Data & Connectivity National Core Study

Sprint 5: 4 February – 3 March 2021

OUTCOMES

What were our aims for Sprint 5?

1. To respond to National Core Study (NCS) emerging data and connectivity needs:
   - **Vaccine research data infrastructure** – national Minimum Vaccine Dataset onboarded and available for approved researcher use within Trusted Research Environments (TREs) across the four nations. Access to rapid feed acute care data for vaccine research.
   - After significant work to address all the Information Governance requirements, **the ONS and NHS Digital joint health data asset** is now available to access via the Innovation Gateway**.** This is the first time that administrative and health data has been available as a joint data asset for research.
   - **Genomic data and other key clinical data assets:** including enabling greater access and linkage to genomic data (e.g COG-UK) and COVID 19 clinical characterisation data (ISARIC 4C).
   - **Conduct data and connectivity mapping exercise** across **National Core Studies, NIHR Urgent Public Health studies and SAGE** sub-groups such as SPI-M.

2. To **engage researchers to use and enrich the Data and Connectivity data infrastructure**:
   - Supporting successful researchers from the joint HDR UK, ONS and UKRI rapid funding call apply for data access and develop patient and public involvement and engagement (PPIE) plans for their COVID-19 research projects via the HDRUK Innovation Gateway
   - Gather initial researcher and TRE **user feedback questionnaire** on ease of use of the Gateway and Data Access Requests process to inform current and future improvements.

3. To make **core datasets relevant to COVID response available** to approved researchers in **secure cloud-based Trusted Research Environments (TREs)** in England, Northern Ireland, Scotland, and Wales:
   - Continue to work towards making sprint 4 datasets available:
     - Vaccine data
     - Viral genome data (COG-UK)
     - Serology datasets (Pillar 3 and priority Pillar 4 studies)
     - Intensive care data
     - Symptom Tracker: Zoe App data

   - Continue progress to build plan to **link secondary datasets before the end of March 2021**:
     - Tracing data
     - Shielded people list
     - Wastewater

- o Prescriptions
- o Mental health
- o Emergency care
- o 999 records/ambulance and 111 data
- o Maternity and neonatal
- o Mobility data

4. To make all these priority COVID datasets **discoverable and accessible through a single "shop window", the Health Data Research Innovation Gateway.** This includes having metadata visible for all datasets and creating one aligned data access form for researchers to use to request data.

5. To implement a targeted programme of **communication and engagement** to inform sprint activities, increase researcher engagement and to improve patient and public involvement and understanding.

What has been achieved in Sprint 5?

1. Responding to National Core Study (NCS) data and connectivity needs
Continuing to enable the research data infrastructure for COVID-19 Vaccination research

a) **Vaccine data (Events and Adverse Reactions) now flowing** to national TRE delivery partners in England and Wales; researchers can now start to work on answering urgent vaccine research questions by requesting access to these critical datasets through the HDR Innovation Gateway Progress continuing with availability and access to vaccine data in Scotland and Northern Ireland.

b) The HDRUK convened **Vaccine Data Infrastructure Group** (chaired by Professor Sir Munir Pirmohamed) continues to map vaccine data flows and linkages required for research, identify key research questions and enable data access.

c) In Sprint 5, the group identified a need for rapid feed of hospital admissions data to support the roll out of the C-19 Vaccination programme. A rapid funding call launched to invite applications from teams to run a feasibility study to establish a data flow giving researchers rapid access to hospital admissions data. An additional **task and finish group** was also set up in this Sprint to compare and align the **definitions and analytical methods** being used for C-19 vaccine research.

2. Engaging with researchers to increase use of data infrastructure

a) **Enabling data enabled trial recruitment via Test and Trace:** We worked with AstraZeneca to roll out this approach to support the **STORMCHASER and TACKLE trials**. In the first 2 days of implementing this approach with **AZ TACKLE trial ~30% of those contacted** via Test and Trace call showed interest, **and 4 participants were recruited (10% of UK total).** AZ have reported global interest in this approach.

b) The 12 research projects awarded in our rapid funding call enable research which **uses and enriches the data within the Data and Connectivity National Core Studies all listed** on the HDRUK Innovation Gateway.

c) They have submitted **19 Data Access Requests (DAR) covering >70 datasets** via the Gateway. **3 projects had DARs approved within this Sprint.**

d) We have collected **user feedback** from the researchers and TRE delivery partners to drive forward improvements in the Gateway and DAR process and will use this feedback to drive forward improvements to the Gateway and the DAR process.

e) The next phase of user feedback will concentrate on the experience of researchers within the TRE environment.

f) The studies will use data from across the 4 nations and will make their **own data accessible for others**

> **Scottish National Data Safe Haven :** NHS Scotland has a rapid COVID data access request approval process and COVID related research is being prioritised. eDRIS has created a dedicated team of research co-ordinators and analysts to work on development of database to support COVID-19 research and to triage COVID projects.

3. Making Data Available

Data has been linked and made available during the sprint:

a) **ONS NHS Digital Joint Health Data Asset (Census-Mortality-HES)** now available to request from the ONS via the Innovation Gateway. Work now underway to prioritise additional linkages to this data asset (e.g. primary care data)

b) **Test and Trace data now linked to Pillar 1 testing data within the ONS** and undergoing final quality assurance prior to being made available for wider research use

c) **Vaccine Events and Adverse Reactions data** flowing into **ONS Secure Research Service (SRS), NHS Digital and SAIL,** accessible soon for approved researcher use. Progress being made in Scotland and Northern Ireland to enable access.

d) Data from the **Covid-19 Clinical Information Network (CO-CIN)** available via **Scottish National Data Safe Haven** for research use, which includes a sub-set of **CO-CIN and COG-UK linked data**. CO-CIN data now also being linked with Pillar 1 and 2 testing data in England and work underway to link to **PHOSP-COVID data** and **vaccine data.**

e) **Pillar 4 serology data** available in ONS SRS (and linked to genome testing data for positive C-19 infection survey participants) and in NHS Digital and SAIL.

> **SAIL:** Welsh Immunisation Service data is flowing into SAIL daily and is being used by the Wales TAG research group as well as being prepared for use on the relevant [HDRUK ONS Rapid funded projects](#)

f) There is progress with work across all the TRE partners to onboard and link **COG-UK genomic data, with additional resource now identified to progress this.**

g) All the TRE delivery partners have **built a roadmap to ensure the newly identified datasets added in previous Sprints can be onboarded and linked by the end of March 2021**. Many of these datasets are already available via the Gateway.

> **Office for National Statistics:** NHS Digital have approved the sublicense agreement for ONS to share the HES data (which is linked to Census and Mortality) for wider research use via the HDR UK gateway.

4. **Making Data Discoverable and Accessible**
   a) **Enhancements to meta-data:**
      - **72** (+10 from sprint 4) National Core Study priority datasets are now listed on the Gateway with detailed technical metadata for **69** (+13 from Sprint 5)
      - **Northern Ireland Honest Broker Service** have signed agreement to enable researchers to access and analyse data remotely using the UK SeRP platform

> **Northern Ireland Honest Broker Service:** Agreement issued and returned from key Data Controllers to enable research to be conducted remotely using the HBS UK SeRP tenancy. Current researchers have received initial correspondence for project migration to the platform, NCS research will now be accessed via the portal. All other direct research applications to the NIHBS will be offered remote access as standard by for July 2021.

5. **Communicating and Engaging with partners, researchers, patients, and the public**

   a) Worked **with AstraZeneca to conduct rapid PPIE consultation on acceptability of data enabled recruitment strategies** for urgent COVID-19 clinical trials. Consultation conducted with >105 members of the public, and directly shaped and informed the trial recruitment strategies for TACKLE and STORMCHASER trials

   b) During Sprint 5 we continued delivery of the **communications and engagement plan** to provide progress and updates on Data and Connectivity work. This included a **video interview with Kamlesh Khunti** on key risk factors of COVID-19 and ethnicity. As well as a [blog post](#) on the impact of involvement so far.

   c) Gathered public and patient input into a new vaccine study with over 45 responses to help shape new funding application.

   d) Lay members reviewed **lay summaries for each of the 12 rapid funded project**s now published on the [HDRUK website](#).

   e) Lay members provided feedback on the development of a summary version of the concept note for building the UK's COVID-19 vaccine data infrastructure for research.

   f) Initial consultation with advisory group lay members reviewing and refreshing the Data and Connectivity HDRUK webpages.

   g) Collections now launched for all TRE delivery partners; providing an **openly accessible summary of the capabilities, software, and compute available in each TRE** and helping to **inform the researcher data access process:**

      - [Scottish National Safe Haven](#)
      - [ONS Secure Research Service](#)
      - [NHS Digital Trusted Research environment for England](#)
      - [Northern Ireland Honest Broker Service TRE](#)
      - [Secure e-Research platform / SAIL databank](#)

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

1. **The rapid** funding call has driven a sharp increase in **Data Access Requests (DARs) for NCS datasets via the Innovation Gateway**. This, along with the user feedback results, has **identified issues with delays in accessing data for urgent COVID research**, **complexity of data access request paperwork**, and **additional requirements** which need to be enabled (researcher accreditation, remote access IT requirements). Key areas for **improvement to the Gateway** on finding and requesting data are now being taken forward and we will work with our delivery partner TREs to help to further streamline the process.

2. There is an urgent **need for a more rapid feed of hospital admission data to monitor roll out of the vaccination programme**. Current national data feeds are incomplete, lack sufficient detail or do not provide fast enough data access to researchers. Working in close partnership with colleagues in NHS and NHSD we are looking at innovative options for national data feeds, but there is a critical need to rapidly

establish whether complementary data flows e.g., at regional level, add value. Our launched funding call will hopefully identify a team who can run a feasibility study to establish a rapid data flow.

3. As we move to the 'Recovery' phase of the COVID response there is a **demonstrable** need for wider **access and linkage to administrative datasets such as education, employment benefit and occupation for research.** Our mapping exercise with the **NCS Longitudinal Health and Wellbeing and NCS Transmission and Environment** have highlighted this and we will work closely with them, ADRUK and ONS to deliver wider access to these data sets in future Sprints.

4. Enabling **rapid recruitment to urgent COVID clinical trials via data enabled approaches** has again shown to be highly efficient and deliver rapid results; especially when **directly informed by public feedback on acceptability of this approach.**

5. **Finalising data sharing agreements,** and the timeframe required to resolve these remains a risk to our ability to achieve sprint deliverables at pace. We will continue to work closely with our TRE delivery partners and stakeholders to help facilitate and support progress.

Outline for Sprint 6

Sprint 6 Goal & Scope: By 31st March
Onboarding and linkage of vaccines and genomic data into TREs, progress of rapid funded projects, data asset needs mapping

Deliverables

1. **Health data infrastructure for C19 vaccine research** – a) national vaccine data onboarded and available for access, vaccine methods b) feasibility assessment call launched for rapid acute care data

2. **Health data infrastructure for genomic data and other sovereign data assets** enable greater access, linkage and availability for access of viral and host genomic data assets e.g (COG-UK, GenOMMIC, ISARIC 4C)

3. **Researcher funding call** – data access requests approved, Milestone 2 review, second phase of feedback (TRE user experience) and first phase informed improvements commenced, focus on system integration with TREs

4. **Preparing for next Phase of NCS Data and Connectivity** roadmap planned for subsequent Sprints, contracting with delivery partners

5. **Data asset mapping exercise continued** to obtain data asset and linkage needs of NCS, NIHR UPH studies and SAGE. Implement HDRUK DAR prioritization process when needed with NHSD.

6. **Finalize publication of delivery partner DPIAs and TRE DEA accreditation**

7. **PPIE/Communications** – Infographic/ visualisation of health data, commercial trials, rapid funding call projects, development of medium longer term strategy, lay summaries (data assets, linkage methodology)

8. **Completion of Sprint 1, 2, 3, 4 and 5 activities:**

    • Core **priority datasets** available to approved researchers in secure cloud-based TREs with demonstrable linkage
        – **Serology testing data** (in collaboration with C-CONNECT)
        – **Intensive care data** - e.g. ICNARC and SICSAG
        – **CO-CIN routine health linkages** for Wales and NI

Annex: Current Status of Health Data Research (SAGE Report 9th March 2021)

Openly available here

Research topics with new insights generated in last 2 weeks

Health data research outputs on COVID-19 continues to grow, now reaching 1,177 (+4) non-peer-reviewed pre-prints & 120 (+3) published papers.

Topic	Insights from ongoing studies (links provide further details):
Surveillance & Epidemiology	• Using primary care and summary hospital records from >17 million people, it was found that rates of events such as heart attacks, strokes and blood clots on the lungs etc during the first wave were reduced in the general population, but were highly likely to occur in people with COVID-19. The complex relationship between COVID-19 and clinical events requires further research. • A new resource linking health data for >54 million people from GP records, hospital data, death records, COVID-19 lab tests and data on medications dispensed from pharmacies, is now accessible in the NHS Digital Trusted Resource Environment – the world's largest single population-based cohort available for research with many applications including enabling long-term COVID-19 follow up.
Immunity & Vaccines	• New analyses of Public Health England data linking testing, vaccination and hospitalisation data show both the Pfizer-BioNTech and Oxford-AstraZeneca vaccines significantly reduce severe COVID-19 in older adults. Data demonstrate protection against severe disease, including the UK variant, though some vaccinated people went on to develop COVID-19 in the early days following vaccination indicating that other precautions should be maintained (particularly in the 3 weeks following vaccination). • Recent results from Public Health England's SIREN (SARS-CoV-2 Immunity & Reinfection EvaluatioN) study in healthcare workers indicate that the Pfizer-BioNTech vaccine prevents both symptomatic and asymptomatic COVID-19 infection, including the UK variant. This is especially important given that reducing transmission in hospitals and care homes will help to protect vulnerable people. • Recent research extrapolating US occupational survey data from O*Net suggests most healthcare workers globally are at a high risk of COVID-19 infection, regardless of their role – indicating that healthcare workers should be equally prioritised for vaccination.
Longitudinal health & wellbeing	• An observational study using historic and recent NHS hospital episode data from 12.5 million patients, found surgical procedures were reduced by 33.6% in England and Wales during the pandemic – equating to a deficit of >6 months pre-pandemic surgical activity that is likely to continue to grow. • Analyses of real-time hospital and administrative data detected marked reductions in cardiovascular disease referral, diagnosis, and treatment across China, Italy, and England. Models suggest this reduction is likely to lead to substantial excess deaths in people with cardiovascular disease and those with more than one long term condition.
Transmission & Environment	• SARS-CoV-2 antibodies and current infection data from >12,000 participants across 121 primary and secondary schools collected as part of the ONS Schools Infection Survey in late 2020 found school infection and staff antibody levels are similar to community levels. Data from future rounds of this survey will be useful in identifying trends, particularly with schools reopening. • A provisional report from Virus Watch, a community cohort study in England and Wales, found that people living in overcrowded homes (including those with less than one room per occupant) have a greater risk of COVID-19 infection compared to under-occupied households. This work underscores the need for COVID-19 public health interventions to consider household overcrowding, and to improve housing for health. • Behavioural monitoring data from 3,884 participants of the CoMix study suggests that contact between people in England was gradually reduced by local and national restrictions during summer and autumn 2020. Work from home orders appeared to be most effective at reducing contacts, whereas the 10pm bar and restaurant closures had no appreciable effect.
Clinical Trials	• The RECOVERY trial has closed recruitment to its colchicine treatment arm, as it has not been found to have a sufficiently significant effect in patients hospitalised with COVID-19. Still, the widely used gout drug will be investigated as a treatment for early-stage COVID-19 as part of the Platform Randomised trial of Interventions against Covid-19 In older peoPLE (PRINCIPLE) trial, to see whether it has a significant effect on less severe, non-hospitalised cases of COVID-19. For the colchicine arm, PRINCIPLE

Data & Connectivity National Core Study: COVID-19 dataset availability – 9 March

Vaccine data now available for Wales and England and will soon be available for Scotland.   Viral genome data also soon to be available for Scotland.

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)
C-19 vaccine data collection	To be made available shortly. Accepting applications now	Vaccines Events & Adverse Reactions - DARs extract only, soon in TRE	Data being transferred to TRE, available shortly.	COVID Vaccination Dataset	Vaccine Management System Operational – Data to be transferred to Data warehouse
COG-UK viral genome	Awaiting data flow from Public Health England	Awaiting data flow from Public Health England	Subset linked to COCIN data. Remaining data being transferred, available shortly.	Awaiting data flow from Public Health Wales	Governance agreed, automation of data flow to PHA in progress
Pillar 1 COVID-19 Testing Data	To be linked to Test and Trace data	COVID19 Second Generation Surveillance System (SGSS)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID19 Test Results	COVID antigen testing Pillar 1
Pillar 2 Testing data (UK Gov)	To be linked to Test and Trace data	COVID19 UK Non-hospital Antigen Testing Results (Pillar 2)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID19 Test Results	Missing results prior to 26 Apr Data quality issue
Primary Care	GPES linked to census, mortality and hospital data for internal access only	GPES extract – 98% practice coverage, large subset of codes (4bn items) Community Prescribing	Albasoft ESCRO GP Extraction* Prescribing Information System	80%+ coverage of full longitudinal record, with 100% coverage for COVID codes	Enhanced Prescribing Database as proxy
Secondary Care	Census-Mortality-HES linked data asset now available (ONS/NHSD)	HES SUS via DARS extract only, available in TRE soon	100% coverage	100% coverage	Admissions & Discharges
Personal Demographic Service	Internal use only	100% coverage (via DARS extract only.	100% coverage	100% coverage	
Death registry	Provisional Monthly Extract & linked Census and death occurrence	100% coverage Civil Registrations Deaths	100% coverage	100% coverage	
C-19 Infection Survey (CIS)		N/A	Awaiting DEA accreditation	Awaiting decision on data access	Awaiting decision on data access
COVID-19 Clinical Information Network (CO-CIN)	Being linked to 2011 census	Data for English COCIN participants available in Scottish Nation Data Safe Haven	Limited metadata. Includes English linked data, and COGUK/COCIN data asset	Awaiting decision on data access	Discussions ongoing to collect data in NI
Census 2011	Household structure	N/A	N/A	N/A	N/A
Covid Opinions Survey		N/A	N/A	Awaiting decision on data access	N/A
Business Impact of Covid Survey	c. 5,000 businesses	N/A	N/A	Awaiting decision on data access	N/A
Labour Force Survey	40,000 households, 100,000 individuals	N/A	N/A	Awaiting decision on data access	N/A
Intensive Care data	HES Critical Care, ICNARC	HES Critical Care Reviewing IG to share ICNARC.	SICSAG (updated weekly)	ICNARC COVID weekly, ICNARC quarterly all admissions and critical care routine data (CCDS) monthly	Critical Care Minimum dataset to be acquired
Pillar 3 Testing data (NHS labs)	Available in COVID Infection Survey			COVID19 Test Results	
Pillar 3 Testing data (iELISA)	N/A	COVID19 UK Non-hospital AntibodyTesting Results (Pillar 3)			Data to be validated
Other Pillar 4 Testing data	VIVALDI, REACT II				
ZOE Symptom Study App Data				UK wide (unlinked) Wales (linked)	

KEY

- 1. Custodian engagement
- 2. Dataset available in secure Trusted Research Environment
- 3. Linkages established to other priority datasets (within TRE)
- 4. Datasets available for COVID-19 research via Gateway

Further information about **Data & Connectivity** can be found **here**. Including the recently published **Sprint 4 report.**

Linked census, mortality and hospital episode statistics now available as a joint data asset between ONS and NHSD

| 4

Status of COVID-19 projects using the data – 9 March 2021

13 new research projects underway, taking the total number of active research projects over 300. 15 data access requests covering 63 data assets now submitted via the Gateway for the 12 rapid funded projects. 3 of 12 project requests approved so far.

# of COVID-19 Projects by 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	13 (+2)	38 (-2)	38 (+2)	95 (-)	4 (-)	188 (+2)
- a/w researcher	12 (+1)	28 (+1)	25 (+3)	56 (-1)	2 (-)	123 (+3)
- a/w data custodian	1 (+1)	10 (-2)	13 (-1)	39 (+1)	2 (-)	66 (-1)
Submitted for Information Governance approval	5 (+4)	10 (+7)	6 (-1)	0 (-1)	0 (-)	21 (+9)
Approved but not yet active	4 (+1)	1 (-3)	9 (+2)	3 (-)	0 (-1)	17 (-1)
Active research taking place	16 (+3)	95 (+4)	69 (+4)	123 (+2)	1 (-)	304 (+13)
Active Number of Researchers	179	52 (+9)	194 (+4)	224 (+18)	1	650
Average time from application to active research	Not yet available	Not yet available	9.33 days	3 days	Not yet available	

Participation in key UK wide studies :

- PRINCIPLE: 4,507 participants (+5% in last 2 weeks, with continuing data flow of Pillar 2 COVID +ve test results to support recruitment)
- RECOVERY: 38,703 participants across 180 active sites (+3% in last 2 weeks, and +2 active site)
- CO-CIN (ISARIC4C)
  - 191,645 Tier 0 (case report) (+3% in last 2 weeks)
  - 649 Tier 1 (single sample)
  - 1,631 Tier 2 (serial sampling)
- GENOMICC: 11,595 participants (+4% in last 2 weeks) across 212 ICUs with a total of 5091 intensive care beds
- COVID-19 ZOE symptom study: 4,620,305
- COG-UK: 332,804 viral genomes sequenced

Data & Connectivity National Core Study projects now underway

15 data access requests covering 63 data assets from 12 studies now submitted via **HDRUK's Innovation Gateway.** 3 of 12 project requests approved so far. Examples include :

- **Ethnicity and COVID-19: investigating the determinants of excess risk** led by Professors Khunti and Yates, University of Leicester
- **COVID-19 Vaccines Pharmacovigilance (DaC-VaP)** led by Professor Sheikh, University of Edinburgh
- **Real-time Assessment of Community Transmission (REACT) Study 2 - COVID antibody (Data & Connectivity NCS)** led by Professor Elliot, Imperial College London

Data Access Registers
For more information on the active projects:

- **ONS Secure Research Service:** List of accredited researchers and research projects under the Research Strand of the Digital Economy Act)
- **NHS Digital:** Register of approved data releases (includes all access)
- **Scotland :** Public Benefit and Privacy Panel approvals
- **SAIL Databank:** COVID-19 projects listed on gateway
- **NI Honest Broker Service:** Projects currently being carried out.