

Article

Coronavirus (COVID-19) Infection Survey technical article: analysis of populations in the UK by risk of testing positive for COVID- 19, September 2021

Analysis of populations in the UK by risk of testing positive for COVID-19 from the Coronavirus (COVID-19) Infection Survey. This survey is being delivered in partnership with University of Oxford, University of Manchester, Public Health England and Wellcome Trust.

Contact:
Rhiannon Yapp, Zoë Willis and
Joel Jones
Infection.Survey.Analysis@ons.
gov.uk
+44 1633 560499

Release date:
27 September 2021

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [Overview](#)
3. [Model 1: the core model: populations by risk of COVID-19 in the UK](#)
4. [Model 2: the screening model: populations by risk of COVID-19 in the UK](#)
5. [Model 3: the behaviours model: populations by risk of COVID-19 in the UK](#)
6. [Coronavirus \(COVID-19\) Infection Survey data](#)
7. [Collaboration](#)
8. [Glossary](#)
9. [Data sources and quality](#)
10. [Related links](#)

1 . Main points

- People who had received one or two doses of a coronavirus vaccine were less likely to test positive for coronavirus (COVID-19) in the fortnight ending 11 September 2021.
- People living in a household of three or more occupants were more likely to test positive for COVID-19 in the fortnight ending 11 September 2021.
- Those in younger age groups were more likely to test positive for COVID-19 in the fortnight ending 11 September 2021.
- People who never wore a face covering in enclosed spaces were more likely to test positive for COVID-19 in the fortnight ending 11 September 2021.
- Those who reported socially distanced contact with 11 or more people aged 18 to 69 years outside their household were more likely to test positive for COVID-19, in the fortnight ending 11 September 2021.

2 . Overview

About this article

This technical article presents the methods and results of analysis to identify characteristics of people who are more likely to test positive for coronavirus (COVID-19) in specific periods of time. It screens the different characteristics of people sampled in our Coronavirus (COVID-19) Infection Survey (CIS) who have and have not tested positive and uses a statistical model to assign risk to each of these characteristics.

In this article, analysis is focused on data from 29 August to 11 September 2021. This provides a snapshot of characteristics associated with testing positive for COVID-19 during this time. However, characteristics may not have the same association with testing positive in analyses of previous or future weeks. Therefore, we also present analyses over a longer period from 14 March to 11 September 2021.

This article contains the technical details of the methods used and the interpretation of results. This analysis will be updated regularly following this article and the main findings will be available in one of our [regular publications](#).

About this analysis

This analysis uses SARS-CoV-2 real-time reverse transcriptase polymerase chain reaction test (RT-PCR) results from nose and throat swabs of participants from the Office for National Statistics (ONS) Coronavirus (COVID-19) Infection Survey (CIS).

The CIS is a large household survey monitoring current COVID-19 infections within the community population in the UK. Community in this instance refers to private residential households and it excludes those in hospitals, care homes and/or other institutional settings. Participants were asked about demographics, living environment, behaviours, work, and vaccination uptake. Further information on the study design can be found in [COVID-19 Infection Survey: methods and further information](#).

The models included in this analysis were fit at the UK level and cover all ages included in the survey (two years and over). A positive result refers to a participant testing positive at least once over the period of analysis. The period of analysis is 14 days if at least 150 survey participants test positive in this period; otherwise, a 28-day period is used.

More detailed information about the [methods used for this publication](#) can be found in a recent [pre-print paper](#).

This analysis is based on three regression models which identify the characteristics that have the greatest effect on the likelihood of testing positive for COVID-19. The three models systematically build up from the first model. This approach enables us to isolate the effects of specific characteristics and behaviours which increase the risk of an individual testing positive.

We started with Model 1, the “core” model, which produces the most accurate representation of the effect of the core demographic characteristics on the likelihood of testing positive for COVID-19 prior to adding in further characteristics. Model 1 provides a model specification, which allows us to test the effect of each of the main demographic factors, on our full sample, before adding any additional variables. This core model is then used as a base for Model 2, which tests whether other factors that change over time independently affect the likelihood of testing positive. Examples of these factors include characteristics of work or school, or vaccination history.

Model 2 allows us to identify other predictors of positivity that were not included in the core model, while still controlling for the core demographic characteristics. These core demographic characteristics from Model 1 are included as controls to make sure the analysis adjusts for these factors. For this reason, when looking at the results from Model 2, we do not draw conclusions about the core demographic characteristics from Model 1.

Similarly, we use the specification from Model 2 as a base for Model 3, where we look at variables that are sensitive to changes – these are behavioural variables. This model allows us to identify behaviours that are related to positivity, while controlling for all the core demographic characteristics and the screened characteristics retained from Model 2. Similarly, when looking at the results from Model 3, we do not draw conclusions about the core demographic characteristics from Model 1 or the screened characteristics from Model 2.

Our first model, Model 1 (the core model), predicts the likelihood of an individual testing positive based on general demographic characteristics in order to help identify broad groups where infections are persisting or arising. This model contains eight core demographic characteristics:

- sex
- ethnicity
- age
- geographical region
- urban or rural classification of their address
- deprivation percentile
- household size
- whether the household was multigenerational

Associations between the risk of testing positive for COVID-19 and these core variables are estimated and included in all model outputs. The model also tests for interactions between the core demographic characteristics, retaining interactions that are significant at a 0.001 threshold.

We then built upon Model 1 resulting in Model 2, the screening model. This includes the core demographic characteristics from Model 1 and incorporates other characteristics individually to identify other factors associated with testing positive for COVID-19. It first gives us the impact of each newly included variable on the likelihood of testing positive while controlling for the core characteristics.

Examples of characteristics screened include work or school status, ability to social distance at work or school, patient-facing healthcare or social-care roles, work sector, travel to work, contact with patients/clients at work, current health status including COVID-19 vaccination, smoking status, travel abroad and whether participants have been in contact with care homes or hospitals. Work or school status is included as an additional control in all models testing other factors associated with work or school. Subsequently we include in Model 2 only those screened characteristics that predict whether someone tests positive for COVID-19, independently of the other characteristics, using a technique called backward elimination with a significance threshold of 0.05. The sample is slightly smaller here as cases with missing responses to the variables used are removed.

Finally, Model 3 (the behaviours model) adds behaviour variables to the core demographic characteristics from Model 1 and the screened characteristics that were kept in Model 2. Subsequently, this model examines behaviours individually to identify further characteristics associated with testing positive for COVID-19.

As with Model 2, it gives us the impact of each included variable on the likelihood of testing positive while controlling for all other variables in the model. Subsequently we include in Model 3 only those characteristics that predict whether someone tests positive for COVID-19, independently of the other characteristics, using backward elimination with a significance threshold of 0.05. A participant's behaviour(s) may change as a result of testing positive before the survey visit, for example, through a test taken in the national testing programme. To minimise this affecting our results, in this model, we take the maximum of each behaviour reported for the individual in the 35 days prior to the survey visit. To ensure that behaviour(s) prior to the positive test are reported, only participants who have at least one negative test between 10 and 35 days prior to the CIS visit were included in this model. This means that this model uses a smaller sample, including only those participants who have responses in a certain window before their visit while testing negative, so we can impute past behaviour and minimise reverse causality effects.

To investigate the effects of characteristic(s) over time we also present 14-day or 28-day results over time for Models 1, 2 and 3 in the [accompanying dataset](#). When looking at the core model (Model 1) over time we use main effect only models, with no interactions included. For Models 2 and 3, each variable is included individually on top of the established core model for each period – the screening characteristics and behaviours are modelled individually at this stage, in separate models, so that we can see the effect of all variables in each period while controlling for the core variables.

3 . Model 1: the core model; populations by risk of COVID-19 in the UK

This analysis uses data from 29 August to 11 September 2021, with 1,889 participants testing positive out of 167,288 participants from across the UK included in the model.

Figure 1 presents findings from Model 1 in the fortnight ending 11 September 2021.

- People living in a household of three or more occupants were more likely to test positive for coronavirus (COVID-19) than single occupancy households.
- People living in less deprived areas were less likely to test positive for COVID-19 than those in more [deprived areas](#).

Results for other characteristics included in the model can be found in the [accompanying dataset](#).

Figure 1: People in the UK living in a household of three or more occupants were more likely to test positive for COVID-19

The likelihood of testing positive for coronavirus (COVID-19) on nose and throat swabs by core demographic characteristic, UK, 29 August to 11 September 2021

Notes:

1. Deprivation is based on an index of multiple deprivation (IMD) score or equivalent scoring method for the devolved administrations, from 1 which represents most deprived up to 10 which represents least deprived. The odds ratio shows how a 1 unit increase in deprivation score, which is equivalent to 10 percentiles, effects the likelihood of testing positive for COVID-19.
2. An odds ratio indicates the likelihood of an individual testing positive for COVID-19 given a particular characteristic/variable. When a characteristic/variable has an odds ratio of one, this means there is neither an increase nor a decrease in the likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio greater than one indicates an increased likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio less than one indicates a decreased likelihood of testing positive for COVID-19 compared with the reference category.

Download the data

[.xlsx](#)

Figure 2 presents the predicted probability of testing positive for COVID-19 by age and by countries of the UK and region of England in the fortnight ending 11 September 2021.

The effect of age of individuals on the likelihood of testing positive for COVID-19 varied by countries of the UK and region of England in the fortnight ending 11 September 2021. People in the younger ages were more likely to test positive for COVID-19 across the UK in the fortnight ending 11 September 2021. In particular, those aged between 15 and 25 years were most likely to test positive. However, the amount by which younger people were at higher risk was greater in some areas, particularly Scotland and the South West, and smaller in others, particularly in London and the North West.

These are predicted probabilities that give the trend across years of age calculated at a reference value for other factors (male, White ethnicity, household size of one, non-multigenerational household, living in a major urban area, 50th deprivation percentile) and should not be used as population positivity rates. Levels of positivity in the population may be different because of other factors not included.

Figure 2: COVID-19 positivity rates were highest in the younger ages across the UK

The likelihood of testing positive for coronavirus (COVID-19) on nose and throat swabs by single year of age, by UK countries and England regions, 29 August to 11 September 2021

Notes:

1. Probabilities are subject to uncertainty, given that a sample is only part of the wider population. The model provides 95% confidence intervals around the probabilities.

Download the data

[.xlsx](#)

Populations at risk over time

Our [accompanying dataset](#) presents findings from Model 1 over time from 14 March to 11 September 2021.

Key findings are:

- People living in households with three or more occupants were more likely to test positive for COVID-19 than single occupancy households in most of the 14-day or 28-day periods from 14 March to 22 May 2021 and from 18 July to 11 September 2021.
- Females were less likely to test positive for COVID-19 than males in four out of five 14-day periods between 20 June and 28 August 2021.
- People living in less deprived areas were less likely to test positive for COVID-19 than those in more deprived areas from 14 March to 22 May 2021 and in four out of five periods from 4 July to 11 September 2021.
- Associations with testing positive for COVID-19 varied by countries of the UK and region of England since 14 March 2021.

Positivity rates were lower over June and July 2021 so the effects of different characteristics on an individual's risk of a positive test were less apparent.

4 . Model 2: the screening model; populations by risk of COVID-19 in the UK

This analysis uses data from 29 August to 11 September 2021, with 1,883 participants testing positive out of 166,296 participants from across the UK included in the model.

The number of participants included is smaller than for Model 1 because of the restriction criteria explained in the 'About this analysis' section in the [Overview](#).

Figure 3 presents findings from Model 2 in the fortnight ending 11 September 2021.

- People who had received one or two doses of a coronavirus vaccine (Astra Zeneca, Pfizer or Moderna) were less likely to test positive for coronavirus (COVID-19) than those not vaccinated.
- People who had had coronavirus (COVID-19) previously (had a positive swab in the survey or the England Test and Trace programme more than 120 days ago) were less likely to test positive for COVID-19 than those with no previous COVID-19 infection.
- People taking regular lateral flow tests were more likely to test positive for COVID-19 than individuals who do not. This has been seen consistently since 14 March 2021. Regular lateral flow tests are recommended for people whose work or other factors put them at high risk of getting COVID-19. Therefore, this association may reflect this underlying risk.

Other findings presented in Figure 3 may be more likely to be chance findings that are not seen consistently and should be viewed with caution.

- Those who report being impacted a lot by disability were less likely to test positive for COVID-19 than non-disabled individuals. This was only found on two other occasions since 14 March 2021.
- People who only vape were more likely to test positive for COVID-19 than non-smokers. This was only found on one other occasion since 14 March 2021.

Results for other characteristics included in the model can be found in the [accompanying dataset](#).

Figure 3: People in the UK who had received one or two doses of a coronavirus vaccine were less likely to test positive for COVID-19

The likelihood of testing positive for coronavirus (COVID-19) on nose and throat swabs by screened characteristic, UK, 29 August to 11 September 2021

Notes:

1. The core demographic characteristics from Model 1 are included as controls to make sure the analysis adjusts for these factors. We do not draw conclusions about the core demographic characteristics in this model.
2. An odds ratio indicates the likelihood of an individual testing positive for COVID-19 given a particular characteristic/variable. When a characteristic/variable has an odds ratio of one, this means there is neither an increase nor a decrease in the likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio greater than one indicates an increased likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio less than one indicates a decreased likelihood of testing positive for COVID-19 compared with the reference category.

Download the data

[.xlsx](#)

Populations at risk over time

Our [accompanying dataset](#) presents findings from Model 2 over time from 14 March to 11 September 2021.

Key findings are:

- People who had received one or two vaccine doses were less likely to test positive for COVID-19 compared with those not vaccinated in almost all periods from 14 March to 11 September 2021.
- People who had had COVID-19 previously were less likely to test positive than those with no previous COVID-19 infection from 20 June to 11 September 2021.
- People not working were less likely to test positive for COVID-19 compared with those working, from 6 June to 14 August 2021.
- People working in healthcare were less likely to test positive for COVID-19 in six out of eight periods between 25 April and 28 August 2021.
- People working in hospitality were more likely to test positive for COVID-19 in four out of five periods between 20 June and 28 August 2021.

5 . Model 3: the behaviours model; populations by risk of COVID-19 in the UK

This analysis uses data from 29 August to 11 September 2021, with 1,128 participants testing positive out of 114,700 participants from across the UK included in the model.

The number of participants included is smaller than for Models 1 and 2 because of the restriction criteria explained in the 'About this analysis' section in the [Overview](#).

Figure 4 presents findings from Model 3 in the fortnight ending 11 September 2021.

- People who never wore a face covering in enclosed spaces were more likely to test positive for coronavirus (COVID-19) than those who always wear a face covering in enclosed spaces.
- People who reported socially distanced contact with 11 or more people aged 18 to 69 years outside their household were more likely to test positive for COVID-19 than those reporting no socially distanced contact with adults aged 18 to 69 years outside their household.

Results for other characteristics included in the model and results for earlier time periods can be found in the [accompanying dataset](#).

Figure 4: People who never wore a face covering in enclosed spaces and people who reported socially distanced contact with adults aged 18 to 69 years were more likely to test positive for COVID-19

The likelihood of testing positive for coronavirus (COVID-19) on nose and throat swabs by screened behaviour, UK, 29 August to 11 September 2021

Notes:

1. The core demographic characteristics and screened characteristics from Models 1 and 2 are included as controls to make sure the analysis adjusts for these factors. We do not draw conclusions about the core demographic characteristics or screened characteristics in this model.
2. The category 'face coverings in enclosed spaces: not needed' refers to individuals reporting not leaving their home (for example, shielding) and so face coverings were not needed.
3. An odds ratio indicates the likelihood of an individual testing positive for COVID-19 given a particular characteristic/variable. When a characteristic/variable has an odds ratio of one, this means there is neither an increase nor a decrease in the likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio greater than one indicates an increased likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio less than one indicates a decreased likelihood of testing positive for COVID-19 compared with the reference category.

Download the data

[.xlsx](#)

6 . Coronavirus (COVID-19) Infection Survey data

[Coronavirus \(COVID-19\) Infection Survey technical article: analysis of populations in the UK by risk of COVID-19](#)

Dataset | Released 27 September 2021

Analysis of populations in the UK by risk of COVID-19 from the Coronavirus (COVID-19) Infection Survey.

7 . Collaboration

This Coronavirus (COVID-19) Infection Survey analysis was produced by the Office for National Statistics (ONS) in collaboration with our research partners at the University of Oxford. Of particular note are:

University of Oxford – Emma Pritchard and Sarah Walker

8 . Glossary

Confidence interval

A confidence interval gives an indication of the degree of uncertainty of an estimate, showing the precision of a sample estimate. The 95% confidence intervals are calculated so that if we repeated the study many times, 95% of the time the true unknown value would lie between the lower and upper confidence limits. A wider interval indicates more uncertainty in the estimate. Overlapping confidence intervals indicate that there may not be a true difference between two estimates.

For more information, see our [methodology page on statistical uncertainty](#).

Odds ratio

An odds ratio indicates the likelihood of an individual testing positive for COVID-19 given a particular characteristic or variable. When a characteristic or variable has an odds ratio of one, this means there is neither an increase nor a decrease in the likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio greater than one indicates an increased likelihood of testing positive for COVID-19 compared with the reference category. An odds ratio less than one indicates a decreased likelihood of testing positive for COVID-19 compared with the reference category.

Deprivation

Deprivation is based on an index of multiple deprivation (IMD) score or equivalent scoring method for the devolved administrations, from 1, which represents most deprived, up to 10, which represents least deprived. The odds ratio/coefficient shows how a one unit increase in deprivation score, which is equivalent to 10 percentiles, affects the likelihood of testing positive for COVID-19.

Multigenerational household

A household was classed as multigenerational if it included individual(s) aged school Year 11 or younger and individual(s) aged school Year 12 to age 49 years and individual(s) aged 50 years or over.

9 . Data sources and quality

Our methodology article provides further information around [the survey design and how we process data](#). A recent [pre-print paper](#) provides further information on the [methods presented in this publication](#).

A recent pre-print paper provides [further information, including strengths and limitations of the analysis presented in this publication](#). More information on the strengths and limitations of the data, data uses and users is available in the [Coronavirus \(COVID-19\) Infection Survey QMI](#) and the Coronavirus (COVID-19) Infection Survey [statistical bulletin](#).

10 . Related links

[Coronavirus \(COVID-19\) Infection Survey, UK](#)

Bulletin | Updated weekly

Estimates for England, Wales, Northern Ireland and Scotland. This survey is being delivered in partnership with the University of Oxford, University of Manchester, Public Health England and Wellcome Trust.

[Coronavirus \(COVID-19\) Infection Survey: characteristics of people testing positive for COVID-19 in England](#)

Bulletin | Updated fortnightly

Characteristics of people testing positive for COVID-19 from the Coronavirus (COVID-19) Infection Survey, including antibody data by UK country, and region and occupation for England. Antibodies data published before 3 February 2021 are available in this series.

[Coronavirus \(COVID-19\) Infection Survey: antibody and vaccination data for the UK](#)

Article | Updated fortnightly

Antibody and vaccination data by UK country and English regions from the Coronavirus (COVID-19) Infection Survey. This survey is being delivered in partnership with the University of Oxford, University of Manchester, Public Health England and Wellcome Trust.

[COVID-19 Infection Survey: methods and further information](#)

Methodology article | Updated 24 August 2021

Information on the methods used to collect the data, process it, and calculate the statistics produced from the Coronavirus (COVID-19) Infection Survey.

[Coronavirus \(COVID-19\) Infection Survey QMI](#)

QMI | Released 16 July 2021 Quality and Methodology Information for the Coronavirus (COVID-19) Infection Survey (CIS), detailing the strengths and limitations of the data, methods used, and data and users.

[Coronavirus \(COVID-19\) Infection Survey technical article: waves and lags of COVID-19 in England, June 2021](#)

Technical article | Released 29 June 2021

Data about the waves and lags of coronavirus from the Coronavirus (COVID-19) Infection Survey. This analysis has been produced in partnership with the University of Oxford.

[Coronavirus \(COVID-19\) Infection Survey technical article: analysis of reinfections of COVID-19: June 2021](#)

Technical article | Released 29 June 2021

Data about reinfections from the Coronavirus (COVID-19) Infection Survey. This analysis has been produced in partnership with the University of Oxford.

[Coronavirus \(COVID-19\) Infection Survey technical article: analysis of positivity after vaccination](#)

Technical article | Released 17 June 2021

Data about positivity after vaccination from the Coronavirus (COVID-19) Infection Survey. This analysis has been produced in partnership with the University of Oxford.

[Coronavirus \(COVID-19\) Infection Survey: characteristics of people testing positive for COVID-19 in England](#)

Article | Released 20 May 2021

Characteristics of people testing positive for COVID-19 from the Coronavirus (COVID-19) Infection Survey, including antibody data by UK country, and region and occupation for England. Antibodies data published before 3 February 2021 are available in this series.