medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Risk factors, symptom reporting, healthcare-seeking behaviour and adherence to public health guidance: protocol for Virus Watch, a prospective community cohort study

Andrew Hayward[1], Ellen Fragaszy[2,3], Jana Kovar[1], Vincent Nguyen[1,2], Sarah Beale[1,2], Thomas Byrne[2], Anna Aryee[2], Pia Hardelid[4], Linda Wijlaars[4], Wing Lam Erica Fong[2], Cyril Geismar[1,2], Parth Patel[2], Madhumita Shrotri[2], Annalan M D Navaratnam[1,2], Eleni Nastouli[5,6], Moira Spyer[6,], Ben Killingley[7,8], Ingemar Cox[9], Vasileios Lampos[9], Rachel A McKendry[9], Tao Cheng[11], Anne M Johnson[14], Susan Michie[13], Jo Gibbs[14], Richard Gilson[14], Alison Rodger[14,15], Robert W Aldridge[2]

Affiliations

1 Institute of Epidemiology and Health Care, University College London, London, UK

2 Centre for Public Health Data Science, Institute of Health Informatics, University College London, UK.

3 Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London, UK.

4 UCL Great Ormond Street Institute of Child Health, London, UK.

5 Department of Population, Policy and Practice, UCL Great Ormond Street Institute of Child Health, London, UK.

6 Francis Crick Institute, London, UK.

7 Health Protection and Influenza Research Group, Division of Epidemiology and Public Heath, University of Nottingham School of Medicine, Nottingham, United Kingdom.

8 University College London Hospital, London, United Kingdom.

9 Department of Computer Science, University College London, London, UK.

10 London Centre for Nanotechnology and Division of Medicine, London, UCL.

11 SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK.

12 Centre for Population Research in Sexual Health and HIV, Institute for Global Health, London, UK.

13 Centre for Behaviour Change, University College London, London, UK.

14 Institute for Global Health, University College London, London, UK.

15. Royal Free London NHS Foundation Trust, London, UK.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Corresponding Author:

Andrew Hayward, Institute of Epidemiology and Health Care, University College London, London, WC1E 7HB, UK

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Introduction: The Coronavirus (COVID-19) Pandemic has caused significant global mortality and impacted lives around the world. Virus Watch aims to provide evidence on which public health approaches are most likely to be effective in reducing transmission and impact of the virus, and will investigate community incidence, symptom profiles, and transmission of COVID-19 in relation to population movement and behaviours.

Methods and analysis: Virus Watch is a household community cohort study of acute respiratory infections in England & Wales and will run from June 2020 to Sept 2021. The study aims to recruit 42,500 people, including 12,500 from minority ethnic backgrounds, for an online survey cohort. Nested within this larger study will be a sub-cohort of 10,000 individuals, including 3,000 people from minority ethnic backgrounds. This cohort of 10,000 people will have full blood serology taken between October 2020 and January 2021 and repeat serology between May 2021 and August 2021. Participants will also post self-administered nasal swabs for PCR assays of SARS-CoV-2 and will follow one of three different PCR testing schedules based upon symptoms.

Ethics and dissemination: This study has been approved by the Hampstead NHS Health Research Authority Ethics Committee. Ethics approval number – 20/HRA/2320. We are monitoring participant queries and using these to refine methodology where necessary, and are providing summaries of our preliminary findings to inform public health action by working through partnerships with our study advisory group, Public Health England, NHS and Government Scientific Advisory panels.

Keywords: COVID-19; cough; fever; diagnostic testing capacity; UK; cohort study

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

The Coronavirus disease 2019 (COVID-19) pandemic has caused millions of deaths and impacted lives around the world with the closure of schools, workplaces, and limitations on freedom of movement. Vaccines and effective scalable treatments for COVID-19 are being discovered, but whilst these are still being approved and further refined, we will need to rely on other measures to stop the spread of COVID-19. We will also require studies to examine their effectiveness as they are implemented across England and Wales.

COVID-19 transmission in the UK has started to increase since the end of August 2020. Governments, including those of the UK devolved nations, are adopting a wide range of control measures to limit the spread of infection. These include isolation of people with COVID-19 symptoms and their household contacts, widespread testing and contact tracing, digital contact tracing using mobile phone apps, broad social distancing measures, and local control measures. Environmental cleaning, hand hygiene and face mask use are also advised.

Much of our current knowledge of COVID-19 comes from observations at the more severe end of the disease spectrum, in hospitalised patients and individuals who die having tested positive for the disease.[1–3] Although large-scale studies of prevalence of PCR positive infection and seroprevalence have been established, there is currently limited information on symptom profiles through the course of illness in non hospitalised populations, children, social and behavioural risk factors for infection, strength and duration of immunity, household and community transmission risk, and population behaviours during periods of wellness and illness (including social contacts, use of public spaces, testing behaviours, isolation, mask use, hand and respiratory hygiene). This information can only be gathered accurately through prospective large-scale community cohorts. Our experience of the MRC/Wellcome Flu Watch study[4,5] and ESRC Bug Watch[6] study has allowed us to rapidly establish a national household cohort study of 42,500 individuals.

Virus Watch aims to provide evidence on which public health approaches are most likely to be effective in reducing the spread and impact of the virus and will investigate

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

community incidence, symptom profiles, and transmission of COVID-19 in relation to population movement and behaviour.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

Study design and setting

Virus Watch is a household community cohort study of acute respiratory infections in England and Wales covering the second and potential subsequent waves of the COVID-19 pandemic. The study period will be 1st June 2020 to 30th Sept 2021. The study aims to recruit 42,500 individuals, including 12,500 from minority ethnic backgrounds for an online survey cohort (study 1). Nested within this larger study will be a sub-cohort of 10,000 individuals (study 2), including 3,000 people from minority ethnic backgrounds. Participants in this laboratory sub-cohort will be selected based on their geographical distance away from one of our blood taking clinics; either a 10km radius from a clinic in cities, or a 20km radius in rural areas. Participants will be balanced to be representative of the UK population for sex, age and region. Figure 1 provides an overview of the study design.

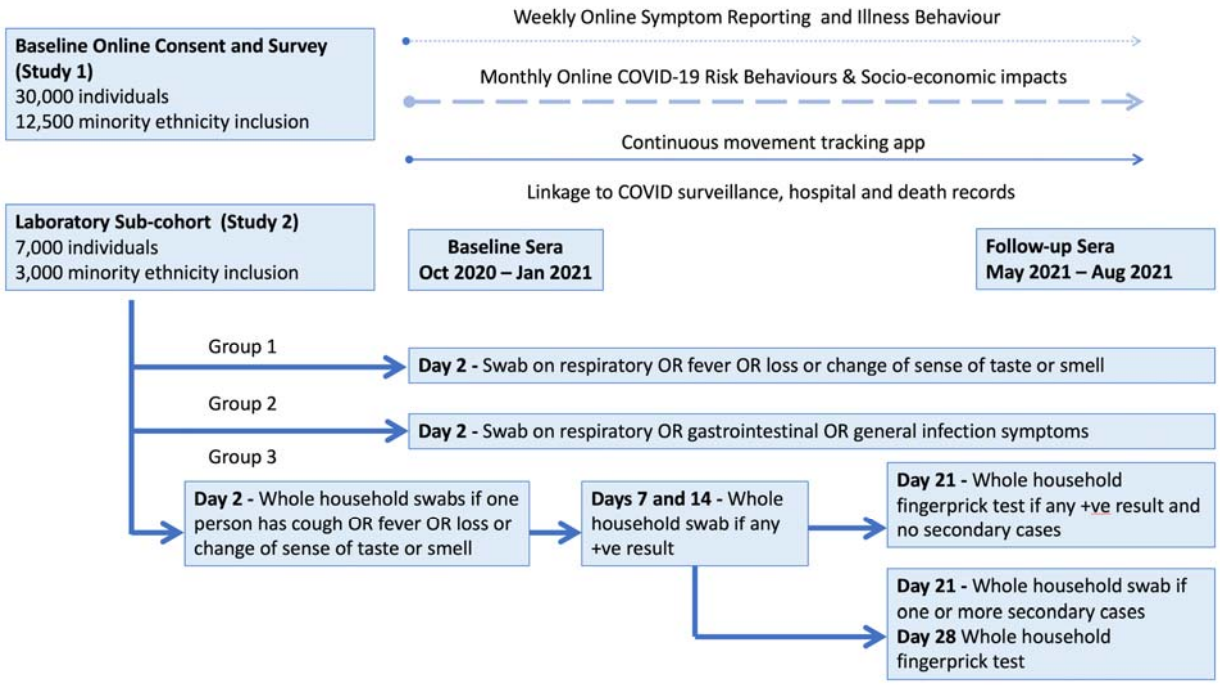

Figure 1. Overview of cohort recruitment and data collection for the Virus Watch household community cohort study.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Households self-select into the study if they live in England and Wales and all members of a household need to consent to take part in the study to meet our inclusion criteria. Households need to have an internet connection on a phone, tablet or computer, email, and, up until the end of November 2020, at least one adult household member that can read English. From December 2020 onwards, online surveys will be translated into multiple languages. A household is defined as one or more people (not necessarily related) whose usual residence (4 days/week or more) is at the same address. These householders share cooking facilities, a living room or sitting room or dining area.

Primary outcomes

Study 1: Online Survey Cohort

1. Incidence of respiratory infection symptoms, including COVID-19 disease case definitions.
2. Effectiveness and impact of recommended COVID-19 control measures including testing, isolation, social distancing, respiratory and hand hygiene measures on risk of respiratory infection.
3. Frequency of adherence with public-health recommendations for these control measures.
4. Proportion of community infections that result in hospital admissions and death.

Study 2: Laboratory testing sub-cohort

1. Incidence of PCR confirmed COVID-19.
2. Incidence of PCR confirmed COVID-19 in those with non-respiratory presentations.
3. Incidence of hospitalisation among PCR confirmed COVID-19 cases.
4. Proportion of individuals with SARS-CoV-2 antibodies acquired through natural infection to pandemic coronavirus.
5. Proportion of individuals with cross-reacting antibodies to seasonal coronaviruses acquiring (or not) SARS-CoV-2.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

6. Household secondary attack rates.
7. Protective effect of antibodies on infection and re-infection as well as the severity and spectrum of presentation.

Recruitment

We will use the Royal Mail Post Office Address File to generate a list of residential address lists from which households can be sampled and sent Virus Watch recruitment postcards to. The proposed initial sample design is a single-stage stratified probability sample where implicit stratification is employed to benefit from the precision gains that stratified sampling can bring. Within each region, residential addresses are sorted by (a) quintiles of Index of Multiple Deprivation 2019 (IMD), (b) within quintiles by Local Authorities, (c) postcodes and (d) address. We will perform this in the 9 Government Office Regions of England as well as Wales (10 study regions in total).

We will assess recruitment rates and the representativeness of this initial sample following the mail out of 50,000 postcards. If recruitment is lower than expected or under-representative of the national population, we will redesign our recruitment campaign to include a range of methods in order to build the cohort. This mixed recruitment strategy will be flexible and use a variety of methods including social media, study leaflet drops, text messaging, personalised letters and incentives. Social media adverts will be used to inform individuals about the study and direct them to our website http://ucl-virus-watch.net/ where they can read the participant information sheets and consent to taking part. Digital invitations will also be created for sharing via WhatsApp. Text messages and postal letters inviting patients from their General Practitioner clinics will be organised via Local Clinical Research Networks.[7] We will also work with trusted community partners and religious organisations to promote recruitment into the study.

In order for a household to be enrolled, they will require an internet connection (Wi-Fi, fixed or on a mobile phone), email address, and all household members must agree to take part. Households will nominate a lead householder who will submit study questionnaires. Up until December 2020, the lead householder will need to be able to

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

read English to support other household members in survey completion. From December 2020 onwards, online surveys will be translated into multiple languages and this will no longer be an inclusion criteria. A household is defined as one or more people (not necessarily related) whose usual residence (4 days/week or more) is at the same address. These householders share cooking facilities, and may share a living room or sitting room or dining area if available. Households with more than six members will not be eligible for the study - this criteria was set due to limitations of the REDCap survey infrastructure which did not function correctly when attempting to work with household sizes of greater than six during our pilot testing of the survey.

Virus Watch is powered for our primary aims in study 2 and the estimation of population-level symptomatic COVID-19 attack rate over time. Based on an estimated clinical attack rate of 30% of whom 20% need hospitalisation, and 0.5% die we expect the following number of outcome events in our cohort of 10,000 individuals in study 2: 3000 COVID-19 illnesses, 600 hospitalised cases, and 15 deaths. At one month into the outbreak we would be able to detect a 1.7-fold greater risk of disease in a population subgroup that constitutes 1/5 of the population, and by 2 months the detectable relative risk would be only 1.2. At one month we could detect a 4% hospital admission rate amongst cases with 95% CI of 0.5-6.8, and by 2 months the confidence intervals would narrow to 3.1-4.1. We have used estimates of the expected number of events over time to provide an indication of the fact that the cohort is sufficiently large to provide valuable information through the course of the pandemic. Sample size calculations have been informed by a realistic assessment of what we can achieve based on our previous experience[4,6]. For the serology cohort of 3000 people from minority ethnic backgrounds we assume a modest design effect (DE) due to household and geographical clustering, and 500 participants for six different minority ethnic backgrounds to enable the measurement of a cumulative incidence of 10% with 95% confidence intervals of 3% by each minority ethnic group.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Participant materials and incentives

Participant information sheets will be held on our study website. In order to participate, the whole household must take part. Each adult participant will need to read through study information, and provide online informed consent for themselves and any children they are legally responsible for. Children aged 6-9 and 10-15 years respectively will also be asked to read through age specific study participant information sheets and provide online informed assent. For children aged 5 and under, parents/guardians will consent on their behalf. Copies of translated consent questions will be provided where possible for those unable to read English. Informed consent data will be securely stored in UCL's Data Safe Haven which has been certified to the ISO27001 information security standard and conforms to NHS Digital's Data Security and Protection Toolkit. Local study teams will re-consent participants face to face, prior to undertaking blood sampling, and adult participants in study 2 will be offered a £10 voucher to reimburse travel costs. We will seek ethical approval for the use of recruitment incentives if levels of recruitment are lower than expected.

Data collection and follow-up

Study 1: Online Survey Cohort

The online survey cohort will collect data and follow up participants through six different sources. Survey data will be collected using Research Electronic Data Capture (REDCap) electronic data capture tools hosted on the UCL Data Safe Haven.[8] REDCap is a secure, web-based application for research studies. The UCL Data Safe Haven provides a technical solution for storing, handling and analysing identifiable data. It has been certified to the ISO27001 information security standard and conforms to NHS Digital's Data Security and Protection Toolkit.

1) Baseline Survey. The Lead Householder will be asked to complete an online baseline survey for each member of their household. Information collected includes:

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

demographics, occupation, income, ethnicity, country of birth, year of entry to UK, chronic medical conditions, medications, pregnancy status, vaccines, mode of transport to work, any previous contact with someone with COVID-19, previous symptoms of COVID-19-like illness and infection-prevention behaviours such as social distancing and hand hygiene.

2) Illness Surveys. Participants will be followed-up weekly via an email with a link to an illness survey. This is a weekly survey of the presence or absence of symptoms that could indicate COVID-19 disease including respiratory, general infection symptoms or gastrointestinal symptoms. During illness, prospective daily symptom recording, quality of life, health seeking behaviour (NHS 111, GP in person, GP by phone, A&E, Pharmacy, Hospital), treatments, and NHS investigations will be recorded. This survey will also include any respiratory and hygiene measures, self-isolation, activities and social contact, travel and face mask use. The survey includes questions to the household on activities undertaken in the week prior to symptom onset. The weekly survey will also be used to capture test results received from outside the study and requests to self isolate eg. via the UK Test-Trace-Isolate system.

3) Monthly Surveys. A number of questions will be asked every month. The monthly surveys also provide flexibility to ask additional questions (eg. behavioural changes) to reflect any new government directives on social distancing, testing, contact tracing, and vaccine delivery. Core questions will also allow us to follow up reasons for any non-response in a given month- (e.g. because of illness, hospitalisation or holiday). We will also ask about online health information seeking, social distancing, including recent (week before) contacts, activities, places visited and hand & respiratory hygiene. We will ask about finances, employment, and mental health to see how the COVID-19 response is affecting participants' wellbeing and ability to work. We will ask about access to healthcare for non–COVID-19 health problems to explore the indirect health impacts of the pandemic. We will ask about any COVID-19 PCR or antibody test results performed outside the study and not already reported through baseline surveys. We will ask about influenza vaccine uptake and COVID-19 vaccination intentions and uptake.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

4) Data Linkage. NHS Digital will undertake quarterly data linkage between cohort 1 and Hospital Episode Statistics (HES), which includes admitted patient and critical care episodes, outpatient department bookings, and emergency care contacts. This linkage will also include Office for National Statistics mortality data and virology testing data routinely collected by Public Health England, Public Health Wales, and the Department of Health and Social Care through 'Pillar 1' (testing in hospital patients and health and care workers) and 'Pillar 2' (community testing). These data sources will be linked to the cohort using name, NHS numbers, dates of birth and postal address. Identifying variables will be removed before the linked data are transferred back to UCL for analysis. These data linkages will continue for up to 5 years after the end of the study as we anticipate COVID-19 will become a recurring winter infection and we wish to understand its impact on health services in subsequent years. These linkage studies will identify any participants that have been admitted to hospital or died due to causes that could be directly or indirectly linked to the COVID-19 pandemic. Indirect causes include those related to limitations in healthcare access during the pandemic. Reductions in the use of routine health services will also be monitored via linkage to HES data.

5) Geo-location Tracking. All adult participants will be asked about optional consent to use a secure geo-location tracking app (Tracker for ArcGIS) installed on their mobile phone for the duration of the study.

6) Home antibody finger prick tests. 5000 members of the online cohort who are not part of the Laboratory testing sub-cohort (including 2500 minority ethnic and 2500 White British people) will be offered home finger prick antibody testing kits as soon as available after the first wave of the pandemic and after the second wave of the pandemic.

Study 2: Laboratory testing sub-cohort

All participants agreeing to take part in the main cohort (study 1) will be asked to provide consent to be contacted and invited to participate in one of the three laboratory testing sub-groups. This will enable a cohort of 10,000 individuals selected from the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

main cohort of 42,500 individuals to be maximally representative of the population of England and Wales. All participants taking part in study 2 will be asked to use the national test, trace and isolation system in addition to providing samples as part of Virus Watch.

Study 2 will consist of three groups that will follow different schedules of antibody testing and nasal/throat swabs for PCR testing.

Group 1 (*n*=7000):
With data from this group we aim to identify infection in those with a wide range of respiratory symptoms. Participants will be asked to submit a nose/throat swab if they experience two consecutive days of: fever (>37.8), feeling feverish, or new persistent cough, or loss or altered sense of smell or taste (COVID-19 suspected case definition), or shortness of breath, or ear pain or change in hearing, or sore throat, or sneezing, or blocked nose, or runny nose, or wheeze or sinus pain or congestion (other respiratory manifestations).

Group 2 (*n*=1000):
This group aims to identify the importance of non-respiratory presentations. Participants will be asked to submit a self-taken nasal/throat swab for PCR identification of COVID-19 and other respiratory viruses if:

- Either - two consecutive days of respiratory symptoms (e.g. cough, runny nose, sneezing, shortness of breath, sore throat, blocked nose, sinus pain or congestion, ear pain or change in hearing, wheezing, loss of or altered sense of taste or sense of smell).
- OR – two consecutive days of gastrointestinal symptoms (e.g. diarrhoea/loose stools, abdominal pain, nausea or vomiting, loss of appetite).
- OR - two consecutive days of general infection symptoms (e.g. feeling feverish, having a high temperature, feelings of severe unexplained tiredness, generalised muscle or joint aches)

Group 3 (*n* = 2000):

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

This group aims to identify the extent of household transmission. Participants will be asked to submit a nose/throat swab if they experience two consecutive days of cough or fever or loss of sense of taste or smell. Household contacts of the index case will also be asked to submit a swab on the same day whether or not they have symptoms.

If any of the swabs indicate SARS-COV-2 infection, all household members will be asked to repeat the swab on Day 7 and Day 14. If there are no new SARS-COV-2 cases in the household arising from swabs on Day 7 and 14 (assumed secondary cases) then all household members will be asked to undertake a home finger prick antibody test on Day 21. If there is one or more secondary cases in the household then the entire household will be asked to take an additional swab on day 21 and then undertake the fingerprick antibody tests on Day 28.

End of follow-up

Online participant follow-up will end in May 2021 although depending on the progression of COVID-19 we may ask participants to continue in the study for longer. Participants will be sent an exit survey via email which will also ask participants for permission to be contacted for involvement in future related research. Participants will be contacted to arrange a second blood sample collection from May 2021. Follow-up through data linkage with Hospital Episode Statistics and Mortality data will continue for 5 years after the end of the study.

Laboratory testing

Antibody testing

Study 2 will be using two different types of antibody tests. First, full blood serology will be taken between October 2020 and January 2021. We will use experienced health care professionals, including research nurses from the NIHR Clinical Research Networks.[9] Depending on local circumstances, visits to participants' homes to take blood may also be arranged. Children aged 15 years or less can opt out of having their

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

blood taken but will be offered a finger prick antibody test conducted by a healthcare worker instead. All participants from laboratory group 3 will additionally be offered a fingerprick antibody test at the same time as blood taking. From May 2021 until September 2021, we will invite all participants back for full blood tests or, for children who do not wish to have a full bleed, healthcare worker-delivered finger prick based antibody tests.

Families of children who have not been able to attend for a blood test, or for a healthcare worker-delivered finger prick antibody test, will be provided with postal kits to perform these at home. We also plan to use finger prick antibody testing where local clinics are no longer able to undertake full blood tests due to COVID-19 travel restrictions. Extremely clinically vulnerable participants will be sent home fingerprick tests instead of being asked to provide a serological sample.

Virus detection

Participants will post swab samples for PCR assays of COVID-19, and subsequent testing for influenza virus, seasonal coronavirus, rhinovirus and respiratory syncytial virus (RSV). When COVID-19 is identified we will also undertake whole genome sequencing of the virus. Samples for COVID-19 diagnostics will be handled and processed according to the NHS and UCL guidance on sample handling during the COVID-19 pandemic.

COVID-19 PCR and serology results will be returned to participants via text and email message systems. These messages will include links to official support, information and advice from NHS and PHE as well as advice on how to interpret results based on current evidence. In laboratory group 3, where positive test results will trigger further testing of the household, the results email will also include details explaining the additional testing requests.

We are currently under a legal duty to notify Public Health England (PHE)/National Institute of Health Protection/Public Health Wales of positive COVID-19 PCR test results. The legal duty to report positive COVID-19 test results will be met by using

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

existing reporting systems used by the laboratories undertaking PCR swab testing for the study. Positive COVID-19 samples will be stored for later sequencing (separate funding to be sought).

Statistical analysis

Our primary analyses during the winter 2020/21 will focus on estimating age-specific weekly rates of symptoms and PCR-confirmed COVID-19 illness and hospitalisation. For this analysis we will use appropriate regression models that account for clustering by household and we will explore the use of stratification or weighting of the sample by age and region as necessary to give nationally representative estimates. Weekly rates will be expressed per 100,000 person-weeks for ease of comparison with national surveillance data.

We will examine the proportion of the population infected during the first wave (e.g. Feb 2020 to Sept 2020) and second and potentially future pandemics waves. We will estimate the percentage of the population infected by calculating age and wave-specific rates of serological infection and PCR-confirmed disease per 100 person-seasons. A person-season will be defined by the epidemic curve in the cohort and therefore rates will account for differential follow-up time during each epidemic peak. In these analyses we will examine risk factors for infection, disease, disease severity and disease transmission.

We will estimate the proportion of serologically confirmed SARS-CoV-2 infections leading to symptomatic disease. First, we will calculate age-adjusted attributable rates of illness due to infection (subtracting rates of respiratory illness in non seroconverters from those in seroconverters). Second, we will measure the proportion of seroconverters with PCR-confirmed COVID-19. Analyses plans will be developed prior to conducting all analyses.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Whilst the study is being conducted, we will produce early, preliminary results for participants, the general public and policy makers in order to inform the public health response to the pandemic. These analyses will be reactive to the epidemiological circumstances and are therefore not defined in this protocol.

Modelling

We will build on our experience of working with PHE, Google, and Microsoft to use anonymous national or subnational aggregate web search engine data[10,11] to monitor the spreading of the disease. We will use our study data as ground-truth to train real-time disease prevalence estimation algorithms. We will annotate GPS tracking data into standard categories including time at work and home, social venues, supermarkets, hospitals, GPs, and transport mode for incorporation in classical epidemiological analyses. Integrating the linked survey data, we will develop multi-level spatio-temporal transmission models predicting the impact of various social distancing strategies.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Patient and public involvement

Due to the urgent nature of this study, we did not involve participants in its original design. We have previously conducted PPI to support similar community cohort studies of acute infections using similar methodologies. We have engaged the Young Persons' Advisory Group for research at Great Ormond Street Hospital to provide feedback on our Children's Participant Information Sheets. We will provide opportunities for survey participants to comment on survey methodology at the first monthly survey and consider revisions based on this. At the baseline survey, and each month, we will ask participants what questions are important to them (in relation to COVID-19 epidemiology and response), and what research questions they would like us to answer. We are also monitoring participant queries through our study email address and using these to refine methodology where necessary. We have worked with the Race Equality Foundation and Doctors of the World in advising on the inclusion of people from minority ethnic backgrounds in Virus Watch and have set up an advisory group to inform the ongoing design and dissemination of health equity aspects of Virus Watch. This advisory group (consisting of lay members of the public, community leaders, charities and policy organisation) will guide our health equity analyses and steer us on its implications for people, communities and policy.

Data sharing and access

We aim to share aggregate data from this project on our website and via a “Findings so far” section on our website - https://ucl-virus-watch.net/. We will also be sharing individual record level data with personal identifiers removed on a research data sharing service such as the UK Data Archive. In sharing the data we will work within the principles set out in the UKRI Guidance on best practice in the management of research data. https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/. Access to use of the data whilst research is being conducted will be managed by the Chief Investigators (ACH and RWA) in accordance with the principles set out in the UKRI guidance on best practice in the management of research data. It is the intention that the data arising from this research will initially be collected, cleaned and validated

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

by the UCL research team and once this has been completed will be shared for wider use. We aim to make subsets of the data more rapidly available both on our study website and via the public facing dashboard during the ongoing phase of data collection. In line with Principle 5 of the UKRI guidance on best practice in the management of research data, we plan to release data in batches as they become available or as updated results are published. Individual record data linked using NHS Digital will not be shared, only aggregated results. HES and mortality data may be obtained from a third party and are not publicly available. These data are owned by a third party and can be accessed by researchers applying to the Health and Social Care Information Centre for England. We will put analysis code on publicly available repositories to enable their reuse.

Ethics

This is a national study that has been approved by the Hampstead NHS Health Research Authority Ethics Committee. Ethics approval number – 20/HRA/2320. The study is compliant with the requirements of General Data Protection Regulation (2016/679) and the Data Protection Act (2018). All investigators and study site staff will comply with the requirements of the General Data Protection Regulation (2016/679) with regards to the collection, storage, processing and disclosure of personal information, and will uphold the Act's core principles.

Contributors

Roles: Conceptualization (AH, EF, JK, PH, EN, BK, IC, VL, RAMcK, TC, AMJ, SM, JG, RG, AR, RWA) Investigation, Methodology (All authors), Project Administration (AH, EF, JK, VN, SB, TB, AA, PH, LW, WLEF, CG, PP, MSh, AMDN, EN, MSp, RWA) , Writing – Original Draft Preparation (All Authors), Software (VN, TB, SB, RWA), Resources (AH, EF, JK, PH, EN, BK, IC, VL, RAMcK, TC, AMJ, SM, JG, RG, AR, RWA), Writing – Review & Editing (All Authors).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Funding

The research costs for the study have been supported by the MRC Grant Ref: MC_PC 19070 awarded to UCL on 30 March 2020 and MRC Grant Ref: MR/V028375/1 awarded on 17 August 2020. The study also received $15,000 of advertising credit to support a pilot social media recruitment campaign on 18th August 2020.

Competing interests

ACH serves on the UK New and Emerging Respiratory Virus Threats Advisory Group. AMJ was a Governor of Wellcome Trust from 2011-18 and is Chair of the Committee for Strategic Coordination for Health of the Public Research.

1 Official UK Coronavirus Dashboard. https://coronavirus.data.gov.uk/ (accessed 30 Nov 2020).

2 Statistics. Statistics » COVID-19 Daily Deaths. https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-daily-deaths/ (accessed 30 Nov 2020).

3 Coronavirus (COVID-19). https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases (accessed 30 Nov 2020).

4 Fragaszy EB, Warren-Gash C, Wang L, *et al.* Cohort Profile: The Flu Watch Study. *Int J Epidemiol* 2017;**46**:e18. doi:10.1093/ije/dyv370

5 Hayward AC, Fragaszy EB, Bermingham A, *et al.* Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. *Lancet Respir Med* 2014;**2**:445–54. doi:10.1016/S2213-2600(14)70034-7

6 Smith CM, Conolly A, Fuller C, *et al.* Symptom reporting, healthcare-seeking behaviour and antibiotic use for common infections: protocol for Bug Watch, a prospective community cohort study. *BMJ Open* 2019;**9**:e028676. doi:10.1136/bmjopen-2018-028676

7 Clinical Research Network. https://www.nihr.ac.uk/explore-nihr/support/clinical-research-network.htm (accessed 30 Nov 2020).

8 Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;**42**:377–81. doi:10.1016/j.jbi.2008.08.010

9 Clinical Research Network. https://www.nihr.ac.uk/explore-nihr/support/clinical-research-network.htm (accessed 30 Nov 2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.15.20248254; this version posted December 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

10 Wagner M, Lampos V, Cox IJ, *et al.* The added value of online user-generated content in traditional methods for influenza surveillance. *Sci Rep* 2018;**8**:13963. doi:10.1038/s41598-018-32029-6

11 Lampos V, Majumder MS, Yom-Tov E, *et al.* Tracking COVID-19 using online search. arXiv [cs.SI]. 2020.http://arxiv.org/abs/2003.08086v10