UK Biobank

Coronavirus Serology Study: Sample collection waves 1-6

Version 1.2

www.ukbiobank.ac.uk
January 2023

This documentation was prepared by UK Biobank's Health Data Group.

UK Biobank coronavirus serology study sample collection waves 1-6 overview document

Background and rationale

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). At the start of the pandemic, coronavirus testing in the UK population was focused on patients with severe symptoms who were being admitted to healthcare settings. As such, the full spectrum of the disease (including the numbers of mild and asymptomatic cases that did not need medical intervention) was unclear. Furthermore, the nature of human-to-human transmission of the virus was not well understood.

Over a period of six months, UK Biobank collected monthly blood samples and symptom data (via a short questionnaire) from a convenience sample of its participants (as well as from their children and grandchildren aged over 18) to enable an assessment of the extent of previous infection with SARS-CoV-2 (by measuring blood antibodies) in different locations and age groups across the UK.

Study design

Design: Prospective population-based convenience sample of UK Biobank participants and their adult (i.e. aged over 18) children and grandchildren.

Setting: Community – UK Biobank population.

Data collection: Each month, participants were asked to provide a sample of capillary blood using a microsampling finger-prick device, and to complete a short paper-based or online questionnaire about symptoms they may have had. At the end of the study, participants were also asked to complete an online questionnaire assessing possible risk factors for SARS-CoV-2 infection from the start of the first national 'lockdown' in March 2020 until the end of December 2020.

Study duration: Participants consented to provide data on a monthly basis over a period of six months during 2020.

Study timeline:

* Samples were generally processed within 24 hours of receipt by the UK Biobank laboratory. Some samples were received by the UK Biobank laboratory after the cut-off date of 04/12/2020 for analysis by the TDI lab. Some samples were therefore received by UK Biobank but not analysed (n=45).

Sample size: The study aimed to recruit approximately 20,000 individuals, comprising approximately 10,000 existing UK Biobank participants and approximately 10,000 adult children or grandchildren of participants who lived at a different address (in order to determine the extent of infection across all age groups and in different areas of the country).

Participant selection: Participants were selected from a pool of consented individuals, stratified by region, socioeconomic status, age and sex.

Inclusion criteria

- The participant was a current participant of UK Biobank, or an adult child or grandchild of a UK Biobank participant (aged over 18);
- The participant was willing and able to give informed consent for participation in the study;
- The participant was willing to be contacted by UK Biobank;
- The participant had a valid email address;
- The participant was resident in mainland Great Britain.

Exclusion criteria

- The child or grandchild of the UK Biobank participant was under 18 years old;
- The participant had an invalid email address;
- The participant was resident in non-mainland Great Britain or overseas, owing to the logistics of distributing kits overseas.

Methods

Recruitment (see Figure 1): UK Biobank identified a sub-sample of UK Biobank participants who then received an email invitation to participate in the study. The first 100,000 invited participants were identified based on the postcode area of their home address, and their age and sex, with the aim of obtaining a pool of participants with a wide geographic spread to recruit from, and a balanced representation of different characteristics with respect to sex, age and socioeconomic status. Approximately 180,000 further participants who had previously told UK Biobank that they had a child/children were later invited to maximise registration from relatives (younger age groups). The email invitation provided brief information about the study, weblinks to the study information sheet and a list of frequently asked questions, and details of how to contact UK Biobank if participants had further questions. Other UK Biobank participants, for example those who had not provided a valid email address, were also able to register an interest in the study via the UK Biobank participant website but were asked to provide an email address.

If interested, participants were directed to the UK Biobank participant website in order to check their contact details, provide consent to provide a blood sample (with no feedback of results) in their own home and answer a brief questionnaire about potential symptoms of COVID-19. They were also asked to indicate if they would be willing to forward an email invitation from UK Biobank to their children and/or grandchildren (if applicable).

Consented participants received an acknowledgement email to thank them for registering their interest and to let them know they would be notified if they were selected to take part.

If a UK Biobank participant had adult children and grandchildren who they thought might agree to participate, UK Biobank sent the participant a similar acknowledgement email plus an invitation from UK Biobank to forward to these relatives. This invitation email for relatives provided brief information about UK Biobank and the specific purpose for inviting them to join the study. It provided weblinks to the study information sheet, a list of frequently asked questions, and details of how to contact UK Biobank if they had further questions.

If adult children and grandchildren were interested in participating, they were asked to visit a webpage to provide personal details (i.e. name, date of birth, address, phone number, email address, sex at birth and ethnicity) and to give consent to join the study. This consent included linkage to their medical and other health-related records and for their data and samples to be donated to UK Biobank to be used for the purposes of COVID-19 research. Participants and their children/grandchildren who agreed to take part were also asked to confirm that that they would do so on the understanding that there would be no feedback of results.

Relatives who signed up were then sent an email asking them to confirm their email address in order to complete the consent process. Once their email address was confirmed, an acknowledgement was sent by email thanking them for signing up and informing them that they would be notified if they were required to take part.

This recruitment process generated a pool of consented individuals from which UK Biobank selected as nationally representative a sample with respect to participants' geographic location as possible (n=20,203) for inclusion in the study. Eight study participants withdrew from UK Biobank prior to data being made available to researchers.

Cohorts of 5,000 participants were selected on a weekly basis for four weeks, with each cohort asked to provide samples every month for six months. Cohorts were stratified with respect to:

a. geographic location (approximately proportional to the UK population and to cover urban/rural areas);
b. age (in ten year bands) and sex (with equal numbers in each group);
c. socio-economic status defined using their Townsend score based on the output area of their postcode (with equal numbers within each tertile of the UK population based on the 2011 Census);
d. random selection of only one person per household (as co-infection rate was 80-100%).

Individuals who were not selected to take part received an acknowledgement email thanking them for their offer of help.

Individuals who were selected were sent an email notifying them that they would receive their first blood sampling pack within the next week, with a further email sent prior to the despatch of each subsequent sampling kit. An email and SMS were sent a week later asking them to take the sample and post it back that same day. A reminder email was sent to participants who had not returned a sample one week after being asked to take their sample. On receipt of a sample, participants were sent a thank you SMS.

Participants were free to opt out of receiving a blood sampling kit at any point and to withdraw their consent at any time.

Mailing blood sampling kits: One day prior to kit despatch, participants received an email informing them that their blood sampling kit was being despatched. UK Biobank despatched kits to participants using Royal Mail.

The kit packaging contained a copy of the instructions for use and a letter telling them how to collect and return their blood sample. The covering letter also included a hyperlink to a video showing how to take the sample, and to further information and the study frequently asked questions (FAQs). A short symptom questionnaire to be completed on the day that the sample was provided was also enclosed, although participants were encouraged to complete this online from wave 5.

Data collection

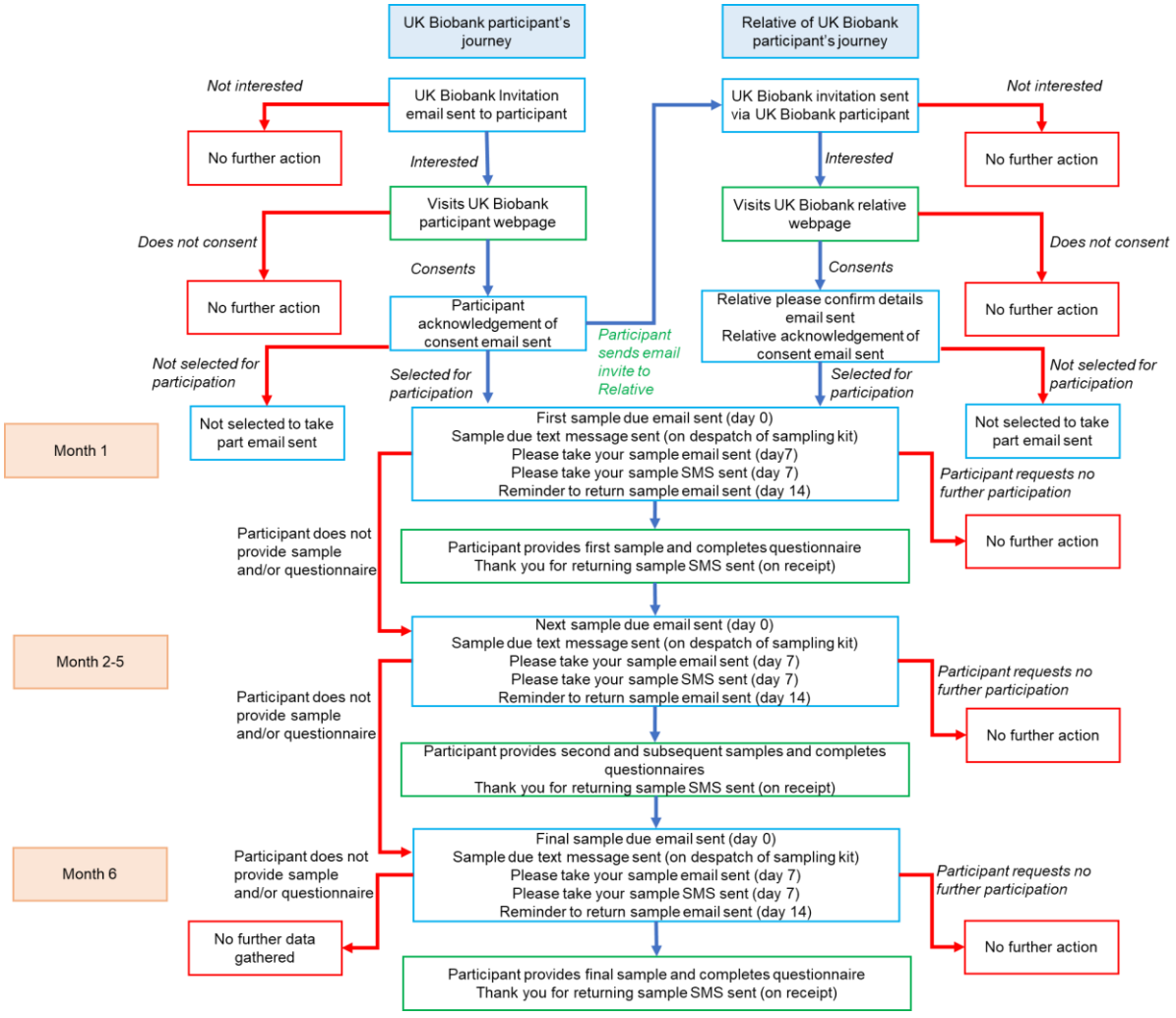
i. Blood samples: A capillary blood collection kit to take 0.5ml of blood (a tenth of a teaspoon) using a microsampling finger-prick device and blood collection tube was sent to participants on a monthly basis.

Participants were asked to return their samples and all lancets (i.e. used and unused) using the packaging provided in the kit to the UK Biobank laboratory on the day of collection. The samples were processed to separate the blood components, and the plasma was then shipped to an external laboratory where assays were conducted for IgG antibodies to the spike (S) protein.

No test results were provided to study participants.

ii. Symptom data: Participants were also asked to complete a short questionnaire about symptoms of COVID-19 infection that they may have had over the past month (since they last provided a sample for this study). This was completed solely on paper during waves 1-4 but participants were encouraged to complete this online from wave 5.

Figure 1: Study flowchart



ii. Online questionnaire to assess risk factors for SARS-CoV-2 infection: An online questionnaire was administered to study participants, with the aim of assessing possible risk factors for SARS-CoV-2 infection between the start of the first national 'lockdown' in March 2020 until December 2020. Participants received an email inviting them to complete the questionnaire. This email included a personalised weblink to the questionnaire homepage and contact details for the UK Biobank Participant Resource Centre, whose staff were trained to provide information and reassurance about the questionnaire. A reminder email was sent two weeks later to those participants who had not responded, and to participants who had a partially complete questionnaire.

Data cleaning

Prior to being made available, data collected from the study have undergone a small amount of cleaning. The data cleaning process involved the following steps:

- the removal of data relating to an invalid antibody test result;
- a small proportion of participants (0.3%) returned more than six samples across waves 1-6 (due to samples being collected using both an initial and replacement kit from the same wave). In such cases, the result from first sample received at the laboratory has been retained, and the result of the subsequent sample from the same wave has been discarded;
- the removal of invalid dates;
  - In the sample processing data (category 993), a self-reported sample collection date (field 28008) was defined as invalid if it met either of the following criteria: (i) it fell before the date the corresponding kit was sent to the participant; (ii) it fell after the sample was received by UK Biobank;
  - A single case where the recorded date the test sample was received by UK Biobank (field 28009) occurred after the date it was received at the laboratory that conducted the assay has also been defined as invalid. Note that the date the results of the symptom questionnaire corresponding to this test sample were returned by post (field 28032) has also been defined as invalid in the symptom questionnaire data (category 992);
  - The self-reported date that symptoms were first experienced (field 28030) and self-reported date of a positive COVID-19 test (field 28031) have been defined as invalid if they occurred after the completed symptom questionnaire was received at UK Biobank;
  - Dates defined as invalid have been replaced with an encoded value;
- The de-duplication of samples reported to have been collected on the same day. In cases where a participant returned two different samples that they reported collecting on the same day, the result of the sample taken using the first kit despatched has been retained, and the result of the other sample has been discarded. In each case the serostatus of the retained result was the same as the discarded result.

Researchers should be aware that the number of samples returned is, in many cases, different from the number of surveys returned for a given participant. This is because

participants did not always return a completed survey with their blood sample, because one or more of their test results were void, or, because they returned completed surveys from both the initial and replacement kits but only returned a sample from one of the kits. Hence, the results of the laboratory testing cannot be linked to the survey responses based on the instance index. However, some of the surveys returned by post can be linked to the sample they were returned with based on the date the envelope they were delivered in was received at UK Biobank (this corresponds to field 28009 in the sample processing data and field 28032 in the symptom survey data).

It should also be noted that a small number of cases exist in the symptom questionnaire data where the date the participant specified first experiencing the symptoms reported (field 28030) occurred before December 2019 (which is the date participants were asked to recall from for the wave 1 survey). These cases only relate to hand-written questionnaire responses returned by post, so may be the result of typographical errors, or due to the participant misunderstanding the question asked.

Accessing the data

Data collected from existing UK Biobank participants are available from the Data Showcase in category 993, and can be accessed in the same way as other standard data-fields, either as part of a main dataset downloaded from the Data Showcase or on the Research Analysis Platform. Data collected from the relatives of UK Biobank participants have been made available from the Returns Catalogue, and can only be accessed using the ukblink utility (see the Data Access Guide for guidance on how to download datasets from the Returns Catalogue).

The data are provided in separate datasets, in tab-separated format, for each of the different types of data collected throughout the study. See the table below for the Return ID associated with each dataset.

Returns Catalogue entries for relatives' data

Dataset	Return ID
Population characteristics	Return 2061
Sample processing for wave 1-6	Return 2063
Symptoms questionnaire for waves 1-6	Return 2064
Exposure questionnaire	Return 2066

As in the main UK Biobank dataset, the relatives datasets are rectangular datasets with one row per participant. The column headers are specified in the following format:

<field_id>.<instance_index>.<array_index>

See the Data Showcase for information about the Data-Codings and Instancings used for each field.