welsh Government Technical Advisory Group (and comparable work has been taking place in Scotland and England).

Addressing a major challenge that multimorbidity is measured in many ways, the team also undertook the largest review of its kind to analyse different definitions and recommend a set of common measures. The paper, published in Lancet Public Health, represents a major milestone in moving to standard definition for research.

National Multi-omics Consortium

The project aims to create a platform that will allow researchers to interrogate the multi-omic data in order to better inform disease aetiology and prediction.

Led by Adam Butterworth, The consortium is successfully bringing together a dozen cohorts (c.800,000 participants) from around the UK that have common information – genetic, genomic, linkages to EHRs and multi-omic. This work has contributed to phenome-wide association studies to identify >100 causal effects between 65 proteins and >50 disease phenotypes, with implications for prioritising therapeutic targets. By combining population-based studies across 174 metabolites and 85,000 participants that identified genetic regulators of human metabolism and health, we have provided proof-of-concept for cross-platform ‘omics data integration.

So far the work has led to a set of powerful research outputs and insights:

- A paper published in Nature Genetics combining data from several of the cohorts while using different platforms for capturing metabolites.
- A paper in Nature Medicine has resulted from the consortium’s work in linking analytes with health outcomes – specifically looking at metabolomics platforms versus hospital derived phenotypes.
- A paper Nature Medicine, following work to use proteomic and transcriptomic data to identify causal factors related to COVID-19 outcomes.

Work has also taken place on using genetic data from two cohorts to predict blood cell traits with the aim of developing a reproducible predictor that could be taken into other cohorts.

Read more about how our national multi-omics consortium is informing disease aetiology and prediction.

Looking ahead, the consortium has also laid foundations for infrastructural innovation to support research that integrates multiomic data and electronic health records, including an innovative Trusted Research Environment (TRE) for multiomics analysis with the Health Informatics Centre at the University; enhancing services already in use by multiple cohorts at Swansea University’s Secure eResearch Platform; and piloting cloud-based software for cross-cohort analyses through collaboration with DNAnexus; and piloting secure TREs for genomic data analysis in large-scale population cohorts through core-funding from the Wellcome Sanger institute.

Supported by an NIHR AI award, the consortium is also developing AI-systems for genetic blood group typing, automated stocking of blood of different types, and precision matching of patients to blood units in order to transform the quality and efficiency of blood matching, reduce complications of blood transfusions, and improve clinical care for patients.



Read more, including how this project links with our “Understanding Causes of Disease” National Priority

Data Insights In a Pandemic

HDR UK learning – attracting and developing talent



As the national institute for health data science, a key part of our mandate is to develop the next generation of health data scientists and provide high-level training for those seeking to advance in the sector.

10k

target of trained health data scientists over five years

During the year we made substantial progress with our talent and training programmes. To date, we have provided training for 6,074 health data against our target to have trained 10,000 health data scientists in five years.

A key driver of success (especially given the accelerated transition to digital and virtual learning as a result of the pandemic), has been the ability to increase our visibility and reach relevant audiences for through digital marketing, an enhanced web presence, social media promotion, a dedicated newsletter, as well as a new identifiable brand identity for the programme.

Short courses

In June 2020, there were 80 attendees at our summer school. We anticipate that by the end of our first five years as an institute, over 250 people will have attended our summer schools and other short courses, such as a health information engineering bootcamp for 30 delegates, planned for next year.

Masters

Our seven newly created master's programmes, delivered in partnership with six leading universities across the four nations (Exeter, London School of Hygiene and Tropical Medicine, Leeds, Queens University Belfast, Bristol and Cambridge), commenced in September 2020 with 76 students.

The programmes provide the best possible foundation for careers in health data science and will be delivered across four cohorts (2020-2024), provide over 300 talented students the opportunity to pursue a master's degree and career in health data science.

In addition, we have also funded 10 master's students in Scotland via DataLab, Scotland's Innovation Centre for Data and AI.

PhD programmes

Our flagship HDR UK/Turing Wellcome PhD programme, led by Christopher Yau, delivered in partnership with 7 leading universities will provide studentships for circa 32 people over 8 years (2020-2028). The first cohort of 7 high-calibre students commenced the PhDs in September 2020. The second cohort will comprise 11 PhD students, including 2 funded by the BHF Data Science Centre, our partnership programme with the British Heart Foundation.

“It has been great to get to know the other fellows, learn about their research and how they are doing things. Learning how to apply some of their methods to my own data has been really invaluable.”

*Joram Posma,
Summer School attendee and HDR UK fellow*

Our future focus is enriching the programme through a bespoke industry-led leadership development, growing industry investment in internships, and further embedding the programme within HDR UK strategic priorities.



Christopher Yau
Director of HDR UK-Turing
Wellcome PhD Programme

Fellows

The HDR UK Fellowship Programme concluded in February 2021 and has developed a new generation of 46 leaders in the field. Our 46 UKRI Innovation Fellows and UK HDR Rutherford Fellows, all received expert-led high-level career development through the Fellows Incubator run in partnership with the National Institute for Health Research (NIHR) Academy. They also pursued their own research agendas, built up new networks and established fresh collaborations with other groups or individual researchers.

With the three-year fellowships complete, there is clear evidence of their contribution to advancing knowledge through high quality research, in guiding policy, engaging a wide audience, and increasing recognition. Due to the programme, Fellows have:

- leveraged £3.6 in further funding for every £1 invested
 - from different sectors and internationally.
- had their publications cited over 9,000 times.
- been recognised for their achievements 39 times – as keynote speakers, through honorary positions, research prizes, and honours such as L’Oréal UNESCO Women in Science fellowship.

- reported over 100 collaborations spanning 20 countries and including representatives from academia, hospitals, industry, and more.
- influenced policy 31 times, both nationally and internationally, including through membership of key advisory committees such as SAGE and SPI-M.
- reported 143 engagements including presentations, working groups, press releases etc., of which 44% were international.

Fellows’ publications are typically cited around five times more than others in same field. One paper on the Early dynamics of transmission and control of COVID-19: a mathematical modelling study has been cited over 1,100 times.

A digital learning platform for the health data science community

In May we launched HDR UK Futures – our new virtual learning platform which provides the core infrastructure for all of our training. It is a key enabler for our training portfolio allowing us to build training at scale and to structure, manage, track and deliver learning activities and content both online and face to face. It also supports our commitment to facilitate the community to collaborate regularly, to share information, improve their skills, and actively work on advancing health data science.

To-date 520 people have registered for the platform, which contains a library of carefully curated on-demand online resources that will be regularly refreshed and updated. There are 15 live modules with a further 100+ in production.

Internships

In addition to the [Black Internship Programme](#), HDR UK has also directly provided two year-long internships within the central HDR UK team, supporting the development of the Innovation Gateway, and providing interns with a multitude of new skills and practical experience which will position them well for a permanent job in the sector.

520

people registered for a
digital learning platform
for the health data science
community

Looking forward

Plans for next year include:

Research and development – we will commission and publish a white paper to research the health data science skills gaps and opportunities for the UK, aiming to provide a helpful resource for policy makers, as well as support the further development of our training programme.

Apprenticeships – a new partnership with Cambridge Spark, a sector leader in apprenticeships, will provide an 18-month AI Apprenticeship Academy for Health - offers organisations in the health sector (both NHS and industry) a work-based route for upskilling staff with a masters-equivalent qualification in AI. The pilot for this programme begins in September with a cohort of 20 apprentices and will act

as a test-case for widening involvement and collaboration with other apprenticeship providers at all levels from level 4 (foundation degree) to level 7 (Master's equivalent).

Placements – from September 2021, our PhD students will also be offered industry and NHS placements as part of their doctoral programme providing them with the opportunity to expand and diversify their skillset, establish new networks, and gain valuable experience of work outside academia. We have also been awarded a grant to provide 30 Wellcome Biomedical Vacation Scholarships from 2022-2026. With a particular emphasis on attracting students from diverse backgrounds and non-Russell Group universities, these 6-week internships will be provided in partnership with the University of Birmingham.

Collaboration – we will continue to build and cement relationships with partner organisations across the NHS, industry and academia, such as Health Education England, NHS Digital Academy and the Association of Pharmaceutical Analysts, to support our ambition to apply skills across a vibrant landscape.

Case study

A FELLOWS CASE STUDY – KRISHNARAJAH NIRANTHARAKUMAR



Krish Nirantharakumar

An HDR UK Fellowship helped unlock over £5 million in research funding.

A three-year HDR UK fellowship allowed Krish Nirantharakumar, a theme lead for health informatics at the Institute for Health Informatics at the University of Birmingham, to make vital connections with researchers and organisations across the UK and beyond – including through the HDR UK summer school.

It also provided him time to pursue his own research, including adding additional features to a tool called DExtER (data extraction for epidemiological research), applying AI to the understanding and management of multimorbidity (for which he has received a grant of £2.5m), and advancing the work of MuMPreDiCT – a collaboration of eight universities, across the four home nations, to improve care for pregnant women with two or more health conditions (for which he received a grant of £2.9m).

UNITING HEALTH DATA



We are uniting the UK's health data to enable discoveries that improve people's lives.

Uniting health data

UK Health Data Research Alliance



Collaboration in practice

To help address important health challenges faced in the UK through trustworthy research and innovation, HDR UK convenes the UK Health Data Research Alliance (the “Alliance”) to establish best practice and accelerate responsible access and use of health data.

In a year that has united science to deliver some of the greatest collaborative research and discoveries of our lifetime, the Alliance has been a driving force for this approach, bringing together multiple health and research organisations to work in partnership to create a national, federated and co-ordinated approach to health data research infrastructure.

This year has seen continued progress across key workstreams:

- Promoting participation across the sector; encouraging research organisations and custodians to join and build the partnership

- Aligning data standards and quality across the UK, including metadata standards
- Developing a common approach to Trusted Research Environments (TREs) to enhance the researcher experience and provide confidence to data custodians
- Engaging practitioners, patients, and the public (including improving the “plain English” explanations of our work)
- Supporting the development of the Gateway as the platform for researchers to discover and request access to health datasets

Making the case for Trusted Research Environments (TREs)

The UK Health Data Research Alliance (the ‘Alliance’) is committed to an approach to data access based primarily around Trusted Research Environments (TREs); with appropriate robust and independent accreditation, monitoring and auditing.

“Working with HDR UK will allow us to make more viral genomic data available to researchers, so that maximum benefit to public health can be derived from this important resource.”

Sharon Peacock, Executive Director and Chair at COG-UK



Sharon Peacock

In July 2020, the Alliance published a comprehensive Green Paper, making the case to use on TREs which, by design, protect the privacy of the individuals’ data they hold, while facilitating the large-scale data analysis which will lead to discoveries and insights that improve healthcare.

Adopting an approach using a network of TREs across the UK would be a way of addressing public concerns and enhancing public confidence in the use of health data for research. The Alliance will continue to make the case for TREs as a vital component of the ecosystem, working on common standards and processes for them to be most effective.



[Find out why the COVID-19 Genomics UK Consortium \(COG-UK\) joined the UK Health Data Research Alliance](#)

Streamlining data access

The COVID-19 pandemic highlighted the existing need to accelerate access to health data to ensure better and faster research. One solution advanced by the Alliance in the last 12 months has been to work together to standardise and align data access processes across organisations.

In a pilot scheme launched this year, streamlining data access across some data custodian organisations and using a standard data access request form has already made it easier for both researchers and custodians to submit and respond to access requests. Working towards a fully digital, automated end-to-end system has also made the process more efficient for data custodians and researchers.

The universal adoption of the Office for National Statistics’ Five Safe framework provides a gold standard guide for both researchers and custodians to meet the principles of transparency, safety and privacy throughout the data use cycle, from research concept inception to data analysis and insight dissemination. Commitment from all custodians to this common framework is crucial to design standard processes and a proportionate approach.

Building alignment across custodians is helping to enable easier and faster access to UK health and social care data for research in the long term. It also demonstrates the members’ commitment to open access to health data for use in research and ultimately leading to discoveries that will benefit patients.

Growing our Alliance

More than 50 of the UK’s NHS, charity and research organisations have joined the Alliance since its formation in February 2019.

During the year, the BHF Data Science Centre, Imperial College Neonatal Data Analysis Unit, Oxford University Hospitals NHS Foundation Trust, COG-UK and Pathlake, University of Dundee’s Health Informatics Centre, ICNARC, Imperial College Healthcare NHS Trust, HIC Dundee, Leeds Teaching Hospitals NHS Trust, QResearch, Oxford Health, The Renal Association, National Consortium of Intelligent Medical Imaging, NIHR Clinical Research Network, Generation Scotland, Public Health Scotland, University Hospitals Bristol and UK Brain Bank Network joined the Alliance.

Five safes



Looking ahead

Work to optimise data access processes is continuing and we are now prioritising provision of automatic built-in validation of processes where possible and integration of external systems with the Gateway, to enable seamless interaction with existing custodian systems.

Our long-term vision in this specific area is to enable researchers to submit data access requests to

multiple data custodians via the same form and allow virtual data access committees to review applications online.

We are developing public Data Use Registers based on standards created by the community and members of the Alliance, showcasing information of how data listed in the Gateway are used, thus ensuring transparency and increase understanding of the purposes for data use.

As Trusted Research Environments become an increasingly important part of our landscape, we will continue to work with partners to develop the common standards necessary for them to be a consistent and effective tool.



Find out why the COVID-19 Genomics UK Consortium (COG-UK) joined the UK Health Data Research Alliance.

Uniting health data

12 months of the Gateway – The UK’s portal for data discovery and tools for research



Despite progress in technology, the discovery of and safe access to health datasets remains a critical challenge for researchers.

In 2020, the Health Data Research Innovation Gateway (Gateway) was launched as a cornerstone of the UK Life Sciences Industrial Strategy. For the first time, it offers a single front door to discover UK health data, on a journey to streamlined and harmonised access management, across the four nations of the UK. This year, the Gateway has reached over 1,100 registered users, 16,000 monthly searches and information about over 640 UK and international datasets listed in its ‘library’.

This covers a breadth of data including primary care, secondary care, palliative care, biobanks and research cohorts. The Gateway is becoming the “go to” place for researchers to discover and request access to UK health datasets, with the added benefit of giving much-needed transparency to the UK public on what data is available and how it is used. The Gateway is at the heart of making open science a reality; so far, 150 tools, 269 projects, and 1146 papers, are available and these are growing daily.

The Gateway has been a key enabler in responding to COVID-19 by providing the UK research community with a single platform to access information about datasets relevant to the pandemic and a means to request access to these. During the year, access to key datasets was accelerated, including those relating to vaccine effectiveness and identifying potential risk factors associated with the vaccines. It also continued to support research on other priority areas, including cancer and heart disease.

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

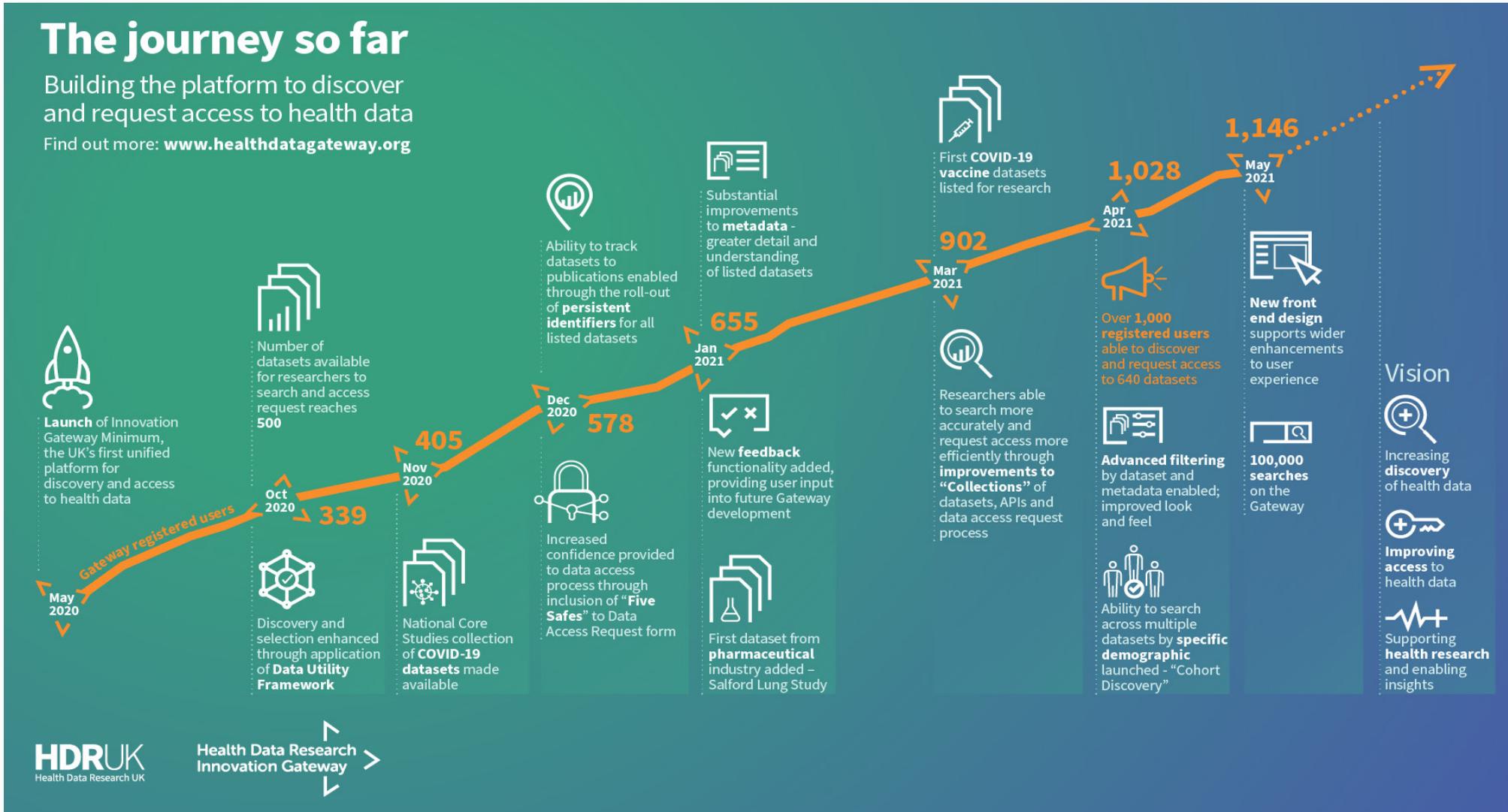
David Leather, Medical VP, GSK

640+

UK and international datasets reached this year by the Gateway



www.healthdatagateway.org/



Uniting health data

12 months of the Gateway – The UK’s portal for data discovery and tools for research *continued*



Highlights from this year include:

- Launch of a Cohort Discovery Tool with our partnership with CO-CONNECT to enable, for the first time, researchers to search across datasets to find cohorts of patients with specific, defined characteristics; opening up huge potential for increased discovery
- Integration of a Common Data Access Request form management process that is compliant with the Office for National Statistics’ Five Safes framework
- Rapid onboarding of information about over 70 datasets relevant for COVID-19 research, which are being safely accessed by the Government-led National Core Studies
- The Gateway became a key part of the fortnightly COVID-19 reports submitted by HDR UK to the government’s Scientific Advisory Group for Emergencies (“SAGE”).

HDR UK’s Public Advisory Board and other patient representatives are involved in the shaping and development of the Gateway. Their input has challenged developments that may cause concern to the public and has helped develop content about the Gateway that is accessible to the public. As with many other organisations, the pandemic also forced a transformation on our ways of working.

It has facilitated 265 requests to access health datasets, in particular actively supporting the government’s National Core Studies into COVID-19 (“Data and Connectivity”).

Uniting health data

BHF Data Science Centre



Launched in January 2020 as a partnership between HDR UK and the British Heart Foundation, the BHF Data Science Centre enables responsible, ethical research that combines the power of advanced analytic methods with the UK's large-scale and diverse cardiovascular data.

The BHF Data Science Centre works with patients, the public, clinicians, researchers and NHS organisations to help them carry out research into the causes, prevention and treatment of all diseases of the heart and circulation.

Creating the UK's largest linked health data research asset

A significant initiative this year has been the CVD-COVID-UK consortium, a NIHR-BHF flagship project which was set up to understand the relationship between COVID-19 and cardiovascular diseases.

Patients with cardiovascular disease are at increased risk of developing COVID-19 and of poor outcomes of COVID-19, such as admission to hospital or intensive care, or of dying. This could be due to cardiovascular conditions themselves, their risk factors, medications, or combinations of these.

At the outset of the pandemic, 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.

To solve this challenge, the BHF Data Science Centre 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 96% of the population of England (>54m people), with similar linked data made available for Scotland and Wales (>8m people). Through the CVD-COVID-UK consortium, over 200 researchers from 41 UK research organisations, are for the first time analysing linked health data on over 60 million people to address COVID-19 related research questions.

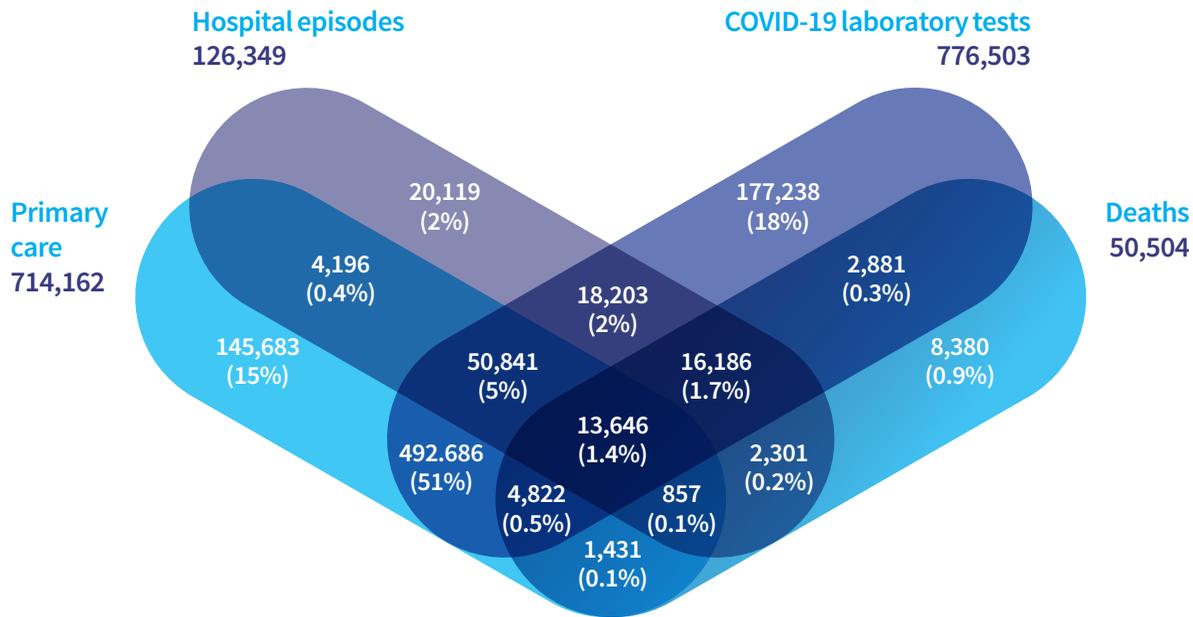
The ability to link different types of health data across the UK population provides a more complete and accurate picture of the impact of the virus on patients with diseases of the heart and circulation than has been possible before now. It also provides the data to understand whether patients with COVID-19 are more likely to go on to develop diseases of the heart and circulation, such as heart attack and stroke.

All protocols, code and analysis is available in the public domain via the BHF Data Science Centre webpages, the Gateway GitHub repository, and open access publications.



[Read more about this here](#)

Data sources reporting person-level data on confirmed or suspected COVID-19 diagnoses between January and October 2020 (n=959,470). Numbers indicate distinct people with a confirmed or suspected COVID-19 diagnosis.



Exploring ‘False Positive’ tests

Testing for COVID-19 infection in populations where active infection is very uncommon will result in some false positive test results. This is where people who do not have infection receive a positive test result, when it should be negative.

In August, Cathie Sudlow, Director of the BHF Data Science Centre, developed a simple interactive tool to help people to understand the problem of incorrect test results.



[View the interactive tool here](#)

“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



Cathie Sudlow

IMPROVING HEALTH DATA



Improving the quality of existing health data resources produces more insights, reduces the barriers to potentially life-saving research and supports innovation.

Improving health data

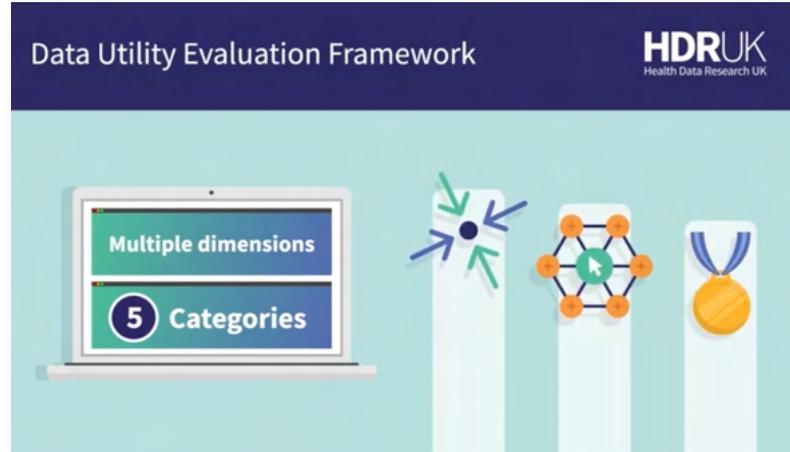
Providing researchers with the tools they need



As we work with our partners to unite the vast quantity of health datasets that exist across the UK and make them safely and responsibly available for health research, there becomes a growing need to be able to define, categorise and curate that data.

As the organisation responsible for uniting health data in the UK, we have developed a Data Utility Framework as a way to do this. This is the first time this has been done and represents a huge development in the way we think about datasets. It aims to help researchers from academia, healthcare and industry to make more sense of the increasing data resources available to them.

The new framework shows the usefulness of data for research. It has over 100 datasets evaluated against it and is now integrated into the HDR Innovation Gateway. To support users in identifying in advance whether a dataset would be suitable for their specific purposes, we developed the framework in partnership with users and data custodians.



[Watch our animation
“Transforming the usefulness of data”](#)



[Read the paper on the Data Utility Framework in the BMJ Journal of Health & Care Informatics](#)

This year, the framework was used to track the improvements in data of accessible through the Health Data Research Hubs. Throughout the pandemic, the Hubs utilised their rich data sources to support areas of analysis that included infection rates, vaccine roll-out, and to understand the impact of COVID-19 on specific disease areas. Collectively the Hubs drove significant improvements in data, guided by our data utility framework, particularly in the richness of metadata, pathway coverage, and data management processes.

The Hubs were tasked with improving the data that they hold to enhance its research potential. To review the data improvement over the past 18 months, we created the new and innovative data utility framework to gather baseline information at the Hubs’ first milestone (December 2019) and compared this with improvements made by the second milestone (March 2021). This was the first opportunity to use the framework not just for guiding users but for reviewing the improvements made to datasets.

Other organisations in the UK, such as NICE and NHSX, are looking to adopt the framework, and it is currently being tested internationally.

Improving health data

Building real-world partnerships through the Health Data Research Hubs



This year, the seven original Health Data Research Hubs reached their second milestone – at 18 months since inception – where they were recognised for the quality improvement of their datasets, the impact of their data, and engagement with patients and the public.

Since launching, the Hubs have made 157 datasets – from genomic data to clinical and public health data – discoverable on the Gateway. They have delivered over 300 projects involving hundreds of partners, over 20,000 meaningful patient and public interactions, and 2,300 training activities.

Our report *Improving UK Health Data: Impacts from the Health*

Data Research Hubs shows how the Hubs have informed UK policy decisions on the effectiveness of COVID-19 vaccines, created tools to improve clinical decision-making in the management of patients with vascular disease, and supported research in cancer, heart disease and hospital care pathways by linking routinely-collected data.

Highlights from the Hubs include:

- **NHS DigiTrials** enabled the RECOVERY trial – the world’s largest successful clinical trial on COVID-19 therapeutics to date. They delivered health outcomes datasets and smart trial design to determine the success of Dexamethasone in reducing mortality for patients with severe COVID-19
- **PIONEER** led rapid analysis of real time data to discover an increase in venous thromboembolic events (VTE) associated with COVID-19

– **BREATHE**, through the citizen science COVID-19 Symptom Study, in partnership with ZOE and Kings College London, urgently provisioned data on COVID-19 symptoms to a range of users, which fed into local and national responses to the pandemic and supports 54 live projects

– **Discover-NOW** linked COVID-19 testing data with all haematology, biochemistry, immunology, microbiology and therapeutic monitoring results for the 2.5m patient records available via their hub

– **DATA-CAN**’s analysis of real-time hospital cancer services was shared with the UK’s four chief medical officers, the National Clinical Director for Cancer for England and SAGE, contributing to decisions to restart cancer services

“The Hub’s role as a facilitator between the NHS, academia and industry partners to provide a level playing field and a safe environment to experiment is critical to the success of innovative and transformational work.”

Ross Stone, Global Programme Director, Healthcare, AstraZeneca

– **INSIGHT** provided the first reliable estimates of the scale and severity of the vision loss arising from delays in treatment for newly-diagnosed ‘wet AMD’ during the COVID-19 period, informing NHS providers on strategies to optimise care of patients during service recovery

– **Gut Reaction**’s key partner, the IBD Registry, developed a COVID-19 IBD Risk Tool in just eight days, to allow patients to self-assess their risk of being infected with COVID-19.



Dr Elizabeth Sapery
Director of PIONEER Hub and Reader in Acute and Respiratory Medicine at University of Birmingham



Read the summary or full report on the Impact of the Health Data Research Hubs



WATCH how the Hubs have supported the UK’s response to the COVID-19 pandemic: film

ONE INSTITUTE



We deliver our strategy through an inclusive, team-oriented ethos – bringing together NHS, universities, research institutes, industry and charities – built on our values of transparency, optimism, respect, courage and humility.



Tackling the underrepresentation of Black people in the science community



The Black community is heavily underrepresented in the science, technology, engineering and maths (STEM) community, with only 65 Black and 310 mixed and other individuals making up the total of 10,560 science professors in the UK.

The COVID-19 pandemic also highlighted the disparity of COVID-19 infections and its effects on different ethnic groups. In December 2020, we launched an internship programme to create better representation within the health data science communities by creating more opportunities for those groups to be involved in this vital work.

Through partnerships with the UK Health Data Research Alliance and the 10,000 Black Interns initiative – which is designed to transform the prospects of young Black people in the UK – we recruited 54 interns, including over 30 women, to the programme.

The programme is providing paid work experience to future data scientists, with the internships taking place across 25 of our partner organisations. These include the BHF Data Science Centre, Big Data Institute, National Institute for Health Research Clinical Research Network, prominent national charities, data science hubs housed at a number of higher education institutes, university hospitals and NHS trusts such as Guy’s and St Thomas’, The Royal Wolverhampton and Barts Health.

“We are really looking forward to welcoming the 54 interns this summer to the exciting field of health data science. It is wonderful that so many opportunities have been created by organisations from the UK Health Data Research Alliance. I hope that for many of these interns, this will be the start of a long and successful career in health data science.”

Caroline Cake, CEO at HDR UK



“I have a passion for science, technology and data; with this internship I hope to gain a real insight on what a data scientist does and their responsibilities. I want to see first-hand the importance of real-world data and how AI technology allows us to analyse it. I plan to use this experience to motivate me to achieve my goals which is to have a career within health data science and a direct impact within the healthcare sector.”

Maria Johnson, intern at HDR UK’s Hub for Respiratory Health BREATHE

Starting in Summer 2021, these six-week placements include activities such as conducting research and developing health data analysis, learning about the latest approaches and technologies in the field, collaborating with teams and communicating results and ideas to stakeholders.



HDR UK Black Internship Programme

One Institute

Our Board



Sarah Harper

Sarah Harper, CBE, is Professor of Gerontology at the University of Oxford, a Fellow at University College, and the Founding Director of the Oxford Institute of Population Ageing. Sarah served on the Prime Minister's Council for Science and Technology, which advises the Prime Minister on the scientific evidence for strategic policies and frameworks. In 2017, she served as the Director of the Royal Institution of Great Britain. Sarah is a Director and Trustee of the UK Research Integrity Office.



Alex Markham

Sir Alex Markham is Professor of Medicine and Director of the MRC Medical Bioinformatics Centre in Leeds. He has made various contributions to medical science, is a physician and pathologist, and trained initially in medicinal chemistry. His commercial experience includes cancer drug development and introduction of DNA Fingerprinting for medico-legal applications (Queen's Award for Technological Achievement in 1990).



Lucy Neville-Rolfe

Baroness (Lucy) Neville-Rolfe DBE CMG was appointed a Life Peer in 2013 and served as a minister in the Business and Culture departments (2014-16) and as Commercial Secretary to the Treasury from December 2016 to June 2017. She is Chairman of Assured Food Standards and a non-executive director of Capita plc. Lucy worked at Tesco PLC (1997–2013), serving on the plc Board from 2006.



James O'Shaughnessy

A Member of the House of Lords since 2015, James most recently served as Parliamentary Under Secretary for Health. As Minister he led efforts to improve innovation within the NHS and to grow the UK's life science & healthtech industries. His responsibilities included the Life Science Industrial Strategy, delivering a new pricing scheme with the pharmaceutical industry, chairing the National Genomics Board, and the digital transformation of the NHS. James was also responsible for preparing the NHS and wider health and care sector for Brexit.



Jim Smith

Professor Sir Jim Smith FMedSci FRS is Director of Science at Wellcome and Senior Group Leader at the Francis Crick Institute. A globally recognised developmental biologist, Professor Smith's research interests include inductive interactions in vertebrate development and applying the principles of developmental biology to stem cell differentiation. His previous roles include Director of the Wellcome Trust/Cancer Research UK Gurdon Institute and Director of the MRC National Institute for Medical Research (NIMR), where he oversaw its transition to the Francis Crick Institute. More recently, he has been Deputy CEO and Chief of Strategy at the Medical Research Council, and Director of Research and Group Leader at the Francis Crick Institute.



Graham Spittle

Dr Graham Spittle, CBE, was appointed as IBM's Chief Technology Officer and Vice President, Software Group in January 2010, having joined IBM in 1985. Graham's other posts have included Member of Prime Minister's Council for Science and Technology, Chair of the Technician Council and Chair of the UK Technology Strategy Board.



Janet Thornton

Professor Dame Janet Thornton is a Senior Research Scientist at the European Molecular Biology Laboratory – European Bioinformatics Institute on the Wellcome Trust Genome Campus at Hinxton, near Cambridge, UK. She was Director of the Institute from 2001-2015. EMBL-EBI provides core bioinformatics data and services and performs bioinformatics research. Janet’s research is focused on proteins, especially their structure, function and evolution. She is a computational biologist, working at the interface of biology with physics, chemistry, and computing. She has published more than 500 scientific papers, was elected to the Royal Society in 1999 and is a Foreign Associate of US National Academy of Sciences.



David Zahn

David Zahn is Head of European Fixed Income at Franklin Templeton Investments, a global investment management firm, and has worked in asset management for over 20 years. He undertakes voluntary work in the healthcare and education sectors and participates on several charitable boards including The Health Foundation where he is Chair of the Investment Committee. He is a Trustee of the Medical Research Foundation.

David has had extensive experience in volunteering through the CFA Institute, a global not-for-profit association of investment professionals that awards the Certified Financial Analyst (CFA) designation.



Bill Boa

Bill Boa is the independent member of the HDR UK Audit and Risk Committee. Bill is a Fellow of the Institute of Chartered Accountants in England and Wales (ICAEW) and an Associate of the Chartered Institute of Public and Finance Accountants. Bill established and is principal of his own practice regulated by the ICAEW. Bill worked for the NHS for over 20 years and was Chief Finance Officer of Barts Health Group, Cambridge University Hospital NHS Trust, and Great Ormond Street Hospital for sick children NHS Foundation Trust amongst others.

We would like to thank Jonathan Montgomery and Frances Rawle for their significant contribution to HDR UK as Board members. Jonathan stepped down in March 2021 and Frances stepped down in April 2021.

One Institute

Financial summary



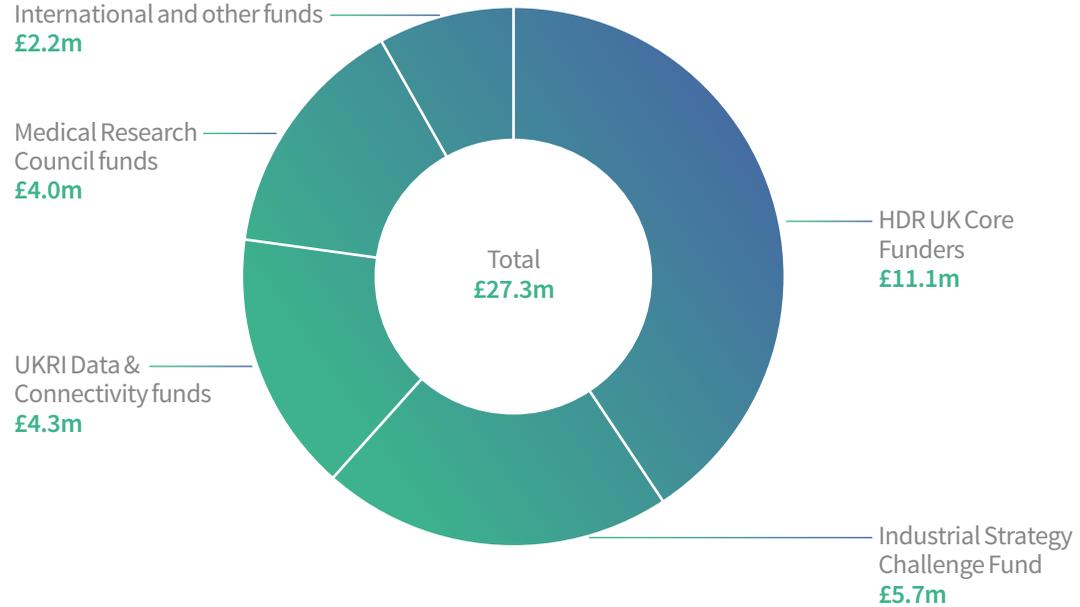
Our funds support long-term scientific and research studies, training and infrastructure that contribute to data science at scale, support our One Institute approach and deliver long-term impact for the health of patients and populations across the UK.

HDR UK's activity across our strategic priority areas has continued to generate significant interest from funders. We also raised thematic funding to support specific programmes of work, including our work to support the research response to COVID-19.

Funds used for expenditure in 20/21 were awarded by HDR UK Core Funders, Industrial strategy challenge, UKRI Data and Connectivity, Medical Research Council, International funders and others. We also received complementary, matched, leveraged and in-kind funding which contributed to HDR UK's outcomes during 2020/21.

In the final part of the year, we concluded the work for Health Data Research UK to become a registered charity (no. 1194431), which came into effect on 12 May 2021.

Source of funds to support 2020/21 expenditure



Health Data Research UK is the national institute for health data science and registered charity (no. 1194431), funded by UK Research and Innovation, the Department of Health and Social Care in England and equivalents in Northern Ireland, Wales and Scotland, and leading medical research charities.

www.hdruk.ac.uk

Funded by



www.hdruk.ac.uk/about-us/funders/