

HDRUK
Health Data Research UK



HSC Public Health
Agency
Research and Development



Public Health
Scotland

NHS
Digital

SAIL DATABANK

UK Health Data
Research Alliance

COVID-19 Health Data Research

15 September 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



COVID-19 Health Data Research recommendations – 15 September 2020

Health data research insights on COVID are continuing, with 110 research questions, 150 (+6) projects active within the national data Trusted Research Environments (TRE), 179 (-1) in development, and 723 pre-print publications and 44 published papers.

Incremental progress across all recommendations. Actions are being incorporated into national core study - data and connectivity planning.



5 Recommendations endorsed by SAGE on 11 June

Progress on SAGE actions identified on 11 June

<p>1 All swab & antibody testing programmes data to be securely linked and used for research. Requires unparalleled cooperation across all four nations between NHS organisations, PHE, data custodians, academic endeavours, and technology partners, whilst building public trust.</p>	<p>SAGE ACTION: HDR UK to work with partners to plan and create a serology and testing data research asset that is linkable to other data sources. PROGRESS: <i>No change to previous report:</i> Iteration ongoing with UKRI/NIHR panel following feedback requesting a more focused resubmission from the collaborative team.</p>
<p>2 Further research, undertaken collaboratively with international partners where appropriate, should address why BAME groups have a higher rate of severe COVID-19 outcomes. This will help to target the best interventions and inform the response to future public health crises.</p>	<p>PROGRESS: Further insights being generated (see next page). UK Health Data Alliance Data Officers Group met during week beginning 7th Sept and agreed first steps to enact the Board commitment to improve consistency and quality of ethnicity coding and enable data use to increase the representativeness of research studies.</p>
<p>3 Enhance data capture on patients and staff in care homes, in particular interconnections between settings, to enable research on health, transmission and outcomes. Clarify appropriate use of national Trusted Research Environments for consolidation of relevant care home COVID-19 data.</p>	<p>PROGRESS: <i>No change to previous report:</i> Outcome of initial quality reviews on the NHS Digital adult social care management coronavirus status data collection is expected mid-September. The outcome of these reviews will determine how/if the dataset will be available for research. Establishing a broader research response still requires more accessible individual-level data for care home staff, residents and domiciliary care (care within own home) sector.</p>
<p>4 Accelerate access to restricted national datasets, since lack of availability is holding back crucial research.</p>	<p>PROGRESS: Community prescribing data now available via the standard NHS Digital Data Access Request Service (DARS). England Testing data and CHES remain available to priority studies to request but not yet via routine access requests. This recommendation will be integrated into the national core studies data and connectivity work from September.</p>
<p>5 Commission large scale collaborative analyses of the long-term impacts of health and social care changes during the COVID-19 lockdown on major diseases. This will require access to linked data from a range of sources (including from COVID-19 laboratory tests, primary and secondary healthcare, death registries, disease-specific audit/registry data). In addition, linkages to cross sectoral data beyond health will be essential to understand the wider impacts of COVID-19.</p>	<p>SAGE ACTION: HDR UK to work with ONS and others to accelerate linkage of cross-sectoral datasets. PROGRESS: HDR UK continues to work with national core study leads to identify and prioritise datasets and linkages to support priority research questions as part of the data and connectivity study programme, which awaits formal approval. BHF Data Science Centre: 17 cardiovascular analysts from 6 institutions actively working via access to linked datasets in the NHS Digital TRE, including primary care GPES data which includes over 4 billion journal data entries across population of over 56 million representing 97.5% of GP practices in England.</p>

Priority research questions with new insights generated this week – 15 September 2020

Health data research on COVID-19 continues to grow, now reaching 723 (non peer-reviewed) pre-prints & 44 published papers

Priority research questions

Insights from ongoing studies (links provide further details):

1. Understanding & protecting vulnerable groups (RQ 22, 32, 36, 62, 102)

- A systematic review and meta-analysis showed that [Black, Asian and Hispanic ethnic groups are at increased risk of SARS-CoV-2 infection; Black individuals may be more likely to require ICU admission for COVID-19; and although risk of mortality was similar across ethnicities among hospitalised patients, it was increased among Asian and Mixed ethnic groups in the general population.](#)
- A study - based on data from an online survey designed to measure symptoms of burnout, depression, anxiety and distress - aimed to predict healthcare worker burnout during the COVID-19 pandemic. [Independent predictors of burnout were found to be age \(younger staff more at risk\), redeployment to a new working area, working with patients with confirmed COVID-19 infection, and being female or having a previous history of depression.](#) One implication of this is that evaluation of existing psychological support interventions is required with targeted approaches to ensure support is available to those most at risk.
- A retrospective genomic epidemiology study analysed COVID-19 infection dynamics in care homes in the East of England. As well as finding that [residents had a significant burden of COVID-19 infections and high mortality, it also identified: a\) larger viral clusters that were consistent with within-care home transmission; as well as b\) multiple clusters per care home, suggesting independent acquisitions.](#)
- A study looked at the impact of delays diagnosing and treating colorectal cancer, as a result of interruptions to cancer care due to COVID-19. Using data made available through DATA-CAN, the health data research hub for cancer, the team found that [delays to treatment of 2-6 months were estimated to result in 650-2250 attributable deaths](#) (in a group of c10,000 patients).

2. Understanding immunity, testing reliability & prevalence (R01, 50, 95, 102, 51, 54, 55, 104)

- A modelling study using data from the Bug Watch community cohort study showed that [current national COVID-19 testing capacity is likely to exceed demand due to baseline cough and fever alone.](#) This suggests that a high proportion of people with symptoms request tests, and that, in order to meet this high predicted demand, testing capacity must be scaled up immediately.
- [Blood and physiological measurements are better collected as early as possible to contribute to improved predictive utility.](#) This was concluded from a study that used a novel artificial intelligence method (ensemble learning) to synergise seven multinational prediction models to develop a robust and high-performing prediction model. This ensemble model consistently outperformed any single models in all aspects validated.
- An international study compared the COVID-19 pandemic geographically and over time in Europe, using numbers of deaths, crude rates and adjusted mortality trend ratios. It found that [the UK, Italy, France and Spain registered the highest mortality in Europe.](#) The study concluded that, in understanding and managing the pandemic of COVID19, comparable international data is a priority and makes recommendations around how to do so.

Patient and Public Voice Feedback

The Panel reiterated the urgent need to focus on access to key datasets, particularly those related to testing, care home and domiciliary care.



19 COVID-19 taskforce calls with **89** clinical and health data research leaders engaged



1408 academic, industry and NHS participants in COVID-19 Slack channel with 10 sub-channels



110 health data research questions identified – 42 prioritised



723 COVID-19 pre-print publications



Click [here](#) for a link to the full prioritised list of questions, status, and prioritisation process

COVID-19 dataset availability and status of projects using the data – 15 September 2020

Important developments in scale and diversity of data being accessed for research: Census data onboarded in Scotland and Wales; metadata for the Welsh Testing data available via Gateway; and GP data for population of over 56 million being used in England TRE.



Health Data Research UK

KEY

- Data flows specified but not yet agreed
- Data flows agreed but not yet available for linkage
- Fully available

KEY UK WIDE PROJECTS:

[RECOVERY](#)

[CO-CIN \(ISARIC 4C\)](#)

[COG-UK](#)

[CARDIOVASCULAR CONSORTIUM](#)

[COVID-19 symptom study](#)

[GENOMICC](#)



Datasets available for COVID-19 research via national TREs for Wales, Scotland and England

1. Census 2011 Data onboarded for urgent priority COVID related research under strict use and permissions (Scotland and Wales)

Core COVID-19 Datasets available for linkage	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)
Primary Care				
Pillar 1 COVID-19 Testing Data				
Pillar 2 Testing data (UK Gov)	N/C Delayed: Was expected Aug			Missing results prior to 26 Apr
Pillar 3 & 4 Testing data	N/C No confirmed date		Pillar 3 available	N/C - Pillar 3 expected Sep
Community Prescribing				
Critical Care (CHESS, ICNARC, SICSAG)	N/C CHESS expected in DARS Sep			N/C - Under review
Personal Demographic Service				
Secondary Care	SUS only - not HES			
Death registry				
Census 2011				

2. First report with a net reduction of 'in development' projects as requests progress to active research.

# of COVID-19 Projects by TRE stage (change from previous report)	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)	Total
In development	43 (+1)	42 (+6)	88 (-6)	6 (-)	179 (-1)
- a/w researcher	(-)	Not available	62 (-11)	6 (-)	N/A
- a/w data custodian	11 (+5)	Not available	26 (+3)	0 (-)	N/A
Submitted for IG approval	4 (-2)	8 (+3)	1 (+1)	0 (-)	13 (+2)
Approved but not yet active	2 (-)	0 (-)	0 (-)	2 (-)	4 (-)
Active research taking place	44 (+3)	37 (-)	69 (+3)	0 (-)	150 (+6)

3. Full metadata for Pillar 1, 2 & 3 testing in Wales is now available on health data research gateway.

4. NHS Digital new active research includes cross-sector data linkage: Assessing the impact of the COVID-19 pandemic on vulnerable children: the DHSC-ECHILD-COVID study that uses both health and education data and will be available via the ONS Secure Research Service TRE.

5. Cardiovascular consortium now accessing over 4 billion journals across population of over 56 million (93.5% England coverage) through NHS Digital TRE. NHSD has also published [management information on coverage](#) of the General Practice Extraction Service (GPES) Data for pandemic planning and research.

NOTES

- N/C – No change
- TRE - Trusted Research Environment
- IG - Information Governance
- DPN – Data Provision Notice
- CHESS - COVID-19 Hospitalisations in England Surveillance System
- SICSAG - Scottish Intensive Care Audit Steering Group
- HES – Hospital Episode Statistics
- SUS – Secondary Uses Service