OX79 Coronavirus Record Linkage Project - QResearch-ICNARC COVID-19 Collaboration

Collaboration between Intensive Care National Audit and Research Centre/University of Oxford Nuffield Department of Clinical Neurosciences Critical Care Research Group and QResearch, Department of Primary Health Sciences, University of Oxford

April 2019

RESEARCH PROGRAM PROTOCOL
# Table of Contents

1 Version history .............................................................................................................. 3
2 Study Investigators .......................................................................................................... 3
   2.1 Joint Chief Investigators .......................................................................................... 3
   2.2 Co-investigators ....................................................................................................... 3
3 Introduction ..................................................................................................................... 3
4 Aims and objectives ........................................................................................................ 5
5 Vision statement ............................................................................................................. 5
6 Summary of potential impact .......................................................................................... 6
7 Methods and analysis ..................................................................................................... 6
   7.1 Study design ............................................................................................................. 6
   7.2 Study population ...................................................................................................... 6
   7.3 Putative modulators of the severity of COVID-19 infection ...................................... 6
   7.4 Risk factor identification for severe disease .............................................................. 7
   7.5 Outcomes ................................................................................................................ 8
      7.5.1 Short term outcomes for ICU patients ............................................................... 8
      7.5.2 Longer term outcomes for ICU patients ........................................................... 8
      7.5.3 Outcomes for patients in primary care ............................................................... 8
      7.5.4 Other Direct and Indirect effects ....................................................................... 9
   7.6 Definition of exposure .............................................................................................. 9
   7.7 Datasets ................................................................................................................... 9
      7.7.1 Primary care datasets: ...................................................................................... 9
      7.7.2 Intensive Care dataset: ICNARC Case Mix Programme .................................... 10
      7.7.3 National datasets: ............................................................................................ 10
      7.7.4 Additional national dataset: ............................................................................. 11
   7.8 Linkage ...................................................................................................................... 11
   7.9 Statistical analysis and data handling ..................................................................... 11
      7.9.1 Descriptive analysis .......................................................................................... 11
      7.9.2 Statistical modelling ......................................................................................... 12
      7.9.3 Sample size considerations .............................................................................. 13
      7.9.4 Machine learning analyses .............................................................................. 13
8 Study oversight .............................................................................................................. 14
9 Patient engagement ...................................................................................................... 14
10 References .................................................................................................................... 14
1 Version history

<table>
<thead>
<tr>
<th>Version</th>
<th>date</th>
<th>author</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6</td>
<td>09.04.2020</td>
<td>JHC on behalf of study investigators</td>
<td>Version approved by Qresearch science committee 09.04.2020 and published</td>
</tr>
</tbody>
</table>

2 Study Investigators

2.1 Joint Chief Investigators

- Professor Peter Watkinson
- Professor Julia Hippsley-Cox, Professor of Clinical Epidemiology & General Practice
- Professor Kathy Rowan, ICNARC

2.2 Co-investigators

- Professor Duncan Young
- Professor Carol Coupland, Professor of Medical Statistics in Primary Care, University of Nottingham
- Stephen Gerry, Centre for Statistics in Medicine, University of Oxford
- Professor David Clifton, Professor of Clinical Machine Learning, University of Oxford
- Professor David Harrison, Professor of statistics, ICNARC

3 Introduction

The majority of UK residents will likely have been infected with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the end of the current pandemic(1). Many of these will require hospital treatment. A significant number of the more severely affected patients will require treatment for respiratory failure on an intensive care unit (ICU) and a substantial proportion of these will die.

Reports suggest some very commonly used drugs may exacerbate or attenuate the severity of COVID-19 disease(2–4). In both cases, plausible mechanisms are proposed, but robust evidence of potential harm or benefit is limited. The drugs most widely discussed are those acting on the pulmonary angiotensin (ACE) pathway. SARS-CoV-2 binds to angiotensin converting enzyme 2 (ACE2) receptors in the lower respiratory tracts of infected patients to gain entry into the lungs. Angiotensin converting enzyme inhibitors (ACEI) and angiotensin II receptor blockers (ARBs) are hypothesised to alter the severity of disease by causing upregulation of the ACE2 receptors in patients who take these medications. Other drugs that may also modulate severity by the same mechanism are non-steroidal anti-inflammatory drugs (NSAIDs)(5) and thiazolidinediones (drugs used for type 2 diabetes mellitus).
These are drugs used to treat common, chronic conditions. Around 14% of the entire adult population in England take anti-hypertensive medications, and around 5% receive medication to treat diabetes mellitus(6). The prevalence increases with age, making usage particularly common in those at risk of for severe COVID-19 infections. In many cases drugs from a different class could be used instead. If these drugs are increasing the risk of severe infection, they represent one of the few modifiable risk factors for severe COVID-19 infection. However, currently professional bodies have issued statements to reassure patients that they should keep taking chronic medications, to avoid harm from patients self-withdrawing treatment(7,8). Medical and research communities need rapid large-scale accumulation of data on the outcomes of patients who develop COVID-19 infection whilst taking these drugs to allow appropriate risk assessment and clinical decision making for these patient groups.

There are also several drugs with suspected anti-viral activity to COVID-19 in common use in primary care patients (including hydroxychloroquine(3,9), used in rheumatoid arthritis and lopinavir-ritonavir(4,10), used in human immunodeficiency virus 1 treatment). Finally, there are many patients on immune suppressive therapies that may either increase the risk of severe illness by preventing the body’s response to infection, or attenuate the hyperinflammation syndrome associated with COVID-19 disease so preventing severe disease(11). The incidence of severe disease in patient groups taking these medications urgently needs to be established to guide both their management and investigation of COVID-19 treatment strategies.

The Intensive Care National Audit and Research Centre’s (ICNARC) case mix programme (CMP) is already being used to provide up to date information for critical care clinicians on the admission characteristics and outcomes of all patients with severe COVID-19 infection treated on an ICU in England, Wales and Northern Ireland. Patients with clinical or laboratory-proven COVID-19 infection are flagged near real time on the reporting system. However, the dataset does not contain the information on prior long-term medication and chronic disease necessary to assess the effects of medications on COVID-19 infection severity. This information is contained within the QResearch databases. QResearch is a large consolidated database derived from the anonymised general practice health records of over 35 million patients from approximately 1500 UK general practices using the EMIS health clinical information system over a 20-year period. QResearch is already linked to mortality data, cancer registry and hospital episode statistics (HES) with ongoing linkages undertaken.

The linkage of the existing QResearch-HES-mortality data to ICNARC CMP data will enable research to look at the association between prior medication and outcome for patients most severely affected by COVID-19 infections. We will provide useful knowledge that patients, GPs and intensive care doctors can use to reduce the risk of severe COVID-19 infection within this pandemic.

In addition to the main aim of the study, we anticipate that the resulting linked database will also provide a large scale, representative invaluable resource which can be used to undertake other research projects on the natural history of COVID-19 as well as the direct and indirect effects of the pandemic on multiple outcomes. In order to illustrate this and
highlight the additional benefits of the new data linkage, we have included some example questions in section 6.5.4. We anticipate that each of these additional questions will have a dedicated analysis plan which will be scientifically reviewed by the QResearch committee as appropriate.

4 Aims and objectives

By undertaking this ICNARC CMP/QResearch linkage we will provide a unique high-quality resource which can undertake the following analyses at scale and at speed:

- Cohort analyses of the associations of chronic medications with admission to, and discharge status from an Intensive Care Unit following COVID-19 infection.
- The prevalence of chronic diseases in patients admitted to ICUs with COVID-19 infection.

Acutely these analyses will provide:

- Identification of candidate drugs for switching to other similarly active agents during the COVID-19 pandemic.
- Identification of high-risk patients in critical and primary care

Subsequently, the combined dataset will provide a valuable resource to understand COVID-19 infection, from primary care to critical care and back.

5 Vision statement

The primary objective is to determine if there are associations between chronic use of drugs:

- acting on the angiotensin system, or
- with potential COVID-19 anti-viral activity, or
- causing immune modulation

and severe COVID-19 virus infection.

The information will be used by GPs and others to advise their patients on continuing or stopping medications during this pandemic and to prioritise drugs for inclusion in ongoing randomised controlled trials.

Along with determining the primary objective, a further short-term objective will be to identify patient factors associated with severe disease. Although not modifiable these will assist risk stratification and could allow high risk patients to be identified electronically from general practice records for tailored advice and/or medication reviews.
The middle-term objective is to assemble a data linkage platform to study all documented COVID-19 infections including those admitted to hospital but who did not go to ICU. We will obtain details of patients with a positive COVID-19 PCR test through Public Health England (agreements proceeding at the time of writing) and when added to the GP or HES records (hence becoming available in QResearch).

The long-term objective is to enable studies investigating health, medication and health resource use before and after critical illness. This work was planned before the COVID-19 epidemic, which will accelerate development of a long term rich research resource.

### 6 Summary of potential impact

The global disease burden is obviously huge in a pandemic. As of 08/04/2020 there were 4690 patients treated in ICUs with COVID-19 infections. This is expected to increase exponentially in the short term.

If the candidate drugs are adversely associated with COVID-19 infection it would be possible to swap patients to alternates. Similarly, if drugs with protective associations can be established, these will become candidates for inclusion in large-scale studies. We already have links with the major studies in the UK so that identified drugs could be added.

### 7 Methods and analysis

#### 7.1 Study design

Record linkage cohort study using routinely collected healthcare data and data collected for comparative audit.

#### 7.2 Study population

There are two main study populations of interest:

(i) All adults and children registered with practices contributing to QResearch (primary care population) including those with a suspected or confirmed recorded diagnosis of COVID-19.

(ii) The subgroup of adults and children within this population who were admitted to ICU with COVID-19.

#### 7.3 Putative modulators of the severity of COVID-19 infection
There are three immediate medication groups of interest. The first is by far the largest and most important group.

1. **Drugs that may modulate the severity of COVID-19 infection**: These include angiotensin converting enzyme inhibitors (ACEi), angiotension II receptor blockers (ARBs), thiazolidinediones and non-steroidal anti-inflammatory drugs (NSAIDS). These drugs increase the number of cell-surface ACE2 receptors, which COVID-19 uses exclusively to gain entry into type 2 pneumocytes in the lung (2). There are competing theories as to whether this increases or decreases the severity of COVID-19 infection (2,12). Similarly different drugs, even within the same class may have opposing effects (13).

2. **Drugs with potential COVID-19 anti-viral activity**: These include hydroxychloroquine (used to treat diseases such as rheumatoid arthritis and systemic lupus erythematosis) and lopinavir-ritonavir (a human immunodeficiency virus 1 treatment). Both have shown limited evidence of in-vitro efficacy and even less evidence of in-vivo effectiveness (3,4,9,10).

3. **Drugs causing immunomodulation**: These include drugs in common use in primary care such as inhaled and systemic corticosteroids, and less common agents such as methotrexate, selective cytokine blockers such as tocilizumab and infliximab and JAK inhibition agents such as baricitinib(11). These latter agents are used in diseases such as inflammatory bowel disease, rheumatoid arthritis and ankylosing spondylitis. Anti-psychotic drugs have also been implicated (14,15).

### 7.4 Risk factor identification for severe disease

Severe disease will be defined pragmatically as those patients who are admitted to hospital as an emergency, admitted to ITU or who die including those in primary care with a suspected or confirmed recorded diagnosis of COVID-19. We will look at the association of patient variables with severity of infection in the primary care population. In those admitted to hospital and those admitted to ICU we will look at the association of patient variables with death. These patient factors are not modifiable but will assist risk stratification and could allow high risk patients to be identified electronically from general practice records for tailored advice and/or medication reviews.

Target variables include the following (the list may expand as further information on COVID-19 becomes available):

1. Demographic variables including age, sex, ethnicity, Townsend Score, BMI, smoking, alcohol, region of the UK.
2. Selected relevant chronic diseases including but not limited to COPD, rheumatoid arthritis, ankylosing spondylitis, inflammatory bowel disease, cystic fibrosis, asthma, diabetes mellitus, epilepsy, hypertension, cardiovascular disease, cancer, severe mental illness, chronic liver disease and Parkinson’s disease.
4. Frailty variables in patients as measured by the QFrailty score(16).
7.5 Outcomes

We will undertake analyses to examine prognosis of patients admitted to ICU.

7.5.1 Short term outcomes for ICU patients

Our primary outcome measure is vital status at ICU discharge (alive or died)

The secondary short term outcome measures are:
1. Vital status at acute hospital discharge (alive/dead at discharge from an acute care hospital, using the hospital containing the ICU as the acute care hospital)
2. Duration of advanced respiratory support (days)
3. Duration of advanced cardiovascular support (days)
4. Duration of advanced renal support (days)
5. Length of ICU stay (for repeat ICU admissions in the same acute hospital admission the total days on ICU will be used)
6. Days of acute hospital care after ICU discharge (for repeat ICU admissions in the same acute hospital admission the total days not on ICU should be used)

7.5.2 Longer term outcomes for ICU patients

We will undertake analyses to look at the following longer-term outcomes for patients admitted to an ICU

- Vital status (alive/dead) at 30 and 90 days after ICU admission.
- All cause and COVID-related specific mortality at 6 and 12 months after ICU admission.

7.5.3 Outcomes for patients in primary care

We will undertake analyses to look at the following longer-term outcomes for patients in the primary care cohort.

1. Emergency admission
2. ICU admission
3. All-cause mortality and cause-specific mortality including COVID-related mortality

We will determine the risks of these outcomes at 30 and 90 days, 6 and 12 months after the study entry date. We will compare risks of these outcomes between patients exposed to the therapies of interest (section 6.3) and unexposed patients accounting for age, sex and comorbidities (see section 6.9).
7.5.4 Other Direct and Indirect effects.

We will also determine the late effects of COVID-19 as the pandemic progresses as outlined below. For each of these, a more detailed statistical analysis plan will be developed and submitted to the QResearch scientific committee as an amendment.

For example, there may be direct clinical consequences (such as pulmonary fibrosis in an otherwise healthy COVID survivor) or indirect respiratory consequences (such as increased consultation/prescription rates for COPD).

Other clinical outcomes of interest include cardiovascular outcomes and mental health consequences (depression, anxiety, self-harm etc). By the time the pandemic wanes there will be more knowledge of likely clinical and non-clinical sequelae.

Finally there may be indirect health consequences relating to changes in access to or delivery of health services. For example, from 23rd March 2020 when the ‘lockdown’ was announced by the government, the majority of consultations in primary care started being delivered by telephone or video consultation without direct face-to-face contact with patients. This may affect how and when diagnoses are made. For example, it could also result in delayed or inaccurate diagnosis or exacerbate health inequalities if, for example, certain subgroups of the population (e.g. the elderly, deprived or ethnic groups) are less able to access services in this way.

7.6 Definition of exposure

Our main exposure variables of interest are medications prescribed prior to hospital admission prioritising those highlighted in section 6.3.

We will classify patients as exposed to a medication on ICU admission if they had two or more prescriptions for the medication in the 6 months before admission. If further therapies are identified as potentially increasing or decreasing the risk of adverse prognosis associated with COVID-19 infections, these will be included as additional analyses.

In the analyses of the primary care cohort we will assess the medication exposures at the study entry date.

7.7 Datasets

7.7.1 Primary care datasets:

7.7.1.1 QResearch

QResearch is a large consolidated database derived from the anonymised health records of over 35 million patients in the UK. The data currently come from approximately 1500 general practices using the EMIS Health clinical computer system. The practices are spread throughout the UK and include data from patients who are currently registered with the
practices as well as historical patients who may have died or left. Historical records extend back to 1989 with linked data on all practices since 1998. QResearch is linked to secondary care data for use in ethical medical research.

7.7.1.2 SystmOne TPP (if available)
Ethical approval has been obtained to expand the system to include data from consenting practices who use SystmOne TPP (around 40% of the population in England) should these data become available.

7.7.2 Intensive Care dataset: ICNARC Case Mix Programme
The ICNARC Case Mix Programme (CMP) is a database that contains data on all admissions to adult, general critical care units (intensive care and combined intensive care/high dependency units) in England, Wales and Northern Ireland. The CMP is listed in the Department of Health’s ‘Quality Accounts’ as a recognised national audit by the National Advisory Group on Clinical Audit & Enquiries (NAGCAE) for ‘Acute’ care. The database goes back to 1994 and currently contains data on 1.8 million admissions. The data has very high completeness and is extensively error-checked. At the Department of Health’s request ICNARC began expedited reporting of all COVID-19 infected patients and receives both an initial notification of an admission with COVID-19 infection and a rapid update when the patient is discharged from the ICU.

The ICNARC dataset is a rich clinical dataset which includes variables which are not available on the HES Critical Care Minimum Dataset. A list of the fields requested is contained in the embedded excel workbook along with the rationale. The field contain details of the severity of illness on admission to the ICU, severe co-morbidities and body habitus as measures of severity of infection and vulnerability to infection. The data set also contains variables that can be used to measure treatment intensity and hospital resource use, as well as a range of non-lethal outcome measures. As part of the de-identification process, the QResearch-HES-ICNARC linked data has been de-identified according to the ICO standard on anonymisation. In particularly, the linked dataset does not include patient identifiers nor information on the identity of the general practices, hospital or ICU.

7.7.3 National datasets:
- Hospital Episode Statistics (England) including Inpatient, Outpatient, A&E and critical care
- Civil Registrations (England and Wales mortality data) with up to 15 causes of death
- Patient Episodes Data Wales (ICU data only currently)

The applicants have a proven track record in linking these national datasets to the ICNARC CMP dataset and linkage already exists in QResearch.
**7.7.4 Additional national dataset:**

COVID-19 is a notifiable disease and as such all labs are required to submit positive test results to Public Health England (PHE) who therefore hold a “master” list of patients with a laboratory confirmed diagnosis of COVID-19. It is also running the COVID-19 Hospitalisation in England Surveillance System (CHESS) of patients requiring high dependency or intensive care support. However this dataset is substantially less detailed than the ICNARC CMP dataset. CHESS is also not linked to the primary healthcare data, so will not provide the medication data required. We are currently exploring the permissions required to access the COVID-19 master list, along with the any coverage, data completeness and related data quality issues. Whilst there are limitations to the PHE dataset because testing is not widespread or systematic, testing coverage may improve during pandemic. If this “master” list were linked to QResearch and HES, and ICNARC CMP a comprehensive overview of the effect of COVID-19 on the nation’s health could be obtained. New knowledge obtained through this extension would be of particular benefit in managing future COVID-19 outbreaks.

**7.8 Linkage**

Importantly Prof Hippisley-Cox, who organised the QResearch database, built in a system for generating a linked, anonymised common dataset from QResearch and a second database. The software is called OpenPseudonymiser ([www.openpseudonymiser.org](http://www.openpseudonymiser.org)) and is available as an open-standards open source software tool kit. It has already been implemented by EMIS, NHS Digital, Civil Registrations, Public Health England and TPP along with a large number of other NHS organisations. Following Confidential Advisory Group consultation, this process does not require Health Research Authority Confidentiality Advisory Group (CAG) s251 approval as neither party hold identifiable data of the other’s patients. This considerably speeds up governance approvals allowing a linkage study with the CMP dataset to occur rapidly.

**7.9 Statistical analysis and data handling**

**7.9.1 Descriptive analysis**

We will undertake a descriptive analysis of the patients in the study population with COVID-19 admitted to an ICU by demographic characteristics, comorbidity and use of medications of interest, using means, medians and proportions as appropriate.

We will undertake descriptive analyses for patients in the primary care study population and in the subset with suspected or confirmed COVID infection to determine the proportions admitted to hospital and to ICU and describe these according to patient demographics, comorbidities and use of medications of interest prescribed prior to admission.
7.9.2 Statistical modelling

We will use a variety of statistical methods to test a priori hypotheses in relation to the risks and benefits of the intercurrent medications identified above.

These are as follows according to the specific outcomes:

7.9.2.1 Outcomes in ICU cohort

- Vital status at ICU discharge (alive/died), vital status at acute hospital discharge (alive/died): robust Poisson regression to estimate risk ratios for mortality according to medication use
- Duration of advanced respiratory support, duration of advanced cardiovascular support, duration of advanced renal support, length of ICU stay, days of acute hospital care after ICU discharge: negative binomial regression to account for likely overdispersion (with zero-inflated versions if improvement in model fit)
- Vital status (alive/dead) at 30 and 90 days after ICU admission: robust Poisson regression to estimate risk ratios for mortality according to medication use
- All cause and cause-specific mortality at 6 and 12 months after ICU admission: Cox proportional hazards models to estimate hazard ratios accounting for censoring

7.9.2.2 Outcomes in primary care cohort

- Emergency admission, ICU admission, mortality at 30 and 90 days after study entry date: robust Poisson regression to estimate risk ratios according to medication use
- Emergency admission, ICU admission, mortality at 6 and 12 months after study entry date: Cox proportional hazards models to estimate hazard ratios accounting for censoring

The ICU cohort analyses will account for clustering by ICU, and the primary care cohort analyses will account for clustering by general practice.

The main analyses will be multivariable analyses adjusting for a range of potential confounders including demographic characteristics (section 6.4), comorbidities (including all indications for the medication of interest), and other prescribed medications. The analyses will examine interactions between the medications and patient characteristics such as age and sex to assess whether there are differences in medication risks according to these factors. Assumptions of models (e.g. proportional hazards assumption) will be checked. Multiple imputation will be used to replace missing data as a sensitivity analysis. We will use P < 0.01 (2-tailed) to determine statistical significance.

To further reduce indication biases, additional analyses will be carried out (where numbers permit) for example (i) restricted to patients with hypertension or heart failure to directly compare risks for ACEI and ARBs with other antihypertensive drugs and (ii) restricted to patients with diabetes to directly compare risks for thiazolidinediones with other diabetes drugs.
To examine the robustness of the analysis approach we will also use other methods such as instrumental variables, propensity score adjustment, propensity score matching, and case matching. Propensity scores will be calculated and incorporated into the models to account for patient demographic variables, comorbidities and use of other medications. For the case matching approach, a patient with the drug treatment of interest is matched with one or more similar patients without the treatment of interest based on patient characteristics (e.g. age, sex, deprivation, comorbidities). It is likely the “Genmatch” algorithm will be used for efficient, semi-automated matching(17). The outcomes will then be compared across the two groups generated by the matching.

7.9.3 Sample size considerations

For the primary outcome of vital status at ICU discharge then a cohort of 2684 ICU patients is needed to detect a risk ratio of 1.25 with 90% and 1% two-sided significance, assuming 50% mortality in the unexposed group and that 10% of patients are exposed to the medication of interest. If 5% are exposed the sample size required is 5079, for 2% exposed it is 12,302 and for 1% exposed it is 24,353. Analyses will only be carried out when the number of accumulated ICU patients reach these orders of magnitude depending on the proportion exposed to medication to reduce the likelihood of a type 2 error (i.e. failure to detect a genuine effect).

It is expected that some analyses will be repeated over time as the number of patients with COVID-19 infections admitted to ICUs increases and some analyses will be extended to examine associations within subgroups as numbers increase. The main analyses will be restricted to patients admitted to ICUs to estimate risks of adverse outcomes according to treatments prescribed prior to admission. A priori we will focus on the drug groups already identified in the literature listed above, particularly those which are more commonly prescribed. We will consider whether to make formal adjustment to account for multiple testing or to select a more restrictive p value e.g. p < 0.001.

7.9.4 Machine learning analyses

Results obtained using the modelling approaches described above will be compared with results from machine learning (ML) equivalents, including those based on (i) conventional ML approaches, such as random forests and support vector machines; and (ii) advanced ML approaches, such as deep learning architectures and Bayesian non-parametric models. Such methods typically offer non-linear approaches, permitting the effect of potentially complex interactions between covariates to be modelled; over-fitting (i.e., low bias, high variance) will be assessed using best-practice methods, including cross-validation (for simpler models) and architecture regularisation (for more complex models).
We will evaluate the use of time-series analysis ML approaches, such as Bayesian Gaussian processes, for capturing through-time structure within the input data, acting as additional inputs to the models described above (compared to analogous models omitting time-series inputs).

We will evaluate the use of phenotypical clustering of patient data using (i) conventional clustering algorithms, such as k-means; and (ii) modern clustering approaches, such as deep clustering, in which dimensionality reduction (i.e., reducing the number of input variables to an informative subset) and clustering are performed simultaneously. We will test the hypothesis that the fit of modelling approaches described previously can be improved by constructing models specific to the clusters that are discovered.

8 Study oversight

The data advisory groups for ICNARC, QRResearch (covering TTP) and the Kadoorie Centre for Critical Care research and Education will be required to approve data handling. We will convene a Study Oversight Group. Two independent PPI representatives (see below) with prior experience in similar research areas have joined the oversight group. We will also recruit an independent clinician with expertise in this area. Ethics approval for this has been obtained.

9 Patient engagement

Non-academic public members (RL and VKH) will join the study oversight committee. These individuals were chosen because, whilst not academics, they do have a clear understanding of data science, University procedures, and governance principles. They are the PPI representatives in the Wellcome/DoH funded HAVEN project and continue to be involved with ongoing use of project data for secondary research.

10 References