

## Lessons from COVID-19: How can we put the UK at the forefront of health data science?

People engaged in health data research from academia, NHS, policy making, charities and industry contributed to a roundtable discussion on 20 July 2020 convened by NHS England and HDR UK to discuss *How can we put the UK at the forefront of health data science?*

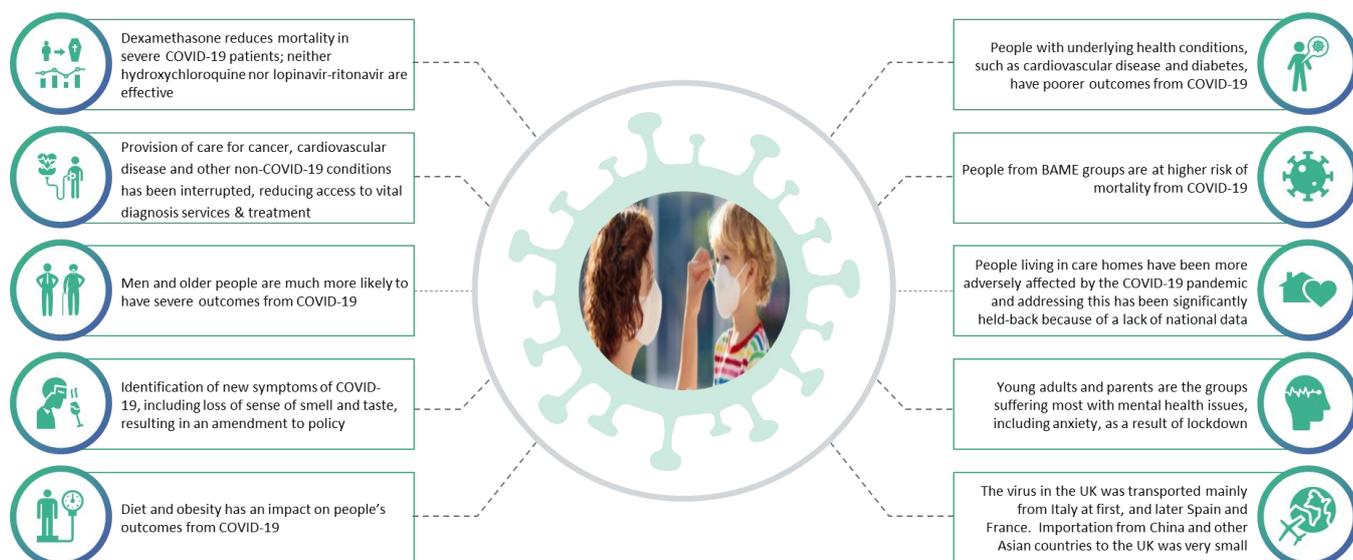
This was a lively, informative, first conversation with the group, reflecting on lessons from the initial response to the COVID-19 pandemic and on the actions that should be taken as a result. The outputs from this discussion will be shared with funders and system leaders to inform the design of UK health data research capabilities to respond effectively to subsequent COVID-19 waves, future pandemics, and also to improve health data research for other diseases.

This document summarises the outputs from the discussion, chaired by Caroline Cake and David Prior, and a recording of the discussion is [available here](#). The discussion was also informed by an open online survey and the detailed responses are [provided here](#).

### Question 1: In relation to health data science, what have we learnt from COVID-19?

Through COVID we have seen the public benefits that can be achieved from linking health data research and clinical delivery, e.g., the Dexamethasone findings via the RECOVERY trials.

Figure 1: Ten things that we now know about COVID-19 because of health and cross-sectoral data research



We also found that public engagement and information governance can be done fast and effectively, and that data-driven decision making for operations is necessary...and possible.

Lessons from this experience include the need for:

**Clear research questions (and therefore tight scope):** Barriers and blockages between clinical teams, researchers, data custodians and policy makers fell away when people were working together on clear, pressing questions to meet an overarching urgent and combined mission.

**Urgency, supported by emergency regulations:** the Control of Patient Information (COPI) notices<sup>1</sup> enabled more rapid access to data to solve urgent questions to inform critical care delivery and government decision making. However, despite this many researchers still found it difficult to access and link datasets.

**Transparency:** Public engagement can be done fast and done well and people seem willing to have data used when they know why and where their data is going. However, saying you're going to be transparent isn't the same as being transparent, for example, more open data and code needs to be shared to be used for rapid analysis of UK health data, and published records are needed of who is accessing the data.

**Established infrastructure:** Where health data infrastructures and relationships were already in place, it was possible to respond to the pandemic at pace. Where there were gaps, for example with social care data, it was much more difficult.

**Scale and linkage:** We will learn most from large scale studies where molecular data and social care data were linked with real-time health data. National, and international coordination is required to get the right sample size for meaningful insights. However, data access is still really hard, especially when linkage is required or to access the more detailed electronic health records in secondary care.

**Improved quality of data** - showing the need to improve digital maturity in NHS and social care settings.

**Action to tackle inequalities:** We learnt from the data who was at risk, but not how to protect them. Blockages and barriers that may have been overcome in health have not been overcome in social care data. There was also a disconnect between local and national - those caring for populations need to have the information.

## **Question 2: Over the next five years, what opportunities could there be for UK health data science?**

These lessons from COVID-19 suggest that there are considerable opportunities for UK health data science to:

**Ask better questions:** We need to focus on great questions and specific use cases for health data science. We can then work back from the question, to the data to the analytics and pull together the data we need for these questions.

**Run larger, multidisciplinary studies:** The benefit of scale, multi-disciplines, and international partnerships is well evidenced from COVID. This requires collaboration between different communities with an understanding of different skills, shared language, mutual respect of differences, and a breadth of rich data

---

<sup>1</sup> <https://www.nhs.uk/covid-19-response/data-and-information-governance/information-governance/copi-notice-frequently-asked-questions/>

source from across health and social care boundaries to look properly at all determinants of health and help plans for the future.

**Use real-time data to innovate:** Data allows us to think differently about how we provide care, for example we can increase care enabled at home, or even create virtual hospitals. We can upload and integrate people's self-tracking information, of which there are probably millions (from their smart-phone step counts to fitbits to blood pressure readings, etc). We can increase participatory health where the person (personalized medicine) has the tools (possibly apps and wearables) that will collect huge amounts of data. We also need to innovate in how we use and improve data, automating how we collect and curate data to make it usable.

**Align investments around delivery and impact:** The sense of urgency created by COVID can be brought to other priorities, for example mental health and obesity. This requires a focus on delivery and demonstration of impact. We need to learn from where delivery and impact are being achieved, and where they are not, so that we invest in what works.

**Build the infrastructure, resource and capacity:** Let's invest in the Trusted Research Environments, aligned to the 5 Safes, so that we have trustworthy infrastructure in place. We need agile data scientists who have been working in the infrastructure in multidisciplinary teams for a number of years so that we can use the data effectively.

**Reach a constructive balance between commercial and public interests:** We need UK PLC to thrive, for private companies to develop successful businesses within the UK, for financial investment to flow into the NHS, and for the public to understand and trust how this is happening.

**Address inequalities and representativeness:** The future approaches need to ensure diverse groups of people, particularly vulnerable populations, will benefit from the insights. Vulnerable groups and seldom heard groups need to be involved in innovative and proactive ways. Social care data needs to be improved. But we also need to design tools and approaches that support groups who are concerned about their data being used, to avoid growing inequalities of health and care for these groups. We also need to consider how representative are the people who are accessing data and using the data.

**Increase transparency:** We need to get over being scared about transparency and we need to communicate who is using data and for what purposes so that people can collaborate and so that the public can understand how data is being used. A statement of how our data is being used, like a bank statement, that creates a normalised channel for public information and engagement.

**Tell people about the benefits to the public from health data research:** We need to develop great use cases for health data for things that matter to people. Let's use the fact that people are interested and eager to find out more about health data research because of COVID and engage and inform them in what can be achieved. Patients with chronic diseases and cancer are better informed, we need to broaden out the wider public's understanding. We also need to involve people (not just patients) as contributors and make them feel involved in a way they may not if they are not currently a patient.

**Establish sustainable regulation:** We can't rely indefinitely on emergency regulations and need to replace them with sustainable policies and structures to move away from the traditional "computer says no"

experience that many researchers and innovators have experienced. We need to shout about why we need better data access, and in a safe way that is understandable to the public.

**Be better coordinated:** Health data is a confusing and fragmented landscape. We need

- greater integration of data across health and care
- clearer signposting, building on the process of the HDR UK prioritisation/funnel.
- within the constraints of grant applications, it will be useful to know who is interested in certain areas and what research is ongoing so that collaboration can occur
- interoperability between different trusted research environments
- to encourage data custodians to join the HDR Alliance
- to keep working on making it easy to find data and improve data quality
- clear & consistent frameworks for access, contracting, governance

**Be international.** The NHS provides one of the most complete health records for a population globally, and our ability to use this to enable real world evidence to inform and perform clinical studies should be world beating. Other large countries can't link longitudinal data to the same extent, the UK needs to step up to the plate to enable this at scale. UK Biobank is a good example, providing good and clear access, international collaboration, and clear ownership.

### **Question 3: Where are we now? And what needs to happen for us to meet the opportunities you identified?**

Health data science has a public profile now that it has never had before. We have seen new relationships forged between organisations, nationally and internationally. But all of this is still new and could easily be lost again – we need to avoid letting practices slip back to the way they were. Let's:

- Work together to have a **joined-up plan** and a **unified voice** across health data in the UK, so that it is less confusing for the public, researchers and innovators to engage in.
- **Broaden the community** to include policy makers (translating science into policy), patients and the public (e.g. when looking at at-risk groups, those groups need to be involved), charity sector who have existing disease registries and networks, incorporate clinician voices through Societies and Associations to ensure questions addressed also affect clinical delivery as well as epidemiology, and ensure all four nations are represented.
- Prioritise effort around **use cases** that have a **clear public benefit**, including those with industry and with international partners. Focusing data access on answering specific questions will help to further build public trust.
- Have a clear strategy for what happens after COPI and for how to translate the benefits achieved through COVID, given that COPI notes are purely for COVID research. **Information Governance** has been seen as a barrier, how instead do we position it as an enabler of public confidence? So that we can nationally make decisions once, have a consistent approach across local and regional decision makers, and establish precedent and expedited processes for certain use cases.
- Introduce a Digital Duty of Care that empowers Local Authorities to access the range of data 'locked up' in **social care** providers. Require digital maturity amongst registered providers of social

care and, if need be, encourage the Treasury to facilitate low-cost loans and/or utilise the NHS's ability to procure at scale to reduce associated costs for providers. Consider co-producing data-driven tech, also, with and for recipients of social care (and their families) - there are far more of them than there are local authority funded individuals. Incentivise data-driven innovation in social care through a dedicated fund - we don't need as much data about the system as is but about 'care services as could be' - the asymmetry between health and social care will persist unless and until we do different.

- Put in place the **resources** to do this work well, with truly open competitive funding for proven delivers.

This is a summary of the roundtable discussion held on 20 July 2020, chaired by Caroline Cake, Health Data Research UK, and David Prior, NHS England and NHS Improvement.