# COVID-19 Health Data Research

09 June 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
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

# COVID-19 Health Data Research recommendations – 09 June 2020

**HDRUK**
Health Data Research UK

**The volume of health data research on direct and indirect effects of COVID continues to grow rapidly, fuelled by improvements to dataset availability, particularly where mature research data infrastructures are established (e.g. Wales, CPRD & Discover-NOW). In contrast, in areas without an established approach to making national data available for research (e.g. in social care) progress is slowed as the effort is going into establishing new linkages, collaborations etc.**

As this is a weekly report, our recommendations are largely unchanged week on week, changes from last week are underlined below. We are now moving to fortnightly reporting and will next report on the 23rd June.

1. Ensure data gathered from all swab & antibody testing programmes can be securely linked and used for research. Including directly supporting collaborative programmes that build on unparalleled cooperation across all four nations between NHS organisations, PHE, data custodians, academic endeavours, and technology partners, whilst building public trust.

2. Ensure that further research, undertaken collaboratively with international partners where appropriate, addresses why BAME groups appear to have a higher rate of severe COVID-19 outcomes.  Including understanding whether BAME groups are more likely to contract COVID-19 and/or have an increased risk of severe outcomes once infected.  This will help to target the best interventions and inform the response to future public health crisis.

3. Enhance data capture on patients and staff in care homes, in particular interconnections between settings, to enable in-depth research on health, transmission and outcomes. Provide clarity on appropriate use of national Trusted Research Environments for consolidation of relevant care home COVID-19 data.

4. Accelerate access to currently restricted national datasets, including CHESS* –  this important data will not be fully available to researchers via the NHS Digital Data Access Request Service until the 22nd June, holding back research.

5. Commission large scale analyses of the long-term impacts of health and social care changes during the COVID-19 lockdown on major diseases, involving researchers, frontline clinical teams and disease registry experts. 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 on all vulnerable populations.

\* CHESS - COVID-19 Hospitalisation in England Surveillance System

# Health data research on COVID-19 continues to grow, now reaching 100 pre-print publications

## Priority research questions with new insights generated this week – 09 June 2020

**HDR**UK
Health Data Research UK

| Priority research questions | Insights from ongoing studies (links provide further details): | SAGE Recommendation |
|---|---|---|
| **1. How do we understand population immunity & improve testing reliability** (Immunology & seroprevalence R01, 50) | • Pilot completed of a new, more rapid, PCR test for COVID-19 that can be used at the point of care. Shorter time from test to result allows **more rapid triage and patient movement** to safe and appropriate isolation wards. | Ensure data gathered from all swab & antibody testing programmes can be securely linked and used for research. This will require unparalleled cooperation across all four nations between NHS organisations, PHE, data custodians, academic endeavours, and technology partners, whilst maintaining public trust. |
| **2. Why do BAME groups have an increased risk of severe COVID-19 outcomes** (RQ34)? | • Largest study to date (on 73,000 patients) on link between ethnicity and COVID-19 in UK, found that, compared to White COVID-19 patients, those with **BAME** background: <br>   • Have higher rates of COVID-19 prevalence at least **partially due to where they live, deprivation and occupational exposure** <br>   • Are **younger**, and have a **higher burden of comorbidity, particularly cardiovascular and endocrine diseases.** | Ensure that further research, undertaken with international partners if possible, addresses <u>why</u> BAME groups appear to have a higher rate of severe COVID-19 outcomes. Including understanding whether BAME groups are more likely to contract COVID-19 and/or have an increased risk of severe outcomes once infected. <br> Ensure that large representative datasets (e.g. CHESS as used in this study) are made fully available to researchers as soon as possible to enable further rapid insights. |
| **3. How do we best understand and protect vulnerable groups?** (RQ 22, 32, 36, 62) & inform an effective phased lockdown release: <br> - Risk prediction <br> - Social & mental health | • Severe COVID-19 is strongly associated with past medical history across all age groups, as determined by a new, more robust, risk classifier – enables **more accurate ID of individuals most in need of shielding until epidemic is over.** <br> • Emerging evidence that an increase in domestic abuse related to lockdown conditions, highlighting an urgent need for **linkage of datasets between police and health records datasets to identify individuals at risk**. <br> • Modelling has shown that reopening of schools will result in increased mixing and infection amongst children and the wider population, although the opening of schools alone is unlikely to push the value of R above one. However caveats to this exist, such as regions being closer to the critical threshold that would lead to a growth in cases, and combined impact of all lockdown changes. <br> • Diabetes is an independent prognostic factor for mortality in people with COVID-19 requiring HDU or ICU treatment. | Further develop, extend and utilise open "risk calculators", symptom trackers and surveys, to better communicate risk and more targeted public health messaging and actions. <br><br> Directly endorse the use of trusted research environments, to enable an open approach to health research data access safely and securely. |
| **4. RECOVERY Trial** (RQ18) | • It has been concluded that hydroxychloroquine is ineffective in the treatment of hospitalised patients with COVID-19. Further details can be found here. | The pace and scale demonstrated by RECOVERY trial has been enabled by the NHS DigiTrials Health Data Hub. It is recommended that timely data linkage is facilitated and replicated in all other clinical studies. |

**12** COVID-19 weekly taskforce calls with **66** clinical and health data research leaders engaged

**1322** academic, industry and NHS participants in COVID-19 Slack channel with 10 sub-channels

**62** volunteers in HDR UK's COVID-19 Public & Patient Group

**106** health data research questions identified – 36 prioritised

**100** COVID-19 pre-print publications

Click here for a link to the full prioritised list of questions, status, and prioritisation process

# SAIL Databank (Wales) leading the way across all dimensions, highlighting the benefits of having mature infrastructure in place from the outset.

## COVID-19 dataset availability and status of projects using the data – 09 June 2020

1. **Progress with Pillar 2 testing data but set back within England in terms of critical care data.**

2. **Risk assessment of restricted re-opening of Northern Ireland Safe Haven to be carried out this month.**

| 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 | Available 15-Jun | Approval process agreed | | |
| Pillar 1 COVID-19 Testing Data (NHS/Public Health) | | | | |
| Pillar 2 Testing data (UK Gov) | Expected 22-Jun | Linkage work underway | Available by 15 Jun | |
| Community Prescribing | Expected July | | | |
| Critical Care (CHESS, ICNARC, SICSAG) | NHSD contracted staff only | | | Options under review |
| Personal Demographic Service | | | | |
| Secondary Care | | | | |
| Death registry | | | | |
| **Pillar 3 Testing data - antibody** | Data flows being specified across all 4 nations | | | |

3. **SAIL databank has around two-thirds of all active research across the 4 national TREs. Almost 50% of SAIL pipeline active (2 x the average rate across Eng, Sco and NI)**

| # of COVID-19 Projects by TRE stage (change from previous week) | England (NHS Digital Data Processing Service) | Scotland (National Data Safe Haven) | Wales (SAIL Databank) | Northern Ireland (Honest Broker Service) | Total |
|---|---|---|---|---|---|
| In development | 44 (+5) | 30 (+3) | 60 (-2) | 3 (-) | 136 (+6) |
| Submitted for IG approval | 5 (-1) | 5 (+2) | 0 (-) | 0 (-) | 10 (+1) |
| Approved but not yet active | 2 (-1) | 0 (-) | 1 (-2) | 0 (-) | 3 (-3) |
| **Active research taking place** | **10 (+2)** | **19 (-)** | **57 (+7)** | **0 (-)** | **86 (+9)** |

## 4. Significant active research is taking place using data assets complementary to the national capability

- Eleven approved research projects using **CPRD** resources, including latest Clinical contact with health services for mental illness and self-harm before, during and after the COVID-19 pandemic. See full list here.
- **Discover-NOW**, the health data research hub for real world evidence, is supporting projects to access North West London COVID-19 data repository. Latest request is for data to support BAME analysis by ONS. Full list of 18 projects (as at 8 June) available via the tracker.

**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

**NOTES**

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