

## Community prevalence of SARS-CoV-2 in England during April to September 2020: Results from the ONS Coronavirus Infection Survey

Koen B Pouwels, PhD<sup>1,2\*†</sup>, Thomas House, PhD<sup>3,4\*</sup>, Emma Pritchard, MSc<sup>2,5</sup> Julie V Robotham, PhD<sup>6</sup>, Paul J Birrell, PhD<sup>6,7</sup>, Andrew Gelman, PhD<sup>8</sup>, Karina-Doris Vihta,<sup>2,5</sup> Nikola Bowers,<sup>9</sup> Ian Boreham,<sup>9</sup> Heledd Thomas, MSc<sup>9</sup>, James Lewis, BA<sup>9</sup>, Iain Bell, BSc<sup>9</sup>, John I Bell, MD<sup>10</sup>, John N Newton, FRCP<sup>11</sup>, Jeremy Farrar, PhD,<sup>12</sup> Ian Diamond, PhD<sup>9</sup>, Pete Benton, MSc<sup>9</sup>, Ann Sarah Walker, PhD<sup>2,5,13,14</sup> and the COVID-19 Infection Survey team

\* contribution considered equal

† Correspondence to [koen.pouwels@ndph.ox.ac.uk](mailto:koen.pouwels@ndph.ox.ac.uk)

<sup>1</sup> Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK

<sup>2</sup> The National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford, Oxford, UK.

<sup>3</sup> Department of Mathematics, University of Manchester, Manchester, UK

<sup>4</sup> IBM Research, Hartree Centre, Sci-Tech Daresbury, UK

<sup>5</sup> Nuffield Department of Medicine, University of Oxford, Oxford, UK

<sup>6</sup> National Infection Service, Public Health England, London, UK

<sup>7</sup> MRC Biostatistics Unit, University of Cambridge, Cambridge Institute of Public Health, Cambridge, UK

<sup>8</sup> Department of Statistics, Columbia University, New York, NY, United States

<sup>9</sup> Office for National Statistics, Newport, UK

<sup>10</sup> Office of the Regius Professor of Medicine, University of Oxford, Oxford, UK

<sup>11</sup> Health Improvement Directorate, Public Health England, London, UK

<sup>12</sup> Wellcome Trust, London, UK

<sup>13</sup> The National Institute for Health Research Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

<sup>14</sup> MRC Clinical Trials Unit at UCL, UCL, London, UK

See Acknowledgements for the Coronavirus Infection Survey team

## Abstract

**Background:** Decisions regarding the continued need for control measures to contain the spread of SARS-CoV-2 rely on accurate and up-to-date information about the number of people and risk factors for testing positive. Existing surveillance systems are not based on population samples and are generally not longitudinal in design.

**Methods:** From 26 April to 19 September 2020, 514,794 samples from 123,497 individuals were collected from individuals aged 2 years and over from a representative sample of private households from England. Participants completed a questionnaire and nose and throat swab were taken. The percentage of individuals testing positive for SARS-CoV-2 RNA was estimated over time using dynamic multilevel regression and post-stratification, to account for potential residual non-representativeness. Potential changes in risk factors for testing positive over time were also evaluated using multilevel regression models.

**Findings** Between 26 April and 19 September 2020, in total, results were available from 514,794 samples from 123,497 individuals, of which 489 were positive overall from 398 individuals. The percentage of people testing positive for SARS-CoV-2 changed substantially over time, with an initial decrease between end of April and June, followed by low levels during the summer, before marked increases end of August and September 2020. Having a patient-facing role and working outside your home were important risk factors for testing positive in the first period but not (yet) in the second period of increased positivity rates, and age (young adults) being an important driver of the second period of increased positivity rates. A substantial proportion of infections were in individuals not reporting symptoms (53%-70%, dependent on calendar time).

**Interpretation** Important risk factors for testing positive varied substantially between the initial and second periods of higher positivity rates, and a substantial proportion of infections were in individuals not reporting symptoms, indicating that continued monitoring for SARS-CoV-2 in the community will be important for managing the epidemic moving forwards.

## Funding

This study is funded by the Department of Health and Social Care. KBP, ASW, EP and JVR are supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with Public Health England (PHE) (NIHR200915). AG is supported by U.S. National Institute of Health and Office of Naval Research. ASW is also supported by the NIHR Oxford Biomedical Research Centre and by core support from the Medical Research Council UK to the MRC Clinical Trials Unit [MC\_UU\_12023/22] and is an NIHR Senior Investigator. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, Department of Health, or PHE.

## **Research in context**

### **Evidence before this study**

Unprecedented control measures, such as national lockdowns, have been widely implemented to contain the spread of SARS-CoV-2. Decisions regarding the continued need for social distancing measures in the overall population, specific subgroups and geographic areas heavily rely on accurate and up-to-date information about the number of people and risk factors for testing positive. We searched PubMed and medRxiv and bioRxiv preprint servers up to 6 June 2020 for epidemiological studies using the terms “SARS-CoV-2” and “prevalence” or “incidence” without data or language restrictions. Most studies were small or had only information about current presence of the virus for a small subset of patients, or used data not representative of the community, such as hospital admissions, deaths or self-reported symptoms. Large population-based studies, such as the current study, are required to understand risk factors and the dynamics of the epidemic.

### **Added value of this study**

This is the first longitudinal community survey of SARS-CoV-2 infection at national and regional levels in the UK. With more than 500,000 swabs from more than 120,000 individuals this study provides robust evidence that the percentage of individuals from the general community in England testing positive for SARS-CoV-2 clearly declined between end of April and June 2020, followed by consistently low levels during the summer, before marked increases end of August and September 2020. Risk factors for testing positive varied substantially between the initial and second periods of higher positivity rates, with having a patient-facing role and working outside your home being important risk factors in the first period but not (yet) in the second period, and age (young adults) being an important driver of the second period of increased positivity rates. Positive tests commonly occurred without symptoms being reported.

### **Implications of all the available evidence**

The observed decline in the percentage of individuals testing positive adds to the increasing body of empirical evidence and theoretical models that suggest that the lockdown imposed on 23 March 2020 in England was associated, at least temporarily, with a decrease in infections. Important risk factors for testing positive varied substantially between the initial and second periods of higher positivity rates, and a substantial proportion of infections were in individuals not reporting symptoms, indicating that continued monitoring for SARS-CoV-2 in the community will be important for managing the epidemic moving forwards.

## Introduction

Since severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) started causing severe respiratory illness in Wuhan, China, in late 2019,<sup>1</sup> the virus has had a drastic impact worldwide. As of 20 September, there have been over 30.6 million confirmed cases and 950,000 deaths reported to the WHO.<sup>2</sup> Control measures, such as national lockdowns, have been widely implemented to contain the spread of the virus in a, at least temporarily successful,<sup>3-5</sup> attempt to prevent the collapse of healthcare systems and even larger numbers of deaths among those infected with SARS-CoV-2. Although such measures are important for control of the pandemic, they also affect the economy, unemployment rates, and global supply-chains.<sup>6,7</sup> Politicians continuously make the difficult decision between continuing strict control measures or relaxing them in some way that would be safe enough from a public health perspective yet beneficial more broadly across society.

Importantly, early detection of population subgroups driving new increases in infections is crucial to decisions around potential tailoring interventions or messaging without having to implement drastic measures affecting the whole society.

There are several reasons why risk factors may vary over time. First, behaviour and contact patterns of subgroups change over time without intervention, e.g. students starting university. Adherence to non-mandatory infection prevention measures may reduce more over time among subgroups with a low risk of COVID-19 related hospital admission and death than those that are more vulnerable. Moreover, subgroups that have been disproportionately affected in a first wave may have acquired sufficient immunity and may have better access to effective measures that reduce the risk of infection making them less likely to acquire a new infection during a second wave.

Here, we use data from the Office of National Statistics (ONS) Coronavirus Infection Survey (CIS). This large national survey with more than 500,000 swab results to 19 September is designed to be representative of the target population, offering a unique opportunity to identify risk factors that are driving recent new increases in the positivity rate, as well as investigating the proportion of individuals testing positive that do not report symptoms, potential false-positivity rate, and other factors that can directly inform policy around COVID-19 related control measures. We used Bayesian dynamic multilevel regression and poststratification (MRP) to account for any residual unrepresentativeness, a potential problem often ignored with surveillance data.

## Methods

Data were collected between 26 April and 19 September 2020 from individuals from randomly selected private households from address lists and previous ONS surveys to provide a representative sample of the population of England. Only individuals aged 2 years and older living in private households were eligible for inclusion in the survey. If one or more individuals from a household agreed to participate, a visit by a study worker was arranged and individuals were asked about any symptoms and contacts, together with information about their gender, age, ethnicity and occupation. These questions were collected directly by the study worker. The study worker provided instructions on how to self-swab the nose and throat, which has been shown to be comparable or even more sensitive than swabs performed by healthcare workers.<sup>9</sup> Parents/carers took swabs from children under 12 years old. The nose and throat self-swabs were couriered directly to the UK's national Lighthouse laboratories at Milton Keynes (National Biocentre) (from 26 April) and Glasgow (from 16 August), where the samples were tested using identical methodology for the presence of SARS-CoV-2 (3 gene targets) using reverse transcriptase polymerase chain reaction (RT-PCR) as part of the national testing programme.<sup>10</sup> After the first visit, participants were asked whether they were willing to participate in further follow-up visits: every week for the first 5 weeks of the study, and

then monthly thereafter. The study protocol is available at <https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/protocol-and-information-sheets>.

The project has been reviewed and given ethical approval by South Central - Berkshire B Research Ethics Committee (20/SC/0195).

#### *Trend in proportion of positive tests over time*

We analysed the proportion of the private-residential population testing positive for SARS-CoV-2 from nose and throat swabs over time using Bayesian dynamic Multilevel Regression and Poststratification (MRP).<sup>11,12</sup> MRP was used to correct for any residual non-representativeness in terms of age, sex and region. In several empirical and simulation studies MRP was superior at both the national and regional levels compared to classical survey weighted and unweighted approaches, including when using small sample sizes.<sup>11-16</sup> Partial pooling through the use of random effects in the multilevel model ensures stable estimates can be obtained for subnational levels from relatively small samples that would be problematic using more traditional survey-weighting approaches.<sup>11-16</sup> MRP consists of two steps. First, a multilevel regression model is used to generate the outcome of interest as a function of (socio)demographic and geographic variables. Next, the resulting outcome estimates for each demographic-geographic respondent type are poststratified by the percentage of each type in the actual overall population.<sup>11</sup>

We used a Bayesian multilevel generalised additive regression model to model the swab test result (positive/negative) as a function of age, sex, time and region. Because there were very few missing values ( $\leq 1\%$ ) in these factors, we restricted all analyses to observations with non-missing data. A complementary log-log link was used due to the ability to interpret regression coefficients as arising from an infection process with varying levels of exposure (see Supplementary File).<sup>17</sup> MRP models with random effects for individual participant and/or household nested within region did not converge. Therefore MRP models were run with only a random intercept for region, without a random intercept for participant and/or household. However, a model with only one participant sampled from each household gave similar results with somewhat wider 95% credible intervals mainly due to the smaller sample size (Figure S1). Time, measured in days since the start of the study (26 April 2020), was modelled using thin plate splines and allowed to vary by region. We set  $k$ , the number of basis functions, to control the smoothness of the fitted function.<sup>18</sup> We use a normal prior with location set to 4 for the standard deviation of the smooth. Very similar results were obtained when using different values for  $k$  (Figure S2A) or different priors for the standard deviation of the smooth (Figure S2B). Subsequently, we poststratified the resulting positivity estimates for each demographic-geographic respondent type by the percentage of each type in the overall population and in each region. This analysis was performed using the `rstanarm` package in R version 3.6.1.<sup>19</sup>

#### *Time-varying risk factors*

To assess whether particular subgroups were more likely to test positive for SARS-CoV-2 viral RNA during the first wave in England we performed a multilevel regression analysis (without poststratification) on the data between 26 April and 28 June 2020 including variables on which we did not post-stratify: work location, having a job that directly involved patients/care-home residents, ethnicity, household size, and number of children in the household. Given the short timescale included, and the fact that questions were not always asked at every visits, we carried non-missing data forward and backwards to adjacent visits with missing data. After this, there were very few

missing values ( $\leq 1\%$ ) so we again restricted all analyses to observations with non-missing data only. Results are shown in Table S1 and suggest that confounding plays a limited role.

During the summer positivity rate remained low and approximately constant, but more recently started to increase again. We evaluated using Bayesian dynamic multilevel regression models to what extent different factors were potentially driving this recent increase. Given that age appeared to be such a strong factor in driving the increase (see Results), all other factors were subsequently stratified by age (<35 and 35+ years). We evaluated the same factors as for the first wave (data up to 28 June) in generalised additive models with thin-plate splines that varied by each level of the factor of interest. These models additionally included a random intercept for region to account for any regional differences. As it was not possible to fit all factors with these time interactions in one model, and given the limited evidence of confounding (Table S1), we fitted separate models for each factor of interest.

#### *Presence of symptoms among those testing positive*

We evaluated the number of positive tests where the participant reported that they had symptoms around the visit (same visit or visit before or after) or not using the same MRP model as for the overall positivity rate.

To assess the impact of potential false positive tests we classified each positive into 3 categories:

- i) 'Higher evidence'; two or three genes detected (irrespective of cycle threshold (Ct) value).
- ii) 'Moderate evidence'; single gene detections if a) the Ct value was <97.5<sup>th</sup> percentile of 'higher evidence' positives (<34) or b) there was a higher pre-test probability of infection, i.e. any symptoms at or around the test (visit before or after) or reporting working in patient-facing healthcare role or resident-facing care home role.
- iii) 'Lower evidence'; all other positives, which by definition were all in asymptomatic individuals not having patient- or resident-facing roles with a single gene detected with Ct  $\geq 34$ .

## **Results**

Between 26 April and 19 September 2020, in total, results were available from 514,794 samples from 123,497 individuals, of which 489 were positive overall from 398 individuals in 342 households. The sample was broadly representative in terms of age, sex, and region (Figure S1). Small under/overrepresentation of certain groups, such as individuals from London being slightly underrepresented, was corrected for using dynamic MRP with poststratification performed on a daily basis. Positivity rates dropped to consistently low levels during the summer before increasing markedly again end of August and September (Figure 1).

Similar patterns were observed for the positivity rate when the participant reported symptoms or not (Figure 2A). In July there was a slight temporary increase in the positivity rate without symptoms, while the positivity rate with symptoms remained approximately constant. The modelled percentage of positives with reported symptoms around the test was lowest around mid-July with 30% and highest around the end of the study period at 46%. The rate of lower evidence positives remained

approximate stable over the study period, although again with a small increase during the summer, while the sharp increase around September 2020 was almost entirely due to an increase in higher evidence positives (Figure 2B), suggesting that people may have become infected with lower viral loads and fewer symptoms during the summer, but with higher viral loads in September potentially leading to a higher proportion of cases with symptoms.

Positivity rates showed marked regional differences, with increases in late August-September largely occurring in the North of England (Figure 1). We explored whether other factors besides region may be underlying the observed sharp increase in September 2020 using dynamic generalised additive mixed models without poststratification since population distributions were unknown for most of these factors. The most important factor was age, with earlier and greater increases apparent in younger adults (Figure 3) as also evident from results from a model categorising age (Figure S4). Importantly, there was clear diffusion of risk from initial increases in younger age groups at lower risk of hospitalisation and death, into older ages at higher risk.

While working outside your home and in patient-facing healthcare roles were a clear risk factor during the initial period of high positivity, as was contact with hospitals (26 April to 28 June, Table S1), there was no evidence that those working outside their home, working in patient-facing roles or with hospital contact were driving increases after the summer (Figure S5-7). Whilst non-White ethnicity was also associated with greater positivity rates during the initial period (Figure S8), increases after the summer were greatest in White individuals under 35 years (Figure 4). In contrast, there were different trends in the most recent period dependent on household composition, where increases were more apparent in households with children of both primary and secondary school age and households consisting of one person age between 17 and 34 than in other categories (Figure 5).

## Discussion

Here we demonstrate substantial changes over time in the percentage of people in private-residential households in the community in England testing positive for SARS-CoV-2, with an initial decrease between end of April and June 2020, followed by consistently low levels during the summer, before marked increases end of August and September 2020. Our estimates have been regularly updated and shared with the UK Government and Scientific Advisory Group for Emergencies (SAGE) sub-group Scientific Pandemic Influenza sub-group on Modelling (SPI-M) to directly inform decisions about potential changes to the current alert level or relaxation of certain restrictions.

Notably, we found that a substantial proportion (53%-70%, dependent on calendar time) of individuals that tested positive did not report any symptoms on the day of the visit or at visits before or after the swab was taken.

The Bayesian dynamic multilevel generalised additive models are useful tools for monitoring the effect of different factors on positivity rates over time. In particular, they show that the epidemic restarted in young people, and that factors associated with an increased risk of testing positive during the initial high-prevalence period in April-May 2020, such as working outside the home and having a job with direct patient contact were not important drivers of the recent increase after the summer.

While false-positives may be a concern when prevalence is low, the low prevalence at the end of June (0.05%) is also reassuring, since it indicates that the specificity of the test used in the national UK programme is very high. A test specificity lower than 99.95% would lead to observed positivity rates above 0.05%, even in the purely hypothetical situation that the virus was not circulating in June. In theory, the percentage of tests that are false-positives might be expected to be relatively stable over time. Interestingly, the number of lower-evidence positives, i.e. with a low pre-test probability and a single gene detected with Ct value  $\geq 34$ , and hence more likely to be false positives, increased slightly during the summer. This may be partly a consequence of a genuinely increase in the number of infections with low levels of virus during this period.

## Comparison with other studies

Our finding of a temporal reduction in the percentage of people in the community infected with SARS-CoV-2 in England between April and June 2020 is consistent with reductions observed in the weekly number of excess deaths and COVID-19 hospitalisations.<sup>20,21</sup> An important advantage of our current population-based study is that it can detect increases in the positivity rate potentially earlier and more systematically than surveillance based on confirmed cases, hospital admissions or deaths (<https://coronavirus.data.gov.uk/>). This is likely especially the case when new increases initially occur in a subgroup of the population that is at low risk of hospitalisation and death, but does contribute to transmission including if asymptomatic,<sup>22</sup> as observed after the summer with the increase in positives among young adults. Furthermore, interpretation of changes in incidence and positivity rate from tests that are taken for contact tracing or clinical cases is likely confounded by substantial changes in testing practice over time. Our study is based on a representative sample of the population, with further correction for residual non-representativeness using MRP, thereby preventing difficulties with interpretation due to changes in testing practice.

There are a few other studies that aimed to assess the prevalence of SARS-CoV-2 infection in the general population. A repeated cross-sectional population-based study from England also found a similar decline in the prevalence among the general population between 1 May and 1 June.<sup>23</sup>



Another cross-section from that study showed also an increase in in the prevalence in September.<sup>24</sup> Among individuals that tested positive in that study, the percentage reporting no symptoms varied between 50% and 81% between the different cross-sections.<sup>24</sup> A study from Vo, an Italian town with a population of 3275 individuals, Lavezzo et al. surveyed 85.9% and 71.5% of the inhabitants at the start and end of lockdown of the town respectively, with an initial infection prevalence of 2.6% (95% confidence interval (CI) 2.1-3.3%) and 1.2% (95% CI 0.8-1.8%) 14 days later.<sup>25</sup> The percentage of those who tested positive that did not report any symptoms was 41.0% (95% CI 29.7-53.2%) in the first survey and 44.8% (95% CI 26.5-64.3%) in the second survey.

As part of a larger study from Iceland, a randomly selected sample of inhabitants aged between 20 and 70 years of age were invited to participate in a survey.<sup>26</sup> By 4 April, 2283 (33.7%) of those invited had participated, with 13 testing positive (0.6%; 95% CI 0.3%-1.0%). However, it is unclear to what extent the participation rate led to a sample that was not representative of the Icelandic population and no correction for population representativeness was performed. Among a larger sample, including participants recruited via an open invitation, which may bias the sample towards people with symptoms, 57% of individuals testing positive reported having symptoms, although 29% of individuals testing negative also reported having symptoms.

While the studies from Iceland and Vo found that around 40-45% of those with a positive test did not report any symptoms,<sup>25,26</sup> this percentage is higher in our study. This may be partly due to differences in respondents, chance (given considerable uncertainty in all studies), differences in definitions of symptoms, over/under-reporting of symptoms or false-positive tests.

Furthermore, the similar trends over time for positives with and without symptoms suggests that there truly may be more asymptomatic cases than reported in some other studies.<sup>27</sup> A recent meta-analysis of studies focusing on close contacts of confirmed COVID-19 cases suggested that only 17% (95% CI 14%-20%) of infected individuals are asymptomatic.<sup>27</sup> However, by informing participants that they were recently in close contact with a confirmed COVID-19 case, individuals may be more likely to think that they have experienced symptoms, resulting in recall bias and overestimating the true prevalence of symptoms among a representative sample of infected persons. Although we may have underestimated the true prevalence of symptoms among SARS-CoV-2 cases in the community, partly due to asking about current symptoms at visits through 23 July (meaning that very transient symptoms only occurring between visits would have been missed) and symptoms in the last 7 days thereafter, our study adds to the growing evidence that a substantial proportion of SARS-CoV-2 in the community may be asymptomatic.<sup>28,29</sup>

### **Limitations of this study**

An important limitation of this study is that the number of people in the community that test positive is low, limiting power and leading to relatively large uncertainty around estimates, and meaning that our multilevel regression model was not able to incorporate likely correlation within households. However, sensitivity analyses suggested that within-household clustering did not have a large impact on our results.

Furthermore, while we adjusted for potential non-representativeness in terms of age, sex and region, there may be other factors for which we do not have detailed information about population distributions that also are associated with testing positive. If these are over- or under-represented in

the survey this could have resulted in some residual bias. We did forwards and backwards imputation for missing data, reflecting the relatively short timescales of the study.

Another limitation is that, in the absence of a true gold standard, we do not know the test sensitivity and specificity, making it difficult to assess what the true prevalence is. However, as detailed above the true specificity is likely very close to 100%. The data cannot inform about the test sensitivity without providing a very informative prior on the true prevalence.<sup>30</sup> However, this should not affect trends in prevalence over time.

## Conclusions

The percentage of individuals from the community in England testing positive for SARS-CoV-2 clearly declined between 26 April and 28 June 2020, remained approximately stable for much of the summer before increasing again from the end of August through September. Important risk factors for testing positive varied substantially between the initial and second periods of higher positivity rates, and a substantial proportion of infections were in individuals not reporting symptoms, indicating that continued monitoring for SARS-CoV-2 in the community will be important for managing the epidemic moving forwards. Bayesian dynamic Multilevel Regression and Poststratification (MRP) is a powerful tool which could be used more widely to ensure population-representativeness of surveillance estimates.

**Contributors:** The study was designed and planned by SW, JF, JB, JN, IB, ID, PB, KBP, and JVR. KBP, TH, EP, KDV, NB, IB, HT, JL and ASW contributed to the statistical analysis. KBP drafted the manuscript and all authors contributed to interpretation of the data and results and revised the manuscript. KBP and ASW are the guarantors of the study. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work. The corresponding authors attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

**Declaration of interest:** We declare no competing interests.

**Data sharing:** Data sharing: De-identified study data are available for access by accredited researchers in the ONS Secure Research Service (SRS) for [accredited research purposes](#) under part 5, chapter 5 of the Digital Economy Act 2017. For further information about accreditation, contact [Research.Support@ons.gov.uk](mailto:Research.Support@ons.gov.uk) or visit the [SRS website](#)

## Acknowledgements

**COVID-19 Infection Survey Team:**

**COVID-19 Infection Survey Team:**

**Office for National Statistics:** Iain Bell, Ian Diamond, Alex Lambert, Pete Benton, Emma Rourke, Stacey Hawkes, Sarah Henry, James Scruton, Peter Stokes, Tina Thomas.

**Office for National Statistics, Analysis:** John Allen, Russell Black, Heather Bovill, David Braunholtz, Dominic Brown, Sarah Collyer, Megan Crees, Colin Daglish, Byron Davies, Hannah Donnarumma, Julia Douglas-Mann, Antonio Felton, Hannah Finselbach, Eleanor Fordham, Alberta Ipser, Joe Jenkins, Joel Jones, Katherine Kent, Geeta Kerai, Lina Lloyd, Victoria Masding, Ellie Osborn, Alpi Patel, Elizabeth Pereira, Tristan Pett, Melissa Randall, Donna Reeve, Palvi Shah, Ruth Snook, Ruth Studley, Esther Sutherland, Eliza Swinn, Heledd Thomas, Anna Tudor, Joshua Weston.

**Office for National Statistics, Secure Research Service:** Shayla Leib, James Tierney, Gabor Farkas, Raf Cobb, Folkert van Galen, Lewis Compton, James Irving, John Clarke, Rachel Mullis, Lorraine Ireland, Diana Airimitoiaie, Charlotte Nash, Danielle Cox, Sarah Fisher, Zoe Moore, James McLean, Matt Kerby.

**University of Oxford, Nuffield Department of Medicine:** Ann Sarah Walker, Derrick Crook, Philippa C Matthews, Tim Peto, Emma Pritchard, Nicole Stoesser, Karina-Doris Vihta, Alison Howarth, George Doherty, James Kavanagh, Kevin K Chau, Stephanie B Hatch, Daniel Ebner, Lucas Martins Ferreira, Thomas Christott, Brian D Marsden, Wanwisa Dejnirattisai, Juthathip Mongkolsapaya, Sarah Hoosdally, Richard Cornall, David I Stuart, Gavin Screaton.

**University of Oxford, Nuffield Department of Population Health:** Koen Pouwels.

**University of Oxford, Big Data Institute:** David W Eyre.

**University of Oxford, Radcliffe Department of Medicine:** John Bell.

**Oxford University Hospitals NHS Foundation Trust:** Stuart Cox, Kevin Paddon, Tim James.

**University of Manchester:** Thomas House.

**Public Health England:** John Newton, Julie Robotham, Paul Birrell.

**IQVIA:** Helena Jordan, Tim Sheppard, Graham Athey, Dan Moody, Leigh Curry, Pamela Brereton

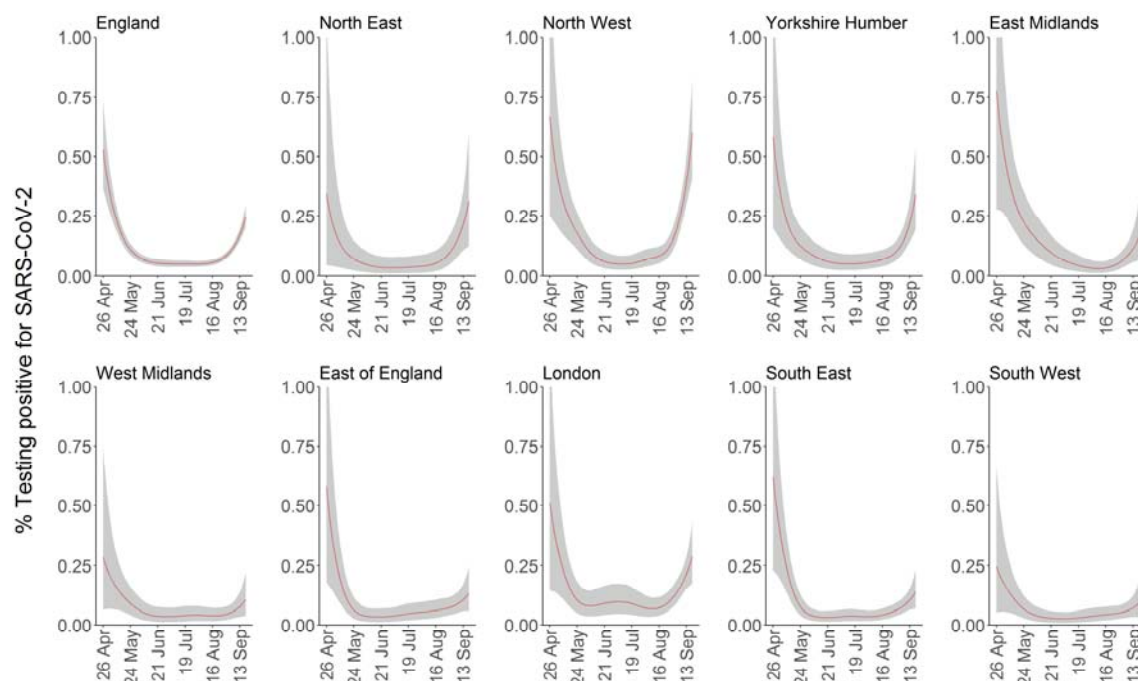
**National Biocentre:** Ian Jarvis, Kirsty Howell, Bobby Mallick, Phil Eeles.

**Glasgow Lighthouse Laboratory:** Jodie Hay, Harper Vansteenhouse.

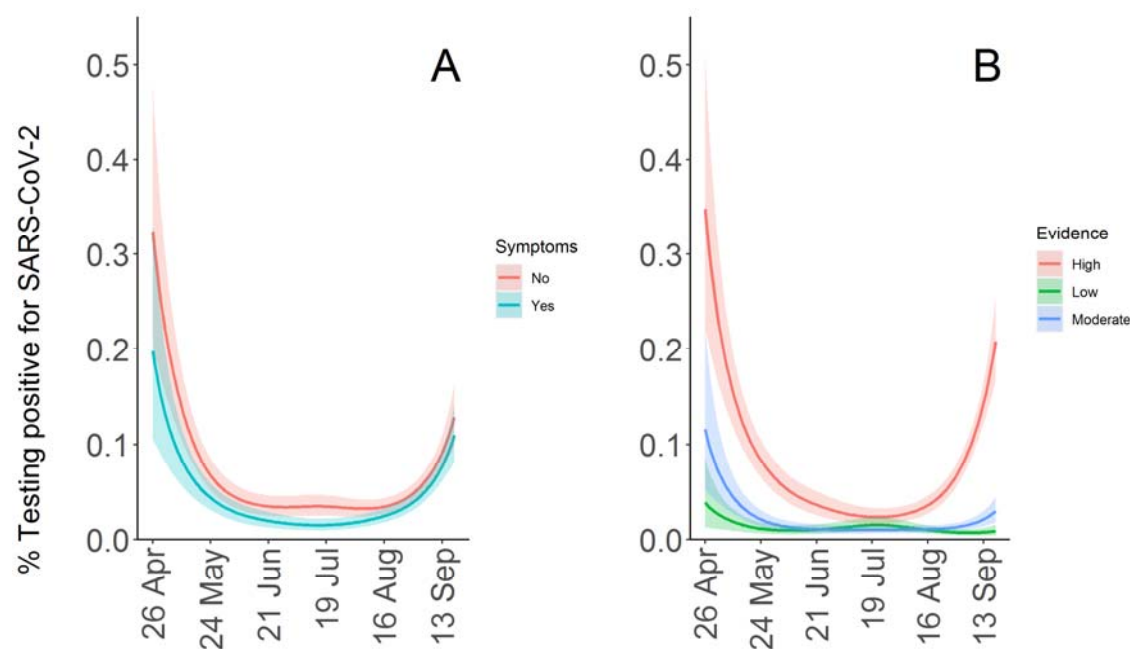
## References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497-506.
2. WHO. Coronavirus disease (COVID-2019) situation report. Weekly Epidemiological Update. 21 September 2020. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200921-weekly-epi-update-6.pdf?sfvrsn=d9cf9496\\_6](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200921-weekly-epi-update-6.pdf?sfvrsn=d9cf9496_6).
3. Salje H, Kiem CT, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. *Science* 2020;eabc3517
4. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med* 2020;18(1):124.
5. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health* 2020. doi: [https://doi.org/10.1016/S2468-2667\(20\)30133-X](https://doi.org/10.1016/S2468-2667(20)30133-X)
6. Guan D, Wang D, Hallegatte S, Davis SJ, Huo J, Li S, et al. Global supply-chain effects of COVID-19 control measures. *Nat Hum Behav* 2020; doi: 10.1038/s41562-020-0896-8.
7. U.S. Department of Labour. Unemployment Insurance Weekly Claims Report. 4 June 2020. <https://www.dol.gov/sites/dolgov/files/OPA/newsreleases/ui-claims/20201165.pdf>
8. HM Government. COVID Alert Levels. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/884352/slides\\_-\\_11\\_05\\_2020.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/884352/slides_-_11_05_2020.pdf)
9. Kojima N, Turner F, Slepnev A, Demig L, Kodeboyina S, Klausner JD. Self-collected oral fluid and nasal swabs demonstrate comparable sensitivity to clinician collected nasopharyngeal swabs for COVID-19 detection. *MedRxiv* 2020; doi: <https://doi.org/10.1101/2020.04.11.20062372>
10. Office of National Statistics. COVID-19 Infection Survey Pilot. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation>
11. Gelman A, Little TC. "Poststratification into Many Categories Using Hierarchical Logistic Regression. *Survey Methodology* 1997;23(2):127-35.
12. Gelman A, Lax J, Phillips J, Gabry J, Trangucci R. Using multilevel regression and poststratification to estimate dynamic public opinion. 2018. [http://www.stat.columbia.edu/~gelman/research/unpublished/MRT\(1\).pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/MRT(1).pdf)
13. Downes M, Carlin JB. Multilevel regression and poststratification as a modeling approach for estimating population quantities in large population health studies: a simulation study. *Biom J* 2020; 62(2):479-91.
14. Warshaw C, Rodden J. How should we measure district-level public opinion on individual issues? *J Politics* 2012; 74(1): 203-19.
15. Kennedy L, Gelman A. Know your population and know your model: using model-based regression and post-stratification to generalize findings beyond the observed sample. *ArXiv* 2020; 1906.11323v2.
16. Si Y, Trangucci R, Gabry JS, Gelman A. Bayesian hierarchical weighting adjustment and survey inference. *ArXiv* 2020; 1707.08220.
17. McCullagh, P. Regression models for ordinal data. *Journal of the Royal Statistical Society B* 1980;42(2):109-42.
18. Wood, SN. Thin plate regression splines. *J R Stat Soc Series B Stat Methodol* 2003;65: 95-114.

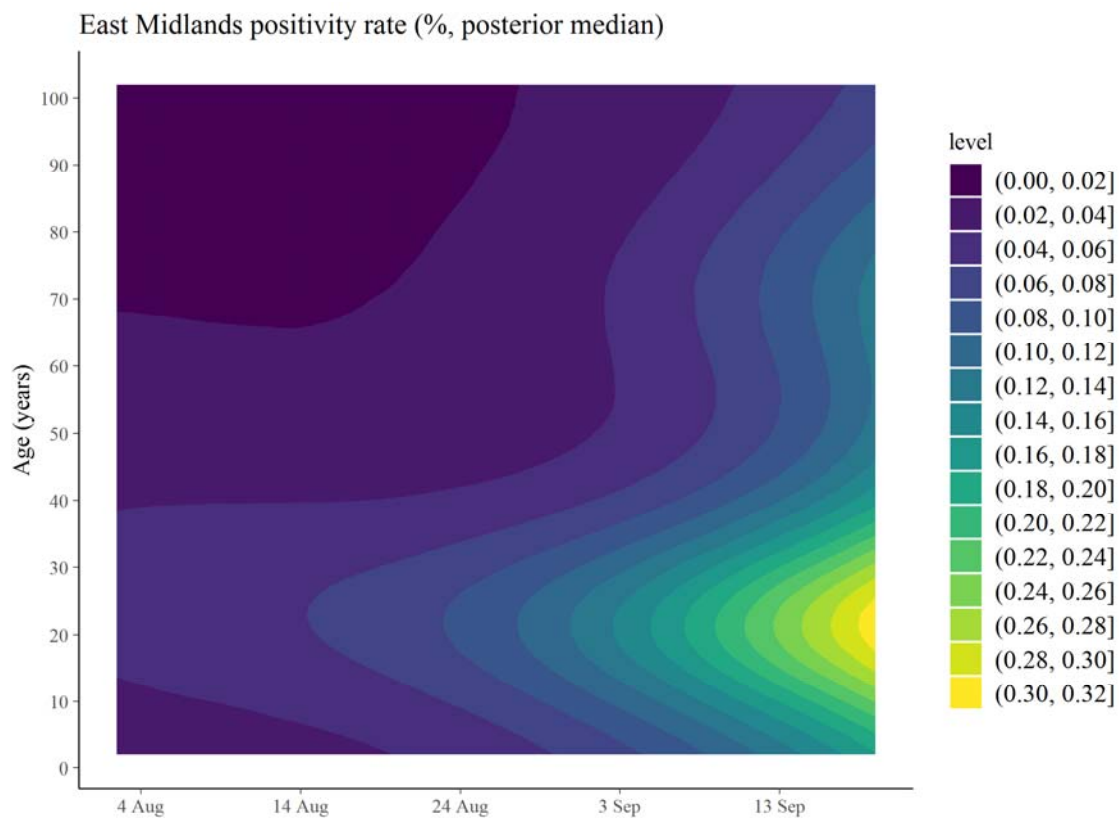
19. Goodrich B, Gabry J, Ali I, Brilleman S. "rstanarm: Bayesian applied regression modeling via Stan." 2020; R package version 2.19.3, <https://mc-stan.org/rstanarm>.
20. Office of National Statistics. Deaths registered weekly in England and Wales provisional. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending22may2020>
21. Public Health England. Weekly coronavirus disease 2019 (COVID-19) surveillance report summary of COVID-19 surveillance system. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/888254/COVID19\\_Epidemiological\\_Summary\\_w22\\_Final.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/888254/COVID19_Epidemiological_Summary_w22_Final.pdf)
22. Chau NVV, Thanh Lam V, Thanh Dung N, Yen LM, Minh MNQ, Hung LM, et al. The natural history and transmission potential of asymptomatic SARS-CoV-2 infection. *Clin Infect Dis*. 2020; doi:10.1093/cid/ciaa711.
23. Riley S, Ainslie KEC, Eales O, Jeffrey B, Walters CE, Atchinson CJ, et al. Community prevalence of SARS-CoV-2 virus in England during May 2020: REACT study. *MedRxiv* 2020; <https://doi.org/10.1101/2020.07.10.20150524>.
24. Riley S, Ainslie KEC, Eales O, Walters CE, Wang H, Atchison C, et al. High prevalence of SARS-CoV-2 swab positivity in England during September 2020: interim report of round 5 of REACT-1 study. *MedRxiv* 2020; <https://doi.org/10.1101/2020.09.30.20204727>.
25. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo. *Nature* 2020;584:425-9.
26. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med* 2020;382:2302-15
27. Byambasuren O, Cardona M, Bell K, Clark J, McLaws M, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *MedRxiv* 2020; doi: <https://doi.org/10.1101/2020.05.10.20097543>
28. Poletti P, Tirani M, Cereda D, Trentine F, Guzzetta G, Sabatino G, et al (2020). Probability of symptoms and critical diseases after SARS-CoV-2 infection. *arXiv* 2020; [2006.08471](https://arxiv.org/abs/2006.08471)
29. [Pollan M, Perez-Gomez B, Pastor-Barriuso R, Oteo J, Hernan MA, Perez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain \(ENE-COVID\): a nationwide, population-based seroepidemiological study. \*Lancet\* 2020; doi: \[https://doi.org/10.1016/S0140-6736\\(20\\)31483-5\]\(https://doi.org/10.1016/S0140-6736\(20\)31483-5\).](https://doi.org/10.1016/S0140-6736(20)31483-5)
30. Lewis IF, Torgerson PR (2012). A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. *Emerg Themes Epidemiol* 2012;9(1):9.



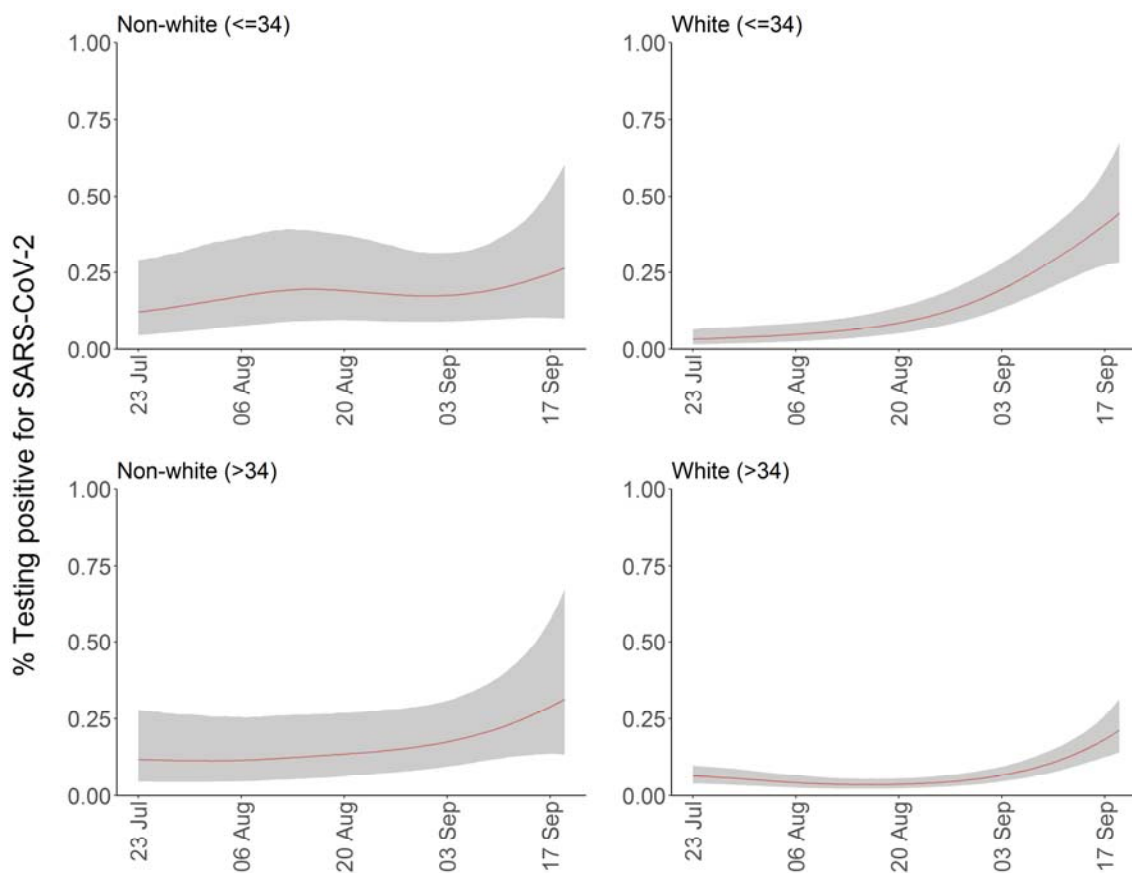
**Figure 1.** Percentage of population living in private households testing positive for SARS-CoV-2 over time in England and the 9 regions of England. Shaded areas are 95% credible intervals.



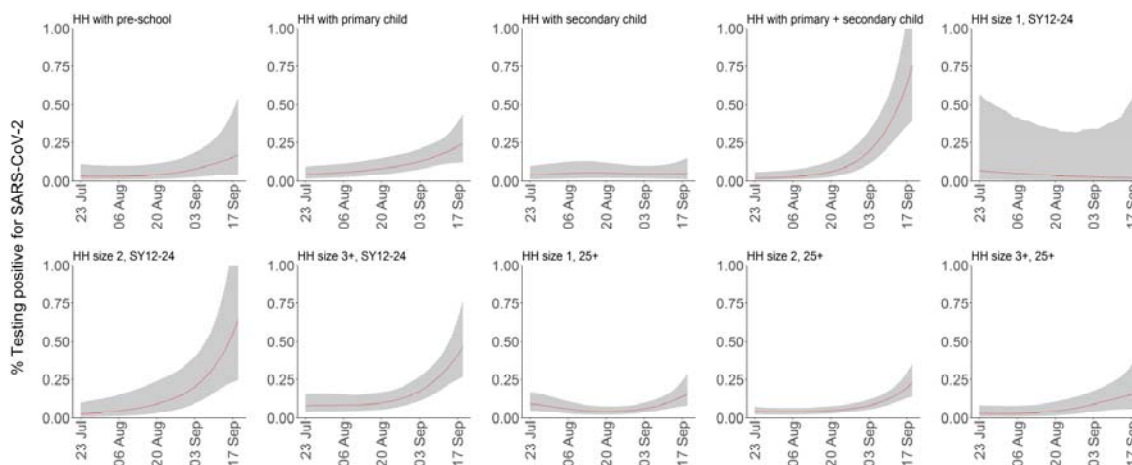
**Figure 2. A:** Percentage of population living in private households testing positive for SARS-CoV-2 with and without reporting symptoms; **B:** Percentage of population living in private households testing positive for SARS-CoV-2 stratified by high, moderate and low evidence positivity. Shaded areas are 95% credible intervals.



**Figure 3.** Modelled estimates (posterior medians) of the distribution of positive SARS-CoV-2 tests across age over time.



**Figure 4.** Percentage of population living in private households testing positive for SARS-CoV-2 stratified by ethnicity and age (<=34 and >34 years of age). Shaded areas are 95% credible intervals.



**Figure 5.** Percentage of population living in private households testing positive for SARS-CoV-2 stratified by their household composition. HH=household; SY=school year; SY12-24 indicates individuals from school-year 12 (aged 16/17) to age of 24 years old. Shaded areas are 95% credible intervals.