COVID-IMPACT-UK [1]

UK-wide linked routine healthcare data to address the impact of health conditions and health-related risk factors on COVID-19 and the impact of COVID-19 on health

Study Protocol
Version 2
21 June 2021

Cathie Sudlow
Director, BHF Data Science Centre

[1] Formerly CVD-COVID-UK

SUMMARY

The COVID-19 pandemic is a major global challenge, whose impacts on population health, healthcare services and the wider economy will be apparent for many years.

Patients with – or at risk of – a range of prior health conditions have an elevated risk of symptomatic infection and mortality. This could be due to the health conditions themselves (e.g., heart disease, diabetes, respiratory disease), their risk factors (e.g., age, raised blood pressure, obesity), medications, or combinations of these. Understanding which patients are affected and why will be a major step towards developing strategies to reduce this excess risk.

Just as important as the effects of pre-existing disease on COVID-19, are the impacts of COVID-19 on a wide range of health outcomes. The *direct* impacts include immediate complications (e.g., respiratory failure, acute cardiac injury, stroke) as well as the potential for increased risk of various physical and mental health conditions in the longer term, through a range of mechanisms, including inflammation, blood clotting risk and other factors, both known and unknown. The nature and extent of these direct effects is far from fully understood.

Crucially, the response by the government and health services to the COVID-19 pandemic has *indirect* impacts on the presentation, diagnosis, management and outcomes of a wide range of disabling and life-threatening diseases, including cardiovascular disease, cancer, mental health disorders and arthritis. To inform government and NHS policy, we need a deeper understanding of these unintended consequences, including the range of conditions affected, variation by age, sex, ethnicity, deprivation and geography, effects on different in- and out-patient services, and response to mitigating actions (e.g., regional and national government advice).

Building on the progress of CVD-COVID-UK, COVID-IMPACT-UK will address these issues through analyses of de-identified, linked, nationally collated, whole population, healthcare datasets across the four nations of the UK. Linked, de-identified data will be accessed and analysed by approved researchers within trusted research environments (TREs) provided by:

(1) SAIL databank for Wales

(2) The National Data Safe Haven in Scotland

(3) NHS Digital in England[2]

(4) The Honest Broker Service in Northern Ireland[3]

[2] Working in partnership with the BHF Data Science Centre as its first and largest client, NHS Digital has now developed a trusted research environment service, which has enabled the CVD-COVID-UK programme and will enable COVID-IMPACT-UK.

[3] Prior to the pandemic, in common with England, Northern Ireland did not have a trusted research environment but is now establishing one in partnership with Swansea University, based on the platform used for the SAIL databank for Wales

OBJECTIVES

To interrogate nationally collated, population level, linked healthcare datasets across the UK population to address the following questions:

1) What are the effects of prior health conditions, their risk factors and medications on susceptibility to and outcomes from COVID-19 disease?
2) What is the direct impact of SARS-CoV-2 infection on acute complications as well as on medium and longer term health conditions?
3) What is the indirect impact of the COVID-19 pandemic and the government and NHS response to it on the presentation, diagnosis, management and outcomes of health conditions?

BACKGROUND

The COVID-19 pandemic is a major global challenge, whose impacts on the population's health, healthcare systems and services and the wider economy will be apparent for many years.

Patients with a range of pre-existing health conditions have a disproportionately elevated risk of symptomatic infection and mortality. The reasons are not fully understood but could be due to health conditions per se (myocardial infarction, diabetes, asthma etc.); their risk factors (e.g., age, ethnicity, blood pressure, obesity, diet and lifestyle) or medications (e.g. antihypertensive medications or statins); or combinations of these. Understanding the drivers of this excess risk (i.e., which patients are affected and why) will be a major step towards developing strategies to reduce it.

Just as important as the effects of pre-existing health conditions, and their risk factors and medications, on COVID-19 disease, are the direct and indirect impacts of COVID-19 on a wide range of health outcomes.

The *direct* impacts include acute complications of SARS-CoV-2 infection, such as acute respiratory compromise and venous thromboembolism. In addition, the inflammatory, pro-coagulant and other effects of COVID-19 are, like influenza and other respiratory virus infections, are associated with increased risk of a wide range of physical and mental health conditions in the short, medium and long term. However, the nature and extent of these direct effects are far from fully understood. These conditions include long-COVID, encompassing an as yet incompletely understood spectrum of longer term effects of SARS-CoV-2 infection on a significant proportion of patients.

Crucially, there are also *indirect* impacts on the presentation, diagnosis, management and prognosis of many diseases resulting from the response by governments and health services to the COVID-19 pandemic. For example, in the UK and elsewhere, the numbers of people attending hospital with acute myocardial infarction and stroke and of those receiving assessment and treatment for cancer declined dramatically during 2020 due to the impact of the pandemic on health service provision and behaviours. A deeper understanding of the nature and extent of these unintended consequences, including the range of conditions affected, variation by patient characteristics (such as age, sex, ethnicity, and deprivation) and geography (both within and between countries), effects on different in- and out-patient services and interventions, including delays in diagnosis caused by cancelled

diagnostic tests, and changes over time in response to mitigating actions (e.g. regional and national government advice), is needed to inform current and future government and NHS policy.

We propose to address these questions by performing analyses of linked, nationally collated healthcare datasets across the four nations of the UK. A transparent, inclusive, multidisciplinary consortium will conduct the research, addressing questions of significant healthcare and health policy relevance. The programme is providing – and will continue to provide - support for the UK Chief Scientific Adviser's National Core Studies programme (in particular the Data and Connectivity and Longitudinal Health and Wellbeing core studies), which coordinates national COVID-19 priority research.

PROGRAMME PLAN

The project started in June 2021 and is currently set to run until 1 June 2023. There are three work packages (WPs) as outlined below:

Work package 1: Coordination, Approvals and Data Access

This work package is led by the BHF Data Science Centre (Director, Prof Cathie Sudlow). It involves:

- the identification of relevant nationally collated datasets across the UK
- the coordination of applications for ethics approval
- the coordination of applications for access to linked data within national trusted research environments
- the coordination of ongoing developments to NHS Digital's TRE in partnership with NHS Digital to support the COVID-IMPACT-UK programme, ensuring future proofing of this development so that it can in due course enable work beyond COVID 19
- the coordination of specialist inputs from data custodians, data scientists with methodological and analytical expertise (statisticians, epidemiologists, health informaticians, bioinformaticians, computer scientists and others) and clinicians
- administrative support for and coordination across all work packages
- the coordination of reporting to SAGE and equivalent bodies in the devolved nations via established HDR UK processes
- the on-boarding of new members to the inclusive and transparent consortium (currently CVD-COVID-UK, to be renamed COVID-IMPACT-UK)
- the coordination of regular consortium update meetings
- the development and documentation of the consortium's ways of working (see Appendix 1)

The relevant generic and disease-specific datasets currently being made available for access, linkage and analysis across the UK is available here: https://www.hdruk.ac.uk/wp-content/uploads/2021/06/210610-CVD-COVID-UK-TRE-Dataset-Provisioning-Dashboard.pdf. The datasets are updated regularly as new data become available (e.g. we have a regular cycle of monthly

updates for all data in the NHS Digital TRE, while the data within the SAIL databank for Wales are updated daily). Data in each environment are accessed by named, approved researchers (certified to have successfully completed safe researcher training – e.g., https://saildatabank.com/application-process/following-approval/#safe-researcher-training) in the national TREs provided by:

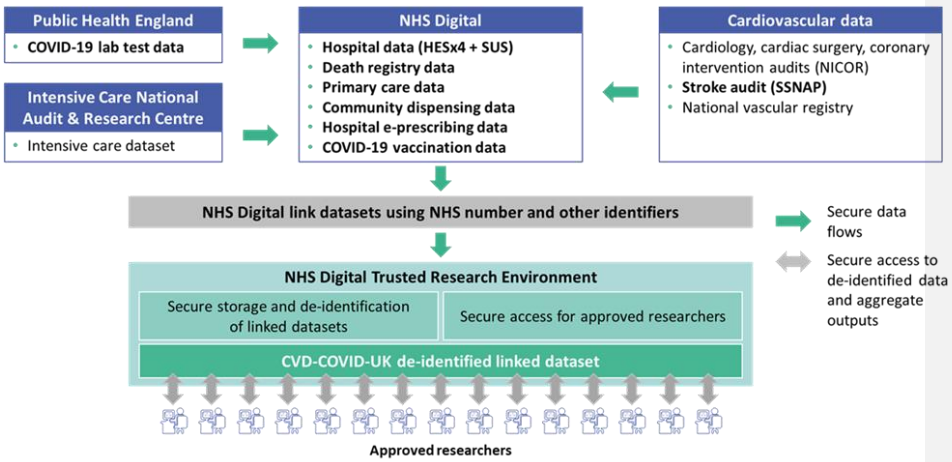
- the SAIL databank for Wales (population around 3 million)
- the National Data Safe Haven in Scotland (population around 5.4 million)
- NHS Digital (who have developed a new trusted research environment to support this programme) for England (population around 57 million)
- and in the future by the Honest Broker Service in Northern Ireland (population around 1.6 million)

Figure 1 shows the flows of data into the NHS Digital TRE co-developed with NHS Digital for this project. Similar data flows are in place for the TREs in the devolved nations (with the exception of Northern Ireland, which lags behind the other UK nations).

The data accessed for research analyses is record level data, which is de-identified and pseudonymised. The organisations providing the TREs strip all direct identifiers (e.g., name, address, exact date of birth, NHS number) from the records and apply a person-specific pseudo-ID to each record so that the datasets can be linked. No direct identifiers are accessed by any researcher working in the COVID-IMPACT-UK consortium. Following similar principles to those already used by the SAIL Databank for Wales (https://saildatabank.com/saildata/data-privacy-security/#secure-access) and the Scottish National Data Safe Haven (https://www.isdscotland.org/Products-and-Services/EDRIS/Use-of-the-National-Safe-Haven/), only summary, aggregate results data can be exported from the TREs by approved researchers, subject to the approval of the organisations providing the TREs. This ensures that no output contains information that could be used either on its own or in conjunction with other data to breach an individual's privacy.

Figure 1. Data flows for Trusted Research Environment for COVID-IMPACT-UK in England

Commented [RP1]: Need to update the figure below to say COVID-IMPACT-UK in the green section



Work package 2: Addressing key research questions

The BHF Data Science Centre has established and coordinates a streamlined system for considering, refining and rapidly approving proposals for research projects that address aspects of one or more of the three key questions laid out in the **OBJECTIVES** section of this protocol (above). Any member of the inclusive and transparent CVD-COVID-UK (soon to be renamed COVID-IMPACT-UK) consortium can contribute to an existing project or submit a project proposal to the Approvals and Oversight Board, which includes representatives from data custodians, researchers and lay members, and is coordinated by the operations team of the BHF Data Science Centre (see Appendix 2 for project proposal form). The Approvals and Oversight Board meets regularly to assess project proposals, to ensure that they fall within the programme's ethical and regulatory approvals and are aiming to access datasets that are appropriate to their research question(s), to identify un-necessary overlap with existing projects, and to provide constructive feedback, in particular from the perspective of the lay members.

An up-to-date list of projects together with their lay summaries is maintained on the CVD-COVID-UK/COVID-IMPACT-UK webpage: https://www.hdruk.ac.uk/projects/cvd-covid-uk-project/. Projects approved as of 21 June 2021 are shown in Table 1.

Given that data custodians in each of the four UK nations will bring together and make available their nationally collated datasets within separate TREs, and given the differences between countries in the datasets available, our general approach for each analytic work package is to develop a master protocol for each project, from which country-specific analysis plans will be developed, aiming for maximum consistency but allowing for country-specific differences in the datasets. Where appropriate, results of country-specific analyses will be combined in meta-analyses.

Analyses based on routinely collected, national healthcare datasets have the advantages of large scale and comprehensive coverage, maximising statistical power as well as inclusiveness and representativeness (e.g. across all age groups, ethnicities, geographies and socioeconomic settings).

A key project of relevance to all others is CCU005, Data management and analysis methods, led by Dr Angela Wood, University of Cambridge. This project coordinates discussions amongst the epidemiological, statistical, informatics and computer science members of the consortium to address data management, curation and analysis challenges encountered across many projects. Through this project, analysts from across the consortium provide user input into the requirements for and development of the newly established NHS Digital TRE for England.

Table 1. Projects approved as of 21 June 2021

Ref	Lead	Title	England/ NHSD	Scotland/ Safe Haven	Wales/ SAIL
CCU001	Jonathan Sterne	Investigating the effects of ACE inhibitors and angiotensin receptor blockers on COVID-19 outcomes	X	X	X
CCU002	William Whiteley	SARS-CoV-2 infection and risk of venous thromboembolism and arterial thrombotic events	X	X	X
CCU003	Ami Banerjee	Direct and indirect effects of COVID-19 in individuals with cardiovascular disease	X	X	X
CCU004	Angela Wood	COVID and Cardiovascular Disease Risk Prediction	X	X	X
CCU005	Angela Wood	Data management and analysis methods	X	X	X
CCU007	Massimo Caputo	Impact of COVID-19 pandemic on Congenital Heart Disease (CHD) patients undergoing cardiac surgery	X		
CCU008	Julian Halcox	Evaluating Impact of COVID-19 on Management of Blood Pressure and Lipids in Patients With and at High Risk of Cardiovascular Disease	X	X	X
CCU010	Claire Lawson	In people with CVD and Covid-19, what is the influence of multi-morbidity on risk of poorer outcomes?	X		
CCU013	Spiros Denaxas	High-throughput electronic health record phenotyping approaches	X	X	X
CCU014	Reecha Sofat	Assessing cardiovascular disease impact through medicines	X	X	X
CCU016	Tim Wilkinson	Cardiovascular and cerebrovascular diseases related to antipsychotic prescribing in patients with dementia during the COVID-19 pandemic	X	X	X
CCU018	Elena Raffetti	COVID infection during pregnancy on cardiovascular disease and related risk factors	X		X
CCU019	Honghan Wu	Identification and personalised risk prediction for severe COVID-19 in patients with rare disorders impacting cardiovascular health	X	X	X
CCU020	Alex Handy	Evaluation of antithrombotic use and COVID-19 outcomes	X	X	X
CCU022	Michael Inouye	Genomics of multimorbidity and CVD associated with susceptibility to COVID-19 infection and complications	X	X	X
CCU023	Reecha Sofat	Repurposing medicines used to treat CVD risk to prevent COVID-19	X		

Work package 3: Patient and public involvement and engagement and communications

This work package is led by Debbie Ringham, Communications and Engagement Manager for the BHF Data Science Centre. Supported by the BHF Data Science Centre core team, she coordinates the contribution of our lay members (currently four) to the CVD-COVID-UK (to become COVID-IMPACT-UK) Approvals and Oversight Board. The lay members review and discuss project proposals with the relevant researchers. This is to ensure that the proposed research meets the interests of patients and the public, to provide input into refining and prioritising the research questions, to address any patient and/or public concerns, and to advise on best approaches to patient and public involvement in each project, including communicating research results to lay audiences. This work package also coordinates communications of the consortium's activities and emerging results through websites, social media and other outlets, and leads on interactions with the press and other media.

RESEARCH OUTPUTS

Through addressing questions about the impacts of health conditions on COVID-19 and the impacts (both direct and indirect) of COVID-19 on a wide range of health outcomes, we expect the outputs of this work to inform public health policy and clinical care, benefiting:

- patients with a history of health problems, risk factors for disease or on medication, who are at increased risk of poor outcomes with COVID-19;
- patients now and in the future who become unwell with COVID-19 and are at risk of short, medium and long term health consequences;
- the population as a whole whose health services are affected by the government, health service and public responses to the COVID-19 epidemic.

Outputs providing these benefits have started to emerge and will continue to be produced throughout the duration of the programme. By their nature, those outputs providing information on the longer term impacts and implications of these for clinical care and public health policy will take longer.

Emerging insights will be reported as appropriate to SAGE and equivalent bodies in the devolved nations as well as to NICE, MHRA, the Joint Committee on Vaccination and Immunisation and others, so helping to drive evidence-based policy decisions for health service providers and clinical professional groups. Outputs will also form the basis of manuscripts for publication in peer-reviewed scientific and medical journals, presentations at national and international scientific and medical professional conferences, and reports aimed at lay audiences, available through relevant websites, including those of Health Data Research UK, the British Heart Foundation and the Longitudinal Health and Wellbeing National Core Study.

WIDER BENEFITS

Whilst being developed with the initial intent of supporting COVID-related research, the TRE for England established in partnership with NHS Digital is being designed to enable its extension to support research beyond COVID-19.

The coordination work for WP1 will provide knowledge about linked health care datasets and routes to their access and so will benefit many other UK-wide initiatives that require linkage to routinely collected healthcare data, including the National Core Studies as well as future initiatives, such as large data-enabled clinical trials and cohort studies.

In addition, initiatives in a wide range of other countries, including Italy, Korea, Scandinavian countries and Canada, are deriving policy-relevant insights from analyses of routine linked datasets. Through the consortium's connections with international consortia and organisations, the research group will engage in the bilateral sharing of emerging results, to learn from others as well as to maximise the international reach and relevance of findings.

CONSORTIUM PRINCIPLES

The consortium is committed to the 'Five Safes' (http://www.fivesafes.org/) and to a transparent and inclusive approach, enabling additional researchers and research groups to join the consortium as the work progresses and evolves.

All analysis plans, protocols and reports arising from this consortium's work are made publicly available via repositories in the BHF Data Science Centre GitHub organisation, relevant HDR UK and institutional websites and open access publications.

Commented [RP2]: Link to this?

All reports for government advisory groups and policy makers, the lay public and academic publications will be written in the name of the collaboration with all relevant individual contributions (coordination, writing, analysis, interpretation etc.) listed.

CURRENT CONSORTIUM MEMBERSHIP

A regularly updated list of the consortium's members is provided here: https://www.hdruk.ac.uk/wp-content/uploads/2021/06/210615-CVD-COVID-UK-Consortium-Members.pdf

Appendix 1: Consortium ways of working

CVD-COVID-UK ways of working v1.1.pdf

Appendix 2: project proposal form

CVD-COVID-UK project proposal form