UK Biobank

Coronavirus Serology Study: Sample collection wave 7

Version 1.0

www.ukbiobank.ac.uk
November 2022

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

UK Biobank coronavirus serology study sample collection wave 7 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 in 2020, UK Biobank collected six monthly blood samples and symptom data (via a short questionnaire) from a convenience sample of 20,203 of its participants and their children and grandchildren aged over 18 to determine the extent of previous infection with SARS-CoV-2 (by measuring blood antibodies) in different locations and age groups across the UK (see Resource 4400 for more information).

In November 2021, a further sample and symptom data were requested from the same sample of UK Biobank participants and their relatives in order to make a comparative assessment of evidence of previous SARS-CoV-2 infection approximately twelve months after the sixth sample collection in October/November 2020.

Study design

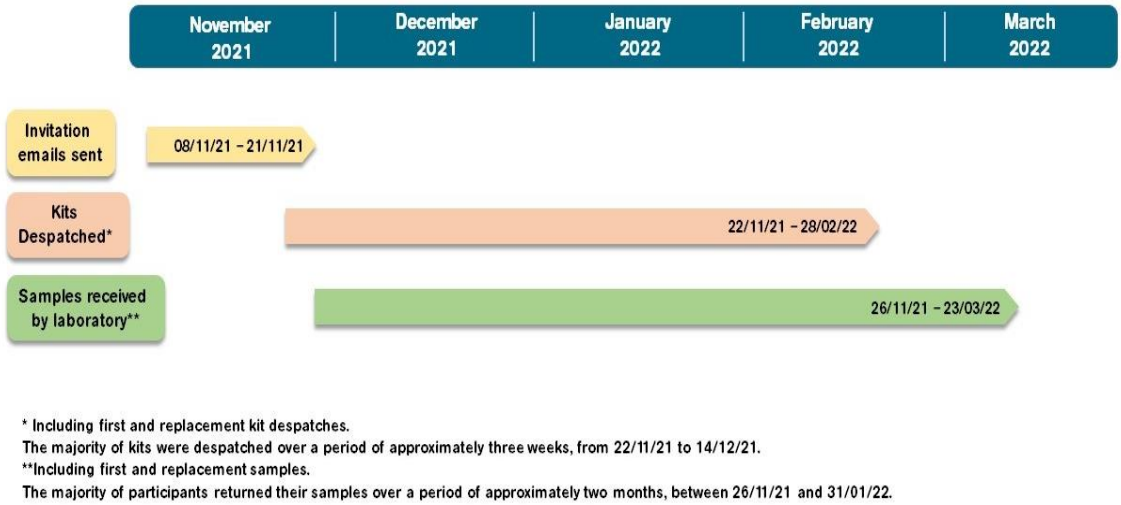
Design: Extension of the UK Biobank seroprevalence study of 20,203 UK Biobank participants and their adult (i.e. aged over 18) children and grandchildren.

Setting: Community – UK Biobank population.

Data collection: Participants received a capillary blood sampling kit (the Thriva coronavirus antibody test) via post. They were asked to provide a sample of capillary blood using a microsampling finger-prick device and to return this sample to a laboratory for testing for IgG antibodies to the nucleocapsid (N) protein (indicative of past infection with SARS-CoV-2). Participants were also asked to complete a short online questionnaire about any symptoms they may have had.

Study duration: Requests to provide a further sample were emailed to participants over a 1-month period. Blood sampling kits were issued and returned over a period of approximately 18 weeks.

Study timeline:



* Including first and replacement kit despatches.
The majority of kits were despatched over a period of approximately three weeks, from 22/11/21 to 14/12/21.
**Including first and replacement samples.
The majority of participants returned their samples over a period of approximately two months, between 26/11/21 and 31/01/22.

Sample size: Approximately 18,500 participants who consented to participate in the UK Biobank SARS-CoV-2 seroprevalence study in 2020, and who had not withdrawn their consent.

Participant selection: The study's inclusion and exclusion criteria are outlined below.

Inclusion criteria

- The participant consented and was selected to take part in the UK Biobank SARS-CoV-2 seroprevalence study in 2020;
- The participant was alive (regardless of health status);
- The participant was willing to be re-contacted;
- The participant had a valid email address;
- The participant was resident in mainland Great Britain.

Exclusion criteria

- The participant consented but was not selected to take part in the UK Biobank SARS-CoV-2 seroprevalence study;
- The participant was dead (as identified through national death registry records or via information from a relative);
- The participant had informed UK Biobank that they no longer wished to receive notifications;
- The participant had an invalid email address;
- The participant was now resident in non-mainland Great Britain or overseas (owing to the logistics of distributing kits overseas).

The participant materials also advised people with certain conditions not to take a blood sample. These were:

- participants with untreated clotting or bleeding disorders;
- participants who had had a recent mastectomy and there was swelling of the arm.

Methods

Recruitment (see Figure 1): Participants meeting the eligibility criteria described above (n=~18,500) were notified by email that UK Biobank would like to obtain a seventh sample to determine SARS-CoV-2 antibody status.

The email notification provided a brief reminder about the aims and objectives of the study, web links to the study information sheet, a list of frequently asked questions (FAQs), and details of how to contact UK Biobank if they had further questions. Participants were asked to visit the UK Biobank participant website to check their contact details.

Participants who were unable or who did not wish to participate were encouraged to decline via a web link to the UK Biobank participant website (provided in the email) to confirm that they would not be participating. Those who phoned the UK Biobank Participant Resource Centre were also able to decline via this route. Participants were given two weeks to check their details or to register a decline before blood sampling kits were despatched.

Individuals were sent an email notifying them that they would receive their blood sampling kit within the next fortnight, along with an SMS which informed them of their participant identification number. An email was sent 28 days later asking them to take the sample and post it back that same day. A further reminder email was sent to participants who had still not returned a sample 14 days after the first reminder email, followed by a final reminder email to return their blood sample, notifying them of the imminent closure of the study. On receipt of a sample, participants were sent a thank you SMS.

If participants returned a blood sample but did not complete the online symptoms questionnaire, reminder emails to do so were sent two days and then seven days after UK Biobank received their test result from Thriva.

Participants were free to withdraw their consent to provide a blood sample at any time.

Mailing blood sampling kits: Up to five days prior to kit despatch, participants received an email informing them that their blood sampling kit was being despatched. Participant addresses were securely transferred to Thriva Ltd (covered by a Data Privacy Agreement), who despatched kits to participants using Royal Mail.

The kit packaging contained a copy of the instructions for use and a letter telling participants how to collect and return their blood sample. The covering letter also included a hyperlink to a video showing how to take the sample (https://www.youtube.com/watch?v=okTozcGMDlU&t=1s), and to further information and the study frequently asked questions (FAQs). Participants were also given details of how to access a short online symptom questionnaire to complete on the day that they provided their sample.

Data collection

i. Blood samples: Participants used a lancet to obtain around 10 drops (0.4-0.6ml) of blood. The FAQs provided clear guidance about the minimum useful volume of blood to return in order to minimise the number of void results.

Participants were asked to return samples and all lancets (i.e. used and unused) using the packaging provided in the kit (a self-seal plastic specimen transport bag, a cardboard box and an outer pre-paid envelope). The envelope was mailed freepost to a nominated Thriva laboratory on the day of collection, and participants were provided with instructions to mail the kits Sunday-Friday only (to ensure prompt delivery).

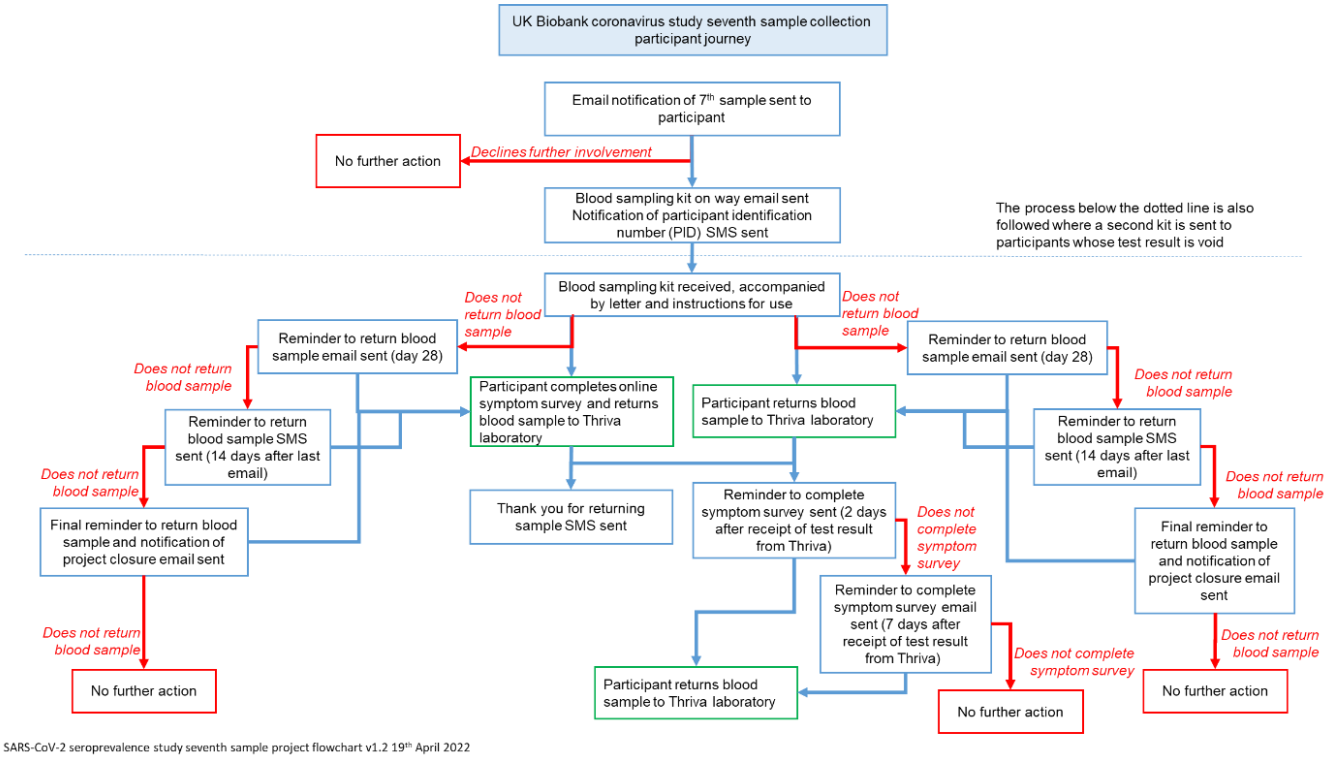
Assays were conducted for IgG antibodies to the nucleocapsid (N) protein (indicative of past infection). Thriva returned test results to UK Biobank on a daily basis. In accordance with the consent provided by participants at the start of the UK Biobank SARS-CoV-2 seroprevalence study, no blood sample test results were communicated to study participants.

If a void test result was received for a participant, UK Biobank informed the participant that the lab was unable to process their sample, and that a new blood sampling kit would automatically be sent to them.

ii. Symptom data: Participants were also asked to complete a short online questionnaire about symptoms of COVID-19 infection that they may have had since November 2020 (the date that we last collected samples for this study).

As part of the symptom survey, participants were asked whether they had ever received a positive PCR test result for COVID-19 and if they responded with 'yes', to provide the date of that result (see Data-Field 28166). The earliest date that was allowed on the online form was 1st January 2020. A few participants reported that they obtained a positive PCR test result on this date and several others also have self-reported dates in early 2020, when PCR testing for COVID-19 was not yet routine in the UK. Therefore, we advise caution when using these dates from early 2020.

Figure 1: Study flowchart



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 result from the Thriva coronavirus (COVID-19) antibody test;
- the de-duplication of more than one valid result (field 28140) per participant. There were few of these and in each case the most recently returned sample was retained;
- the removal of an invalid self-reported antibody test date in the symptom survey (field 28146). The definition of an invalid test date was: (i) it fell on or before the date the first test kit was sent to the participant (field 28143); (ii) it fell on or after the date the first test sample for the participant was received by the laboratory (field 28142). Invalid dates have been replaced by encodings.

Accessing the data

Data collected from existing UK Biobank participants are available from the Data Showcase in category 994, 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 7	Return 2067
Symptoms questionnaire for wave 7	Return 2085

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.