COVID-19 Health Data Research

27 May 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
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 – 27 May 2020

Insights from health data have further accelerated, fuelled by new data sources made accessible in safe environments. In particular, there has been a big increase in requests for symptom tracker data and additional intensive care data has been made available in Wales and progress with primary care data in England. However, national testing data remains an issue.

Recommendations for SAGE based on current health data research insights:

1. (NEW) Ensure data generated from all swab & antibody testing programmes can be securely linked and used for research. This will require **unparalleled cooperation between NHS organisations, PHE, data custodians, academic endeavours, technology partners, whilst maintaining public trust.**

2. Commission independent meta-analysis on ethnicity analyses within the UK and with international studies and ensure that further research is undertaken to understand why BAME groups appear to have a higher rate of severe COVID-19 outcomes. This should be undertaken collaboratively with international partners as much as possible.

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

4. Further develop, extend and utilise open “risk calculators”, symptom trackers (e.g. ZOE app) and surveys, integrated with targeted public health messaging and actions

5. Commission meta-analysis on outcomes across major disease groups to assess the long-term impacts of health and social care changes during the COVID-19 lockdown period, compared with previous 5 years, including research, frontline clinical teams and existing disease registry experts. Create COVID-19 registries for the four nations to provide an ongoing source of data on COVID-19 patients, akin to the national cancer registry. And ensure that government cross-departmental data linkage beyond health is enabled to understand the wider impacts of COVID-19 on all vulnerable populations.
We are seeing rapid growth in the research generated from health data, with a 50% increase in pre-print publications over 2 weeks (from 57 to 86)

Priority research questions, studies & insights – 27 May 2020

<table>
<thead>
<tr>
<th>Priority research questions</th>
<th>Studies already working on this:</th>
<th>Insights</th>
<th>SAGE Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How do we best understand and monitor population immunity to COVID-19? (Immunology &amp; seroprevalence RQ1, 50)</td>
<td>University Hospital Birmingham NHS Foundation Trust (UHB)</td>
<td>Asymptomatic health care workers were tested for current infection and for COVID specific antibodies in their blood (seroconversion) to determine whether they now have immunity. Asymptomatic infection led to variable seroconversion rates depending on department and role of the worker (e.g. high rates in housekeeping and acute medicine; lowest in intensive care &amp; emergency medicine). Several seroprevalence studies are underway, recruiting participants from existing research cohorts, social and health care workers as well studies directed by Public Health England. However, fragmentation of data collection and availability for research analysis remains a significant challenge.</td>
<td>Ensure data generated from all swab &amp; antibody testing programmes can be securely linked and used for research. This will require extensive cooperation and coordination between NHS organisations, PHE, data custodians, academia and technology partners, whilst maintaining public trust. Ensuring the right data is collected is essential.</td>
</tr>
<tr>
<td>2. Why do BAME groups have an increased risk of severe COVID-19 outcomes (RQ34)?</td>
<td>Uni Hospitals Birmingham, Pioneer Hub &amp; DECODIV</td>
<td>Further research has shown that hospitalised COVID-19 patients of BAME background are on average much younger than white patients and have more comorbidities such as higher levels of diabetes and high blood pressure.</td>
<td>Further research, ideally with international partners, to understand why BAME groups appear to have a higher rate of severe COVID-19 outcomes, should be prioritised.</td>
</tr>
<tr>
<td>3. What impact has COVID-19 having on care home patients? (RQ63 et al)</td>
<td>HDR UK North – University of Sheffield &amp; Lancaster</td>
<td>Core insights require the identification of care home residents in routinely collected data. Work is being accelerated to increase the sensitivity of an algorithm which uses address matching and core health data to identify the population, and to validate the algorithm across geographies. Analysis of care home residents in Wales (using the SAIL Databank) found that deaths more than doubled during lockdown, from 4% (2019) to 9% (2020). Ongoing research to analyse patterns of secondary care referral, patient management – cohort, testing, impact on non-COVID care and outcomes.</td>
<td>Enhance data capture on patients and staff in care homes with a focus on resident discharge and referral patterns and staff working patterns. Ensure alignment with existing datasets, providing clarity on appropriate TRE for consolidation of relevant care home COVID-19 datasets. Better communication of individual risk will be vital to the ongoing public health messaging. Recommend that openly available online risk based tools (such as OurRiskCoV) are further developed &amp; integrated into targeted public health messaging.</td>
</tr>
<tr>
<td>4. How do we best understand and protect vulnerable groups? (RQ22, 32, 36, 62) &amp; inform a effective phased lockdown release: - Risk prediction - Social &amp; mental health - Vulnerable groups</td>
<td>UCL, HDR UK London, University of Edinburgh, HDR UK Scotland, KCL, HDR UK London</td>
<td>Analysis of data from children with suspected COVID-19, including those with underlying health conditions, found no significantly increased risk of either contracting COVID-19 or severe complications, apart from those undergoing chemotherapy. Susceptibility to and transmission of COVID-19 amongst children and adolescents, is 56% lower than adults. There is limited evidence that children and young people play a lesser role in transmission of SARS-CoV-2.</td>
<td>Enhance data capture on patients and staff in care homes with a focus on resident discharge and referral patterns and staff working patterns. Ensure alignment with existing datasets, providing clarity on appropriate TRE for consolidation of relevant care home COVID-19 datasets. Better communication of individual risk will be vital to the ongoing public health messaging. Recommend that openly available online risk based tools (such as OurRiskCoV) are further developed &amp; integrated into targeted public health messaging.</td>
</tr>
</tbody>
</table>

10 COVID-19 weekly taskforce calls with 67 clinical and health data research leaders engaged

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

99 health data research questions identified – 32 prioritised

86 COVID-19 pre-print publications
Active research has doubled in the last 2 weeks. Priorities to scale-up data use are unchanged: 1) National testing data 2) continue moving projects from development to IG scrutiny

COVID-19 dataset availability and status of projects using the data – 27 May 2020

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Care</td>
<td>First extract 29/5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pillar 1 COVID-19 Testing Data (NHS/Public Health)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pillar 2 Testing data (UK Gov)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community Prescribing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical Care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Demographic Service</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary Care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death registry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Pillar 3 Testing data - antibody            | Data flows being specified across all 4 nations |

1. Increasing research interest in the national antibody testing programme highlights need to include future data flows as a core dataset available for linkage.

2. There has been a 29% increase in active research taking place via the national TREs. The proportion of projects ‘in development’ is virtually unchanged at 68% of total.

3. The Cardiovascular Consortium is working closely with the national data custodians in preparation for demonstrating the benefits of research at UK-wide scale

- Working with NHS Digital to help to shape the specification for the interim GP data flows for England and prepare for the initial data curation activity once the first extract 'lands' on 28 May.
- Sourcing hospital level data to undertake analysis of COVID-19 on cardiovascular activity to provide early insights similar to work undertaken by DATA-CAN, the health data research hub for Cancer.