

HDRUK
Health Data Research UK



HSC Public Health
Agency
Research and Development



Public Health
Scotland



SAIL DATABANK

UK Health Data
Research Alliance

Office for
National Statistics

COVID-19 Health Data Research

13 October 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

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
Pete Stokes, Office for National Statistics
Members of the HDR UK Public Advisory Board and COVID-19 PPIE Group



COVID-19 Health Data Research recommendations – 13 October 2020

Health data research insights on COVID are continuing, with 111 research questions, 175 (+17) projects active within the national data Trusted Research Environments (TRE), 159 (+1) in development, and 769 pre-print publications (reduced figure due to retracted papers) and 54 published papers.

Incremental progress across all recommendations.

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: Award made by UKRI/NIHR - proposal led by Philip Quinlan, Emily Jefferson and partners will commence 19 Oct.</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. Special Interest Group, to include UK Health Data Alliance Data Officers, set up and initial steps taken to review current landscape and issues around ethnicity coding, towards enacting the Alliance Board commitment to improve consistency and quality of ethnicity coding, enabling data use to increase the representativeness of research.</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: Initial reviews of the NHS Digital adult social care management coronavirus status data collection show rapidly increasing data quality and completeness across CASPA members. CASPA are exploring using an opt-out model to increase provider coverage. Ongoing work to understand how care providers can access and use the data to inform service provision and to identify driver research questions to further explore data utility. <i>Update from previous report:</i> Research insights are being facilitated on a local level through the use of integrated datasets however this is still challenging to achieve on a wider scale.</p>
<p>4 Accelerate access to restricted national datasets, since lack of availability is holding back crucial research.</p>	<p>PROGRESS: England Testing data and CHES available to priority studies to request but not yet via routine access requests and continued issues with missing data. Ongoing engagement with NHS Digital to understand and resolve issues.</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 is working with the national Trusted Research Environments to prioritise datasets and linkages to support national priority COVID-19 research questions. BHF Data Science Centre: <i>No change from previous report.</i> 19 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



Health data research on COVID-19 continues to grow, now reaching 769* (non peer-reviewed) pre-prints & 54 published papers


Topic	Insights from ongoing studies (links provide further details):
Surveillance	<ul style="list-style-type: none"> Analysis from the community survey, REACT-1, reports rapid growth in the pandemic in September 2020, which has led to high prevalence of SARS-CoV-2 virus in England among all regions and age groups, including those age groups at highest risk. Additional hospitalisations and mortality in coming weeks are predicted; re-doubling of public health efforts to return to a declining phase of the epidemic is suggested. A team has evaluated the sensitivity and specificity of individual/combinations of symptoms across different age groups on participants that reported data at least once in the week preceding their COVID-19 test. One potential policy change explored based on the findings would be that delirium should be added to the case definition in patients >85 yrs old; but for adults <85 yrs old they found that cough/fever/loss of smell remained the optimal definition in terms of balancing sensitivity and specificity.
Immunity	<ul style="list-style-type: none"> A study addresses a major issue in identifying immunity to COVID-19 - the ability to distinguish between people infected with SARS-CoV-2 from those with cross-reactive immunity generated by exposure to other coronaviruses. They found that some assays do not distinguish between immunity to COVID-19 vs other coronaviruses and identified a method to distinguish recent infection from pre-existing immunity in exposed populations.
Longitudinal health	<ul style="list-style-type: none"> The impact of the national implementation of social distancing for COVID-19 was assessed on mental health service activity at a large multi-team provider in London. In summary: all grouped liaison, home treatment and community mental health services had substantially reduced face-to-face patient contacts, with variable and only partial compensatory increases in non-face-to-face contacts; many had also reduced their caseloads and numbers of patients in inpatient care had also been substantially reduced; daily numbers of deaths in past and present patients had increased significantly over a relatively short time period.
Transmission	<ul style="list-style-type: none"> The 'Majora' data infrastructure is now developed and deployed, which underpins the UK's ability to rapidly sequence SARS-CoV-2 genomes as part of a national-scale genomic surveillance strategy. This has enabled the COG-UK consortium to rapidly become the leading contributor of SARS-CoV-2 genomes to international databases, generate over 60,000 sequences to date and yield insights into virus transmission. Given the presence of viral RNA in faeces in around 40% of infected persons, next generation sequencing (NGS) of sewage samples can be used to approximate the diversity of SARS-CoV-2 viruses circulating in a community. NGS from routine wastewater testing, was compared to the virus diversity in patients in Netherlands and Belgium. Viruses belonging to the most prevalent clades worldwide were present in both countries; however some clades were not identified in wastewater, while they were present in clinical samples during the same period. 57 unique mutations in the SARS-CoV-2 genome were found in wastewater which have not been described before.


*Reduced figure due to retracted papers

Patient and Public Voice Feedback 


Active patient and public engagement, particularly around care home data use is important. Further accelerated progress to obtain the right data for both mental health and social care research is also needed.

<p>21 COVID-19 taskforce calls with 91 clinical and health data research leaders engaged </p>	<p>1438 academic, industry and NHS participants in COVID-19 Slack channel with 10 sub-channels </p>
--	---

111 health data research questions identified – 42 prioritised 

769* 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 – 13 October 2020

Large increase in active research across three of the four nations (+17 to 175). No significant change to dataset availability. We are working with ONS to include dataset availability and research activity being enabled by the Secure Research Service.

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](#)
- [PHOSP-COVID](#)
- [COVID-19 symptom study](#)
- [GENOMICC](#)

1. Dataset availability within ONS Secure Research Service includes COVID Opinions Survey and Business Impact Survey (both linkable)

2. No significant change to other priority datasets

3. Improvement in dataset discoverability via Innovation Gateway (not shown in table)

Core COVID-19 Datasets available for linkage	Office for National Statistics Secure Research Service	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)
Primary Care	To be onboarded				
Pillar 1 COVID-19 Testing Data	Not yet linked				
Pillar 2 Testing data (UK Gov)	Not yet linked	N/C Delayed: Was expected Aug			Missing results prior to 26 Apr
Pillar 3 & 4 Testing data	Infection survey	N/C No confirmed date		Pillar 3 available	Data flow in place, no tests yet
Community Prescribing					
Critical Care (CHESS, ICNARC, SICSAG)		N/C CHESS expected in DARS Sep			N/C - Under review
Personal Demographic Service					
Secondary Care	HES	SUS only - not HES			
Death registry					
Census 2011					
Covid Opinions Survey					
Business Impact of Covid Survey					

4. Seventeen new projects now active, with a large reduction in number awaiting IG approval.

5. Small increase (+5) in overall number of projects across the pipeline

# of COVID-19 Projects by TRE stage (change from previous report)	Office for National Statistics Secure Research Service	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)	Total
In development	<i>to be added</i>	34 (-3)	34 (+5)	85 (-1)	6 (-)	159 (+1)
- a/w researcher	<i>for</i>	30 (+2)	<i>Not available</i>	58 (-5)	6 (-)	<i>N/A</i>
- a/w data custodian	<i>future</i>	4 (-5)	<i>Not available</i>	27 (+4)	0 (-)	<i>N/A</i>
Submitted for IG approval	<i>reports</i>	4 (-6)	5 (-4)	0 (-1)	0 (-)	9 (-11)
Approved but not yet active		3 (+2)	0 (-)	0 (-4)	2 (-)	5 (-2)
Active research taking place		56 (+6)	42 (+2)	77 (+9)	0 (-)	175 (+17)
<i>Active Number of Researchers</i>	<i>to</i>	<i>be</i>	<i>added</i>	<i>for</i>	<i>future</i>	<i>reports</i>
<i>Average time from application to active research</i>	<i>to</i>	<i>be</i>	<i>added</i>	<i>for</i>	<i>future</i>	<i>reports</i>

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

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