BMJ Open

# Research Evaluation Alongside Clinical Treatment in COVID 19 (REACT COVID 19): An Observational and Biobanking Study



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-043012 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 21-Jul-2020 |
| Complete List of Authors: | Burke, Hannah; University of Southampton Faculty of Medicine<br>Freeman, Anna; University of Southampton Faculty of Medicine,<br>Dushianthan, Ahilanandan; University of Southampton Faculty of Medicine<br>Celinski, Michael; University of Southampton Faculty of Medicine<br>Batchelor, James; University of Southampton Faculty of Medicine<br>Phan, Hang; University of Southampton Faculty of Medicine<br>Borca, Florina; University of Southampton Institute for Life Sciences<br>Kipps, Christopher ; University of Southampton<br>Thomas, Gareth; Faculty of Medicine, Cancer Sciences Unit<br>Faust, Saul; University of Southampton, UK, NIHR Wellcome Trust Clincial Research Facility<br>Sheard, Natasha; University Hospital Southampton NHS Foundation Trust<br>Williams, Sarah; University Hospital Southampton NHS Foundation Trust<br>Fitzpatrick, Paul; University of Manchester, Cancer Biomarker Centre, Cancer Research UK Manchester Institute<br>Landers, Donal; University of Manchester, Cancer Biomarker Centre, Cancer Research UK Manchester Institute<br>Wilkinson, Tom; University of Southampton, Faculty of Medicine |
| Keywords: | COVID-19, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Chronic airways disease < THORACIC MEDICINE, Respiratory infections < THORACIC MEDICINE, RESPIRATORY MEDICINE (see Thoracic Medicine) |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Research Evaluation Alongside Clinical Treatment in COVID 19 (REACT COVID 19): An Observational and Biobanking Study

Burke, H.#[1,3], Freeman, A.#*[1,3], Dushianthan, A.[1,3], Celinksi, M.[1,3], Batchelor, J.[1,2,3,6], Phan, H.[1,4,6], Borca, F.[2,4], Kipps, C.[1,2,3,4], Thomas, G.J. [1,2,3,4], Faust, S.N.[1,2,3,4,5], Sheard, N.[3], Williams, S.[3], Fitzpatrick, P.[7], Landers, D.[7], Wilkinson, T.M.A.[1,3]

# Joint first authors

* Corresponding author: Anna Freeman

Email: a.freeman@soton.ac.uk

Affiliations

1. Faculty of Medicine, University of Southampton.
2. Institute for Life Sciences, University of Southampton.
3. University Hospitals Southampton NHS Foundation Trust.
4. NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust.
5. NIHR Southampton Clinical Research Facility, University Hospital Southampton NHS Foundation Trust.
6. Clinical Informatics Research Unit Faculty of Medicine, University of Southampton.
7. Digital Experimental Cancer Medicine Team, Cancer Biomarker Centre, Cancer Research UK Manchester Institute, The University of Manchester.

Word count: 2,988

# Abstract

## Background

The pandemic outbreak of coronavirus disease 2019 (COVID-19) caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) places immense demand on worldwide healthcare services. Earlier identification of those patients at risk of severe disease may allow intervention with experimental targeted treatments mitigating the course of their disease and reducing demand on critical care services.

## Methods

This prospective observational study of patients tested or treated for SARS-CoV-2 who are under the care of the tertiary University Hospital Southampton NHS Foundation Trust (UHSFT), captures data from admission to discharge from UHSFT. Core demographic and clinical information, as well as results of disease-defining characteristics are captured and recorded electronically from hospital clinical record systems at the point of testing. Manual data is collected and recorded by the clinical research team for assessments which are not part of the structured electronic health care record, for example symptom onset date. Thereafter, participants records are continuously updated during their hospital stay and in their follow up period under UHSFT. In addition, participants over the age of 16 years old are given the opportunity to provide consent for storage of excess clinical sample with optional further biological sampling. These anonymised samples are linked to the clinical data in the REACT platform and stored within a biorepository at UHSFT.

## Discussion

The REACT COVID study captures the natural history of SARS-CoV-2 infection and its clinical response to current interventions, as well as characterising specific phenotypes and endotypes, with the potential to rapidly identify patients for novel treatment trials. The simplicity of data visualisation within the REACT COVID study belies the complexity of the concepts behind this project. The REACT COVID study enables visualisation of detailed clinical data which will allow greater understanding of the natural history of this novel disease through both longitudinal and cross-sectional analysis.

Word count 300

**Key words:** SARS-CoV-2, COVID-19, Longitudinal, Biobank, Clinical Informatics, Clinical Registry, Natural History, Phenotype, Endotype

## Background

The outbreak of coronavirus disease 2019 (COVID-19) caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was declared a pandemic by the World Health Organisation (WHO) on 11th March 2020 (1). In the United Kingdom (UK), 299,251 people have tested positive for SARS-CoV-2, of which 42,153 have died (https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public; correct as of 5pm 16th June 2020) (2). So far, emerging effective treatments such as Remdesivir and dexamethasone have only modest effects on clinical outcomes in a sub-cohort of patients (3, 4) and any vaccines have yet to demonstrate their efficacy. Whilst many patients recover from SARS-CoV-2 infection without need for hospitalisation, a small proportion go on to develop severe disease. The demand on healthcare services, especially critical care, is immense. Earlier identification of those patients at risk of severe disease may allow intervention with experimental targeted treatments mitigating the course of their disease and reducing demand on critical care services.

The need for rapid translation of treatments from development to clinical application is being addressed by Urgent Public Health studies such as ACCORD-2 (5) and RECOVERY trials (6). Despite the urgency to enrol patients in these studies, there are challenges in identifying suitable patients in a fluid, "real-time" clinical environment within a novel disease. The longitudinal REACT COVID database and biosampling study captures the natural history of SARS-CoV-2 infection and its clinical response to current interventions, as well as characterising specific phenotypes and endotypes, with the potential to rapidly identify patients for novel treatment trials.

The Research Evaluation Alongside Clinical Treatment observational and biobanking study of COVID-19 (REACT COVID study) evolved around a collaboration between the University of Southampton (UoS), University Hospital Southampton NHS Foundation Trust (UHSFT) and the digital Experimental Cancer Medicine Team (digital ECMT) and their work to create the REACT (REal-time Analytics for Clinical Trials) platform. The REACT platform is used for the determination of early clinical benefit of new cancer medicines in early phase drug trials and has been adapted by the digital ECMT in collaboration with UHSFT to allow rapid upload and interpretation of clinical data in the context of the COVID-19 pandemic. This study is supported by the National Institute for Health and Research (NIHR) Southampton Clinical Research Facility (CRF) and NIHR Southampton Biomedical Research Centre (BRC), which takes advantage of the expertise and resource in hosting large prospective cohort studies.

**Objectives**

- Development of a real-time "parent" database of well-characterised COVID-19 patients to facilitate a better understanding of the natural history of SARS-CoV-2 infection from pre-hospital through to level 3 care and subsequent course of COVID-19 related complications from discharge through to clinic follow up.
- Cross-sectional phenotypic characterisation of a cohort of COVID-19 patients to describe the disease heterogeneity seen in clinical practice
- Identify potential risk factors for progression to severe disease measured by admission to level 2 or 3 care, increase in respiratory support or death (primary endpoint).
- Use of samples stored in the biorepository (for example blood, urine, sputum) to develop an endotype level understanding of disease clusters.

- Assess the efficacy of current best practice management strategies.
- Identify subgroups of patients for novel treatment strategies in the form of independent formal randomised control trials.

# Methods

## Study Design

The REACT COVID study is a prospective observational study of patients tested or treated for SARS-CoV-2 infection who are under the care of UHSFT. The study is planned as a long-term research project with no limit to enrolment nor any time defined end. Patients are recruited into the study at the point of testing for SARS-CoV-2 at UHSFT. Core demographic and clinical information, as well as results of disease-defining characteristics are captured and recorded electronically from hospital clinical record systems at the point of testing. Manual data is collected and recorded by the clinical research team for assessments which are not part of the electronic health care record, for example symptom onset date. Thereafter, participants records are continuously updated during their hospital stay and for 12 months following discharge from acute admission, in their follow up period under UHSFT. This pragmatic, opportunistic approach to data collection takes full advantage of the electronic health care records and seamlessly links with the REACT platform to reduce data collection burden on the patient, clinician or researcher. In addition, participants over the age of 16 years old are given the opportunity to provide consent for storage of excess sample taken as part of their routine clinical management with optional further biological sampling. These anonymised samples

are linked to the anonymised clinical data in the REACT platform and stored within the Southampton Research Biorepository (SRB), managed by the NIHR CRF at UHSFT.

## Patient and Public Involvement

Patient representatives were involved in and are part of the governance structure of the Southampton Research Biorepository and patient representatives were involved in the design and ongoing management of the WATCH study, upon which much of this observational biobank study was based. The SRB Oversight Committee has a lay representative to work alongside other members and may jointly be consulted in the different matters related to the construction and running of the Biorepository.

## Setting

UHSFT is an 1100 bed tertiary centre and as such, has assessed and admitted a large number of SARS-CoV-2 positive patients.

## Recruitment

All patients under the care of UHSFT who are tested or treated for SARS-CoV-2 are included in the REACT COVID study. A sub-cohort of patients ≥16 years old are given the opportunity to provide consent for storage of excess sample taken as part of their routine clinical management with optional further biological sampling. They are offered the opportunity to provide this at any point during their care under UHSFT. So far UHSFT has admitted 629 patients with confirmed COVID-19, correct as of 17th June 2020. Study enrolment will be continual into the future aligned to the ongoing function of the clinical service with no discrete recruitment target or endpoint.

## Data collection

This study is divided into two parts and is conducted in conjunction with standard clinical care. The database component involves the assimilation of data collected as part of routine clinical care for patients of any age tested for or admitted to UHSFT for suspected SARS-CoV-2 infection. Core demographic and clinical data collected as part of standard clinical care will be extracted from the electronic hospital records. Where required, additional data relevant to COVID-19 clinical course is manually extracted from the clinical notes and uploaded to the database by clinical researchers where automatic electronic export is not possible. Data collected from study participants is collated in the highly secure, contemporary and encrypted data platform, BC|INSIGHT (Microsoft Data Centre, United Kingdom). Data is then uploaded to the REACT platform to allow rapid capture of the natural history of the disease, with clinicians able to visualise trends in patient trajectories and identify patients who may be suitable for approach for intervention studies (summarised in Figure 1.).

Data captured at point of testing for SARS-CoV-2 includes core demographic and clinical data relevant to COVID-19. Subsequent data capture follows clinical disease course and includes information on medication, vital signs, pathological and radiological measurements. This data is captured in real time during the follow up period under UHSFT care, including any period of hospitalisation, and for 12 months after admission to capture virtual follow up or outpatient clinic follow up data (summarised in Figure 2.). Any treatment interventions as part of routine clinical care are also captured (where possible electronically) and these form part of the visualisations in the REACT platform. Data for research purposes will be stored in

an anonymised format, and any data sharing outside of the clinical team caring for the patient at any point in their patient journey will be in an anonymised format.

The majority of data collected is consistent with the data already being collected as part of routine clinical care of patients with COVID-19, therefore seamlessly linking with clinical practice. Clinical data to be captured is listed in Table 1. The REACT database will host a clinically identifiable 'front end' for real time patient management during the acute admission and for up to 12 months post discharge, in which the link anonymised data will be made identifiable, and visible only to the clinical team as determined by a statement of reference. Data for research purposes will be stored in an anonymised format, and any data sharing outside of the clinical team caring for the patient at any point in their patient journey will be in an anonymised format.

| Data variable | Data source | Frequency of capture |
|---|---|---|
| **Demographics** | | |
| Sex at birth (male/female) | Electronic | Once |
| Date of Birth & Age | Electronic | Once |
| Pregnancy status (yes/no) | Electronic | Once |
| Admission weight (kilograms) | Electronic | Once |
| Admission height (metres) | Electronic | Once |
| Ethnicity | Electronic | Once |
| Profession/employment status | Electronic/manual | Once |
| **Disease onset and admission to UHSFT care** | | |
| Symptom onset date | Notes | Once |
| Admission date | Electronic | Once |
| Admission ward & level of care | Electronic | Once |
| Frailty score | Electronic | Once |
| **Co-morbidities & Risk factors** | | |
| Chronic obstructive pulmonary disease | Electronic | Once |
| Asthma | Electronic | Once |
| Interstitial lung disease | Electronic | Once |
| Bronchiectasis | Electronic | Once |
| Hypertension | Electronic | Once |
| Chronic cardiac disease (not hypertension) | Electronic | Once |
| Chronic kidney disease | Electronic | Once |
| Chronic Liver disease | Electronic | Once |
| Chronic neurological disorder | Electronic | Once |
| Metastatic solid tumour | Electronic | Once |
| Malignant neoplasm (including leukaemia & lymphoma) | Electronic | Once |
| Diabetes | Electronic | Once |

| | | |
|---|---|---|
| Obesity (Body mass index >30) | Electronic | Once |
| Acquired immune deficiency syndrome/Human immunodeficiency virus infection | Electronic | Once |
| Rheumatological disorder | Electronic | Once |
| Dementia | Electronic | Once |
| Other diagnoses including immunosuppression | Notes/electronic | Once |
| **Vital signs at admission to UHSFT care** | | |
| Temperature (degrees Celsius) | Electronic | Continuous |
| Heart Rate (beats per minute) | Electronic | Continuous |
| Respiratory Rate (breaths per min) | Electronic | Continuous |
| Systolic blood pressure (mmHg) | Electronic | Continuous |
| Diastolic blood pressure (mmHg) | Electronic | Continuous |
| Oxygen Saturations (%) | Electronic | Continuous |
| Fraction of inspired oxygen (%) | Electronic | Continuous |
| **Symptoms at admission to UHSFT care** | | |
| Anosmia | Notes | Once |
| History of fever | Notes | Once |
| Cough | Notes | Once |
| Shortness of breath | Notes | Once |
| Myalgia | Notes | Once |
| Arthralgia | Notes | Once |
| Fatigue/Malaise | Notes | Once |
| Headache | Notes | Once |
| Altered consciousness/confusion | Notes | Once |
| Cold symptoms (Sore throat, rhinitis, ear pain) | Notes | Once |
| Wheezing | Notes | Once |
| Chest pain | Notes | Once |
| Seizures | Notes | Once |
| Abdominal pain | Notes | Once |
| Diarrhoea | Notes | Once |
| Vomiting | Notes | Once |
| Conjunctivitis | Notes | Once |
| Skin Rash | Notes | Once |
| Ulcers | Notes | Once |
| Lymphadenopathy | Notes | Once |
| **Laboratory results** | | |
| Haemoglobin (g/L) | Electronic | Daily |
| White blood cell count ($10^9$/L) | Electronic | Daily |
| Neutrophil count ($10^9$/L) | Electronic | Daily |
| Lymphocyte count ($10^9$/L) | Electronic | Daily |
| Eosinophil count ($10^9$/L) | Electronic | Daily |
| Platelet count ($10^9$/L) | Electronic | Daily |
| International normalized ratio | Electronic | Daily |
| Alanine aminotransferase (units/L) | Electronic | Daily |
| Aspartate transaminase (units/L) | Electronic | Daily |
| Bilirubin (µmol/L) | Electronic | Daily |
| Alkaline phosphatase (units/L) | Electronic | Daily |
| Sodium (mmol/L) | Electronic | Daily |
| Potassium (mmol/L) | Electronic | Daily |
| Urea (mmol/L) | Electronic | Daily |
| Creatinine (mmol/L) | Electronic | Daily |
| Glucose (mmol/L) | Electronic | Daily |
| C-Reactive Protein (mg/L) | Electronic | Daily |
| Lactate Dehydrogenase (units/L) | Electronic | Daily |

| | | |
|---|---|---|
| Ferritin (ng/L) | Electronic | Daily |
| Triglycerides | Electronic | Daily |
| D-dimer | Electronic | Daily |
| Troponin (ng/L) | Electronic | Daily |
| Vitamin D | Electronic | Daily |
| **Arterial Blood Gas measurements**<br>Fraction of inspired oxygen (%)<br>Partial pressure of oxygen (kPa)<br>Partial pressure of carbon dioxide (kPa)<br>pH<br>Bicarbonate<br>Base excess<br>Lactate | Electronic | Daily |
| **Chest X-ray on admission to UHSFT care** | | |
| Are infiltrates present? | Manual | Daily |
| Check all quadrants where infiltrates are present | Manual | Daily |
| Chest X-ray report | | |
| **Computer tomography scan of chest under UHSFT care** | | |
| Ground glass opacification (bilateral? Location e.g. peripheral, basal) | Manual | Daily |
| Consolidation<br>(Bilateral? Location e.g. peripheral, basal) | Manual | Daily |
| Bronchovascular thickening within lesions? | Manual | Daily |
| Computer tomography report | Manual | Daily |
| **Pathogen testing** | | |
| Nasopharyngeal swab (Sars-Cov2) Positive/Negative & viral titre | Electronic | Daily |
| Nasopharyngeal swab respiratory virus panel Positive/Negative & viral titre | Electronic | Daily |
| COVID-19 serology | Electronic | Daily |
| Sputum microscopy culture & sensitivity | Electronic | Daily |
| Blood culture | Electronic | Daily |
| Urine culture | Electronic | Daily |
| **Medication** | | |
| Each individual medication:<br>Name, dose, frequency, start date, stop date | Electronic | Daily |
| **Data to be collected from patients not requiring admission to UHSFT** | | |
| Online survey via my Medical Record | Electronic online | Daily until discharge or symptoms resolve (maximum 28 days) |
| **Level 2/3 (data from MetaVision -** a clinical information system for critical care) | | |
| Invasive/Non-invasive ventilation | Electronic | Hourly |
| Ventilator Mode | Electronic | Hourly |
| Respiratory Rate | Electronic | Hourly |
| Tidal volume | Electronic | Hourly |
| Positive end expiratory pressure | Electronic | Hourly |
| Hours prone | Electronic | Hourly |
| Fraction of inspired oxygen (from ABG) | Electronic | 3 Hourly |
| Partial pressure of oxygen (from ABG) | Electronic | 3 Hourly |
| Renal replacement therapy | Electronic | Daily |
| Procalcitonin | Electronic | Every 2 days |
| proADM | Electronic | Every 2 days |
| Number of inotropes | Electronic | Continuous |
| Steroids (Y/N) | Electronic | Daily |
| Antivirals (Y/N) | Electronic | Daily |

| | | |
|---|---|---|
| Heart rate (bpm) | Electronic | Hourly |
| Blood pressure (mean) | Electronic | Hourly |
| Fluid balance/ 24 hours | Electronic | Daily |
| Temp | Electronic | Hourly |
| Glasgow coma scale | Electronic | Hourly |
| Richmond Agitation-Sedation Scale score | Electronic | Hourly |
| Confusion Assessment Method for the ICU score | Electronic | Hourly |
| **OUTCOMES** | | |
| **Secondary diagnosis** | | |
| Bacterial pneumonia | Manual | Once |
| Acute respiratory distress syndrome | Manual | Once |
| Cytokine Release Syndrome | Manual | Once |
| Acute kidney injury | Manual | Once |
| Acute liver injury | Manual | Once |
| Myocarditis | Manual | Once |
| Cardiomyopathy | Manual | Once |
| Pulmonary Embolus | Manual | Once |
| Stroke | Manual | Once |
| **Treatment during admission** | | |
| Antiviral agent Y/N | Electronic | Once |
| Antibiotic Y/N<br>Route of administration | Electronic | Once |
| Oral Corticosteroid Y/N | Electronic | Once |
| IV Corticosteroid Y/N | Electronic | Once |
| ACE inhibitor/ARB Y/N | Electronic | Once |
| Statin Y/N | Electronic | Once |
| Level 2/3 admission Y/N | Electronic | Once |
| Treatment Escalation Plan Ceiling of Care | Manual | Once |
| Oxygen therapy<br>- date started<br>- duration of treatment | Electronic | Once |
| Non-invasive ventilation<br>- date started<br>- duration of treatment | Electronic | Once |
| Non-invasive ventilation<br>- date started<br>- duration of treatment | Electronic | Once |
| Invasive ventilation<br>- date started<br>- duration of treatment | Electronic | Once |
| Renal replacement therapy<br>- date started<br>- duration of treatment | Electronic | Once |
| Inotropes<br>- date started<br>- duration of treatment | Electronic | Once |
| Plasma exchange Y/N | Electronic | Once |
| IV Immunoglobulin Y/N | Electronic | Once |
| **Destination of discharge from UHSFT care** | | |
| Palliative discharge | Electronic | Once |
| Alive at discharge | Electronic | Once |
| Death | Electronic | Once |
| 28-day mortality | Electronic | Once |
| Readmission within 28 days | Electronic | Once |
| Post COVID-19 complications | Manual | Once |

| Clinic Follow up for COVID-19 complications | | |
|---|---|---|
| Medication changes | Manual | Once |
| Long term oxygen requirements | Manual | Once |
| Pathology results | Manual | Once |
| Chest X-ray and Computer tomography chest scan results | Manual | Once |
| Recorded secondary diagnosis | Manual | Once |
| Exercise tolerance | Manual | Once |

Table 1: Data points to be captured and frequency of capture.

## Biobanking

A sub-cohort of patients ≥16 years old are given the opportunity to provide consent for storage of excess sample taken as part of their routine clinical management with optional further biological sampling (i.e. blood, urine, induced sputum, nasal samples, exhaled breath or bronchial wash samples, see Table 2). They are offered the opportunity to provide this at any point during their care under UHSFT. Care is taken to offer this study to participants after or alongside recruitment to the national NIHR Urgent Public Health Priority observational studies (ISARIC-CPP, DIAMONDS, NIHR Bioresource). These samples are stored within the SRB (managed under the UoS Human Tissue Act Licence) and analysed accordingly to enable biomarker measures related to COVID-19, including genomics, transcriptomics, proteomics, lipidomics and biochemical studies. These measures and the clinical data are reciprocally iterative and inform knowledge as they feed into the clinical care in determining treatment pathways and the response to treatment, as well as the natural history of SARS-CoV-2 infection. Samples are stored as coded samples, without participant names, that can be linked to clinical data (linked anonymised).

| BIOLOGICAL SAMPLE | PROPOSED ANALYSIS |
|---|---|
| Blood | Whole blood sample for genotyping (10 mL EDTA tube), serum (5 x 10 mL serum tube), plasma (10 mL EDTA tube), mRNA (PAXgene) (Total ~80 mL). In addition to these blood samples taken universally, subgroups of participants may be invited to donate a further sample of blood (up to 100mL) for isolation of immune cells for *ex vivo* experimental study and for gene array. Serial sampling may be required (up to once a day), however no more than 210 mL will be taken within a two-week period. |
| Urine | Up to 50 mL of urine may be taken for characterization of biomarkers of disease. |
| Saliva | Up to 5 samples within a 14 day period |
| Nasal samples | Nasal brushing (for microbial carriage analysis and culture of nasal epithelial cells) and nasosorption for mucosal sampling (for inflammatory indices) may be performed as part of the study. Serial sampling may be required, but no more than one nasal brushing and six nasosorption samples will be taken in a two-week period. |
| Exhaled breath samples | Exhaled breath condensate may be collected for assessment of volatile compounds which may influence disease progression. Serial sampling may be required, but no more than two samples will be taken within a 2-week period. |
| Sputum samples | Excess sputum samples collected as part of routine clinical practice may be processed and analysed for presence of microbes as well as inflammatory cell infiltrate, gene expression and other immune/inflammatory indices. |
| Bronchoscopy samples | Non-directed bronchial wash via a suction catheter may be performed in intubated and ventilated patients only. Samples will be processed for microbial analysis, cell culture and molecular biology. Serial sampling may be required, but no more than one sample per day will be taken. |

Table 2: Biological samples to be collected for storage within the Southampton Research

Biorepository (SRB)

## Data management

Data is captured longitudinally, with change over time treated as explicit. After admission, data capture will occur longitudinally in parallel with participant ongoing clinical follow-up for up to 12 months of their discharge following their acute COVID-19 admission. Active participation in the database would end 12 months after the participant was discharged from UHSFT following their acute admission with COVID 19 or if there was no further involvement of UHSFT clinical teams in their care before this time. The Database will continue to aquire data until no further patients are seen at UHSFT with COVID 19 and all patients have been discharged from UHSFT care. Storage of data for research purposes will be in an anonymised format and any data sharing outside of the clinical team will be in an anonymised format. Clinical and research data collected from the cohort in the study are kept in a highly secure contemporary encrypted data platform BC|INSIGHT (within the Clinical Informatics Research Unit, UoS) that was set up in a Microsoft Data Centre in South UK.

## Missing data

The majority of data capture is electronic and therefore we expect missing data to be minimal. For all data other than symptom onset, any missing data will be characterised as 'missing' and excluded from analysis. In keeping with the data capture running in parallel with clinical care rather than via independent research 'visits', the data collection may not include certain variables for all patients if the capture of that information or a sample subset was not clinically indicated.

For symptom onset date, which is utilised for the purposes of analysis as day 0, a valid date is necessary. Data will be extracted manually from the medical records and classified as the following:

1. Clear date
2. Unclear: in this scenario, admission date will be used
3. Nosocomial infection (acquired whilst in hospital): Date of 1st + swab will be used if symptom onset date is not otherwise available

These three groups will be identified within the data set and analysed for significant differences prior to assessment as a whole cohort for a specific outcome.

Analysis of the whole dataset will be both cross-sectional and longitudinal according to specific study questions. Specific datasets can be extracted from the database to investigate specific hypothesis.

This cohort can also be interrogated for inclusion criteria towards additional trials occurring in parallel to the REACT COVID study. Participation in clinical trials will not exclude patients from the cohort, but may omit patients from some analysis depending on the research question.

## Statistical Analysis

A variety of statistical methods will be employed to support a range of assessment needs. Associations to potential risk factors will also be tested via multiple logistic regression analysis. Statistical significance levels will be set at $p<0.05$.

Cluster analysis methodology will also be used to determine the heterogeneous nature of COVID-19 through an unbiased approach. Tests of correlation will be used to ensure that

excessive collinearity of cluster variables does not bias the clustering process. Associations of disease clusters/phenotypes with cluster variables across resulting clusters will use ANOVA for continuous variables and Chi-Square tests for binary variables. Subsequent tests of association to identify patterns of disease morbidity for disease clusters will be conducted. Further tests of association for disease clusters with potential epidemiological and pathophysiological risk factors will also be undertaken.

To facilitate artificial intelligence and machine intelligence learning from this dataset, the following will be undertaken

1. Mapping summary statistical data from existing intensive care unit studies (Observational Health Data Sciences & Informatics (OHDSI) and similar) as well as demographics data, using it to contextualise and enrich the patterns observed for COVID-19 participants at UHSFT.
    a. The contextualisation component would use existing studies as a reference frame to contrast and compare similarities and differences among UHSFT and external patient cohorts.
    b. The enrichment component would use existing studies and demographics as priors to support Bayesian inference over UHSFT data.
2. Co-design and evaluate a Bayesian inference model
    a. With a possible extension of the Bayesian framework into a causal inference setting
    b. Provide a critical analysis of the ethical and safety aspects related to the clinical application of Bayesian inference, i.e. supporting the clinical decision-making process.

3. Other innovative artificial intelligence and machine learning methods will also be explored to identify and predict those patients who have worse outcomes.

## Ethics and dissemination

Ethics approval for the study was obtained from HRA Specific Review Board (REC 20/HRA/2986) for waiver of informed consent for the database only cohort. The database will be conducted in accordance with the principles of good clinical practice (GCP). Database documents (paper and electronic) will be collected and retained in accordance with the General Data Protection Regulations 2018 in a secure location during and after the trial has finished. All essential documents including source documents will be retained for a minimum period of 5 years following the end of the study. Clinical and research data collected in the database will be kept in a highly secure contemporary encrypted data platform BC|INSIGHT that was set up in a Microsoft Data Centre in South UK. Access to the study's data is only given to study team members that have signed the delegation logs. Data access is via an encrypted web service with username/password type authentication. The data held in the server are backed-up weekly using the snapshot technology provided by Microsoft Azure and the retention period of the snapshot storage is 2 years. Storage of data for research purposes will be in an anonymised format and any data sharing outside of the clinical team will be in an anonymised format.

Written informed consent will be obtained from all study participants who agree to biobanking of excess clinical sample +/-additional sampling, as per approvals for amendment to the SRB reviewed by North West Research Ethics Committee (REC 17/NW/0632). For patients who have impaired cognition, informed consent will be sought from their legally

acceptable representative, with retrospective consent sought for those who regain capacity to consent following recovery from acute illness. The findings from this study will be disseminated locally and internationally through manuscript publications in peer-reviewed journals.

## Discussion

SARS-CoV-2 infection is presenting a heterogenous pattern of disease, with varying presentations from asymptomatic carriage, as demonstrated amongst healthcare workers (7) to severe disease with mortality rates of 50% in patients requiring intensive care (8). Emerging understanding is identifying some co-morbidities and risk factors for progression to severe disease, but there is still a wide breadth of severity amongst patients displaying the same risk profile.

The longitudinal, mainly automated data collection of the REACT COVID study enables a granularity of data that has potential to identify early markers of disease severity and the relevance of their trajectory of change on disease progression. The REACT COVID database can be used to rapidly and easily identify participants who may be appropriate for clinical trials and offers visual interpretation of large amounts of data for hypothesis generation.

The database provides a rich dataset to support the biobanking sub-cohort, within which there is the potential to recruit to smaller, focussed basic science studies investigating the mechanisms driving disease processes resulting from infection with SARS-CoV-2.

The REACT COVID study is mindful of the national and international efforts to bring together data from multiple centres in order to enhance the power of studies, be they clinical,

observation, mechanistic, or intervention studies. The study will therefore be aligning itself with the broader UHS and UoS research project of ENACT (Enabling new treatment approaches for COVID-19 Treatment), supported by the Southampton Coronavirus Support Fund, NIHR Southampton BRC and NIHR Southampton CRF.

The ability of the REACT COVID study to capture the granularity of the longitudinal path of clinical care is both a major strength and a potential limitation. There is no strict study 'visit' protocol to follow, with data capture occurring in a more fluid manner. Therefore, there is the potential for a greater number of incomplete data sets, with some patients having a greater amount of available data. This is likely to be reflective of the level of care required, such as in those patients who required intensive care support. Any analysis will allow this, and populations may be sub-cohorted to answer specific questions if the data is only available for a specific group.

## Summary

The simplicity of data visualisation within the REACT COVID study belies the complexity of the concepts behind this project, facilitated through the unique collaboration between UHSFT, the UoS and the digital ECMT. This collaboration possesses the requisite expertise in their respective specialisms to deliver all aspects of this study, alongside the infrastructure provided by through the NIHR Southampton CRF and NIHR Southampton BRC. The REACT COVID study enables visualisation of detailed clinical data and the application of novel artificial intelligence methods, which will allow greater understanding of the natural history of this novel disease through both longitudinal and cross-sectional analysis.

## Abbreviations

| | |
|---|---|
| ACCORD2 | Accelerating Covid-19 Research and Development |
| BRC | Biomedical Research Centre |
| CBC | Cancer Biomarker Centre |
| CIRU | Clinical Informatics Research Unit |
| COVID-19 | Coronavirus disease 2019 |
| CRF | Clinical Research Facility |
| CRUK MI | Cancer Research UK, Manchester Institute |
| digital ECMT | digital Experimental Cancer Medicine Team |
| ENACT | Enabling new treatment approaches for COVID-19 Treatment |
| GCP | Good clinical practice |
| NIHR | National Institute of Health Research |
| REACT | REal-time Analytics for Clinical Trials |
| REACT COVID study | Research Evaluation Alongside Clinical Treatment in COVID-19 |
| REC | Research Ethics Committee |
| RECOVERY | Randomised Evaluation of COVID-19 Therapy |
| ODAR | Office of Development and Alumni Relations |
| OHDSI | Observational Health Data Sciences & Informatics |
| SARS-CoV-2 | Severe Acute Respiratory Syndrome coronavirus 2 |
| SRB | Southampton Research Biorepository |
| UHSFT | University Hospital Southampton NHS Foundation Trust |
| UoM | University of Manchester |
| UoS | University of Southampton |

# Declarations

## Ethics approval and consent to participate

The study design, protocol and patient facing documentation have been approved by North West Research Ethics Committee (REC 17/NW/0632) as an amendment to the National Institute of Health Research (NIHR) Southampton Clinical Research Facility (CRF) - managed Southampton Research Biorepository (SRB). No patient identifiable data is included in this manuscript.

## Consent for publication

Not applicable

## Availability of data and materials

Not applicable

## Competing interests

The authors have no competing interests to declare in relation to this manuscript

## Funding

The REACT platform has been supported by the digital Experimental Cancer Medicine Team (digital ECMT) free of charge. The biobanking subcohort is supported the NIHR Southampton CRF and NIHR Southampton Biomedical Research Centre (BRC) at University Hospital Southampton NHS Foundation Trust (UHSFT) and as part of a broader effort (Enabling new treatment approaches for COVID-19 Treatment - ENACT) by the University of Southampton

(UoS) charity (Office of Development and Alumni Relations - ODAR). In addition the Clinical Informatics Research Unit, UoS has supported infrastructure costs.

## Authors' contributions

HB and AF designed the protocol and drafted the manuscript; AD and MC were involved in protocol design, JB, HP and FB were involved realisation of data extraction, integration, transformation and upload processes, CK, GT and SNF were involved in protocol design for the biobanking subcohort, NS and SW are involved in manual data collection processes, DL, PF, HP and JB were involved in the design and adaptation of the REal-time Analytics for Clinical Trials (REACT) platform and TW was involved in study conception and protocol design. All authors have reviewed the final manuscript.

## Acknowledgements

The authors would like to acknowledge the NIHR Southampton CRF and UHSFT Research Nursing teams for their support in the set-up of this study, Dr Dave Stockley (NIHR CRF SRB Manager) and Dr Alastair Watson for his help in proof reading and preparation of the manuscript for submission.

## References


1. World Health Organisation. WHO Coronavirus Disease (COVID-19) Dashboard. 2020. https://covid19.who.int/?gclid=EAIaIQobChMIz5Xu84OJ6gIVCLrtCh3WlQ8KEAAYASAAEgIohvD_BwE. Accessed 16th June 2020.

2. UK Governement Guidance. 2020. https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public. Accessed 18th June 2020.

3. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19 - Preliminary Report. N Engl J Med. 2020.

4. Oxford University News Release. Low-cost dexamethasone reduces death by up to one third in hospitalised patients with severe respiratory complications of COVID-19. 2020. https://www.recoverytrial.net/files/recovery_dexamethasone_statement_160620_v2final.pdf. 16th June 2020.

5. ACCORD ACcelerating COVID-19 Research and Development. 2020. https://www.nihr.ac.uk/news/national-collaborative-platform-launched-to-fast-track-potential-covid-19-treatments/24678. Accessed 18th June 2020.

6. RECOVERY. 2020. https://www.recoverytrial.net. Accessed 18th June 2020.

7. Rivett L, Sridhar S, Sparkes D, Routledge M, Jones NK, Forrest S, et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. Elife. 2020;9.

8. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. Bmj. 2020;369:m1985.

## Figure Legends

**Figure 1: Data flow between systems and institutions within the REACT COVID database.**

COVID-19, Coronavirus disease 2019; CRUK MI, Cancer Research UK, Manchester Institute; digital ECMT, digital Exp erimental Cancer Medicine Team; REACT, REal-time Analytics for Clinical Trials; SARS-CoV-2, Severe Acute Respiratory Syndrome coronavirus 2; UHSFT, University Hospital Southampton NHS Foundation Trust.

**Figure 2: Data capture for all patients to be included in the database .**

Data variables will be recorded as part of routine clinical care and will be captured as part of the observational study. Data variables listed in italics will be manually captured and added to the database.

ABG, arterial blood gas; ACE, angiotensin converting enzyme; AKI, acute kidney injury; ARB, angiotensin receptor blocker; ARDS, acute respiratory distress syndrome; CHARTS, is an integrated user interface for electronic health care records; COVID-19, Coronavirus disease 2019; CPAP, continuous positive airways pressure; CRP, C-reactive protein; CRS, Cytokine release syndrome; HR, Heart rate; FBC, Full blood count; Obs, Clinical observations; MetaVision, is a clinical information system for critical care; NP, Nasopharyngeal; PEEP, Positive end expiratory pressure; proADM, pro-adrenomedullin; RT-PCR, Reverse transcriptase polymerase chain reaction; RR, Respiratory rate; Temp, Temperature; TV, Tidal volume; Imaging findings from Chest X-ray (CXR), Computed Tomography (CT) or Magnetic Resonance Imaging (MRI); UHSFT, University Hospital Southampton NHS Foundation Trust.

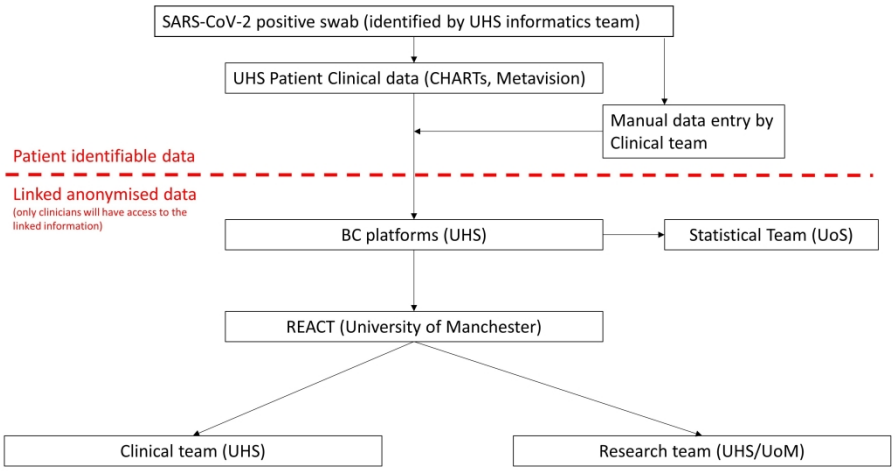


Figure 1: Data flow between systems and institutions within the REACT COVID database. COVID-19, Coronavirus disease 2019; CRUK MI, Cancer Research UK, Manchester Institute; digital ECMT, digital Experimental Cancer Medicine Team; REACT, REal-time Analytics for Clinical Trials; SARS-CoV-2, Severe Acute Respiratory Syndrome coronavirus 2; UHSFT, University Hospital Southampton NHS Foundation Trust.

338x190mm (300 x 300 DPI)

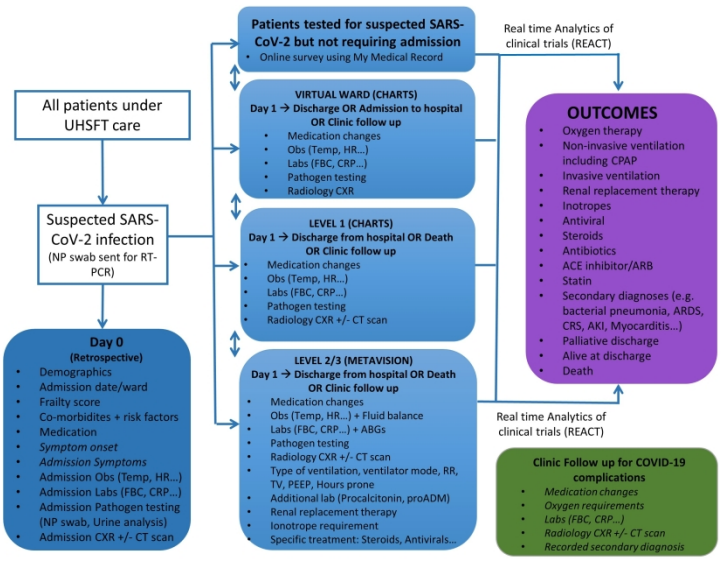


Figure 2: Data capture for all patients to be included in the database .Data variables will be recorded as part of routine clinical care and will be captured as part of the observational study. Data variables listed in italics will be manually captured and added to the database.ABG, arterial blood gas; ACE, angiotensin converting enzyme; AKI, acute kidney injury; ARB, angiotensin receptor blocker; ARDS, acute respiratory distress syndrome; CHARTS, is an integrated user interface for electronic health care records; COVID-19, Coronavirus disease 2019; CPAP, continuous positive airways pressure; CRP, C-reactive protein; CRS, Cytokine release syndrome; HR, Heart rate; FBC, Full blood count; Obs, Clinical observations; MetaVision, is a clinical information system for critical care; NP, Nasopharyngeal; PEEP, Positive end expiratory pressure; proADM, pro-adrenomedullin; RT-PCR, Reverse transcriptase polymerase chain reaction; RR, Respiratory rate; Temp, Temperature; TV, Tidal volume; Imaging findings from Chest X-ray (CXR), Computed Tomography (CT) or Magnetic Resonance Imaging (MRI); UHSFT, University Hospital Southampton NHS Foundation Trust.

338x190mm (300 x 300 DPI)

BMJ Open

# Research Evaluation Alongside Clinical Treatment in COVID 19 (REACT COVID 19): An Observational and Biobanking Study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-043012.R1 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 09-Dec-2020 |
| Complete List of Authors: | Burke, Hannah; University of Southampton Faculty of Medicine<br>Freeman, Anna; University of Southampton Faculty of Medicine,<br>Dushianthan, Ahilanandan; University of Southampton Faculty of Medicine<br>Celinski, Michael; University of Southampton Faculty of Medicine<br>Batchelor, James; University of Southampton Faculty of Medicine<br>Phan, Hang; University of Southampton Faculty of Medicine<br>Borca, Florina; University of Southampton Institute for Life Sciences<br>Kipps, Christopher ; University of Southampton<br>Thomas, Gareth; Faculty of Medicine, Cancer Sciences Unit<br>Faust, Saul; University of Southampton, UK, NIHR Wellcome Trust Clincial Research Facility<br>Sheard, Natasha; University Hospital Southampton NHS Foundation Trust<br>Williams, Sarah; University Hospital Southampton NHS Foundation Trust<br>Fitzpatrick, Paul; University of Manchester, Cancer Biomarker Centre, Cancer Research UK Manchester Institute<br>Landers, Dónal; University of Manchester, Cancer Biomarker Centre, Cancer Research UK Manchester Institute<br>Wilkinson, Tom; University of Southampton, Faculty of Medicine |
| Primary Subject Heading: | Respiratory medicine |
| Secondary Subject Heading: | Infectious diseases |
| Keywords: | COVID-19, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Chronic airways disease < THORACIC MEDICINE, Respiratory infections < THORACIC MEDICINE, RESPIRATORY MEDICINE (see Thoracic Medicine) |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Research Evaluation Alongside Clinical Treatment in COVID 19 (REACT COVID 19): An Observational and Biobanking Study

Burke, H.#[1,3], Freeman, A.#*[1,3], Dushianthan, A.[1,3], Celinski, M.[1,3], Batchelor, J.[1,2,3,6], Phan, H.[1,4,6], Borca, F.[2,4], Kipps, C.[1,2,3,4], Thomas, G.J. [1,2,3,4], Faust, S.N.[1,2,3,4,5], Sheard, N.[3], Williams, S.[3], Fitzpatrick, P.[7], Landers, D.[7], Wilkinson, T.M.A.[1,3]

# Joint first authors

* Corresponding author: Anna Freeman

Email: a.freeman@soton.ac.uk

Affiliations

1. Faculty of Medicine, University of Southampton.
2. Institute for Life Sciences, University of Southampton.
3. University Hospitals Southampton NHS Foundation Trust.
4. NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust.
5. NIHR Southampton Clinical Research Facility, University Hospital Southampton NHS Foundation Trust.
6. Clinical Informatics Research Unit Faculty of Medicine, University of Southampton.
7. Digital Experimental Cancer Medicine Team, Cancer Biomarker Centre, Cancer Research UK Manchester Institute, The University of Manchester.

Word count: 3,254

# Abstract

## Introduction

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) places immense worldwide demand on healthcare services. Earlier identification of patients at risk of severe disease may allow intervention with experimental targeted treatments, mitigating the course of their disease and reducing critical care service demand.

## Methods and analysis

This prospective observational study of patients tested or treated for SARS-CoV-2, who are under the care of the tertiary University Hospital Southampton NHS Foundation Trust (UHSFT), captures data from admission to discharge; data collection commenced on 7/3/2020. Core demographic and clinical information, as well as results of disease-defining characteristics are captured and recorded electronically from hospital clinical record systems at the point of testing. Manual data is collected and recorded by the clinical research team for assessments which are not part of the structured electronic health care record, for example symptom onset date. Thereafter, participant records are continuously updated during hospital stay and their follow up period. Participants >16 years are given the opportunity to provide consent for excess clinical sample storage with optional further biological sampling. These anonymised samples are linked to the clinical data in the REACT platform and stored within a biorepository at UHSFT.

## Ethics and dissemination

Ethics approval was obtained from the HRA Specific Review Board (REC 20/HRA/2986) for waiver of informed consent for the database only cohort, the procedures conform with the Declaration of Helsinki. The study design, protocol and patient facing documentation for the biobanking arm of the study have been approved by North West Research Ethics Committee (REC 17/NW/0632) as an amendment to the National Institute of Health Research (NIHR) Southampton Clinical Research Facility (CRF) - managed Southampton Research Biorepository (SRB). The study of this study will be published as peer-reviewed articles and presented at conferences, presentations and workshops.

Word count 300

## Strengths and limitations of this study

- Close alignment of research and clinical practice in a near real-time manner.
- Longitudinal data collection and sampling opportunities.
- A single centre study with data collection reflective of clinical need rather than to a strict protocolised timeframe.
- Use of novel AI techniques for data analysis.

**Key words:** COVID-19; Health informatics; Protocols & guidelines; Chronic airways disease; Respiratory infections; Respiratory medicine

# Introduction

The outbreak of coronavirus disease 2019 (COVID-19) caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was declared a pandemic by the World Health Organisation (WHO) on 11th March 2020 (1). In the United Kingdom (UK), 1,629,657 people have tested positive for SARS-CoV-2, of which 66,713 have died (https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public; correct as of 5pm 30th November 2020) (2). So far, emerging effective treatments such as Remdesivir and dexamethasone have only modest effects on clinical outcomes in a sub-cohort of patients (3, 4) and any vaccines have yet to demonstrate their efficacy. Whilst many patients recover from SARS-CoV-2 infection without need for hospitalisation, a small proportion go on to develop severe disease. The demand on healthcare services, especially critical care, is immense. Earlier identification of those patients at risk of severe disease may allow intervention with experimental targeted treatments mitigating the course of their disease and reducing demand on critical care services.

The need for rapid translation of treatments from development to clinical application is being addressed by Urgent Public Health studies such as ACCORD-2 (5) and RECOVERY trials (6). Despite the urgency to enrol patients in these studies, there are challenges in identifying suitable patients in a fluid, “real-time” clinical environment within a novel disease. The longitudinal REACT COVID database and biosampling study captures the natural history of SARS-CoV-2 infection and its clinical response to current interventions, as well as characterising specific phenotypes and endotypes, with the potential to rapidly identify patients for novel treatment trials.

The Research Evaluation Alongside Clinical Treatment observational and biobanking study of COVID-19 (REACT COVID study) evolved around a collaboration between the University of Southampton (UoS), University Hospital Southampton NHS Foundation Trust (UHSFT) and the digital Experimental Cancer Medicine Team (digital ECMT) and their work to create the REACT (REal-time Analytics for Clinical Trials) platform. The REACT platform is used for the determination of early clinical benefit of new cancer medicines in early phase drug trials and has been adapted by the digital ECMT in collaboration with UHSFT to allow rapid upload and interpretation of clinical data in the context of the COVID-19 pandemic. This study is supported by the National Institute for Health and Research (NIHR) Southampton Clinical Research Facility (CRF) and NIHR Southampton Biomedical Research Centre (BRC), which takes advantage of the expertise and resource in hosting large prospective cohort studies.

## Objectives

- Development of a real-time "parent" database of well-characterised COVID-19 patients to facilitate a better understanding of the natural history of SARS-CoV-2 infection from pre-hospital through to level 3 care and subsequent course of COVID-19 related complications from discharge through to clinic follow up.
- Cross-sectional phenotypic characterisation of a cohort of COVID-19 patients to describe the disease heterogeneity seen in clinical practice
- Identify potential risk factors for progression to severe disease measured by admission to level 2 or 3 care, increase in respiratory support or death (primary endpoint).

- Use of samples stored in the biorepository (for example blood, urine, sputum) to develop an endotype level understanding of disease clusters.
- Assess the efficacy of current best practice management strategies.
- Identify subgroups of patients for novel treatment strategies in the form of independent formal randomised control trials.



## Methods and analysis

### Study Design

The REACT COVID study is a prospective observational study of patients tested or treated for SARS-CoV-2 infection who are under the care of UHSFT. The study is planned as a long-term research project with no limit to enrolment nor any time defined end, with data collection commencing on 7/3/2020 with admission of the first Sars-CoV-2 patient to UHS. Patients are recruited into the study at the point of testing for SARS-CoV-2 at UHSFT. Core demographic and clinical information, as well as results of disease-defining characteristics are captured and recorded electronically from hospital clinical record systems at the point of testing. Manual data is collected and recorded by the clinical research team for assessments which are not part of the electronic health care record, for example symptom onset date. Thereafter, participants records are continuously updated during their hospital stay and for 12 months following discharge from acute admission, in their follow up period under UHSFT. This pragmatic, opportunistic approach to data collection takes full advantage of the electronic

health care records and seamlessly links with the REACT platform to reduce data collection burden on the patient, clinician or researcher. In addition, participants over the age of 16 years old are given the opportunity to provide consent for storage of excess sample taken as part of their routine clinical management with optional further biological sampling. These anonymised samples are linked to the anonymised clinical data in the REACT platform and stored within the Southampton Research Biorepository (SRB), managed by the NIHR CRF at UHSFT.

## Patient and Public Involvement

Patient representatives were involved in and are part of the governance structure of the Southampton Research Biorepository and patient representatives were involved in the design and ongoing management of the WATCH study, upon which much of this observational biobank study was based. The SRB Oversight Committee has a lay representative to work alongside other members and may jointly be consulted in the different matters related to the construction and running of the Biorepository.

## Setting

UHSFT is an 1100 bed tertiary centre and as such, has assessed and admitted a large number of SARS-CoV-2 positive patients.

## Recruitment

All patients under the care of UHSFT who are tested or treated for SARS-CoV-2 are included in the REACT COVID study. A sub-cohort of patients ≥16 years old are given the opportunity

to provide consent for storage of excess sample taken as part of their routine clinical management with optional further biological sampling. They are offered the opportunity to provide this at any point during their care under UHSFT. So far UHSFT has admitted 629 patients with confirmed COVID-19, correct as of 17th June 2020. Study enrolment will be continual into the future aligned to the ongoing function of the clinical service with no discrete recruitment target or endpoint.

## Data collection

This study is divided into two parts and is conducted in conjunction with standard clinical care. The database component involves the assimilation of data collected as part of routine clinical care for patients of any age tested for or admitted to UHSFT for suspected SARS-CoV-2 infection. Core demographic and clinical data collected as part of standard clinical care will be extracted from the electronic hospital records. Where required, additional data relevant to COVID-19 clinical course is manually extracted from the clinical notes and uploaded to the database by clinical researchers where automatic electronic export is not possible. Data collected from study participants is collated in the highly secure, contemporary and encrypted data platform, BC|INSIGHT (Microsoft Data Centre, United Kingdom). Data is then uploaded to the REACT platform to allow rapid capture of the natural history of the disease, with clinicians able to visualise trends in patient trajectories and identify patients who may be suitable for approach for intervention studies (summarised in Figure 1.).

Data captured at point of testing for SARS-CoV-2 includes core demographic and clinical data relevant to COVID-19. Subsequent data capture follows clinical disease course and includes

information on medication, vital signs, pathological and radiological measurements. This data is captured in real time during the follow up period under UHSFT care, including any period of hospitalisation, and for 12 months after admission to capture virtual follow up or outpatient clinic follow up data (summarised in Figure 2.). Any treatment interventions as part of routine clinical care are also captured (where possible electronically) and these form part of the visualisations in the REACT platform. Data for research purposes will be stored in an anonymised format, and any data sharing outside of the clinical team caring for the patient at any point in their patient journey will be in an anonymised format.

The majority of data collected is consistent with the data already being collected as part of routine clinical care of patients with COVID-19, therefore seamlessly linking with clinical practice. Clinical data to be captured is listed in Table 1. The REACT database will host a clinically identifiable 'front end' for real time patient management during the acute admission and for up to 12 months post discharge, in which the link anonymised data will be made identifiable, and visible only to the clinical team as determined by a statement of reference. Data for research purposes will be stored in an anonymised format, and any data sharing outside of the clinical team caring for the patient at any point in their patient journey will be in an anonymised format.

| Data variable | Data source | Frequency of capture |
|---|---|---|
| **Demographics** | | |
| Sex at birth (male/female) | Electronic | Once |
| Date of Birth & Age | Electronic | Once |
| Pregnancy status (yes/no) | Electronic | Once |
| Admission weight (kilograms) | Electronic | Once |
| Admission height (metres) | Electronic | Once |
| Ethnicity | Electronic | Once |
| Profession/employment status | Electronic/manual | Once |
| **Disease onset and admission to UHSFT care** | | |
| Symptom onset date | Notes | Once |
| Admission date | Electronic | Once |

| | | |
|---|---|---|
| Admission ward & level of care | Electronic | Once |
| Frailty score | Electronic | Once |
| **Co-morbidities & Risk factors** | | |
| Chronic obstructive pulmonary disease | Electronic | Once |
| Asthma | Electronic | Once |
| Interstitial lung disease | Electronic | Once |
| Bronchiectasis | Electronic | Once |
| Hypertension | Electronic | Once |
| Chronic cardiac disease (not hypertension) | Electronic | Once |
| Chronic kidney disease | Electronic | Once |
| Chronic Liver disease | Electronic | Once |
| Chronic neurological disorder | Electronic | Once |
| Metastatic solid tumour | Electronic | Once |
| Malignant neoplasm (including leukaemia & lymphoma) | Electronic | Once |
| Diabetes | Electronic | Once |
| Obesity (Body mass index >30) | Electronic | Once |
| Acquired immune deficiency syndrome/Human immunodeficiency virus infection | Electronic | Once |
| Rheumatological disorder | Electronic | Once |
| Dementia | Electronic | Once |
| Other diagnoses including immunosuppression | Notes/electronic | Once |
| **Vital signs at admission to UHSFT care** | | |
| Temperature (degrees Celsius) | Electronic | Continuous |
| Heart Rate (beats per minute) | Electronic | Continuous |
| Respiratory Rate (breaths per min) | Electronic | Continuous |
| Systolic blood pressure (mmHg) | Electronic | Continuous |
| Diastolic blood pressure (mmHg) | Electronic | Continuous |
| Oxygen Saturations (%) | Electronic | Continuous |
| Fraction of inspired oxygen (%) | Electronic | Continuous |
| **Symptoms at admission to UHSFT care** | | |
| Anosmia | Notes | Once |
| History of fever | Notes | Once |
| Cough | Notes | Once |
| Shortness of breath | Notes | Once |
| Myalgia | Notes | Once |
| Arthralgia | Notes | Once |
| Fatigue/Malaise | Notes | Once |
| Headache | Notes | Once |
| Altered consciousness/confusion | Notes | Once |
| Cold symptoms (Sore throat, rhinitis, ear pain) | Notes | Once |
| Wheezing | Notes | Once |
| Chest pain | Notes | Once |
| Seizures | Notes | Once |
| Abdominal pain | Notes | Once |
| Diarrhoea | Notes | Once |
| Vomiting | Notes | Once |
| Conjunctivitis | Notes | Once |
| Skin Rash | Notes | Once |
| Ulcers | Notes | Once |
| Lymphadenopathy | Notes | Once |
| **Laboratory results** | | |
| Haemoglobin (g/L) | Electronic | Daily |

| | | |
|---|---|---|
| White blood cell count ($10^9$/L) | Electronic | Daily |
| Neutrophil count ($10^9$/L) | Electronic | Daily |
| Lymphocyte count ($10^9$/L) | Electronic | Daily |
| Eosinophil count ($10^9$/L) | Electronic | Daily |
| Platelet count ($10^9$/L) | Electronic | Daily |
| International normalized ratio | Electronic | Daily |
| Alanine aminotransferase (units/L) | Electronic | Daily |
| Aspartate transaminase (units/L) | Electronic | Daily |
| Bilirubin (µmol/L) | Electronic | Daily |
| Alkaline phosphatase (units/L) | Electronic | Daily |
| Sodium (mmol/L) | Electronic | Daily |
| Potassium (mmol/L) | Electronic | Daily |
| Urea (mmol/L) | Electronic | Daily |
| Creatinine (mmol/L) | Electronic | Daily |
| Glucose (mmol/L) | Electronic | Daily |
| C-Reactive Protein (mg/L) | Electronic | Daily |
| Lactate Dehydrogenase (units/L) | Electronic | Daily |
| Ferritin (ng/L) | Electronic | Daily |
| Triglycerides | Electronic | Daily |
| D-dimer | Electronic | Daily |
| Troponin (ng/L) | Electronic | Daily |
| Vitamin D | Electronic | Daily |
| **Arterial Blood Gas measurements**<br>Fraction of inspired oxygen (%)<br>Partial pressure of oxygen (kPa)<br>Partial pressure of carbon dioxide (kPa)<br>pH<br>Bicarbonate<br>Base excess<br>Lactate | Electronic | Daily |
| **Chest X-ray on admission to UHSFT care** | | |
| Are infiltrates present? | Manual | Daily |
| Check all quadrants where infiltrates are present | Manual | Daily |
| Chest X-ray report | | |
| **Computer tomography scan of chest under UHSFT care** | | |
| Ground glass opacification (bilateral? Location e.g. peripheral, basal) | Manual | Daily |
| Consolidation<br>(Bilateral? Location e.g. peripheral, basal) | Manual | Daily |
| Bronchovascular thickening within lesions? | Manual | Daily |
| Computer tomography report | Manual | Daily |
| **Pathogen testing** | | |
| Nasopharyngeal swab (Sars-Cov2) Positive/Negative & viral titre | Electronic | Daily |
| Nasopharyngeal swab respiratory virus panel Positive/Negative & viral titre | Electronic | Daily |
| COVID-19 serology | Electronic | Daily |
| Sputum microscopy culture & sensitivity | Electronic | Daily |
| Blood culture | Electronic | Daily |
| Urine culture | Electronic | Daily |
| **Medication** | | |
| Each individual medication:<br>Name, dose, frequency, start date, stop date | Electronic | Daily |

| | | |
|---|---|---|
| **Data to be collected from patients not requiring admission to UHSFT** | | |
| Online survey via my Medical Record | Electronic online | Daily until discharge or symptoms resolve (maximum 28 days) |
| **Level 2/3 (data from MetaVision -** a clinical information system for critical care) | | |
| Invasive/Non-invasive ventilation | Electronic | Hourly |
| Ventilator Mode | Electronic | Hourly |
| Respiratory Rate | Electronic | Hourly |
| Tidal volume | Electronic | Hourly |
| Positive end expiratory pressure | Electronic | Hourly |
| Hours prone | Electronic | Hourly |
| Fraction of inspired oxygen (from ABG) | Electronic | 3 Hourly |
| Partial pressure of oxygen (from ABG) | Electronic | 3 Hourly |
| Renal replacement therapy | Electronic | Daily |
| Procalcitonin | Electronic | Every 2 days |
| proADM | Electronic | Every 2 days |
| Number of inotropes | Electronic | Continuous |
| Steroids (Y/N) | Electronic | Daily |
| Antivirals (Y/N) | Electronic | Daily |
| Heart rate (bpm) | Electronic | Hourly |
| Blood pressure (mean) | Electronic | Hourly |
| Fluid balance/ 24 hours | Electronic | Daily |
| Temp | Electronic | Hourly |
| Glasgow coma scale | Electronic | Hourly |
| Richmond Agitation-Sedation Scale score | Electronic | Hourly |
| Confusion Assessment Method for the ICU score | Electronic | Hourly |
| **OUTCOMES** | | |
| **Secondary diagnosis** | | |
| Bacterial pneumonia | Manual | Once |
| Acute respiratory distress syndrome | Manual | Once |
| Cytokine Release Syndrome | Manual | Once |
| Acute kidney injury | Manual | Once |
| Acute liver injury | Manual | Once |
| Myocarditis | Manual | Once |
| Cardiomyopathy | Manual | Once |
| Pulmonary Embolus | Manual | Once |
| Stroke | Manual | Once |
| **Treatment during admission** | | |
| Antiviral agent Y/N | Electronic | Once |
| Antibiotic Y/N<br>Route of administration | Electronic | Once |
| Oral Corticosteroid Y/N | Electronic | Once |
| IV Corticosteroid Y/N | Electronic | Once |
| ACE inhibitor/ARB Y/N | Electronic | Once |
| Statin Y/N | Electronic | Once |
| Level 2/3 admission Y/N | Electronic | Once |
| Treatment Escalation Plan Ceiling of Care | Manual | Once |
| Oxygen therapy<br>- date started<br>- duration of treatment | Electronic | Once |
| Non-invasive ventilation<br>- date started<br>- duration of treatment | Electronic | Once |

| | | |
|---|---|---|
| Non-invasive ventilation<br>- date started<br>- duration of treatment | Electronic | Once |
| Invasive ventilation<br>- date started<br>- duration of treatment | Electronic | Once |
| Renal replacement therapy<br>- date started<br>- duration of treatment | Electronic | Once |
| Inotropes<br>- date started<br>- duration of treatment | Electronic | Once |
| Plasma exchange Y/N | Electronic | Once |
| IV Immunoglobulin Y/N | Electronic | Once |
| **Destination of discharge from UHSFT care** | | |
| Palliative discharge | Electronic | Once |
| Alive at discharge | Electronic | Once |
| Death | Electronic | Once |
| 28-day mortality | Electronic | Once |
| Readmission within 28 days | Electronic | Once |
| Post COVID-19 complications | Manual | Once |
| **Clinic Follow up for COVID-19 complications** | | |
| Medication changes | Manual | Once |
| Long term oxygen requirements | Manual | Once |
| Pathology results | Manual | Once |
| Chest X-ray and Computer tomography chest scan results | Manual | Once |
| Recorded secondary diagnosis | Manual | Once |
| Exercise tolerance | Manual | Once |

Table 1: Data points to be captured and frequency of capture.

## Biobanking

A sub-cohort of patients ≥16 years old are given the opportunity to provide consent for storage of excess sample taken as part of their routine clinical management with optional further biological sampling (i.e. blood, urine, induced sputum, nasal samples, exhaled breath or bronchial wash samples, see Table 2). They are offered the opportunity to provide this at any point during their care under UHSFT. Care is taken to offer this study to participants after or alongside recruitment to the national NIHR Urgent Public Health Priority observational studies (ISARIC-CPP, DIAMONDS, NIHR Bioresource). These samples are stored within the SRB (managed under the UoS Human Tissue Act Licence) and analysed accordingly to enable

biomarker measures related to COVID-19, including genomics, transcriptomics, proteomics, lipidomics and biochemical studies. These measures and the clinical data are reciprocally iterative and inform knowledge as they feed into the clinical care in determining treatment pathways and the response to treatment, as well as the natural history of SARS-CoV-2 infection. Samples are stored as coded samples, without participant names, that can be linked to clinical data (linked anonymised).

| BIOLOGICAL SAMPLE | PROPOSED ANALYSIS |
|---|---|
| Blood | Whole blood sample for genotyping (10 mL EDTA tube), serum (5 x 10 mL serum tube), plasma (10 mL EDTA tube), mRNA (PAXgene) (Total ~80 mL). In addition to these blood samples taken universally, subgroups of participants may be invited to donate a further sample of blood (up to 100mL) for isolation of immune cells for *ex vivo* experimental study and for gene array. Serial sampling may be required (up to once a day), however no more than 210 mL will be taken within a two-week period. |
| Urine | Up to 50 mL of urine may be taken for characterization of biomarkers of disease. |
| Saliva | Up to 5 samples within a 14 day period |
| Nasal samples | Nasal brushing (for microbial carriage analysis and culture of nasal epithelial cells) and nasosorption for mucosal sampling (for inflammatory indices) may be performed as part of the study. Serial sampling may be required, but no more than one nasal brushing and six nasosorption samples will be taken in a two-week period. |
| Exhaled breath samples | Exhaled breath condensate may be collected for assessment of volatile compounds which may influence disease progression. Serial sampling may be required, but no more than two samples will be taken within a 2-week period. |
| Sputum samples | Excess sputum samples collected as part of routine clinical practice may be processed and analysed for presence of microbes as well as inflammatory cell infiltrate, gene expression and other immune/inflammatory indices. |
| Bronchoscopy samples | Non-directed bronchial wash via a suction catheter may be performed in intubated and ventilated patients only. Samples will be processed for microbial analysis, cell culture and molecular biology. Serial sampling may be required, but no more than one sample per day will be taken. |

Table 2: Biological samples to be collected for storage within the Southampton Research

Biorepository (SRB)

## Data management

Data is captured longitudinally, with change over time treated as explicit. After admission, data capture will occur longitudinally in parallel with participant ongoing clinical follow-up for up to 12 months of their discharge following their acute COVID-19 admission. Active participation in the database would end 12 months after the participant was discharged from UHSFT following their acute admission with COVID 19 or if there was no further involvement of UHSFT clinical teams in their care before this time. The Database will continue to aquire data until no further patients are seen at UHSFT with COVID 19 and all patients have been discharged from UHSFT care. Storage of data for research purposes will be in an anonymised format and any data sharing outside of the clinical team will be in an anonymised format. Clinical and research data collected from the cohort in the study are kept in a highly secure contemporary encrypted data platform BC|INSIGHT (within the Clinical Informatics Research Unit, UoS) that was set up in a Microsoft Data Centre in South UK.

## Missing data

The majority of data capture is electronic and therefore we expect missing data to be minimal. For all data other than symptom onset, any missing data will be characterised as 'missing' and excluded from analysis. In keeping with the data capture running in parallel with clinical care rather than via independent research 'visits', the data collection may not include certain variables for all patients if the capture of that information or a sample subset was not clinically indicated.

For symptom onset date, which is utilised for the purposes of analysis as day 0, a valid date is necessary. Data will be extracted manually from the medical records and classified as the following:

1. Clear date
2. Unclear: in this scenario, admission date will be used
3. Nosocomial infection (acquired whilst in hospital): Date of 1st + swab will be used if symptom onset date is not otherwise available

These three groups will be identified within the data set and analysed for significant differences prior to assessment as a whole cohort for a specific outcome.

Analysis of the whole dataset will be both cross-sectional and longitudinal according to specific study questions. Specific datasets can be extracted from the database to investigate specific hypothesis.

This cohort can also be interrogated for inclusion criteria towards additional trials occurring in parallel to the REACT COVID study. Participation in clinical trials will not exclude patients from the cohort, but may omit patients from some analysis depending on the research question.

## Statistical Analysis

A variety of statistical methods will be employed to support a range of assessment needs. Associations to potential risk factors will also be tested via multiple logistic regression analysis. Statistical significance levels will be set at $p<0.05$.

Cluster analysis methodology will also be used to determine the heterogeneous nature of COVID-19 through an unbiased approach. Tests of correlation will be used to ensure that excessive collinearity of cluster variables does not bias the clustering process. Associations of disease clusters/phenotypes with cluster variables across resulting clusters will use ANOVA for continuous variables and Chi-Square tests for binary variables. Subsequent tests of association to identify patterns of disease morbidity for disease clusters will be conducted. Further tests of association for disease clusters with potential epidemiological and pathophysiological risk factors will also be undertaken.

To facilitate artificial intelligence and machine intelligence learning from this dataset, the following will be undertaken

1. Mapping summary statistical data from existing intensive care unit studies (Observational Health Data Sciences & Informatics (OHDSI) and similar) as well as demographics data, using it to contextualise and enrich the patterns observed for COVID-19 participants at UHSFT.
    a. The contextualisation component would use existing studies as a reference frame to contrast and compare similarities and differences among UHSFT and external patient cohorts.
    b. The enrichment component would use existing studies and demographics as priors to support Bayesian inference over UHSFT data.
2. Co-design and evaluate a Bayesian inference model
    a. With a possible extension of the Bayesian framework into a causal inference setting

b. Provide a critical analysis of the ethical and safety aspects related to the clinical application of Bayesian inference, i.e. supporting the clinical decision-making process.

3. Other innovative artificial intelligence and machine learning methods will also be explored to identify and predict those patients who have worse outcomes.

## Ethics and dissemination

The study design, protocol and patient facing documentation have been approved by North West Research Ethics Committee (REC 17/NW/0632) as an amendment to the National Institute of Health Research (NIHR) Southampton Clinical Research Facility (CRF) - managed Southampton Research Biorepository (SRB). No patient identifiable data is included in this manuscript. Ethics approval for the study was obtained from HRA Specific Review Board (REC 20/HRA/2986) for waiver of informed consent for the database only cohort. The database will be conducted in accordance with the principles of good clinical practice (GCP). Database documents (paper and electronic) will be collected and retained in accordance with the General Data Protection Regulations 2018 in a secure location during and after the trial has finished. All essential documents including source documents will be retained for a minimum period of 5 years following the end of the study. Clinical and research data collected in the database will be kept in a highly secure contemporary encrypted data platform BC|INSIGHT that was set up in a Microsoft Data Centre in South UK. Access to the study's data is only given to study team members that have signed the delegation logs. Data access is via an encrypted web service with username/password type authentication. The data held in the server are backed-up weekly using the snapshot technology provided by Microsoft Azure and the

retention period of the snapshot storage is 2 years. Storage of data for research purposes will be in an anonymised format and any data sharing outside of the clinical team will be in an anonymised format. Anonymised data may be released to individuals or organisations outside the UK, following approval by the REACT COVID data access committee. A REACT COVID Data Access Management Committee will be established to prioritise and ensure appropriate governance of requests to access linked anonymized clinical data. REACT Committee membership will include: Chief Investigator and Principle Investigators, the scientific lead for REACT and a representative from UHSFT R&D and/ or Academic Health Sciences Centre prioritization committee.

Written informed consent will be obtained from all study participants who agree to biobanking of excess clinical sample +/-additional sampling, as per approvals for amendment to the SRB reviewed by North West Research Ethics Committee (REC 17/NW/0632). For patients who have impaired cognition, informed consent will be sought from their legally acceptable representative, with retrospective consent sought for those who regain capacity to consent following recovery from acute illness. The findings from this study will be disseminated locally and internationally through manuscript publications in peer-reviewed journals.

## Discussion

SARS-CoV-2 infection is presenting a heterogenous pattern of disease, with varying presentations from asymptomatic carriage, as demonstrated amongst healthcare workers (7) to severe disease with mortality rates of 50% in patients requiring intensive care (8). Emerging

understanding is identifying some co-morbidities and risk factors for progression to severe disease, but there is still a wide breadth of severity amongst patients displaying the same risk profile.

The longitudinal, mainly automated data collection of the REACT COVID study enables a granularity of data that has potential to identify early markers of disease severity and the relevance of their trajectory of change on disease progression. The REACT COVID database can be used to rapidly and easily identify participants who may be appropriate for clinical trials and offers visual interpretation of large amounts of data for hypothesis generation.

The database provides a rich dataset to support the biobanking sub-cohort, within which there is the potential to recruit to smaller, focussed basic science studies investigating the mechanisms driving disease processes resulting from infection with SARS-CoV-2.

The REACT COVID study is mindful of the national and international efforts to bring together data from multiple centres in order to enhance the power of studies, be they clinical, observation, mechanistic, or intervention studies. The study will therefore be aligning itself with the broader UHS and UoS research project of ENACT (Enabling new treatment approaches for COVID-19 Treatment), supported by the Southampton Coronavirus Support Fund, NIHR Southampton BRC and NIHR Southampton CRF.

The ability of the REACT COVID study to capture the granularity of the longitudinal path of clinical care is both a major strength and a potential limitation. There is no strict study 'visit' protocol to follow, with data capture occurring in a more fluid manner. Therefore, there is the potential for a greater number of incomplete data sets, with some patients having a greater amount of available data. This is likely to be reflective of the level of care required, such as in those patients who required intensive care support. Any analysis will allow this, and

populations may be sub-cohorted to answer specific questions if the data is only available for a specific group.

The simplicity of data visualisation within the REACT COVID study belies the complexity of the concepts behind this project, facilitated through the unique collaboration between UHSFT, the UoS and the digital ECMT. This collaboration possesses the requisite expertise in their respective specialisms to deliver all aspects of this study, alongside the infrastructure provided by through the NIHR Southampton CRF and NIHR Southampton BRC. The REACT COVID study enables visualisation of detailed clinical data and the application of novel artificial intelligence methods, which will allow greater understanding of the natural history of this novel disease through both longitudinal and cross-sectional analysis.

## Acknowledgements

The authors would like to acknowledge the NIHR Southampton CRF and UHSFT Research Nursing teams for their support in the set-up of this study, Dr Dave Stockley (NIHR CRF SRB Manager) and Dr Alastair Watson for his help in proof reading and preparation of the manuscript for submission.

## Contributorship statement

HB and AF designed the protocol and drafted the manuscript; AD and MC were involved in protocol design, JB, HP and FB were involved realisation of data extraction, integration, transformation and upload processes, CK, GT and SNF were involved in protocol design for the biobanking subcohort, NS and SW are involved in manual data collection processes, DL, PF, HP and JB were involved in the design and adaptation of the REal-time Analytics for Clinical

Trials (REACT) platform and TW was involved in study conception and protocol design. All authors have reviewed the final manuscript.

## Competing interests

The authors have no competing interests to declare in relation to this manuscript

## Funding

The REACT platform has been supported by the digital Experimental Cancer Medicine Team (digital ECMT) free of charge. The biobanking subcohort is supported the NIHR Southampton CRF and NIHR Southampton Biomedical Research Centre (BRC) at University Hospital Southampton NHS Foundation Trust (UHSFT) and as part of a broader effort (Enabling new treatment approaches for COVID-19 Treatment - ENACT) by the University of Southampton (UoS) charity (Office of Development and Alumni Relations - ODAR). In addition the Clinical Informatics Research Unit, UoS has supported infrastructure costs. The support described above was not provided from a specific award or grant.

# References

1. World Health Organisation. WHO Coronavirus Disease (COVID-19) Dashboard. 2020. https://covid19.who.int/?gclid=EAIaIQobChMIz5Xu84OJ6gIVCLrtCh3WlQ8KEAAYASAAEgIohvD_BwE. Accessed 16th June 2020.

2. UK Governement Guidance. 2020. https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public. Accessed 18th June 2020.

3. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19 - Preliminary Report. N Engl J Med. 2020.

4. Oxford University News Release. Low-cost dexamethasone reduces death by up to one third in hospitalised patients with severe respiratory complications of COVID-19. 2020. https://www.recoverytrial.net/files/recovery_dexamethasone_statement_160620_v2final.pdf. 16th June 2020.

5. ACCORD ACcelerating COVID-19 Research and Development. 2020. https://www.nihr.ac.uk/news/national-collaborative-platform-launched-to-fast-track-potential-covid-19-treatments/24678. Accessed 18th June 2020.

6. RECOVERY. 2020. https://www.recoverytrial.net. Accessed 18th June 2020.

7. Rivett L, Sridhar S, Sparkes D, Routledge M, Jones NK, Forrest S, et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. Elife. 2020;9.

8. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. Bmj. 2020;369:m1985.

## Word Count

3,253

## Figure Legends

**Figure 1: Data flow between systems and institutions within the REACT COVID database.**

COVID-19, Coronavirus disease 2019; CRUK MI, Cancer Research UK, Manchester Institute; digital ECMT, digital Experimental Cancer Medicine Team; REACT, REal-time Analytics for Clinical Trials; SARS-CoV-2, Severe Acute Respiratory Syndrome coronavirus 2; UHSFT, University Hospital Southampton NHS Foundation Trust.

**Figure 2: Data capture for all patients to be included in the database .**

Data variables will be recorded as part of routine clinical care and will be captured as part of the observational study. Data variables listed in italics will be manually captured and added to the database. ABG, arterial blood gas; ACE, angiotensin converting enzyme; AKI, acute kidney injury; ARB, angiotensin receptor blocker; ARDS, acute respiratory distress syndrome; CHARTS, is an integrated user interface for electronic health care records; COVID-19, Coronavirus disease 2019; CPAP, continuous positive airways pressure; CRP, C-reactive protein; CRS, Cytokine release syndrome; HR, Heart rate; FBC, Full blood count; Obs, Clinical observations; MetaVision, is a clinical information system for critical care; NP, Nasopharyngeal; PEEP, Positive end expiratory pressure; proADM, pro-adrenomedullin; RT-PCR, Reverse transcriptase polymerase chain reaction; RR, Respiratory rate; Temp, Temperature; TV, Tidal volume; Imaging findings from Chest X-ray (CXR), Computed Tomography (CT) or Magnetic Resonance Imaging (MRI); UHSFT, University Hospital Southampton NHS Foundation Trust.

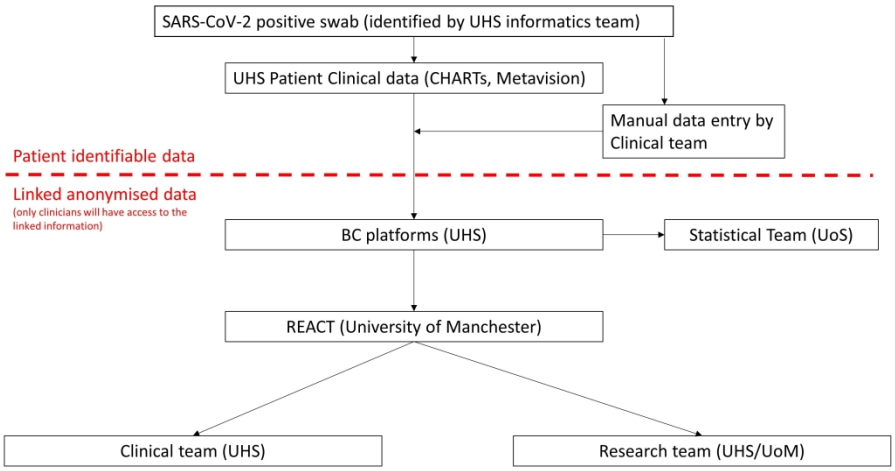


Figure 1: Data flow between systems and institutions within the REACT COVID database. COVID-19, Coronavirus disease 2019; CRUK MI, Cancer Research UK, Manchester Institute; digital ECMT, digital Experimental Cancer Medicine Team; REACT, REal-time Analytics for Clinical Trials; SARS-CoV-2, Severe Acute Respiratory Syndrome coronavirus 2; UHSFT, University Hospital Southampton NHS Foundation Trust.

338x190mm (300 x 300 DPI)

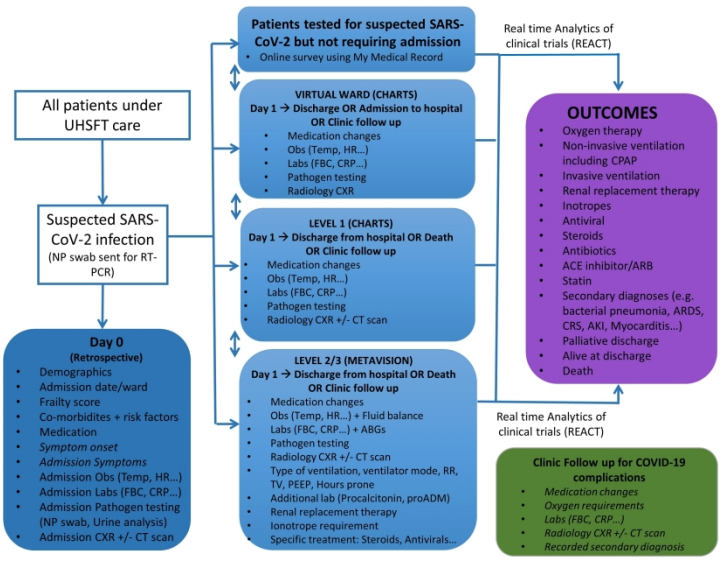


Figure 2: Data capture for all patients to be included in the database .Data variables will be recorded as part of routine clinical care and will be captured as part of the observational study. Data variables listed in italics will be manually captured and added to the database.ABG, arterial blood gas; ACE, angiotensin converting enzyme; AKI, acute kidney injury; ARB, angiotensin receptor blocker; ARDS, acute respiratory distress syndrome; CHARTS, is an integrated user interface for electronic health care records; COVID-19, Coronavirus disease 2019; CPAP, continuous positive airways pressure; CRP, C-reactive protein; CRS, Cytokine release syndrome; HR, Heart rate; FBC, Full blood count; Obs, Clinical observations; MetaVision, is a clinical information system for critical care; NP, Nasopharyngeal; PEEP, Positive end expiratory pressure; proADM, pro-adrenomedullin; RT-PCR, Reverse transcriptase polymerase chain reaction; RR, Respiratory rate; Temp, Temperature; TV, Tidal volume; Imaging findings from Chest X-ray (CXR), Computed Tomography (CT) or Magnetic Resonance Imaging (MRI); UHSFT, University Hospital Southampton NHS Foundation Trust.

338x190mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract Yes |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found Yes-protocol so no results reported |
| **Introduction** | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported Yes |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses Yes |
| **Methods** | | |
| Study design | 4 | Present key elements of study design early in the paper Yes |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection Yes |
| Participants | 6 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up Yes<br>*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed NA<br>*Case-control study*—For matched studies, give matching criteria and the number of controls per case |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable YES |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group Yes |
| Bias | 9 | Describe any efforts to address potential sources of bias yes |
| Study size | 10 | Explain how the study size was arrived at yes |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why protocol study, NA |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses<br>Planned statistical analyises described but protocol paper only |

Continued on next page

| | | |
|---|---|---|
| **Results** | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed Planned numbers reported |
| | | (b) Give reasons for non-participation at each stage Yes |
| | | (c) Consider use of a flow diagram Yes |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders NA |
| | | (b) Indicate number of participants with missing data for each variable of interest |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) Yes-planned follow up |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time Planned reported |
| | | *Case-control study*—Report numbers in each exposure category, or summary measures of exposure |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included NA |
| | | (*b*) Report category boundaries when continuous variables were categorized NA |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period NA |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses Planned analyses reported |
| **Discussion** | | |
| Key results | 18 | Summarise key results with reference to study objectives Yes |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias Protocol paper |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence Protocol paper |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results Protocol paper-will be discussed in context of study findings in subsequent papers |
| **Other information** | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based Yes |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.