

**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

27 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  
Garry Coleman, 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 – 27 October 2020

Health data research insights on COVID are continuing, with 111 research questions, 183 (+8) projects active within the national data Trusted Research Environments (TRE), 157 (-2) in development, and 905 (+136) pre-print publications and 67 (+13) 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 <b>swab &amp; antibody testing programmes</b> 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><b>SAGE ACTION:</b> HDR UK to work with partners to plan and create a serology and testing data research asset that is linkable to other data sources. <b>PROGRESS:</b> Award made by UKRI/NIHR - proposal led by Philip Quinlan, Emily Jefferson and partners commenced on 19 Oct.</p>
<p>2 Further research, undertaken collaboratively with international partners where appropriate, should address why <b>BAME groups</b> 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><b>PROGRESS:</b> Ongoing work via a Special Interest Group, including UK Health Data Alliance Data Officers, 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. Engagement with SAGE Ethnicity sub-group.</p>
<p>3 Enhance data capture on patients and staff in <b>care homes</b>, 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><b>PROGRESS:</b> Initial reviews of the NHS Digital adult social care management coronavirus status data collection show rapidly increasing data quality and completeness across <a href="#">CASPA</a> 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. 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 <b>restricted national datasets</b>, since lack of availability is holding back crucial research.</p>	<p><b>PROGRESS:</b> England Testing data and CHESS available to priority studies to request but not yet via routine access requests and continued issues with missing data. NHS Digital developing testing data product that provides full dataset. Alternatives identified for CHESS (combination of ICNARC and SUS)</p>
<p>5 Commission large scale collaborative analyses of the <b>long-term impacts</b> 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><b>SAGE ACTION:</b> HDR UK to work with ONS and others to accelerate linkage of cross-sectoral datasets. <b>PROGRESS:</b> 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: Currently in the process of adding the NHS BSA Dispensed Medicines dataset to the NHS Digital TRE. <i>No change from previous report:</i> 19 cardiovascular analysts from 6 institutions actively working via access to linked datasets in the 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>

# Research questions with new insights generated in last 2 weeks

Health data research on COVID-19 continues to grow, now reaching 905 (+136) non peer-reviewed pre-prints & 67 (+13) published papers

Topic	Insights from ongoing studies (links provide further details):
<b>Surveillance</b>	<ul style="list-style-type: none"><li>Results from the 5<sup>th</sup> round of observations (between 18<sup>th</sup> Sept- 5<sup>th</sup> Oct) of the REACT-1 COVID-19 swab positivity study has shown a national prevalence of 0.60% and a doubling of virus every 29 days in England, corresponding to an estimated national R of 1.16. <u>This represents approximately 45,000 new infections each day. The highest prevalence is in the North West, Yorkshire and The Humber and the North East.</u></li><li><u>The ability of genomic epidemiology to inform the COVID-19 pandemic response</u> has been underpinned by the development and use of Majora (<a href="https://github.com/SamStudio8/majora/">https://github.com/SamStudio8/majora/</a>), a digital infrastructure to address the challenge of collecting &amp; integrating genomic sequencing data &amp; sample-associated metadata produced across the COG-UK network.</li></ul>
<b>Immunity</b>	<ul style="list-style-type: none"><li><u>Changes in innate immune function have been observed following recovery from COVID-19</u> indicating that immune modulating therapies targeting monocytes &amp; leucocyte migration may be useful in patients with persistent symptoms.</li><li>Markers of myeloid or endothelial cell activation are associated with severe, progressive, and fatal COVID-19 disease indicating a <u>central role for innate immune activation and vascular inflammation in COVID-19.</u></li></ul>
<b>Longitudinal health</b>	<ul style="list-style-type: none"><li>Research has suggested that <u>severe COVID-19 outcomes are likely to be elevated in cancer survivors</u>, particularly haematological cancer survivors, based on prevalence of comorbidities considered risk factors for COVID-19; and risk of severe influenza, as a marker of susceptibility to severe outcomes from epidemic respiratory viruses.</li><li><u>In a young, low-risk population with ongoing symptoms, almost 70% of individuals have impairment in one or more organs four months after initial symptoms of COVID-19</u> – assessed through questionnaires, blood investigations and quantitative magnetic resonance imaging.</li><li>From <u>4182 cases of COVID-19 who logged their symptoms in the COVID Symptom Study app: 13.3% had symptoms lasting &gt;28 days, 4.5% for &gt;8 weeks and 2.3% for &gt;12 weeks.</u> Such “long-COVID” cases were associated with increasing age, BMI and female sex and experiencing more than five symptoms during the first week of illness.</li><li>Data from the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C) study has been used to create the <u>4C Deterioration model</u>, which is able to predict clinical deterioration among adults hospitalised with COVID-19 to inform clinical decision making.</li></ul>
<b>Transmission</b>	<ul style="list-style-type: none"><li>Modelling suggests that <u>in the absence of interventions 16% (2% - 38%) of UK university students could be infected with COVID-19 during the autumn 2020 term.</u> Full adherence to test, trace and isolate measures, were found to lower cumulative infection estimates to 1.4% (0.4% - 5%). Room isolation generated minimal benefits. A one-off instance of mass testing did not drastically reduce the term-long case load or end-of-term prevalence, but regular weekly or fortnightly testing could reduce both measures by &gt;50%, suggesting that adherence throughout the term would suppress unwitting asymptomatic transmission to family and community members at the end of term.</li></ul>
<b>Clinical Trials</b>	<ul style="list-style-type: none"><li>Ongoing progress, but no new research insights published in the last 2 weeks.</li></ul>

## Patient and Public Voice Feedback

Priority area of focus should be to enable & accelerate access to complete datasets on domiciliary (at home) care & further research is needed to better understand the impact of health inequalities, poor mental health and "Long Covid"



**22** COVID-19 taskforce calls with **96** clinical and health data research leaders engaged



**1454** academic, industry and NHS participants in COVID-19 Slack channel with 10 sub-channels



**111** health data research questions identified – 42 prioritised



**905** 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 – 27 October 2020

Recent BMJ Open publication provides excellent explanation of approach taken through SAIL to support Welsh COVID response.

Progress on dataset availability increasingly being captured and shared via Innovation Gateway, and continued shift to active projects.

1. The protocol of the data linkage work undertaken in Wales (through SAIL) to support COVID-19 response has been published in [BMJ Open](#)

2. Further improvements in dataset discoverability via [Innovation Gateway](#) including first ONS and HSC NI datasets

3. Eight new projects now active with small reduction 'in development'

4. Northern Ireland Honest Broker Service physical Data Safe Haven has had to be closed as part of COVID-19 infection control measures.

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	Being used in SRS				
Pillar 2 Testing data (UK Gov)	Being used in SRS	Introduction Ongoing			Missing results prior to 26 Apr
Pillar 3 & 4 Testing data	Infection survey	Pillar 3 - Introduction Ongoing		Pillar 3 available	Data flow in place, no tests yet
Community Prescribing					
Critical Care (CHESS, ICNARC, SICSAG)		Available as an extract			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					

**KEY**

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

**KEY UK WIDE PROJECTS:**

- [RECOVERY](#)
- [PRINCIPLE](#)
- [CO-CIN \(ISARIC 4C\)](#)
- [COG-UK](#)
- [CARDIOVASCULAR CONSORTIUM](#)
- [PHOSP-COVID](#)
- [COVID-19 symptom study](#)
- [GENOMICC](#)

# 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>	30 (-4)	34 (-)	87 (+2)	6 (-)	157 (-2)
- a/w researcher	<i>for</i>	18 (-12)	<i>Not available</i>	61 (+3)	6 (-)	N/A
- a/w data custodian	<i>future</i>	12 (+8)	<i>Not available</i>	26 (-1)	0 (-)	N/A
Submitted for IG approval	<i>reports</i>	6 (+2)	5 (-)	0 (-)	0 (-)	11 (+2)
Approved but not yet active		2 (-1)	0 (-)	0 (-)	2 (-)	4 (-1)
<b>Active research taking place</b>		<b>60 (+4)</b>	<b>46 (+4)</b>	<b>77 (-)</b>	<b>0 (-)</b>	<b>183 (+8)</b>
<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