

# Goldacre Review of Use of Health and Care Data for Research and Analysis

HDR UK Additional Messages: 24 March 2021

## Introduction

Health Data Research UK (HDR UK) welcomes the opportunity to contribute to the Goldacre review of health data for research and analysis. The health data research community have actively participated in the interviews and working groups and the HDR UK Public Advisory Board has also made a submission ([here](#)).

This document summarises those contributions, provides additional inputs, and HDR UK case studies to support the published terms of reference for the Goldacre review. We look forward to working together with the review team and other partners on the implementation of the recommendations to help further the use of health data to enable discoveries that improve people's lives.

## Key points highlighted during HDR UK interviews

These are the points that HDR UK has emphasised through the review interviews to enable a trustworthy health data research ecosystem:

1. **Transparency of use** and **streamlined and harmonised data access** through the **Health Data Research Innovation Gateway**: HDR UK has worked in partnership with NHS and research organisations in the Alliance to make improvements in data access based on the 5 Safes - this needs to be sustained, embedded and extended (particularly for primary care) to enable research benefiting other conditions
2. Sustain and embed **trustworthiness** demonstrated through COVID-19, through effective patient and public engagement and involvement in data research and transparent reporting
3. Commit to a trustworthy ecosystem that enables **many researchers** safely accessing **many datasets** for **many uses** (e.g. Clinical Trials, Understanding Causes of Disease, population health, industry)
4. Include 'health' in the **Digital Economy Act** – enabling linkages between health and non-health data
5. Support and enable the role of **industry** in the ecosystem delivery (e.g. public cloud)
6. Celebrate the benefits of and importance of the **four nations** approach to enable research based on data representing the whole of the UK
7. Support **federation** across Trusted Research Environments, including federated analysis to increase the power and impact of research studies
8. Further develop professionalised, **user orientated** research services and expertise to help researchers and innovators use data for research
9. Enable the ecosystem to scale up for **new and diverse datasets**, beyond the national NHS data, e.g. industry, symptom trackers; and increase representativeness in data to reduce **data poverty**

10. Support and enable research on **local/regional/national/international levels of data**

## Additional points for the Review

### Open and transparent sharing of data and code

- We have shared with you the vision for the [Health Data Research Innovation Gateway](#) as a front door to FAIR data services to enable open collaborative health data science
- This is being developed iteratively but is already enabling many researchers to gain access to >600 datasets at scale across the UK's four nations
- We are seeing usage of the Gateway grow month by month (currently 15,000 users) and is being developed in partnership to ensure we clearly address the user need
- In addition, facilitated by our [Attribution Policy](#), which sets out that our community members should share code through GitHub, we have brought together [145 repositories of code](#) tackling some of the most important challenges in wrangling multi-modal data and generating replicable insights. On our website, we provide an [initial introductory overview and signposting of some of these highlights](#).

### Federation and Interoperability

- Not one organisation will be able to achieve a true transformation and COVID-19 is an important test of our healthcare ecosystem
- We need a federated approach with interoperability as a key consideration for any infrastructure development. This should include people, process, technology and supported with appropriate funding for the infrastructure
- The [UK Health Data Research Alliance](#) is an established organisation of >40 major data controllers across the UK with the purpose of establishing best practice for the ethical use of UK health data for research at scale
- The Alliance co-ordinates the adoption of tools, techniques, conventions, technologies, and designs that enable the use of health data in a trustworthy and ethical way for research and innovation
- To improve awareness and understanding of the different uses of health data for research and to promote transparency, the UK Health Data Research Alliance is undertaking a project to agree standards for data access registers and we expect this work to conclude towards the end of May 2021.

### Developing a community of analysts/data scientists across the health sector

- In Spring 2021, we will be unveiling the next stage of our strategy to train over 10,000 data scientists
- We have sponsored an **open virtual learning environment** with hundreds of CPD courses for anyone wishing to build their skills in data science - targeted at clinicians, health care professionals, students and existing data scientists
- We are planning a series of events throughout 2021 to engage with the wider health data research community, both in the UK and internationally, and are very open to working in partnership
  - **Government Office for Science National Core Studies for COVID-19 Symposium** on 24 June in partnership with UKRI, NIHR, CSO, Wales and Northern Ireland
  - **HDR UK Annual Scientific Symposium** on 23 June; we anticipate running joint sessions with 6-8 organisations and anticipate up to 2000 delegates.

## Engaging the public and patients

- Recognising that while health data used for research and analysis comes from a variety of sources, it all relates to peoples' interaction with the health and care system in some way, whether as a patient, a participant in a trial, or by donating blood/tissue/genomic data etc
- Patients and the public are key partners in our work, through clear communications and engagement to build understanding and create knowledge; being transparent in how data is access and used; and through meaningful involvement in decision making and shaping projects
- During 2020, HDR UK and our partners involved over 18,000 patients and the public in our work, we have published several high impact case studies, and we will continue to grow this important activity
- We would be delighted to continue to work with the NHS and government to build capability.

## Improving data quality

- The seven UKRI funded [Health Data Research Hubs](#) are being assessed against their second key milestone and will report on their progress in April 2021
- The initial evidence suggests a very positive impact of the Hubs on improving the quality of UK health data over the past 12 months and their role in the COVID response
- There is an opportunity to use these impact case studies with you shortly to demonstrate the value of open and transparent sharing of data and code.

## Case studies

### HDR UK Black Internship Programme

Black data scientists are currently under-represented within the health data science community, and this is one of several steps that we are taking to start to address this.

By bringing together the extraordinary variety and breadth of health data science activities across different diseases (from cancer to cystic fibrosis), different organisations (mental health trusts to biobanks), and from different parts of the country (all four nations) – we have worked with the UK Health Data Research Alliance to create an internship programme to attract talented Black people who are considering a career in health data research.

51 interns are participating in the inaugural programme in summer 2021 at NHS Trusts, universities and charities across the UK.

### Data & Connectivity COVID-19 National Core Study

Data and Connectivity is one of six national core studies established by the Government's Office for Science to accelerate answers to key research questions on COVID-19 by enabling streamlined data access and analysis. Set up in October 2020, it is led by HDR UK in partnership with the Office for National Statistics. It builds on the established Health Data Research Innovation Gateway and trusted research environments across the four nations of the UK to ensure data is stored safely and securely, made readily available to researchers and provide services that make the data easier to use.

Over 60 datasets are now linked and available for vital COVID-19 studies, utilising vaccine, genomic, primary care, secondary care, testing and survey data. Using these data for research has shown, for the first time, early indications of the real-world effectiveness in the Pfizer/BioNTech and Oxford/AstraZeneca vaccines in reducing hospital admissions. Faster enrolment of people to clinical trials has been enabled. For example, the use of data for people who had recently tested positive for COVID-19 rapidly accelerated recruitment to the community-based PRINCIPLE Trial. Data & Connectivity has also demonstrated rapid public engagement, including more than 800 members of the public helping to prioritise research questions related to COVID-19 vaccines.

#### **BHF Data Science Centre-NHS Digital TRE- transforming access to England-wide health data for research**

At the start of the COVID-19 pandemic, apart from through special honorary contract arrangements, bona fide researchers could not access national linked healthcare data across the population of England to enable critical research to support healthcare decisions and public health policy.

There were two main reasons for this: there was no national collection of linkable, comprehensive primary care data; and there was no secure, privacy protecting mechanism for researchers to access and conduct population-wide research using national datasets linked across different parts of the health data system (from primary care, hospitals, death registries, laboratories etc).

Working in partnership with NHS Digital, Health Data Research UK and its British Heart Foundation Data Science Centre have changed this, through:

- establishing a new, linkable primary care dataset covering almost the entire population of England (>98% of practices)
- developing a new trusted research environment in partnership with NHS Digital to hold and provide secure, remote, non-exclusive and non-preferential access for research to an unprecedented breadth of linked health data (covering primary care, secondary care, registered deaths, COVID-19 laboratory and vaccination data and cardiovascular specialist audit data) covering the whole population of England
- enabling access to these data for a large consortium conducting a broad programme of COVID-19 related research.

This population-wide resource encompassing a cohort of almost 55 million people in England demonstrates the importance of linking person-level data across health settings to maximize completeness of key characteristics and to ascertain cardiovascular events and Covid-19 diagnoses. Although established initially to support research on COVID-19 and cardiovascular disease to benefit clinical care and public health and to inform health care policy, it can broaden further to enable a very wide range of research.

#### **Health Data Research UK (HDR UK) and the UK Health Data Research Alliance**

42 of the UK's NHS, charity and research organisations have worked as an alliance since 2019 to establish best practice for the trustworthy and ethical use of UK health data for research at scale. The Alliance is convened by HDR UK, the national institute for health data science. It covers the 4 nations of the UK and is focused on data sharing for research across research cohorts, biobanks, medical research charities and NHS Trusts. It has successfully built alignment across data custodians on TREs, diversity/representativeness in data, PPIE, data and metadata standards, and streamlined, and harmonised access management processes.

Datasets from Alliance are discoverable through the Health Data Research Innovation Gateway, with almost 600 datasets, covering up to 54.4 million people.

Working in partnership with over 22,000 public and patient participants, the Alliance, Gateway, Health Data Research Hubs and Research Programmes are enabling a secure, federated approach to data-enabled health and care breakthroughs including in COVID-19 Clinical Trials (RECOVERY, PRINCIPLE, and Tackle); large-scale and rapid whole-genome virus sequencing (The COVID-19 Genomics UK (COG-UK) consortium, including identifying the new, more virulent strain first identified in SE England; mapping of COVID-19 cases in high-resolution and near real-time, directly informed public health policy – most notably in Wales; and 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.

**The Health Data Research Innovation Gateway – the UK’s portal for data discovery and tools for research**

In January 2020, the UK’s Health Data Research Innovation Gateway was launched by HDR UK in partnership with the UK Health Data Research Alliance as a cornerstone of the UK Life Sciences Industrial Strategy. For the first time, this provides a single front door to UK health data, with streamlined and harmonised access management, across the four nations of the UK. The Gateway is at the heart of making open science a reality; so far, almost 600 datasets, 150 tools, 250 projects, 200 education courses and 860 papers, are available and these are growing daily. The datasets cover primary care, secondary care, acute care, palliative care, biobanks, research cohorts and deeply phenotyped cohorts, including COVID-19 vaccine datasets and international datasets hosted by data custodians across all four nations of the UK. The Gateway has more than 13,000 searches per month from across the world and over 220 data access requests via the Gateway in the first year.

Through the Gateway, HDR UK and its partners are championing open access, transparent and trustworthy data science to enable the scientific community to pool expertise, draw insights and increase our collective understanding. All infrastructure source code and related materials are published in [Github](#) using a permissive open-source license, and this contains research software outputs from affiliated research projects that host over 100 repositories and 86 items of research software. The Gateway is [open source](#), supporting researchers to develop their own services, tools and encouraging reuse of the Gateway.