COVID-19 Health Data Research

18 August 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 COVID-19 Patient and Public Group
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 – 18 August 2020

Health data research insights on COVID are continuing, with 110 research questions, 138 (+7) projects active within the national data Trusted Research Environments (TRE), a further 173 (-) in development, and 624 pre-print publications and 36 published papers. Some but limited progress across recommendations 1-3. Actions related to the five recommendations 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
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.	**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:** Iteration ongoing with UKRI/NIHR panel following feedback requesting a more focused resubmission from the collaborative team.
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.	**PROGRESS:** Further insights being generated (see next page).
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.	**PROGRESS**: *No change to previous report:* NHS Digital have released a Data Provision Notice of the <u>adult social care management system coronavirus (COVID-19) status collection.</u> Data quality of initial submissions will be reviewed before confirming 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 for the domiciliary care (care within own home) sector.
4	Accelerate access to **restricted national datasets**, since lack of availability is holding back crucial research.	**PROGRESS**: Priority studies can request access but access via the standard NHS Digital Data Access Request Service (DARS) expected in Aug 20 (Testing data and Community Prescribing) and Sep 20 (CHESS). This recommendation will be integrated into the national core studies data and connectivity work from September.
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.	**SAGE ACTION: HDR UK** to work with ONS and others to accelerate linkage of cross-sectoral datasets. **PROGRESS:** HDR UK is working 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. BHF Data Science Centre: 17 cardiovascular analysts from 6 institutions actively working via access to linked datasets in the NHS Digital TRE, recently including the primary care GPES dataset.

Priority research questions with new insights generated this week – 18 August 2020

Health data research on COVID-19 continues to grow, now reaching 624 (non peer-reviewed) pre-prints[1] & 36 published papers

[1] The number of pre-prints has increased significantly, reflecting a change in methodology with additional use of new pre-print sources with more historical data.

HDRUK
Health Data Research UK

Priority research questions	Insights from ongoing studies (links provide further details):
1. Understanding & protecting vulnerable groups? (RQ 22, 32, 36, 62, 102)	• People living with HIV had nearly three-fold higher risk of COVID-19 death after adjusting for age, sex, deprivation, ethnicity and comorbidities, with some evidence the association was larger among people of black ethnicity. • Research using 2011 Census plus NHS Patient Register records has confirmed a substantially elevated risk of COVID-19 mortality in ethnic minority groups, although this was largely (but not entirely) attenuated by geographical and socio-economic factors including living circumstances, socio-economic factors, occupational exposure, and health status. • A study of healthcare workers & their households has shown that whilst the absolute risk of hospitalisation was low, patient facing healthcare workers and their households had 3- and 2-fold increased risks of COVID-19 hospitalisation. • The higher risk of infection seen in BAME and healthcare workers may explain their increased risk of severe COVID-19. 6% of the population were shown to have antibodies for the virus by the end of June, an estimated 3.4 million people. London had the highest number (13 per cent), while the South West had the lowest (3 per cent). London has also been shown to have increased infection intensity. • Research using UK Biobank data has shown a protective association between physical activity and adverse COVID-19 outcomes independent of age, sex, measures of obesity, and smoking status. • Research also using UK Biobank data linked with national routine data has shown that pre-pandemic psychiatric disorders are associated with increased risk of COVID-19, especially severe and fatal COVID-19. • A study using high geographical precision has shown some evidence of association between exposure to NO2 air pollution & COVID-19 mortality - lower than observed in previous studies. The role of PM2·5 remains more uncertain. • COVID-19 patients, particularly men >60 yrs old with comorbidities (particularly hypertension, diabetes, cardiovascular and cerebrovascular disease) have a higher risk of fatality in East Kalimantan, Indonesia. • COVID-19 mortality rate in patients admitted to Swedish intensive care units appears to be driven by age, baseline disease severity, the degree of organ failure and ICU treatment, rather than preexisting comorbidities.
2. Impact of lockdown release (RQ22)	• 77% of parents of older children (>3yrs) are worried about their child returning to school. 87% of parents with children aged <3 years are concerned about the impact of their child missing social experiences and 73% are worried about when/if their child would return to nursery or school.
3. Understanding immunity, testing reliability & prevalence (R01, 50, 95, 102, 51, 54, 55, 104)	• Modelling using anonymised Google Community Mobility Datasets has shown that it provides a promising low-lag control signal that can be used to infer the spread of COVID-19. • A study of 6 serological assays has demonstrated that even while high specificity can be achieved, more than one test may be necessary to rule out false positives by initial tests, such has been the rule for HIV and hepatitis. A study testing 11 commercial serological tests has confirmed that test sensitivity is dependent on time since symptoms onset & that combined detection leads to better performance than measurement of a single antibody class, regardless of time range. • Modelling has suggested that the higher predisposition of SARS-CoV-2 towards super-spreading events is not due to its additional viral shedding. Rather, a person infected exposes more people within close physical contact, supporting policies that limit crowd size in indoor spaces.

Patient and Public Voice Feedback
There is a need to delve deeper, including around testing, to ensure we better understand risks, outcomes and long-term effects to help us better protect and support populations.

17 COVID-19 taskforce calls with **85** clinical and health data research leaders engaged

1374 academic, industry and NHS participants in COVID-19 Slack channel with 10 sub-channels

110 health data research questions identified – 42 prioritised

624 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 – 18 August 2020

Number of active research projects continues to increase but rate of progress slowing with majority of 'in development' projects awaiting researcher action to progress.

HDR UK — Health Data Research UK

1. **Pillar 3 antibody testing data now expected to be available in Northern Ireland.**

2. **ISARIC CCP data for Scottish Participants available in National Data Safe Haven.**

3. **England HES linkage issues require resolution.**

4. **Seven further research projects now active across the four national TREs.**

5. **Majority of 'in development' projects are awaiting researcher action to progress.**

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	Fully available	Fully available	Fully available	Data flows agreed but not yet available
Pillar 1 COVID-19 Testing Data (NHS/Public Health)	Fully available	Fully available	Fully available	Fully available
Pillar 2 Testing data (UK Gov)	N/C Expected Aug	Fully available	Fully available	Missing results prior to 26 Apr
Pillar 3 & 4 Testing data	No confirmed data	Data flows specified but not yet agreed	Pillar 3 available	Pillar 3 expected within month
Community Prescribing	N/C Expected Aug	Fully available	Fully available	Fully available
Critical Care (CHESS, ICNARC, SICSAG)	N/C CHESS expected in DARS Sep	Fully available	Fully available	N/C - Under review
Personal Demographic Service	Fully available	Fully available	Fully available	Fully available
Secondary Care	SUS only - not HES	Fully available	Fully available	Fully available
Death registry	Fully available	Fully available	Fully available	Fully available

# 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	37 (-2)	35 (-2)	95 (+6)	6 (-)	173 (-)
- a/w researcher	31 (+7)	Not available	72 (+6)	6 (-)	N/A
- a/w data custodian	6 (-9)	Not available	23 (-)	0 (-)	N/A
Submitted for IG approval	5 (-3)	6 (+1)	0 (-)	0 (-)	11 (-2)
Approved but not yet active	2 (-)	0 (-)	0 (-)	2 (+2)	4 (+2)
Active research taking place	**38 (+1)**	**34 (+6)**	**66 (-)**	**0 (-)**	**138 (+7)**

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

6. **Honest Broker Service physical Safe Haven (NI) has re-opened** with restricted access for researchers. Pilot project using the SeRP platform in partnership with HDR-UK Wales/NI and training in the use of LinXmart software has started.

7. **SAIL Databank report that requests for new emergency projects for NHS/government with short turnaround have largely stopped.** Shift to academic groups requesting access for longer term studies which are going through SAIL's normal scoping and governance application processes. August holiday period slowed progress down with no additional active studies.

8. Two new publications for **research enabled by data linkage in Scotland Data Safe Haven**: Risk of hospitalisation with coronavirus disease 2019 in healthcare workers and their households: a nationwide linkage cohort study and COVID-19 – exploring the implications of long-term condition type and extent of multimorbidity on years of life lost: a modelling study . Further details on these publications can be found in slide 3, section 1.

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