SIREN: The impact of detectable anti SARS-COV2 antibody on the incidence of COVID-19 in healthcare workers

Synopsis

<table>
<thead>
<tr>
<th>Study title</th>
<th>Impact of detectable anti-SARS-COV2 on the subsequent incidence of COVID-19 in healthcare workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short title</td>
<td>SIREN (Sarscov2 Immunity &amp; REinfection Evaluation)</td>
</tr>
<tr>
<td>Chief Investigator</td>
<td>Susan Hopkins</td>
</tr>
<tr>
<td>Sponsor</td>
<td>Public Health England</td>
</tr>
<tr>
<td>Funder</td>
<td>Department of Health and Social Care</td>
</tr>
<tr>
<td>Study coordination</td>
<td>Public Health England</td>
</tr>
<tr>
<td>Statistics and design</td>
<td>Andre Charlett</td>
</tr>
</tbody>
</table>

Study management group

- Study design and management: Susan Hopkins, Meera Chand, Colin Brown
- Statistics: Andre Charlett
- Epidemiology: Mary Ramsay
- Virology and serology: Maria Zambon, Tim Brooks
- Data management: Annie Marie O’Connell
- Genomics: Richard Myers
- Information governance: Samantha Organ

Study Design

- Prospective cohort study to determine incidence of new SARS-COV2 infection in seronegative and seropositive healthcare workers

Study participants

- NHS staff working in clinical settings in England

Sample size

- 100,000 healthcare workers

Planned study period

- 24 months

Planned recruitment period

- 3 months with follow-up for up to 18 months; interval analysis of incidence and prevalence monthly and statistical analysis on primary outcome 3 monthly.

Study Protocol version

- 3.1
Contents

SIREN: The impact of detectable anti SARS-COV2 antibody on the incidence of COVID-19 in healthcare workers .1
Synopsis .................................................................................................................................................. 1
Summary ............................................................................................................................................... 3
1.0 Background .................................................................................................................................. 3
2.0 Aim and Objectives ...................................................................................................................... 4
3.0 Methods ........................................................................................................................................ 5
  3.1 Study design ................................................................................................................................. 5
  3.2 Study sites .................................................................................................................................... 5
  3.3 inclusion/exclusion criteria ............................................................................................................ 6
    Inclusion criteria .............................................................................................................................. 6
    Exclusion criteria ............................................................................................................................ 6
  3.4 Sample size .................................................................................................................................. 6
  3.5 Site set up ..................................................................................................................................... 9
  3.6 Recruitment .................................................................................................................................. 9
  3.9 Storage of materials and additional Laboratory testing ............................................................... 11
4 Data management ............................................................................................................................. 12
5. Analysis plan .................................................................................................................................... 13
6. Confidentiality and information governance .................................................................................. 14
7. Quality assurance and research governance .................................................................................. 15
Appendix 1: Amendment History ..................................................................................................... 17
References ............................................................................................................................................ 18
Summary

COVID-19 is causing a global pandemic. This study aims to find out whether healthcare workers who have evidence of prior COVID-19, detected by antibody assays (positive antibody tests), compared to those who do not have evidence of infection (negative antibody tests) are protected from future episodes of infection.

In this study, we will recruit healthcare workers to be followed for at least a year and study their immune response to the virus causing COVID-19, called SARS CoV2. We will do this by collecting data on their history of COVID-19 infection and any new symptoms. All NHS staff who deliver care to patients are being asked to have a nose and throat swab every other week in order to detect mild cases or cases who do not have symptoms. This is the main test that is currently used to detect and diagnose infection. It looks directly for the virus in the nose and throat. Once the infection is cleared, we cannot detect virus in samples. Therefore, we will also ask these individuals to have blood samples taken every other week to determine whether they have antibodies to the infection. These blood samples allow the previous infection to be detected as the response to infection in the body is to produce small particles in the blood called “antibodies”. It takes up to 4 weeks to make enough antibodies to fight the infection. But once someone recovers, antibodies stay in the blood at low levels—this is may help prevent us from getting infected with the same infection again. However, for SARS CoV2 infection we do not know yet if the detection of antibodies protects people from future infection. Through this study we will provide this very important information which will help to understand the future impact of COVID-19 on the population.

1.0 Background

On 31st December 2019, the first cases of infection with a novel coronavirus, subsequently designated SARS-CoV-2, emerged in Wuhan, China. A global pandemic was declared by the World Health Organisation (WHO) on 12th March 2020. By 9th May, there were more than 3.6 million confirmed infections and 240,000 reported deaths globally.

In order to determine the population impact of potential future waves of the pandemic of COVID-19, there is an urgent need to understand whether prior infection and any host immune response provides protection from future reinfection with SARS-CoV2.

Currently, a confirmed diagnosis of infection of SARS-CoV2, relies on laboratory diagnosis of infection using reverse transcriptase polymerase chain reaction (PCR). Diagnostic PCR typically targets the viral ribonucleic acid (RNA)-dependent RNA polymerase (RdRp) or nucleocapsid (N) genes using swabs collected from the upper respiratory tract (nose and throat). However as -PCR only detects acute infection we are unaware of the true population prevalence of infection. While there is increasing evidence that a proportion of the population can be infected and asymptomatic, the true proportion of asymptomatic infection in the population is unknown.
In England, there is a new national NHS policy to screen healthcare workers with RT-PCR at least every other week to determine the incidence of new infection and reduce the risk of healthcare worker (HCW) transmission to other HCW and patients. However, PCR from upper respiratory tract swabs may be falsely negative, due to quality or timing of collection; or due to viral titres in upper respiratory tract secretions peaking in the first week of symptoms, but declining below the limit of detection in patients who present with symptoms beyond this time frame. In individuals who have been infected and recovered, PCR provides no information about prior exposure or immunity. In contrast, assays that reliably detect antibody responses specific to SARS-CoV-2 could contribute both to diagnosis of acute infection (via rises in IgM and IgG levels) and to identify those who have been exposed and recovered with or without symptoms (via persisting IgG). A further research gap is whether the detection of an antibody response to SARS-CoV2 will provide immunity to re-infection. In a study of medical students, reinfection with HCoV-229E was detected in 937 student years of follow-up; neutralizing antibody to 229E was commonly but it did not appear to influence the occurrence of, or likelihood of illness with, reinfection as judged by complement fixation seroconversion. While IgG and IgM to SARS-CoV2 can be effectively used to determine past exposure. It may be that these antibodies do not protect against future infection. The presence of neutralising antibodies may be required as these not only confirm prior infection but also prevent future infection. There are no commercial assays available to detect neutralising antibodies at present but PHE and other university partners are developing in house assays to detect these specific antibodies. We will examine correlates of different types of antibody response and the protective effect against re-infection.

Therefore, this study will answer the key question on whether prior SARS-CoV-2 infection confers future immunity to SARS-CoV2 re-infection in health care workers who have a higher prevalence of infection than the general population; in one study 20% of asymptomatic London HCW tested for SARS-CoV2 were positive on PCR tests over a 5-week period from 23rd March to May 3rd, with incidence peaking at 7.1% and declining 6 fold after 5 weeks. It will also determine whether reinfection is due to a lack of neutralising antibody or waning immunity over time in individuals who are anti-SARS-CoV-2 positive, whether such individuals can be re-infected symptomatically or asymptotically, about the incidence of new infection in individuals without prior exposure and the severity of COVID-19 in initial and subsequent episodes of infection if re-infection occurs. We will also use genomics to determine the viral diversity in healthcare workers at each sampling point and particularly if individuals are re-infected.

2.0 Aim and Objectives

Aim: The overall aim of this study is to determine if prior SARS-CoV-2 infection in health care workers confers future immunity to re-infection.

Objectives:
Primary: To determine whether the presence of antibody to SARS-CoV2 (anti-SARS-CoV2) is associated with a reduction in the subsequent risk of re-infection over short term periods (reviewed monthly) and the next year

Secondary:

1. To estimate the prevalence of SARS-CoV-2 infection in healthcare workers by region, using baseline serological testing at study entry of healthcare workers and symptom history from February 1st 2020 to date of study entry
2. To estimate the subsequent incidence of symptomatic and asymptomatic SARS-CoV-2 infection and determine how this varies over time, using fortnightly PCR testing (combined with any intercurrent testing on development of symptoms)
3. To estimate cumulative incidence of new infections in healthcare workers stratified by age, sex, ethnicity and co-morbidities.
4. To measure the ability to culture viable virus from cases of re-infection diagnosed by PCR and whether those that are persistently positive on PCR are continuing to shed viable virus.
5. To use genomic comparison to determine whether healthcare workers who become PCR-positive for a second time within a defined time frame are experiencing persistent infection or re-infection.
6. To determine whether serological response wanes over time
7. To determine whether there is a relationship between serological response (using enzyme immunoassay detection of IgG) and the presence of neutralising (protective) antibodies
8. To identify serological, demographic or clinical factors that correlate with the presence of neutralising antibodies
9. To investigate the phylogenetic relatedness of SARS-CoV2 viruses causing healthcare worker infections
10. To investigate the relationship between illness severity, demographics and neutralising antibody production

3.0 Methods

3.1 Study design
The overall study design is a prospective longitudinal cohort study of NHS staff at representative acute Trusts across England

3.2 Study sites
Each acute Trust will be asked to recruit invite all staff to participate, with allocated quotas within the cohort at each site to ensure appropriate pro rata representation across doctors, nurses, professions allied to medicine, healthcare assistants, administrative staff in clinical settings, and porters i.e.
professions which have a patient facing role. A total of minimum 40 NHS hospitals will be asked to participate; with a minimum of five hospitals in each of the six NHS regions outside London and oversampling of 10 Trusts in London given the higher incidence of disease, higher density of the population, and higher likelihood of seeing re-introduction of cases. Sites in Northern Ireland, Wales and Scotland will also be asked to participate.

3.3 inclusion/exclusion criteria

Inclusion criteria

- Health care worker who works in a clinical setting where patients are present
- Can provide written consent
- Is likely to be remain engaged with follow up, even if they move to another hospital or Trust site

Exclusion criteria

- Written consent not provided
- Temporary staff member

3.4 Sample size

A simulation approach has been used to estimate the power to detect relative differences between the study cohorts.

It is assumed that a total of 40 acute Trusts will be selected for this study. It is anticipated that in total 100,000 healthcare workers will be recruited; and with an estimated 25% seropositivity of healthcare workers (based on 20% of staff who were asymptomatic and tested positive in one London hospital between 23rd March and 2nd May 2020\textsuperscript{11}), this would provide a cohort of 2,500 seropositive HCWs and 7,500 seronegative HCWs to be recruited to this study.

For the longer term outcomes of the study, the initial sample sizes were calculated on the basis of recruiting 10,000 participants. It has been assumed that a minimum of 250 healthcare workers would be recruited from each selected Trust, with a standard deviation of 50. The proportion of seropositive recruits at each Trust has been obtained from a Gaussian distribution with a mean of 0.25 and standard deviation of 0.05 to reflect expected inter-Trust variation. The cumulative incidence in each Trust in the seronegative cohort has been simulated using Gaussian distributions with means of 0.05, 0.1, 0.2 and 0.3 each with a coefficient of variation of 0.2. This range represents that which is feasible to observe over a 12-month period, given the behavioural and social interventions still being employed during the study to control transmission.

A study duration of 52 weeks has been assumed with the inter-test period of 2 weeks. A total attrition of 35% of HCWs is assumed, unaffected by sero-status and occurring at a constant rate over the 52 weeks of the study. Attrition has been assumed to be independent of the infection process.
A range of cumulative incidence in the seronegative cohort has been used to reflect the immune effectiveness; with units in the simulations being allocated to be infected or not, using a draw from a Bernoulli distribution with p equal to the Trust and cohort specific simulated cumulative infection rate. A simplifying assumption of a constant infection rate over the study period has been used.

For each scenario as set of 200 simulations were performed. For each simulation, the total number if infections and person weeks of follow up was calculated for each cohort in each Trust. This data was analysed using a mixed effects Poisson model, using the natural logarithm of the person weeks as an offset. Power was estimated as the proportion of simulations for which the Wald statistic p value for the estimated incidence rate ratio of the seropositive to seronegative cohorts was less than 0.05. These are presented in Table 1, indicating that there is sufficient power for all but the smallest immune efficacy of 0.1 i.e. a 10% reduction in incidence in the seropositive cohort. Such a small reduction is indicative of an irrelevant level of protection to provide a means of controlling the pandemic via natural herd immunity.

Table 1: Power estimates obtained via simulation for a range of immune effectiveness and cumulative incidence

<table>
<thead>
<tr>
<th>Cumulative incidence in the seronegative at baseline cohort (per 100 HCW) in 12 months</th>
<th>Immune Effectiveness 10%</th>
<th>Immune Effectiveness 20%</th>
<th>Immune Effectiveness 30%</th>
<th>Immune Effectiveness 40%</th>
<th>Immune Effectiveness 50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.15</td>
<td>0.44</td>
<td>0.79</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>0.1</td>
<td>0.20</td>
<td>0.77</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.2</td>
<td>0.53</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.3</td>
<td>0.67</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

To investigate the ability of a study of this size to detect differential immune efficacy in subgroups of healthcare workers, for example those aged over 50s or from a BAME group a further set of simulations were undertaken. It has been assumed that the subgroup of interest is on average 35% of the total cohort size. The cumulative incidence used in the previous simulations has again been used with just those in the subgroup that are seropositive having an absolute reduction in immune efficacy of 5%, for example, if the cumulative incidence in the seronegative cohort is 0.3 (30%) and the immune efficacy in the non-sub group is 0.5 with a cumulative incidence of 0.15 (15%) in this group, the cumulative incidence in the seropositive sub group is 0.2 (0.15+0.05 20%). The interaction between cohort type and subgroup has additionally been added to the model and the estimated power obtained using the proportion of the 200 simulations for which this interaction had a Wald test p value of less than 0.05. Only immune efficacies of 0.3 or greater have been used, and these are presented in Table 2.

Table 2: Power estimates obtained via simulation for a range of immune effectiveness and cumulative incidence for detecting a sero-status by subgroup interaction

<table>
<thead>
<tr>
<th>Cumulative incidence in the seronegative at baseline cohort (per 100 HCW) in 12 months</th>
<th>Immune effectiveness 30%</th>
<th>Immune effectiveness 40%</th>
<th>Immune effectiveness 50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.15</td>
<td>0.79</td>
<td>0.98</td>
</tr>
<tr>
<td>0.1</td>
<td>0.20</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>0.2</td>
<td>0.53</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.3</td>
<td>0.67</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
In order to determine the outcome of immunity in much shorter intervals, by increasing the cohort recruitment to 100,000, we will be able to detect a difference between 0.5% and 0.1% cumulative incidence; even taking the incidence to as low as 0.2% in the seronegative group there is still excellent power of around 94%.

Estimated power for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1

<table>
<thead>
<tr>
<th>alpha</th>
<th>power</th>
<th>N</th>
<th>N1</th>
<th>N2</th>
<th>nratio</th>
<th>delta</th>
<th>p1</th>
<th>p2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>.9403</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.001</td>
<td>.001</td>
<td>.002</td>
</tr>
<tr>
<td>0.05</td>
<td>1</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.002</td>
<td>.001</td>
<td>.003</td>
</tr>
<tr>
<td>0.05</td>
<td>1</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.003</td>
<td>.001</td>
<td>.004</td>
</tr>
<tr>
<td>0.05</td>
<td>1</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.004</td>
<td>.001</td>
<td>.005</td>
</tr>
</tbody>
</table>

If the incidence is lower than this, i.e. 0.05% in the seropositive group and 0.1% in the seronegative the power becomes sub-optimal at around 66%, but is sufficient for large differences.

Estimated power for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1

<table>
<thead>
<tr>
<th>alpha</th>
<th>power</th>
<th>N</th>
<th>N1</th>
<th>N2</th>
<th>nratio</th>
<th>delta</th>
<th>p1</th>
<th>p2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>.663</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.005</td>
<td>.005</td>
<td>.001</td>
</tr>
<tr>
<td>0.05</td>
<td>.9933</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.001</td>
<td>.005</td>
<td>.0015</td>
</tr>
<tr>
<td>0.05</td>
<td>1</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.0015</td>
<td>.005</td>
<td>.002</td>
</tr>
<tr>
<td>0.05</td>
<td>1</td>
<td>1.0e+05</td>
<td>25000</td>
<td>75000</td>
<td>3</td>
<td>.002</td>
<td>.005</td>
<td>.0025</td>
</tr>
</tbody>
</table>
3.5 Site set up

A site initiation visit (SIV) will take place, in person or by telephone, for each study site. During the SIV, a member of the central study team will describe the study methods and discuss with the clinical and research team possible ways to organise recruitment and data collection. An electronic recruitment database will be provided to be completed by the site including all data PHE requires for data linkage and for monitoring the cohort composition. This will include the NHS number, name, date of birth, gender, date of enrolment, occupation, and study number. Study sites will be responsible for securely transferring an updated extract of this data base to PHE each week.

3.6 Recruitment

Each Trust will be provided with a timeline by which enrolment must be complete. Participants are recruited by an all-staff communication requesting volunteers and offering open sessions with members of the local study team.

Screening, eligibility assessment and formation of the cohort at each hospital: For each volunteer, a member of the local study team will review eligibility assessment. The proportions of participants from different occupational groups will be monitored. The cohort will be formed by accepting consecutive eligible volunteers until the quota for each staff group is filled. The site team may choose to send out further communications to support recruitment to under-represented groups. If a particular group’s quota cannot be filled, those places may be used for staff from other groups. The study is observational only and there is no randomisation or allocation to any subgroups.

Staff such as junior doctors may have planned moves to other Trusts within the study period. For these staff, they will be asked to continue to complete the questionnaire and either attend the enrolment hospital for regular swabs and bloods or transfer this function to their new hospital if possible.

All formal consent procedures will be undertaken by the team in each trust and participant will give full informed written consent.

3.7 Data collection and Measurements

3.7.1 Baseline data
Self-completed baseline questionnaire: The participant will self-complete an online survey at enrolment. Baseline data will include demographic/social factors (gender; date of birth; ethnicity; Smoking history); comorbidities including whether immunocompromised; job role including exposure to aerosol-generating procedures; whether had laboratory confirmed COVID-19 since February 1st, 2020, or symptoms compatible with COVID-19. (Appendix 1)

Initial serological assessment: The participant will have a blood taken at baseline which will be separated and aliquoted locally at each Trust. One aliquot will be tested locally for anti-SARS-CoV-2 using the commercial assay in use routinely in the trust. The second aliquot will be stored at -20°C or -70°C and batch shipped to PHE for later analysis using alternative assays to establish a correlate of protection and compare the serological responses across different assays. The composite PHE serological assessment will be used to assign the participant to the seronegative or seropositive cohort at baseline.

3.7.2 Ascertainment of Outcome Measures

Frequency of follow up visits: If possible the site team at each Trust will coordinate each study follow up visit to coincide with the routine PCR screening process being undertaken for all healthcare workers in the Trust. At the initiation of the study it is envisaged this will occur every two weeks; frequency may be altered to be more or less frequent (1 - 4 week intervals) depending on national and regional epidemiology.

PCR screening: This will be undertaken according to the local Trust procedure for screening healthcare workers or through a protocol set up specific for this study, and performed at regular intervals from 1 to 4 weeks and at any point the participant has clinical symptoms that meet local criteria for testing. It may be an administered nose and throat swab, or a self-swabbing nose kit depending on local trust preference. The swab will be tested for SAR-CoV-2 RNA in an accredited laboratory, which may be the clinical laboratory or through the national testing programme, using a polymerase chain reaction (PCR). As this is a clinical sample that requires reporting of the result to each individual HCW, full standard identifiers will be used.

Anti-SAR-CoV-2 testing: Each participant will have a venous blood test on the same day as the RT-PCR screening (within 48 hours is acceptable). The site team will determine the optimal procedure within each Trust, which may be visiting a study nurse, attending the phlebotomy service with a pre-designed form, or similar. A 10 ml venous blood sample will be taken and serum will be separated. The volume required for local testing will be removed and the residual volume will be used to make 1 ml aliquots. One aliquot will be tested immediately using the enzyme-linked immunosorbent assay (ELISA) rolled out as part of the national serological programme for healthcare workers. As this is a clinical sample that requires reporting, full standard identifiers will be used. At the start of the study, sites will be asked to retain the sera from participants who are, or who have previously been, seropositive or PCR positive, and ship this in batches to PHE for further characterisation. Any participant who has been vaccinated against SAR-COV-2 should also have their serum retained from the point of vaccination onwards. This additional PHE serological evaluation will be stopped or adjusted when the study team are of the view that there is sufficient data to predict the relationship of the commercial assay result to the serological
characterisation. A small seropositive sub-cohort will be selected to follow for a detailed assessment of antibody dynamics.

**Reporting of results:** The local PCR and serological results will be reported according to the Trust’s standard occupational health procedures.

**Action on positive results:** If the PCR screen is positive, the participant will follow the national guidance for self-isolation and the Trust guidance for return to work. If the serology assay is positive, the participant will receive advice indicating that this does not mean they are immune to reinfection, and that they must continue to adhere to infection control measures at home and work as usual.

**Follow up questionnaire:** The participant is asked to log in to the online survey system, through a reminder text or email, and fill in the monitoring questionnaire at the same frequency and timing of the PCR and Anti-SAR-CoV-2 testing. The follow up questionnaire will collect information on symptoms consistent with COVID-19 since last study visit, date of onset, duration, laboratory results, contact with definite case of Covid-19 (Appendix 2).

### 3.8 Participant withdrawal

Participants may withdraw at any time and this is explained in the participant information leaflet. On a request to withdraw they will be given the following options:

- Withdraw from any future study visits, but allow their existing materials and data to remain in the study and be tested, and allow the study team to continue to access their results from their routine PCR testing undertaken by the Trust.
- Withdraw from any future study visits, and withdraw the use of their existing materials, but allow the data generated up to the date of withdrawal to remain in the study. Their samples will be destroyed but their record will remain in the study database.
- Withdraw from any future study visits, and withdraw the use of their existing materials, and have their data removed from the study. Their samples will be destroyed and their record removed from the study database.

### 3.9 Storage of materials and additional Laboratory testing

All residual positive swab material, nucleic acid extract, and serum as outlined in the previous sections will be stored at the hospital site and shipped to PHE in monthly batches (or to agreed sequencing laboratories in the case of PCR-positive material). They will be used for confirmatory testing or for serological or viral genomic characterisation. No human DNA genomic investigations will be undertaken.

**Additional serological testing:** Seropositive participants in whom reinfection is identified, plus a cohort of matched non-infected seropositive controls, will have their sera further characterised using additional assays and for the presence of neutralising antibody, to provide hypothesis generating data on mechanisms of protective immunity.
Genomic analysis: All positive samples from participants will be sequenced as part of the routine sequencing of NHS residual samples in COG-UK Consortium laboratories. For participants who have more than one PCR test positive, genomes will be compared where possible to provide evidence to support reinfection or persistent infection as a mechanism. Phylogenetic analysis of SARS-COV2 from healthcare workers, using the study samples and the wider collection of genomes available through the COG-UK Consortium, will also be undertaken as an exploratory analysis into the diversity and spread of SARS-COV2 in healthcare workers.

4 Data management

All study documentation will be stored at each site, either in hard copy in a secure environment and/or in electronic copy in an access-limited location on a Trust server, as decided suitable by the Trust. Information received by PHE as the central study site will be stored securely in access-limited locations on PHE servers.

Study data is as follows:

- Recruitment log, containing personal and demographic information, stored by the Trust
- NHS and Lighthouse laboratory records, identified by personal information – this data (including name, date of birth and postcode) is automatically transmitted to PHE as part of the official notification of infection (under Health Protection Regulations) – it includes both results of RT-PCR tests and serology results performed by the laboratories. The study will access the routine data generated from the national screening programme in the participant Trusts with no additional laboratory requirements. The data collected by PHE will include the sample type, assay used and cycle threshold value for each sample tested.
- Questionnaire and electronic consent data entered by the participants and will include participant identifiable information to minimise the burden on research staff as the sample size has increased. This will be managed using SnapSurvey, a data management system held on a PHE server with end to end encryption.
- Viral genomic data generated by the local or consortium laboratories, identified by the Consortium identifier (pseudonymised with the link to personal information held by PHE under existing surveillance protocol)

All source data will be securely transferred to PHE with the identifiers described above. The transfer will be by secure email (recruitment log), through PHE’s established Second Generation Surveillance System (SGSS; laboratory data), or through end to end encrypted bespoke arrangements (questionnaire and genomic data). The person information is required for PHE to undertake secure data linkage across the sets as well as drawing in data from the Lighthouse laboratories. The database containing personal information will be access-limited to essential PHE staff working in the study team. Once linked to a participant, the data will be pseudonymised using the participant’s study identifier and stored on a
mirror database containing pseudonymised data only. All analysis will be performed by study personnel on the pseudonymised linked datasets.

Access to Data: Direct access will be granted to authorised representatives from the Sponsor for monitoring and/or audit of the study to ensure compliance with regulations. All researchers involved in data linkage have been trained in handling data according to Caldicott guidelines and Section 60 of the Health and Social Care Act. All researchers are aware of the Data Protection Act 1998 and the need to maintain absolute confidentiality.

Data Recording and Record Keeping: The data will be securely held at the National Infection Service, PHE. Data collection, storage and use will be consistent with the procedures described in the NHS Information Governance Toolkit. All databases will be encrypted and appropriately access-restricted.

Data storage: Electronic data will be stored on PHE secure servers and will remain active for the duration of the study. Participants’ identifiable data will then be removed and the data will be archived within the platform to be retained for a period of 5 years.

5. Analysis plan

All enrolled participants will be included in analyses, which will account for clustering by Trust. Analyses will be conducted after each 4-week period to inform the UK’s response to the COVID-19 pandemic. Results will be available to all organisations involved in the research. The study will end by default after 18 months, but by consensus of the study management group and funder may be terminated sooner if findings are sufficient.

There are no formal stopping rules for futility, utility or lack of power. The final decision to terminate the study will be made by the Public Health England and Department for Health and Social Care.

Estimates of both cumulative incidence and incidence density in the seropositive and seronegative cohorts will be obtained using mixed effects models assuming counts of PCR positive have a negative binomial distribution, a log link function, and the natural logarithm of the total number of subjects or the total follow-up time use as an offset, respectively. Inclusion of a binary predictor indicating the sero-status of the cohort into this model will provide estimates of the incidence rate ratio. Trust will be incorporated as a random intercept to account for unmeasured, shared, Trust level factors. To account for a non-constant force of infection, calendar month will be incorporated as an additional random effect. An assessment of the role of factors such as age, gender, ethnicity in immunity will be explored by inclusion of interactions within the model between each and serological status.

While the above analytical approaches provide a “classical” person-years approach to prospective cohort analysis and provide familiar measures of association, it may be inadequate to assessment of immunity provided by seroconversion. As it is expected that seropositivity is likely to confer a degree of short to median term protection for a SARS-CoV-2 infection, multi-state and parametric cure rate models incorporating frailty will be employed. These “survival” type of models provide a more detailed
assessment of factors associated with both short term and longer-term protection from infection, and how immunity may wane over time. Both mixture, explicitly assuming an immune and non-immune group and non-mixture “cure rate” models will be assessed using information criterion to choose which provides a better fit to the observed data. Bayesian approaches to cure rate models with frailty as describe by deSouza\(^\text{17}\) will be employed.

Multi state models explicitly allowing those within an “immune” state to flow into a “susceptible” state as antibodies wane will also be employed. This framework can allow subjects to move from seronegative (susceptible) to seropositive (immune) when infected during the study period. An additional absorbing state will be used for those infected that died. It is also possible to introduce “misclassification” of state into the multi state model, providing an estimate of sensitivity to account for imperfect serological tests. Approach like those proposed by Jackson\(^\text{18}\) will be employed.

Procedure for Accounting for Missing, Unused, and Spurious Data: Analyses will be restricted to complete cases. The PCR test for virus is being used as a diagnostic test and hence has extremely high performance. Sufficient sera will be obtained to re-run the immunological assays in case of initial assay failure. For similar reasons we do not anticipate that spurious data will be obtained.

Procedures for Reporting any Deviation(s) from the Original Statistical Plan: Deviations from the original statistical plan or the statistical analysis plan will be described and justified in the analysis reports.

6. Confidentiality and information governance

The data management plan (Section 4) demonstrates that PHE will need to be able to extract data from multiple Trusts, including staff moving between Trusts, national testing laboratories undertaking clinical diagnostic tests when staff are unwell, and potentially national sequencing consortium laboratories. Data linkage based on a single study identifier is neither practical nor robust in such circumstances. The suggested process for handling identifying data in a secure and appropriate way is therefore as follows:

- On enrolment, participants will be given a unique study identifier (three letters identifying the Trust, plus five digits identifying the participant)
- The Trust will compile a participant database containing the name, date of birth, NHS number, and study identifier of each enrolled participant
- The Trust will retain this for local study management and will transfer a copy of the database securely to PHE using appropriate encryption and will be stored in a limited-access environment on a PHE server. The database will be updated and mirrored securely to PHE every month.
- PHE will use the participant's personal information to retrieve their test results from the national surveillance systems as described in the data management plan. They may also use personal information to request the residual samples for additional processing in the sub studies.
• Positive swabs from the study will be routinely sequenced under the arrangement between the NHS and the COG-UK Consortium. In order to retrieve the genomic data, PHE may use the participant's identifying information to search the COG-UK data hub, which is within PHE, and retrieve the COG-UK study ID. The COG-UK study ID allows PHE to retrieve the viral genomes from that participant's sample from COG-UK (which maintains a pseudonymised genome collection on a non-PHE server). SIREN study team will make a request to the COG-UK team and will not be able to access any other fields in the COG-UK data hub or access the data hub directly.
• Once PHE has retrieved the data from the appropriate systems, it will be uploaded to a study database identified only by the study identifiers, and with all participant identifying information removed.
• This database will be submitted monthly for analysis.
• If a participant moves to another Trust, PHE will be notified through the monthly update of the study log.

7. Quality assurance and research governance

The study may be monitored or audited in accordance with the study protocol and standard operating procedures, GCP and relevant regulations.

Risk assessment: No formal risk assessment is required. The study involves recruiting individuals without symptoms who will be asked to give full informed consent to have swabs taken or self-swab their throat and nose and provide a blood sample taken by a study research nurse or the hospital phlebotomy department. Further participation in follow up visits to collect the same samples is based on consent of the participant. The main burden of participating in the study is the time taken for the study visits (which will be as far as possible at the same time as routine trust mandated screening) or the potential for minimal bruising from blood sampling though this is unlikely with experienced staff taking the samples. There is minimal risk of harm to any patient from participating since it does not include any therapeutic intervention. The surveys do not ask personal or intrusive information. The diagnostic test for the presence of virus from the nose and throat swab will be conducted by an accredited laboratory and will be returned to the participant’s research nurse and HCW through standard hospital practice.

Study monitoring: No GCP monitoring will be undertaken. As described there are minimal risks posed to patients by this observational and non-interventional study. The only study procedures are the completion of the questionnaire and taking of samples.
**Safety reporting:** There are no interventions in this study, and the only procedures are a standard blood draw performed by study HCP and a participant self-swab using a methodology that is being used widely across the country. Therefore, there is minimal safety risk to participants and safety reporting is not applicable.

**Study committees:** Oversight will be provided by a study management group including the investigators named above, representatives of collaborating and participating organisations as appropriate, and chaired by the Chief Investigator.

**Protocol deviations:** A deviation is a departure from the approved study protocol or other study document or process, or from Good Clinical Practice or any applicable regulatory requirement. Any deviations from protocol will be documented in a protocol deviation form and filed in the study master file.

**Serious breaches:** A serious breach is a breach of the protocol or of the conditions or principles of Good Clinical Practice which is likely to affect to a significant degree –

(a) the safety or physical or mental integrity of the trial subjects; or

(b) the scientific value of the research.

In the event that a serious breach is suspected the Sponsor must be contacted within 1 working day. In collaboration with the Chief Investigator, the serious breach will be reviewed by the Sponsor and, if appropriate, the Sponsor will report it to the approving REC committee and the relevant NHS host organisation within seven calendar days.
## Appendix 1: Amendment History

<table>
<thead>
<tr>
<th>Amendment No.</th>
<th>Protocol Version No.</th>
<th>Date issued</th>
<th>Author(s) of changes</th>
<th>Details of Changes made</th>
</tr>
</thead>
</table>
| 1             | 2                    | 20052020    | Susan Hopkins         | Removed Appendix 1 including PIL, Consent and Questionnaires.  
Added table of contents. |
| 2             | 3.1                  | 08062020    | Susan Hopkins         | Modified Consent to be electronic to reduce data burden.  
Increased sample size to determine whether there is immunity at much shorter time intervals to provide UK government with evidence of immunity or not at much faster intervals.  
Modified sample storage to reflect the larger sample size.  
Added that PCR swab samples could form part of screening protocol for Trust or specific to this study and taken by nose and throat swab or nose self sampling.  
Removed RT-PCR and changed to PCR to allow other local platforms to be used. |
References