Data & Connectivity National Core Study

Sprint 1: 5th October – 4th November, 2020

OUTCOMES

What were our aims for Sprint 1?

1. To make **core datasets relevant to COVID response available** to approved researchers
    - Pillar 1 COVID antigen testing data
    - Pillar 2 COVID antigen testing data
    - Primary care data
    - Secondary care data
    - Mortality
    - C-19 Infection Survey, CO-CIN (stretch goal)

2. Data to be made available in **secure cloud-based national Trusted Research Environment (TREs)** in England, Northern Ireland, Scotland, Wales

3. Evidence to show that both the **technical and governance requirements for linkage exist**

4. All above datasets to be **discoverable to researchers through a single "shop window",** the Health Data Research Innovation Gateway

What has been achieved in Sprint 1?

Supporting National Core Studies

1. *National Core Study use cases supported*

    a. **Clinical Trials Therapeutics NCS:** The PRINCIPLE Clinical Trial, led by Chris Butler, has received dedicated support to enable **daily access to positive C-19 test results within 24 hours to increase recruitment.**

        This support has included
        - daily problem solving and management scrums
        - in-depth engagement with the ICO
        - a public survey with over 90 responses
        - engagement with multiple data custodians
        - review by NHS Digital's IGARD

        - NHS Digital's Trial team are now addressing the remaining actions to support data flow

        Provides confidence in concept for future trials that require incident COVID cases

2. Earning Trust - Public, patient and practitioner involvement and engagement (PPPIE)

    a. A consultation on the Data and Connectivity aims, objects and delivery approach has been completed across seven national and **UK-wide patient and public networks** and collected **168 responses**. This feedback has been used to inform ongoing programme development and identify areas for more detailed follow-up.

    b. **Three public contributors** have been involved in regular sprint planning, delivery group meetings and planning and provided oversight of PPPIE activities.

    c. Consultation with **92 people** to inform decisions on methods and process for clinical trial recruitment (see use cases above).

Making Data Discoverable

3. *Data has been made available* during the sprint:

    a. Linked **Census/mortality data**, and **C-19 infection survey data**, available in the Office of National Statistics (ONS) Secure Research Service.

    b. **Test and trace data** from DHSC available on internal ONS systems.

    c. Remote access to NHS Digital secure data environment, providing access to **mortality and Hospital Episode Statistics (HES) data** alongside NHS Digital's General Practice data for 55M residents of England accessible for ONS internal users and available to request access via the [Gateway.](#).

    d. Approval for **linkage of primary and secondary health data for England to CO-CIN (ISARIC 4C)** has been established with data flows from NHS Digital expected to commence by early November.

    e. Continued provision of **extensive national linked data assets for Wales and Scotland** through respective delivery partners.

4. *New linkages that have been created* during the sprint

    a. ONS owned Census 2011 for England and Wales linked to:

        • **ONS owned mortality data for England and Wales**, 2011 to most recent data available

        • **HES data for England from NHS Digital**, 2017/18 to most recent data available

5. *Enhancements to meta-data*, making the data discoverable and useful for researchers

    a. **32 National Core Study priority datasets are now listed on the Gateway** with detailed technical metadata for 22

    b. Datasets collected by the **ONS** and **Northern Ireland Honest Broker Service** are listed on the Gateway, making administrative data and data from across the four nations discoverable to researchers in one place for the first time.

Enabling Data Access

6. **Progress with** *streamlined access management* across datasets

    a. A **Data Access Request form** aligned across all TRE delivery partners has been developed and undergone User Acceptance Testing by all delivery partners.

    b. Current work is focused on confirming requirements to **integrate the form** into the existing systems used by TRE delivery partners to allow live deployment.

    c. **Northern Ireland** have completed a pilot of the SAIL remote access TRE platform and are now in a period of stakeholder consultation to enable next steps.

Establishing the Data and Connectivity National Core Study

7. *Rapid turnaround of contract arrangements*

    a. Five delivery partners have joined the study and participated in Sprint 1. Each delivery partner offers a TRE for research.

    b. Each delivery partner has received a Letter of Intent and draft contracts setting out requirements, ways of working and funding.

    c. Award letter received from UKRI and assurance through GDS commenced.

8. *Building and connecting a delivery community*

    a. Established a regular schedule of **weekly delivery group meetings**.

    b. **>40 attendees from across study delivery partners**, (ONS, Health & Social Care Northern Ireland, NHS Digital, University of Swansea, SAIL Databank, Public Health Scotland, HDR UK / PA Consulting – Health Data Gateway); National Core Study teams and wider system stakeholders (HDR UK Public Advisory Board, ADR UK, JBC, ICO, CO-CONNECT, BHF Cardiovascular Data Science Centre, Genomics England, COG UK, OpenSAFELY, Wellcome Trust).

    c. Meeting materials circulated to **75 attendees and broader parties** to ensure connectivity and awareness.

    d. **Regular interaction** with all NCS teams through delivery group meetings and bilateral touchpoints, to align Data and Connectivity study with needs of all NCS.

What we have learnt for next Sprints and key risks to delivery?

1. **We can achieve a lot in 4 weeks** by working together across a national network of delivery partners.

2. To realise the benefits of the positive progress with the data assets, we **now need to have researchers set up** in the Trusted Research Environments with **research questions** that use this data.

3. The challenge is to communicate the available data asset and **mobilise the research community to respond.**

4. **A very clear overarching NCS narrative** that articulates the public outcomes and benefits that the NCS are collectively working to deliver by 31 March 2021 **is essential**. This will help to guide **prioritisation of linkages, data flows and use** cases.

5. **Legal barriers to data sharing**, and the timeframe required to resolve these, risks our ability to achieve sprint deliverables at pace. We have made important progress on "proof of concept" activities, enabling data sharing from NHS Digital to ISARIC for linkage to the unconsented CO-CIN cohort and supporting recruitment to the PRINCIPLE clinical trial by recontact individual post-COVID testing, however IG and legislative barriers will continue to be monitored and escalated to the Oversight Committee where necessary.

Outline for Sprint 2

Annex: Current Status of Health Data Research (SAGE Report 27th October 2020)

Openly available here

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):
Surveillance	• Results from the 5th round of observations (between 18th Sept- 5th 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. 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. • The ability of genomic epidemiology to inform the COVID-19 pandemic response has been underpinned by the development and use of Majora ((https://github.com/SamStudio8/majora/), a digital infrastructure to address the challenge of collecting & integrating genomic sequencing data & sample-associated metadata produced across the COG-UK network.
Immunity	• Changes in innate immune function have been observed following recovery from COVID-19 indicating that immune modulating therapies targeting monocytes & leucocyte migration may be useful in patients with persistent symptoms. • Markers of myeloid or endothelial cell activation are associated with severe, progressive, and fatal COVID-19 disease indicating a central role for innate immune activation and vascular inflammation in COVID-19.
Longitudinal health	• Research has suggested that severe COVID-19 outcomes are likely to be elevated in cancer survivors, 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. • 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 – assessed through questionnaires, blood investigations and quantitative magnetic resonance imaging. • From 4182 cases of COVID-19 who logged their symptoms in the COVID Symptom Study app: 13.3% had symptoms lasting >28 days, 4.5% for >8 weeks and 2.3% for >12 weeks. 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. • Data from the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C) study has been used to create the 4C Deterioration model, which is able to predict clinical deterioration among adults hospitalised with COVID-19 to inform clinical decision making.
Transmission	• Modelling suggests that in the absence of interventions 16% (2% - 38%) of UK university students could be infected with COVID-19 during the autumn 2020 term. 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 >50%, suggesting that adherence throughout the term would supress unwitting asymptomatic transmission to family and community members at the end of term.
Clinical Trials	• Ongoing progress, but no new research insights published in the last 2 weeks.

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

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

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.

# 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	to be added	30 (-4)	34 (-)	87 (+2)	6 (-)	157 (-2)
- a/w researcher	for	18 (-12)	Not available	61 (+3)	6 (-)	N/A
- a/w data custodian	future	12 (+8)	Not available	26 (-1)	0 (-)	N/A
Submitted for IG approval	reports	6 (+2)	5 (-)	0 (-)	0 (-)	11 (+2)
Approved but not yet active		2 (-1)	0 (-)	0 (-)	2 (-)	4 (-1)
Active research taking place		60 (+4)	46 (+4)	77 (-)	0 (-)	183 (+8)
Active Number of Researchers	to	be	added	for	future	reports
Average time from application to active research	to	be	added	for	future	reports

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