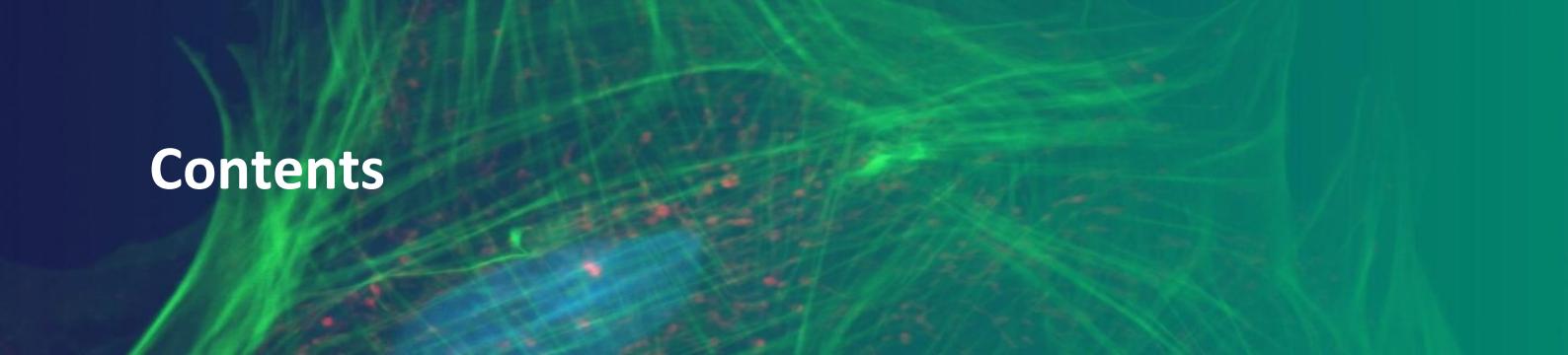


Health Data Science in the COVID-19 Era

Health Data Research UK 'One Institute' Conference 2020

16 June 2020

Contents



Executive summary.....	1
Introduction.....	7
Session takeaways	8
How has health data research in the UK pivoted for COVID-19?.....	8
International approaches to combating the COVID-19 pandemic – how have data and technology been harnessed?	8
Breakout 1: How do we make the UK world-leading for COVID-19 clinical trials?	9
Breakout 2: COVID-19 Data – data definitions and new data types in a crisis.....	11
Breakout 3: Understanding the causes of disease and COVID 19 – connecting molecular science with health records.....	12
Breakout 4: How are Health Data Research Hubs and BHF Data Science Centre supporting COVID-19 Research?	13
Breakout 5: Better Care – role of data at the frontline of the COVID-19 response	14
Breakout 6: Responding to the Public Health Emergency of COVID-19.....	15
Breakout 7: What do national and international health data researchers need in a post-COVID world? .	17
Breakout 8: Relationship of trust: Realising patient and NHS benefits in times of COVID-19	18
Rapid fire early career lightning talks.....	20
Fireside chat: How will data and technology influence the future of clinical and regulatory science?.....	23
Next steps	24
About Health Data Research UK & further information.....	24

Executive summary

"The COVID-19 pandemic represents an unparalleled global emergency. It has demonstrated the absolute necessity for open research and safe data access. The Health Data Research UK conference provided an opportunity to build on this need by sharing new research findings and exchanging international experiences related to the COVID-19 pandemic. Importantly the event highlighted the lessons we have learnt and the key priorities as we move forwards."

Andrew Morris, Director of Health Data Research UK

In December 2019 COVID-19 emerged; relatively little was known about the virus and no one would have predicted the impact it would have on all our lives. Since then it has wreaked havoc across the world with more than 10 million cases across the world and over 42,000 deaths in the UK¹.

Health Data Research UK (HDR UK) and the broader health data community have been at the forefront of the response to the pandemic, providing critical insights to support the fight against the disease. The HDR UK 'One Institute' annual conference showcased the key contribution these efforts have made in informing our response and the important part the work continues to play as we move forwards towards a post-COVID-19 era. Presented in a virtual format, the conference was attended by scientists, researchers, technologists, clinicians, patients and the public from across the UK and internationally.

The event highlighted the breadth, depth, power and expertise of health data research and its ability to answer key gaps in our knowledge, inform our approach to effectively control, treat and prevent the virus, and our strategy to understand and mitigate its long-term impact on society and health.

Patrick Vallance, Chief Scientific Advisor and Co-Chair, Scientific Advisory Committee for Emergencies (SAGE) opened the event



Participants were welcomed with an address from the government's Chief Scientific Advisor, Patrick Vallance, who highlighted the unique scale of the COVID-19 challenge. This was followed by a plenary session led by Andrew Morris (Director, HDR UK) and Caroline Cake (Chief Executive Officer, HDR UK) who examined how HDR UK and the wider health data community have responded to the "call to action" we have all faced.

Attendees then heard from international colleagues in Germany (Andreas Poensgen, Managing Partner, Turgot Venture), Switzerland (Effy Vayena, Chair of Bioethics, Swiss Institute of Technology) and Singapore

¹ Figures correct at the time of writing, 29/06/2020. Number of deaths reflects cases where COVID-19 is listed as cause of death.

(Teo Yik Ying, Dean of Saw Swee Hock School of Public Health, National University of Singapore). The speakers provided key insights into the different strategies used by countries across the world, highlighting the lessons we can learn from others and the potential to translate approaches across country and cultural borders.

<i>Andreas Poensgen, Managing Partner, Turgot Venture</i>	<i>Effy Vayena, Chair of Bioethics, Swiss Institute of Technology</i>	<i>Teo Yik Ying, Dean of Saw Swee Hock School of Public Health, National University of Singapore</i>
		

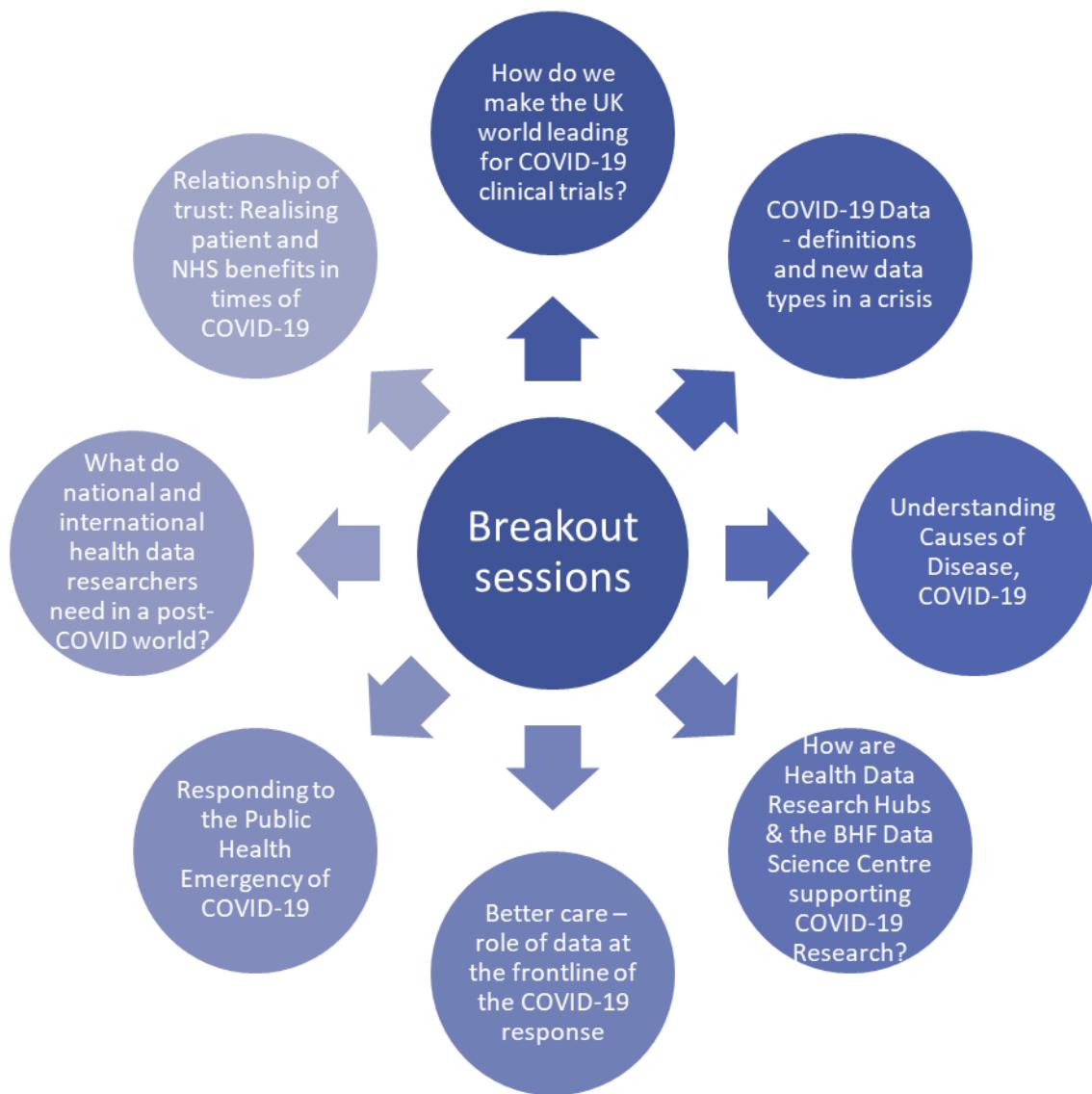
Participants were welcome to attend breakout session, exploring key topics in greater depth. The scope of the sessions reflected the diverse interests, roles and expertise of attendees and included:

- Progress and insights offered by COVID-19 clinical trials and their ability to adapt and deliver at unprecedented pace and scale during the pandemic
- Opportunities provided by novel data sources including the new era of app-based data collection
- Underlying mechanisms and genomics of COVID-19
- The role being played by the HDR UK Hubs in supporting COVID-19 research
- The key contribution health data is making to support the front-line clinical response to the pandemic
- Lessons learnt from the data-driven public health response across the devolved nations; the demand and opportunities to connect national and international datasets as we move forwards in a post-COVID era
- Building a relationship of trust and transparency to realise the benefits of health data research for patients, the public and healthcare system

A highlight of the day featured a session of lightning talks from early career researchers, providing reassurance that the future of health data research is in good hands. The seven rapid-fire talks covered a range of topics from the impact of COVID-19 on cancer, through to investigating the role of vitamin D in COVID-19, using smartphone data to monitor the COVID response and the contribution of socioeconomic and ethnic inequalities to risk of COVID infection.

The day closed with a fireside chat between HDR UK Director Andrew Morris and Amy Abernethy, Principal Deputy Commissioner of Food and Drugs for the US FDA, reflecting on the role of health data in responding to the pandemic, and how the monitoring real world outcomes can influence drug regulation and pricing.

Breakout sessions



The event saw the presentation of the HDR UK annual awards for Team of the Year (HDR UK Sprint Exemplar Project in Rare Diseases, led by John Bradley, COVID-19 themed winning team - HDR UK London COVID-19 Response Team, led by Amitava Banerjee); Impact of the Year (Model to estimate excess deaths from COVID-19, led by Amitava Banerjee) and Most Thought-Provoking Lightning Talk (Real-time hospital cancer data, excess deaths and a question for the Prime Minister – Alvina Lai, UCL).

Despite a broad range of diverse sessions several key themes emerged over the day:

1. Delivery at pace without sacrificing quality:

Speakers repeatedly demonstrated the phenomenal speed of the response to a virus which does not wait for research or policy timelines. Decisions that would normally take months or years have been happening in a matter of hours and have allowed us to seize important opportunities to generate valuable insights and drive progress.

Coupled with this, the pandemic has highlighted the importance of timely information-gathering and analysis – we need data in real-time when we are trying to understand and respond to urgent issues on the ground - today.

However, data must also be accurate and reliable. It is essential not to sacrifice quality in the drive for speed and risk releasing insights which may be mis-leading or inaccurate. The old computing saying “garbage in, garbage out” still holds true today, so we must make sure that we are paying attention to the quality of data, even while we are working at speed and at scale.

2. Collaboration and a willingness to learn from others is key:

As well as highlighting the incredible speed at which the UK data science community has responded to the pandemic, presentations revealed the impressive collaborative nature of the response, bringing together teams across all four nations of the United Kingdom and internationally at a scale that has not been seen before. This willingness to share expertise and insights has proved essential at a regional, national and international level, to ensure we rapidly adopt best practice, avoid duplication and leverage the combined power of data and resources.

Building on this need for collaboration, the conference also saw the announcement of a new International COVID-19 Data Alliance led by HDR UK and funded by the COVID-19 Therapeutics Accelerator. The International Alliance brings together organisations from across the world to create a trusted research environment (Workbench) which will use data at scale to accelerate insights into COVID-19 and speed up the development of treatments.

“We’re starting small, thinking big and moving fast in what we think will be an enduring global endeavour to ensure the international sharing of data in an ethical and trustworthy way.”

Andrew Morris,
Director, Health Data Research UK

“After this hammer, society has to learn to start dancing again – this will be a trial-and-error process.”

Andreas Poensgen,
Managing Partner, Turgot Venture

“The overarching message...was that there is an opportunity for the Covid-19 pandemic to act as a catalyst for the health data science community to establish and further improve adherence to common data standards, shared ontologies and

data sharing infrastructure. Being able to combine related data from different sources is key to addressing data gaps."

Julia Scott,
HDR UK COVID-19 Patient Advisory Group
member

3. The importance of common standards:

COVID has highlighted the need for shared 'languages' and 'dictionaries' for connecting multiple datasets to enable richer insights to be drawn. Speakers highlighted, if we use COVID-19 to establish firm foundations, we can continue to build on these principles, and accelerate our progress towards ensuring interoperability and integration in the future.

4. The need to enable data access and integration:

Critical to our response to COVID-19 is the requirement for pragmatic, secure and efficient resources for shared data access, such as Trusted Research Environments, where researchers can securely access data in a way that ensures privacy and governance requirements are met. Along with this we need the investment of time and skills to develop and connect these TREs safely together in the future, finding ways to enable safe access to integrated datasets to support the health of everyone, everywhere.

"This demonstrated to me just how much work there is to do so health data can be freely and safely exchanged for research and crisis preparedness. The challenges are huge, but not insurmountable."

Colin Wilkinson,
HDR UK Patient Advisory Group
member

5. The role of novel data collection:

In addition to using data to enable health service leaders and policymakers to understand and respond to a fast-moving situation, the effective use of health datasets and app-based technology has enabled the rapid roll out of world class clinical trials and generated significant new insights into the symptoms and progression of COVID-19. How these approaches are embedded and integrated into the health and care system in the longer term will need to be explored but COVID-19 has proved how effectively this can be done.

6. Transparency and trust:

These themes will shape health data research for the future but underlying them all there must be transparency and clear communication to build a relationship of trust. It is essential that the public, patients and practitioners are involved and engaged throughout our work and are clear what information is being used, who has access and how it is helping to shape decisions.

The pandemic has created a cross-disciplinary challenge, requiring, amongst others, epidemiologists, immunologists, social scientists, statisticians, and policy makers to work closely together. The UK's scientific response has been exceptional, with rapid data flows quickly established from health services and other sources, but there still outstanding questions to answer: what protection is provided by antibody-based

immunity against coronavirus? How can we continue to roll-out of rigorous therapeutic and vaccine trials? What role can novel data sources can play as we move forwards?

We must leverage the UK's data science community not just to address the immediate issues around coronavirus but to uncover the wider indirect effects of the pandemic, to identify and mitigate the long-term health impacts of COVID-19, and to support the most vulnerable groups in society that will be hardest hit in the months and years to come.

The past six months have not been easy, but the community building around Health Data Research UK shows that we are moving towards a future where we can realise the benefits of data science to improve health for all. The conference highlighted that there is still a long road ahead, but that we should be optimistic given the great efforts, expertise and skills of the community.



Introduction

Health Data Science in the COVID-19 Era, the annual Health Data Research UK (HDR UK) ‘One Institute’ conference 2020, was held on 16 June. Hosted by HDR UK, this virtual event was designed to showcase the vital impact of health data in combating the global coronavirus pandemic. The packed day of talks and panels from world leaders in the health data field provided an interactive forum to share experiences, drive progress and consider future priorities. The wide-ranging sessions revealed the breadth, depth and speed of the UK health data community’s response to the pandemic, demonstrating a commitment to open science and highlighting how research efforts have pivoted to accelerate insights into all aspects of the pandemic, from policy responses to molecular mechanisms.

This report constitutes a summary of key takeaways from each session. For a full record of the event, please refer to the conference slides video recordings that are available on [HDR UK’s conference webpage](#).

Session takeaways

How has health data research in the UK pivoted for COVID-19?

Speakers:

- *Andrew Morris, Director, Health Data Research UK*
- *Caroline Cake, Chief Executive, Health Data Research UK*

Takeaways:

- The COVID-19 pandemic has triggered an unprecedented response from the health data science community.
- HDR UK has played a vital role co-ordinating the UK's data science response to COVID-19, curating research questions and delivering research insights.
- The COVID-19 response has been underpinned by patient and public engagement and involvement (PPEI), and bound together by HDR UK's values of transparency, optimism, respect, courage and humility.
- A new international collaborative data alliance, convened by HDR UK and funded by the COVID-19 Therapeutics Accelerator, will bring together international organisations and create a new trusted research environment (Workbench) to accelerate insights into COVID-19 and speed up the development of treatments using data at scale.
- Linking datasets remains a challenge, along with larger-scale digital infrastructure
- Health data research will play a vital role in identifying and mitigating the long-term health impacts of COVID-19 and supporting the most vulnerable groups in society that will be hardest hit in the months and years to come.

International approaches to combating the COVID-19 pandemic – how have data and technology been harnessed?

Chair: Rhos Walker, Chief Science Strategy Officer, Health Data Research UK

Panellists:

- *Germany: Andreas Poensgen, Managing Partner, Turgot Venture*
- *Switzerland: Effy Vayena, Chair of Bioethics, Swiss Institute of Technology (ETHZ)*
- *Singapore: Teo Yik Ying, Dean of Saw Swee Hock School of Public Health, National University of Singapore*

Session objectives:

- Understand how Germany, Switzerland and Singapore have responded to COVID-19.
- Understand the role of data and technology in the international response to the crisis.
- Discuss the challenges and opportunities for building public trust in the use of health data for research during a pandemic crisis.

Takeaways:

- Quarantine, contact tracing and social distancing have been used internationally and have proved critical to minimise the impact and spread of coronavirus. This has required combining novel digital approaches, such as contact tracing apps, with more ‘traditional’ human efforts to track and trace infections.
- The German response to COVID-19 had been driven from a regional level, with decentralised decision-making, trusting and celebrating local expertise and solutions - for example, supporting the use of independent testing laboratories rather than focusing on a centralised testing response)
- Switzerland has been able to adapt at speed, setting up scientific oversight and governance groups which enabled a response to the pandemic whilst building and maintaining public trust.
- Singapore has mounted an impressive test, trace and isolate strategy in response to the pandemic, with 40,000 tests per day anticipated by July for a population of around 6 million. However, the country has implemented other, more stringent measures, whose success has been dependent on the national culture. The wider social and behavioural “norms” is a key factor to consider.
- The biggest challenge for all countries is how to safely release lockdown measures and return life to a more sustainable normality.

Breakout 1: How do we make the UK world-leading for COVID-19 clinical trials?

Chairs:

- *Martin Landray, Professor of Medicine and Epidemiology, University of Oxford & Research Director, HDR UK*
- *Richard Torbett, Chief Executive, ABPI*

Panellists:

- *Janet Valentine, Director, Clinical Practice Research Datalink (CPRD)*
- *Mark Toms, Chief Scientific Officer UK, Novartis Pharmaceuticals UK Ltd*
- *Janet Frost, Public & Patient Representative*

Session objectives:

- Share insights from the UK’s leading clinical trials that have delivered at unprecedented pace and scale during the COVID-19 crisis.
- Consider the lessons learnt from trial delivery during and following the COVID- 19 pandemic – how do we retain momentum for the longer-term?
- Explore the needs and opportunities for patients, clinicians, industry and academia - how can we work together to advance UK trial delivery for patient and public benefit?

Takeaways:

- Delivering high quality, robust clinical trials at speed during a pandemic has been extremely challenging.
- The RECOVERY trial selected treatments to test based on pragmatic and scientific rationale, minimised the data-gathering workload for clinicians, and ‘filled in the gaps’ using existing health data records.

- Conventional clinical trial recruitment relies on advertising methods – ‘pre-screening’ patients using anonymised data from the Clinical Practice Research Datalink (CPRD) can significantly speed up the process and make trials more efficient.
- Interoperability is key to future clinical trial success: currently, there are disparate datasets that cannot be interlinked, so data is effectively siloed. Telecommunications and banking have solved this problem, so why can't we solve it for health data? Getting this right will make the UK an attractive location for doing clinical research which benefits our population as a whole.
- While the pace and scale of the COVID-19 response has been impressive, how much of the rapid response would be acceptable and appropriate in a regular setting outside the pandemic? How can we keep the efficiencies and progress without compromising patient safety and data security?

RECOVERY Preliminary result - Hydroxychloroquine

Randomised Evaluation of COVID-19 Therapy

Based on continued review of scientific data, FDA has determined that chloroquine and hydroxychloroquine are unlikely to be effective in treating #COVID19 and therefore we are revoking the emergency use authorization for these drugs: go.usa.gov/xwU2c

5:50 PM · Jun 15, 2020 · Twitter Web App

704 Retweets 277 Likes

The drugs on trial that may be the key to defeating the coronavirus

TOM CLARKE SCIENCE EDITOR

Video report by ITV News Science Editor Tom Clarke

It takes time to develop totally new drugs to treat totally new diseases. But people in intensive care with Covid-19 don't have time.

It's why in hospitals around the world doctors have been experimenting with existing medicines that might have an effect on Covid-19.

Your condition alters... without warning! My Covid-19 experience

Pressure on Government to deliver on ambitious five-point plan against coronavirus

But to get an accurate picture of which ones might offer the most benefit to the most people, the medical community needs to work together.

And that's exactly where the Randomised Evaluation of COVID-19 thERapY (RECOVERY) trial comes in.

How many coronavirus trials are there and how do they work?

Support The Guardian

Hydroxychloroquine does not cure Covid-19, say drug trial chiefs

Major study of thousands of patients led by University of Oxford shows drug is ineffective

Coronavirus outbreak

Sarah Boseley Health editor

Hydroxychloroquine

If you are admitted to hospital, don't take hydroxychloroquine, said Martin Landray, a key chief investigator of the Recovery trial and professor of medicine and epidemiology at Oxford University. "It doesn't work."

Many countries have permitted emergency use of the drug for Covid-19 patients, including the US, UK, France and Australia. In the US, the Food and Drug Administration panel in France, that it was a cure, and the ensuing claims from the public, President Donald Trump took the drug, saying it should be given to patients, and later he was personally taking it to protect himself from the virus.

Q&A

What is hydroxychloroquine?

most viewed

Plans to reopen English schools

Breakout 2: COVID-19 Data – data definitions and new data types in a crisis

Chairs:

- *Richard Dobson, Professor of Biomedical and Health Informatics, Institute of Health Informatics, UCL*
- *Helen Parkinson, Head of Molecular Archival Resources, EMBL-EBI*

Panellists:

- *Paul Clarke, Chief Technology Officer, Ocado*
- *Tim Spector, Professor of Genetic Epidemiology and Director, TwinsUK Registry, Kings College London*
- *Indra Joshi, Director of AI, NHSX*

Objectives:

- *Highlight the false dichotomy of speed and accuracy given the rapid pace of research during COVID-19*
- *Emphasise the critical need for data sharing and a highlight the lack of coordinated metadata standards and context*
- *Celebrate the successes of and the challenge of providing new datatypes for COVID-19 research - real-time symptom tracker, testing data*

Takeaways:

- Healthcare is all too often seen as ‘sick care’, with data only gathered when people are ill rather than well.
- App-based health data, such as the information gathered by the King’s College/ZOE COVID Symptom Study, can make significant contributions to public health and scientific research – the ZOE app data is currently the most requested dataset through the HDR UK BREATHE data hub.
- Interoperability and infrastructure again pose key challenges: we must invest in digital infrastructure, developing shared data standards, ‘language’ and ‘dictionaries’ (ontologies) to create a shared digital commons, meaning that disparate datasets can be linked together for the benefit of society.
- Interconnected health data could be used to create simulation, emulations, visualisations and ‘digital twins’, building *in silico* populations, environments or even humans where we could explore the impact of interventions like policy measures or therapies in order to predict the likely outcomes in the real world.
- It is likely that the health impacts of the diseases will go on much longer, so how do we manage data into the future in order to protect privacy while not losing the valuable progress that has been made in using health data for research?
- As we move forward we need to be mindful of people who are excluded due to lack of access to digital tools such as apps, and the need for data standards in order to effectively combine and link datasets for research.

Breakout 3: Understanding the causes of disease and COVID 19 – connecting molecular science with health records

Chairs:

- *John Danesh, Professor of Epidemiology and Medicine and Head of Department, Public Health and Primary Care, University of Cambridge & Research Director, HDR UK*
- *Adam Butterworth, Reader in Molecular Epidemiology, University of Cambridge*

Panellists:

- *Ewan Harrison, UKRI Innovation Fellow, Wellcome Trust Sanger Institute & COG-UK*
- *Mark Caulfield, Chief Scientist, Genomics England*
- *Aroon Hingorani, Chair of Genetic Epidemiology, UCL*
- *Emily Jefferson, Director, Health Informatics Centre, University of Dundee*

Session objectives:

- Explore viral genome tracking during COVID-19 and using human genomics in relation to the viral tracking effort.
- Examine use of consented multi-omic cohorts to study risk factors for the severity of COVID-19 disease.
- Discuss how the COG-UK effort can be used as an example for future pandemic preparedness.

Takeaways:

- Large-scale sequencing and linking of viral genomic data and host (human) genomes is providing important insights into how the coronavirus is transmitted; finding where outbreaks come from and how they spread; identifying potential vaccine antigens and therapeutic targets; discovering who is most at risk and why some people get more severely ill than others; and understanding how the virus might be evolving over time.
- Linking rich data from public sources to genomic data will provide new insights into this disease to improve care and save lives.
- Existing cohorts are a valuable source of multi-omic data to understand how COVID-19 affects individuals, and a way of reducing some of the methodological biases that can arise in patient studies.
- Trusted Research Environments (TREs, ‘safe havens’) need to be enhanced to support multi-omic datasets and analysis, including importing and exporting tools such as AI algorithms without compromising data security.

Breakout 4: How are Health Data Research Hubs and BHF Data Science Centre supporting COVID-19 Research?

Chairs:

- *Cathie Sudlow, Director, BHF Data Science Centre*
- *Sarah Brooke, Deputy Chair, Public Advisory Board, Health Data Research UK*

Panellists:

- *Amitava Banerjee, Associate Professor in Clinical Data Science & Honorary Consultant Cardiologist, Institute of Health Informatics, UCL*
- *Geoff Hall, Clinical Lead & Deputy Director, DATA-CAN – The Health Data Hub for Cancer - COVID and cancer*
- *Jenni Quint, Chief Clinical Officer & Deputy Director, BREATHE – The Health Data Research Hub for Respiratory Health*

Session objectives:

- Highlight the convening power of the Health Data Research Hubs and BHF Data Science Centre in coordinating UK-wide efforts to provide clinical and policy relevant insights on COVID-19.
- Consider additional at-scale data needs for further insights.
- Consider the trade-off between speed, robustness and quality of findings, including how to engage with the public and patients when acting at speed.

Takeaways:

- The HDR UK BHF Data Science Centre has moved at pace in response to COVID-19, working with data custodians across the four nations and NHS digital to bring together generic and specialist datasets into the first Trusted Research Environment (TRE)² dedicated to cardiovascular research.
- Health data research has revealed the direct and indirect impacts of COVID-19 on patients with pre-existing conditions such as cancer and cardiovascular disease, potentially causing thousands of additional deaths.
- Between 6,000 to 17,000 indirect excess deaths may result from delayed cancer diagnosis or treatment during the pandemic.
- Around 20% of the English population fall into the high-risk category for COVID-19 with the baseline risk varying by the underlying condition and any comorbidities as well as age and sex. Ongoing work is key to identify whether there is any influence of social factors such as age, sex, ethnicity, or socioeconomic status to discover how this phenomenon is impacting equally across society.

"Very interesting selection of speakers, who raised interesting points about the conversations we need to be having with all stakeholders whenever the opportunity arises, and also the importance of robust local systems for responding to emergency health situations."

**Rachel Plachcinski,
Patient Advisory Group member**

² Trusted Research Environments (TREs) TREs provide a secure analytics environment (i.e. a safe setting) where researchers can bring analysis algorithms to the data.

- A personalised online risk predictor proved very popular with the public, gaining about 350,000 users within 24 hours and 1.3 million users within a month, showing that there is an appetite for the public to interact with health data and risk.

Breakout 5: Better Care – role of data at the frontline of the COVID-19 response

Chairs:

- *Simon Ball, Executive Medical Director, University Hospitals Birmingham, NHS Trust & Research Director, HDR UK*
- *Alastair Denniston, Director, INSIGHT – The Health Data Research Hub for Eye Health*

Panellists:

- *Liz Sapey, Director, PIONEER – The Health Data Research Hub for Acute Care & Managing Director of NIHR Clinical Research Facility Birmingham*
- *Ben Goldacre, Director, The DataLab*
- *Sue Mason, Professor of Emergency Medicine, University of Sheffield*

Session objectives:

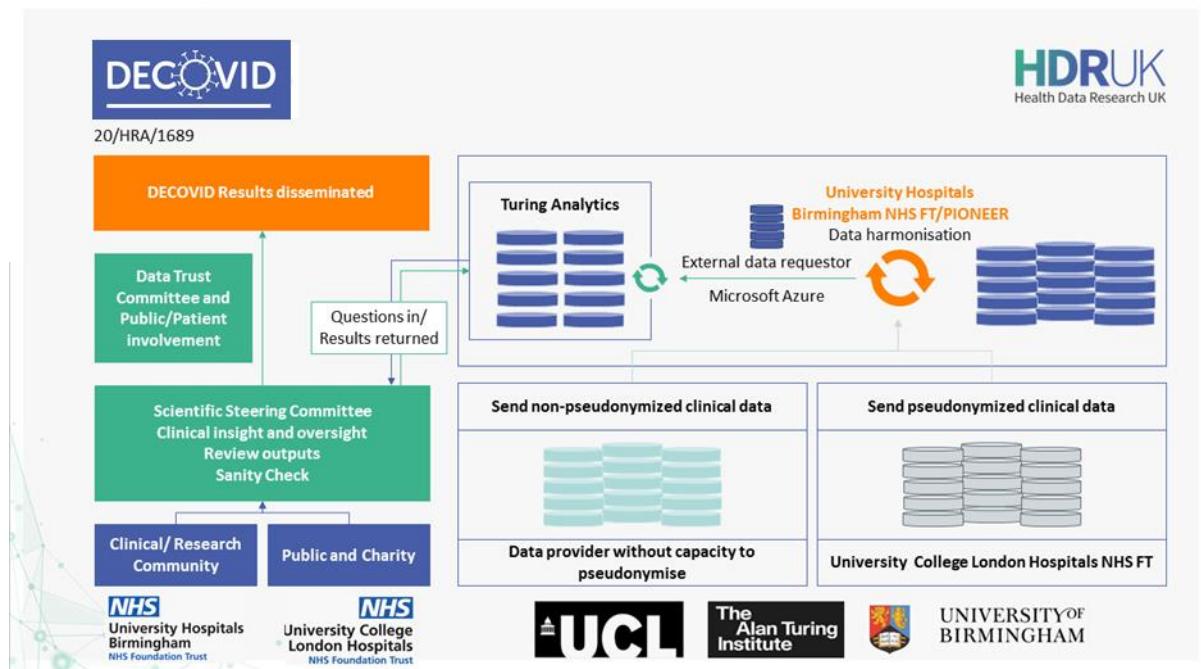
- Demonstrate how the COVID-19 response has exemplified the demand and value of using health data to inform frontline clinical decision-making.
- Celebrate successes of a data-driven response - highlighting rapid innovation which allows continuous improvement and a learning health response in a changing environment.
- Consider areas of unmet need and opportunities for future development – exploring data gaps and approaches to bridge the gap between granular local data and national insights at scale.

Takeaways:

- Patients traverse healthcare services, but their data does not.
- COVID-19 has generated a need to reconfigure hospital service provision to meet patient requirements but also to effectively embed data collection to inform the ongoing response
- Feeding near real-time data, such as positive patient tests, bed capacity and patient flows, into an accessible dashboard has provided an invaluable tool to effectively manage the demands of the pandemic in University Hospitals Birmingham ([DECOVID study](#)).
- The [OpenSAFELY](#) platform provides a model for creating a TRE through which to analyse GP records to understand COVID-19 risks.
- There is a need for more TRE platforms which are embedded in routine electronic healthcare records (EHR) to allow rapid and cheap analysis of us to large amounts of data, with high security, to inform the operational response.
- Nearly 50% of care homes have suffered a COVID-19 outbreak, with care home residents experiencing significantly worse outcomes following hospitalisation. EHRs have provided key data to inform insights in other settings, such as hospitals, but their use is limited in the care home setting.
- There must be Patient, Public and Practitioner/Professional Involvement and Engagement at every level to ensure trust, openness, and accountability in the appropriate use of data.

- This is “a time of turmoil but not of anarchy” – we have repeatedly seen the value of openness, collaboration and transparency to enable a rapid and innovative response to the virus.

Liz Sapey discussed the DECOVID dataset:



Breakout 6: Responding to the Public Health Emergency of COVID-19

Chairs:

- Peter Bradley, Director of Health Intelligence, Public Health England
- Jill Pell, Director, Institute of Health and Wellbeing, University of Glasgow

Panellists:

- Ronan Lyons, Clinical Professor of Public Health, University of Swansea & Research Director, HDR UK
- Helen Colhoun, AXA Chair of Medical Informatics and Life Course Epidemiology, University of Edinburgh, Honorary Consultant in Public Health, NHS Fife
- Julian Flowers, Head of Data Science, Public Health England

Session objectives:

- Demonstrate how insights from routine data, community data and real-time tools are informing the public health response to COVID-19.
- Highlight immediate lessons to inform the ongoing response.
- Identify research questions and data and infrastructure requirements which can enhance the approach to future public health challenges.

Takeaways:

- There are many datasets being leveraged for the first time during the pandemic response, including daily, weekly and monthly data-streams that have been feeding through to the government to inform the policy response across all four devolved nations.
- Multiple datasets are being linked together and being used to answer key questions in areas such as: how many cases there are to date, the trajectory of the epidemic, demographic factors affecting risk, excess mortality, the needs of those in hospital and care homes, the proportion of the population who have had COVID-19, and the risk factors for becoming seriously ill with the disease.
- Overall, 74% of people becoming seriously ill had at least one pre-existing condition compared with 48% of age, sex and GP-matched controls. However, around half of the patients under 40 had one or more listed condition compared with just 8% of controls, challenging the perception that severe COVID-19 risk in younger people is not predictable.
- The pandemic presents significant data science challenges including data management, data access, data linkage and public reporting. We also need to upscale the analytical workforce and build data science skills, as well as maintaining the new connections that have now come between the academic research sector and public health.
- There is a need for more spatial data to understand the epidemic as we move towards relaxing lockdown measures and returning to some semblance of normal life, in order to understand how the virus is being transmitted between groups, people and sectors – including households, schools, hospitals, care homes and other locations – and to look at the impact of measures such as lockdowns and social distancing.

"Overall, we have consolidated our work as we are going along to make it sustainable and do much wider data linkage in the future, building on the lessons that we have learnt and solutions we've created during this time. It would be great to maintain the connections that have now come between the academic research sector and public health."

**Julian Flowers, Head of Data Science,
Public Health England**

Breakout 7: What do national and international health data researchers need in a post-COVID world?

Chairs:

- *Andrew Morris, Director, Health Data Research UK*
- *Emily Jefferson, Director, Health Informatics Centre, University of Dundee*

Panellists:

- *Gerry Reilly, Chief Technology Officer, Health Data Research UK*
- *Steve Kern, Deputy Director, Quantitative Sciences, Bill & Melinda Gates Foundation*
- *Tim Hubbard, Professor of Bioinformatics, King's College London and Associate Research Director, HDR UK*

Session objectives:

- Explore how we coordinate our health data research nationally for international alignment.
- Assess how to develop collaborative frameworks with international efforts - COVID-19 related efforts.
- Discuss how to roll out the principles outlined in the UK Health Data Research Alliance Trusted Research Environment (TRE) Green paper, nationally and internationally?

Takeaways:

- COVID-19 is a global health emergency: emergencies can be catalysts to do new things, but they are not always the most efficient catalysts.
- We must learn how to take principles of what HDR UK has done during the COVID-19 response and turn that into a global model not just for this pandemic, but also future global health challenges we are trying to solve together.
- It is time to change the language and conversation around health data from data ownership to data stewardship (custodianship) and from 'sharing' to 'safe access'. This change is fundamental to build trust and meet changing health data research needs discussed in more detail in the [Health Data Alliance TRE Green paper](#).
- To foster effective international collaborations, we need to bridge the 'archipelago of data islands' where global health data is often missing, of poor quality, siloed, redundant and under-utilised. How do we join it all up safely and ethically for the benefit of all, with trust at the heart?
- There is a need for more TREs that embody the '5 Safes': Safe people, Safe projects, Safe setting, Safe outputs, Safe data. The [Genomics England TRE](#) is an exemplar that meets all of these standards.
- This is now possible thanks to cloud computing and virtualisation, bringing researchers to the data rather than sending the data to researchers. We should also work towards a system of federated trusted research environments to enable research to be done on multiple joined up datasets across the whole country.

"We have an archipelago of data islands, where global health data is often missing, of poor quality, siloed, redundant and under-utilised. How do we join it all up safely and ethically for the benefit of all?"

**Steve Kern, Deputy Director,
Quantitative Sciences,
Bill & Melinda Gates**

- The [Health Data Research Innovation Gateway](#) has a pivotal role to play in increasing the discoverability of COVID-19-relevant datasets, and supporting the sharing of data and tools

Breakout 8: Relationship of trust: Realising patient and NHS benefits in times of COVID-19

Chairs:

- *Jonathan Montgomery, Professor of Healthcare Law, UCL, non-Executive Board member, HDR UK*
- *James O'Shaughnessy, Former Under-Secretary of State, Department of Health and Social Care, non-Executive Board member, HDR UK*

Panellists:

- *Angela Coulter, Chair, HDR UK Public Advisory Board*
- *Phil Booth, Coordinator, medConfidential*
- *Jason McFall, Chief Technology Officer, Privitar*

Session objectives:

- Reflect on successes and lessons in building public trust in use of health data for research and innovation during COVID-19.
- Build on progress already achieved by [Hubs](#) and other members of the [UK Health Data Research Alliance](#) in increased transparency through the commercial framework.
- Identify opportunities to build greater patient and NHS benefit in the next phase of COVID and of HDR UK.

Takeaways:

- It is time to start thinking about what it takes to make it worthwhile for people to place trust in the data initiatives we are inviting them to participate in.
- Patients and the public do not all need to see the same gains coming from our participation, but they do need to feel being involved is worth it, in order to have a fair and mutual relationship.
- There is a ‘silver lining’ to the pandemic in the form of greater public discussion about health research, data and data science, along with a huge increase in interest and willingness to participate in research (for example, the [COVID Symptom Study](#) app).
- COVID-19 is a special case – the rules are being relaxed to speed things up ([COPI notices](#)), but it shouldn’t be an excuse for standards for data privacy to be relaxed, otherwise we risk losing public trust in the long term. We would be wise to build effective data privacy and consent systems and processes for the future rather than rely on this specific situation.
- Technology should be open to scrutiny from broad range of experts - this brings rigour to the process and a move from merely talking about transparency and openness to drive concrete initiatives that improve trust.
- We should think about privacy as ‘public goods’ in terms of health data – consent only exists because of earned trust through collaboration and interactions.

Maintaining public trust ... we have seen too many examples of how not to do it



Muddled communications and dodgy statistics,

Poorly labelled and inadequately explained



Implication that there's a single truth and a single value-free science

"Led by the science"

Lack of acknowledgement of uncertainties and controversies



Failure to demonstrate trustworthiness

Delays in publishing SAGE evidence and membership

Fewer studies with patient/public involvement

Cutting corners on privacy, consent and scrutiny

Rapid fire early career lightning talks

Three-minute talks by early career researchers, technologists and innovators from across the HDR UK community. Speakers were selected by an independent panel following an open competition, featuring high impact, innovative exemplars of work which provides new insights, and progresses HDR UK's mission.

Chairs:

- *Lucy McCloughan, Health Data Programme Manager, University of Edinburgh*
- *Rona Strawbridge, HDR UK Fellow, University of Glasgow*

Assigning unique pseudonymised identifiers to addresses to enable linkage to care homes: evaluation of a novel algorithm – Gill Harper, Queen Mary University of London

- An algorithm has been developed that uses a unique property reference number (UPRN), address matching and anonymised routine medical records to identify residents living in care homes.
- Running the algorithm against a test set of pseudonymised addresses gathered from GP-registered patients in a mixed ethnic community in London showed that 99% of care home residents could be successfully identified.
- The remaining 1% are more likely to be from BAME communities (except Bangladeshi), the very young or very old, men, those most recently registered with a GP, and those living in the least deprived areas.
- The team are now collaborating with Wales and Scotland to validate the algorithm in other settings and provide a gold-standard approach to identify care home residents which will be invaluable to evaluate COVID-19 risk and outcomes in this population.

Vitamin D Concentration and COVID-19 – Claire Hastie, University of Glasgow

- There has been a great deal of attention in the media regarding a possible link between low blood levels of vitamin D and increased risk of COVID-19.
- Studies so far are limited, and evidence is uncertain – but if confirmed, vitamin D supplementation could be a cheap and effective intervention suitable for widespread population use.
- Analysis of UK Biobank data reviewed 449 individuals with confirmed COVID-19 out of ~350,000 participants. Results showed no association between vitamin D concentration in blood and COVID-19 risk, after adjusting for confounders such as obesity and low socioeconomic status (also correlated with low vitamin D).
- Further analysis suggests that vitamin D does not explain any increased risk of COVID-19 seen in BAME populations.

Real-time hospital cancer data, excess deaths and a question for the Prime Minister – Alvina Lai, University College London (WINNER)

- COVID-19 has resulted in direct and indirect effects on cancer patients and services. Individuals with cancer are likely to be at increased risk of COVID-19 infection and show poorer outcomes, but there is an added impact of late diagnosis or delayed treatment for cancer due to the pandemic.

- Hospital data from multiple centres in the UK show that urgent referrals for early cancer diagnosis and chemotherapy attendance drastically reduced during the pandemic.
- Estimates of the impact of the pandemic suggests up to 17,000 excess cancer deaths will occur over the next 12 months. These figures assume of 10% patients with cancer directly affected by COVID-19 and 30% indirectly affected across 24 cancer types.
- A key question results: How can we ensure that cancer patients are continuing to receive optimal care and access to treatment as we move through the pandemic?

The Polygenic Score (PGS) Catalogue: an open database for reproducibility and systematic evaluation – Samuel Lambert, University of Cambridge

- Polygenic risk scores (PGS) are an increasingly widespread tool for predicting genetic risk for a wide range of traits and diseases.
- However, there is a lot of difference about how these scores have been developed and evaluated, leading to confusion about what they mean and how best to use them.
- www.PGScatalog.org is an open database of 200 polygenic scores for 92 traits curated from 82 publications.
- The platform is being used by researchers and clinicians to evaluate the predictive ability of PGSs by linking through to health records.

Ethnic and socioeconomic inequalities in SARS-CoV-2 infection – Claire Niedzwiedz, University of Glasgow

- The work used data from UK Biobank linked to COVID-19 test results from Public Health England to review the probability of being tested for COVID-19 and the risk of testing positive both in general and in hospital (i.e. having more severe disease).
- Factors influencing the chances of testing were considered, including ethnicity socioeconomic status, country of birth, occupation, household size, pre-existing conditions and lifestyle.
- Black and south Asian groups were x3 time more likely to test positive, with Pakistani ethnicity at highest risk within the south Asian group. These ethnic groups were more likely to be hospital cases compared to the white British.
- Socioeconomic disadvantage and poorer pre-existing health do not explain all the increased risk.
- We need to determine why this increased risk occurs and ensure the health system is responsive to the needs of ethnic minority groups as we move forwards.

Monitoring diet more accurately to aid healthy living in a post-COVID-19 society – Joram Posma, Imperial College London

- Obesity is linked to increased risk from COVID-19, helping people to achieve or maintain a healthy weight is likely to be associated with better health during and following the pandemic, however many diet studies are hampered by relying on self-reported data
- The research team have developed a method of analysing diet based on metabolites in urine. These data also capture information on underlying genetics and microbiome.

- The approach was tested in a small clinical trial with participants receiving one of four standardised diets. Urine metabolite score proved highly effective in predicting which diet participants were following.
- The approach is currently being tested in people at risk of cardiovascular disease to evaluate what they are eating and allow personalised dietary advice to be provided.

Using smartphones and wearable devices to monitor behavioural changes during COVID-19 – Shaoxiong Sun, King's College London

- During the pandemic there is a need to understand how people are behaving on a local level in response to interventions such as lockdown.
- RADAR-base offers an open-source mobile health platform and provides a readout of behaviour such as travel outside the home, phone use, sleep and social app interactions.
- 1062 participants were recruited to test the app in Italy, Spain, Denmark, the Netherlands and the UK.
- The app data showed significant changes in behaviour during lockdown for all activities except total sleep duration. Changes included leaving home and travelling less and increased phone activity.

Fireside chat: How will data and technology influence the future of clinical and regulatory science?

Amy Abernethy, Principal Deputy Commissioner of Food and Drugs, the FDA, USA and Chair, HDR UK International Advisory Board



Speakers:

- *Amy Abernethy, Principal Deputy Commissioner of Food and Drugs, the FDA, USA and Chair, HDR UK International Advisory Board*
- *In conversation with Andrew Morris, Director, Health Data Research UK*

Takeaways:

- ‘Real world’ health data (sourced from outside conventional clinical trials) and developing technologies bring exciting, new capabilities that have great potential to support clinical research and regulation.
- Bigger, linked datasets are powerful, but we still need to develop approaches to use them safely and effectively to answer important questions and implement common languages to ensure interoperability. National and international collaboration is essential to successfully answer this demand.
- This work is hard, these methods are challenging to develop and implement, and, in the case of COVID-19, efforts are set against a backdrop of fast-moving research, constantly changing situation and evolving evidence base.
- Irresponsible or inadequate application of real-world data in the response to COVID-19 (for example retracted papers about hydroxychloroquine based on flawed datasets) can hamper efforts of the wider research and policy community and risk losing public trust.
- Real-world data should not just be used to provide the evidence to bring drugs to the market, but its’ continuous evaluation provides a key opportunity to identify drugs that are not effective and justify their removal from the market.
- Algorithm-based products and their regulation presents many new challenges. We need to consider how to future-proof the regulation of digital healthcare products and ensure the approach is reviewed, updated, and improved over time.
- The UK present a unique opportunity due to the close interactions between the National Health Service and research community. By investing in people and skills for the future, linking biological data and clinical datasets and ensuring efficient translation from research to the clinic, the UK can act as a global leader in this space.

Next steps

- Please refer to the [conference page on our website](#) for more detailed information, including session recordings, slides, prize winners and the full agenda.
- Over the coming weeks, we will share case studies of selected work discussed in our the conference in the [Opinion section of our website](#).
- To find out more about any of the work discussed over the day, please contact events@hdruk.ac.uk and our team will be happy to help.

About Health Data Research UK & further information

Health Data Research UK

Health Data Research UK is the national institute for health data science in England, Wales, Scotland and Northern Ireland. Its mission is to unite the UK's health data to enable discoveries that improve people's lives. It is a not-for-profit public benefit company 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.

Visit www.hdruk.ac.uk for more information.

Health Data Research Innovation Gateway

The Health Data Research Innovation Gateway is a portal to find and request access to UK health datasets controlled by members of the UK Health Data Research Alliance. The first phase of the Gateway provides detailed descriptions (meta-data) of these datasets, which researchers can search, browse and request access to health data. It does not hold or store any patient or health data. It aims to increase transparency around accessible datasets and processes associated with their access. The Gateway is part of the UK-wide capability for the responsible use of health data for research and innovation, funded by the [Industrial Strategy Challenge Fund \(ISCF\)](#).

Visit www.healthdatagateway.org for more information and to test out the Gateway.

UK Health Data Research Alliance

The UK Health Data Research Alliance is an independent alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale. Convened by Health Data Research UK, the Alliance develops and co-ordinates the adoption of tools, techniques, conventions, technologies, and designs that help researchers to answer some of the most difficult questions and address the most important health challenges faced in the UK. The Alliance is part of

the UK-wide capability for the responsible use of health data for research and innovation, funded by the [Industrial Strategy Challenge Fund \(ISCF\)](#).

Visit www.ukhealthdata.org for more information.

Get in touch

Follow us on Twitter: [@HDR_UK](#)

Follow us on [LinkedIn](#)

General enquiries: enquiries@hdruk.ac.uk

Events enquiries: events@hdruk.ac.uk

Media enquiries: media@hdruk.ac.uk

Subscribe to our [mailing list](#)