Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus Infection Survey

Koen B Pouwels, Thomas House*, Emma Pritchard, Julie V Robotham, Paul J Birrell, Andrew Gelman, Karina-Doris Vihta, Nikola Bowers, Ian Boreham, Heledd Thomas, James Lewis, Iain Bell, John I Bell, John N Newton, Jeremy Farrar, Ian Diamond, Pete Benton, Ann Sarah Walker, and the COVID-19 Infection Survey Team†*

Summary

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

Methods Samples were collected from individuals aged 2 years and older living in private households in England that were randomly selected from address lists and previous Office for National Statistics surveys in repeated cross-sectional household surveys with additional serial sampling and longitudinal follow-up. Participants completed a questionnaire and did nose and throat self-swabs. The percentage of individuals testing positive for SARS-CoV-2 RNA was estimated over time by use of dynamic multilevel regression and poststratification, to account for potential residual non-representativeness. Potential changes in risk factors for testing positive over time were also assessed. The study is registered with the ISRCTN Registry, ISRCTN21086382.

Findings Between April 26 and Nov 1, 2020, results were available from 1 191 170 samples from 280 327 individuals; 5231 samples were positive overall, from 3923 individuals. The percentage of people testing positive for SARS-CoV-2 changed substantially over time, with an initial decrease between April 26 and June 28, 2020, from 0·40% (95% credible interval 0·29–0·54) to 0·06% (0·04–0·07), followed by low levels during July and August, 2020, before substantial increases at the end of August, 2020, with percentages testing positive above 1% from the end of October, 2020. Having a patient-facing role and working outside your home were important risk factors for testing positive for SARS-CoV-2 at the end of the first wave (April 26 to June 28, 2020), but not in the second wave (from the end of August to Nov 1, 2020). Age (young adults, particularly those aged 17–24 years) was an important initial driver of increased positivity rates in the second wave. For example, the estimated percentage of individuals testing positive was more than six times higher in those aged 17–24 years than in those aged 70 years or older at the end of September, 2020. A substantial proportion of infections were in individuals not reporting symptoms around their positive test (45–68%, dependent on calendar time.

Interpretation Important risk factors for testing positive for SARS-CoV-2 varied substantially between the part of the first wave that was captured by the study (April to June, 2020) and the first part of the second wave of increased positivity rates (end of August to Nov 1, 2020), 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 COVID-19 pandemic moving forwards.

Funding Department of Health and Social Care.

Introduction

Since severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) started causing severe respiratory illness (COVID-19) in Wuhan, China, in late 2019,[1] as of Nov 1, 2020, there have been nearly 46 million confirmed cases and 1·2 million deaths reported to WHO.[2] Control measures (eg, national lockdowns) have been widely implemented to contain the spread of SARS-CoV-2 in an (at least temporarily successful)[3–5] attempt to prevent the collapse of health-care systems and even more deaths. Although such measures are important for control of the COVID-19 pandemic, they also affect the economy, unemployment rates, and global supply chains.[6,7] Politicians continuously make difficult decisions between continuing strict control measures or relaxing them in ways that would be safe enough from a public health perspective yet beneficial more broadly across society.[8] Importantly, early detection of population subgroups driving new increases in infections is crucial to potentially tailor interventions or messaging without having to implement drastic measures affecting the whole society.

Lancet Public Health 2021; 6: e30–38

Published **Online** December 10, 2020 https://doi.org/10.1016/S2468-2667(20)30282-6

*Contributed equally

†Members listed at the end of the report

Health Economics Research Centre, Nuffield Department of Population Health (K B Pouwels PhD), **The National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford** (K B Pouwels, E Pritchard MSc, K-D Vihta PhD, Prof A S Walker PhD), **Nuffield Department of Medicine** (E Pritchard, K-D Vihta, Prof A S Walker), **Office of the Regius Professor of Medicine** (Prof J I Bell MD), **and The National Institute for Health Research Oxford Biomedical Research Centre** (Prof A S Walker), **University of Oxford, Oxford, UK; Department of Mathematics, University of Manchester, Manchester, UK** (T House PhD); **IBM Research, Hartree Centre, Sci-Tech, Daresbury, UK** (T House); **National Infection Service** (J V Robotham PhD, P J Birrell PhD) **and Health Improvement Directorate** (Prof J N Newton FRCP), **Public Health England, London, UK; Medical Research Council (MRC) Biostatistics Unit, University of Cambridge, Cambridge Institute of Public Health, Cambridge, UK** (P J Birrell); **Department of Statistics, Columbia University, New York, NY, USA** (Prof A Gelman PhD); **Office for National Statistics, Newport, UK** (N Bowers MSc, I Boreham MSc, H Thomas MSc, J Lewis BA, I Bell BSc, Prof I Diamond PhD,

P Benton MSc); **Wellcome Trust, London, UK** (J Farrar PhD); **MRC Clinical Trials Unit at University College London, London, UK** (Prof A S Walker)

Correspondence to: Dr Koen B Pouwels, Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
koen.pouwels@ndph.ox.ac.uk

Research in context

Evidence before this study

Unprecedented control measures, such as national lockdowns, have been widely implemented to contain the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Decisions about the continued need for physical distancing measures in the overall population, specific subgroups, and geographical areas heavily rely on accurate and up-to-date information about the number of people infected and risk factors for testing positive for SARS-CoV-2. We searched PubMed and the *medRxiv* and *bioRxiv* preprint servers up to Nov 15, 2020, for epidemiological studies using the terms "SARS-CoV-2" AND "prevalence" OR "incidence", without data or language restrictions. Most studies were small, had information about current presence of SARS-CoV-2 in only a few patients, or used data not representative of the community (eg, hospital admissions, deaths, or self-reported symptoms). Large population-based studies are needed to understand risk factors and dynamics of the COVID-19 pandemic.

Added value of this study

Our study is one of the largest longitudinal community surveys of SARS-CoV-2 infection at national and regional levels. With more than 1 000 000 swabs from almost 300 000 individuals, this ongoing study provides robust evidence that the percentage of individuals from the general community in England testing positive for SARS-CoV-2 clearly declined between the end of April and June, 2020, followed by consistently low levels during July and August, 2020, before substantial increases at the end of August, 2020. Risk factors for testing positive for SARS-CoV-2 varied substantially between the first and second waves (April 26 to June 28, 2020, *vs* end of August to Nov 1, 2020) of higher positivity rates. Having a patient-facing role and working outside your home were important risk factors in the first wave of high positivity but not as of Nov 1, 2020, in the second wave, and age (young adults, particularly those aged 17–24 years) was an important driver of the second wave of increased positivity rates. Positive tests were reported without symptoms being reported in roughly half to two-thirds of cases.

Implications of all the available evidence

Our longitudinal survey showed that community supervised self-swabbing RT-PCR-based surveillance is achievable and practical. The survey could serve as a model for other countries and potential future pandemics. The recorded decline in the percentage of individuals testing positive for SARS-CoV-2 adds to the increasing body of empirical evidence and theoretical models suggesting that the national lockdown in England, which was imposed on March 23, 2020, was associated—at least temporarily—with a decrease in infections. Important risk factors for testing positive for SARS-CoV-2 varied substantially between the first and second waves 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 COVID-19 pandemic moving forwards. Using multilevel regression and poststratification to account for potential residual non-representativeness of the sample, this survey provided early warnings that specific regions (eg, the North West region of England) were probably going to experience increases in hospital admissions and deaths.

There are several reasons why risk factors might vary over time. First, behaviour and contact patterns of subgroups change over time without intervention (eg, students starting university). Adherence to non-mandatory infection prevention measures can reduce more over time among subgroups with a low risk of COVID-19-related hospital admission and death than among those who are more vulnerable. Moreover, subgroups of people who have been disproportionally affected in a first wave of SARS-CoV-2 infection might have acquired sufficient immunity and 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.

See Online for appendix

Here, we use data from the Office for National Statistics (ONS) Coronavirus Infection Survey. This ongoing, large national survey is designed to be representative of the target population, offering a unique opportunity to identify risk factors that are driving new increases in the SARS-CoV-2 positivity rate and investigating the proportion of individuals testing positive for SARS-CoV-2 who do not report symptoms, the potential false-positive rate, and other factors that can directly inform policy around COVID-19-related control measures. We used Bayesian dynamic multilevel regression and poststratification to account for any residual unrepresentativeness, a potential problem often ignored with surveillance data.

Methods

Participants

Data were collected between April 26 and Nov 1, 2020, from individuals in private households randomly selected from address lists and previous ONS surveys to provide a representative sample of the population of England (details on sampling design in appendix p 1). Individuals aged 2 years and older who were living in private households were eligible.

The survey has been reviewed and given ethics approval by South Central–Berkshire B Research Ethics Committee (20/SC/0195).

Procedures

If one or more individuals from a household agreed to participate, a study worker visited the household and

directly collected information from individuals about any symptoms (current until July 23, 2020, then in the past 7 days before the visit) and contacts, together with demographic information. The study worker provided instructions on how to self-swab the nose and throat and monitored the self-swabbing, which is comparable to or more sensitive than swabs done by health-care workers.[9] Parents or carers took swabs from children younger than 12 years.

Nose and throat self-swabs were couriered directly to the UK's national Lighthouse laboratories (National Biocentre in Milton Keynes from April 26, 2020; and Glasgow from Aug 16, 2020) where samples were tested as part of the national testing programme. Identical methodology was used to test for the presence of SARS-CoV-2 genes for nucleocapsid protein (N), spike protein (S), and ORF1ab using RT-PCR.[10] We used the TaqPath RT-PCR COVID-19 kit (Thermo Fisher Scientific, Waltham, MA, USA), which was analysed using UgenTec Fast Finder 3.300.5 (TagMan 2019-nCoV assay kit V2 UK NHS ABI 7500 v2.1; UgenTec, Hasselt, Belgium). The assay plugin contains an assay-specific algorithm and decision mechanism that allows conversion of the qualitative amplification assay PCR raw data from the ABI 7500 Fast into test results with little manual intervention. Samples are called positive in the presence of at least one gene (N, ORF1ab, or both) but could be accompanied by the gene for S protein (ie, one, two, or three gene positives). The gene for S protein is not considered a reliable single gene positive (as of mid-May, 2020; personal communication, National Biocentre, Milton Keynes, UK).

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 then optional monthly visits thereafter. The study protocol and questionnaires are available online.

Statistical analysis

The survey was designed to test 150 000 people every 2 weeks across England in October, 2020, to provide 15 000–20 000 individuals in each of the nine governmental office regions, giving approximately a 0·1%, 0·2%, and 0·5% margin of error on 0·1%, 0·5%, and 2% prevalence, respectively.

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.[11,12] This method was used to correct for any residual non-representativeness in terms of age, sex, and region. In several empirical and simulation studies, multilevel regression and poststratification was superior at both the national and regional levels compared with classic survey weighted and unweighted approaches, including when using small sample sizes.[11–16] 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.[11–16] Multilevel regression and poststratification consists of two steps. First, a multilevel regression model is used to predict the outcome of interest as a function of (socio) demographic and geographical variables. Second, the resulting outcome estimates for each demographic–geographical respondent type are poststratified by the percentage of each type in the actual overall population.[11]

We used a Bayesian multilevel generalised additive regression model to model the swab test result (positive or negative) as a function of age, sex, time, and region. We did not poststratify for other factors (eg, ethnicity) because reliable estimates in the target population were not available and a model for the full period did not converge with ethnicity in the model (divergent transitions). A model including ethnicity for the most recent 7 weeks did converge and showed similar estimates and trend as did the main model (appendix p 9). Besides the nine regions in England, we also took into account that specific local authorities within regions (boosted areas) were purposefully oversampled at the end of July, 2020 (appendix p 1). Because there were very few missing values (≤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 (appendix p 2).[17] Multilevel regression and poststratification models with random effects for individual participant, household, or both nested within a region did not converge. Therefore, these models were run with only a random intercept for region (including separate levels for boosted areas within a region—eg, Yorkshire & The Humber non-boosted and Yorkshire & The Humber boosted), without a random intercept for participant, household, or both. 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 (appendix p 10). Time, measured in days since the start of the study (April 26, 2020), was modelled with thin-plate splines and allowed to vary by region. We set k, the number of basis functions, to 10 to control the smoothness of the fitted function.[18] We used a normal prior with location set to 4 for the SD of the smooth. Very similar results were obtained when using different values for k or different priors for the SD of the smooth (appendix p 11). Subsequently, we poststratified the resulting positivity estimates for each demographic–geographical respondent type by the percentage of each type in the overall population and in each region.

Because the effect of potential risk factors might change over time, and it was not feasible in terms of run time and available central processing unit to fit a model

For the **study protocol and questionnaires** see https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey

with a much more flexible thin-plate spline for the entire period (April 26 to Nov 1, 2020), we also ran a multilevel regression and poststratification model using data for the most recent 7 weeks. This analysis was done with the *rstanarm* package in R, version 3.6.1.

For more on the ***rstanarm* package** see https://mc-stan.org/rstanarm

Time-varying risk factors

To assess whether particular subgroups were more likely to test positive for SARS-CoV-2 during the first wave of increased positivity in England we did a multilevel regression analysis (without poststratification) on the data between April 26 and June 28, 2020, including variables on which we did not poststratify—ie, work location, having a job that directly involved patients or care-home residents, ethnicity, household size, and number of children in the household (appendix p 8). In view of the short timescale included and the fact that questions were not always asked at every visit, we carried non-missing data forwards and backwards to adjacent visits with missing data. After this process, there were very few missing values (≤1%) so we again restricted all analyses to observations with non-missing data only.

We evaluated to what extent different factors were potentially driving recent increases in the SARS-CoV-2 positivity rate. Since age seemed to be a strong factor in driving the increase, all other factors were subsequently stratified by age (<35 years and ≥35 years). We evaluated the same factors as for the first wave (to June 28, 2020) 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. Because it was not possible to fit all factors with these time interactions in one model, and in view of the limited evidence for confounding (appendix p 8), we fitted separate models for each factor of interest.

Presence of symptoms among those testing positive

To assess the number of positive tests for which participants reported either symptoms around the time of the visit (same visit, or visit before or after) or no symptoms around the time of the visit, we used the same multilevel regression and poststratification model as for the overall positivity rate. To assess the effect of potential false-positive tests, we classified each positive test into three categories:[19] higher evidence, with two or three genes detected (irrespective of cycle threshold [Ct] value); moderate evidence, with single gene detection if the Ct value was less than the 97·5th percentile of higher evidence positives (<34) or if there was a higher pretest probability of infection (ie, any symptoms at or around the test [visit before or after] or reporting working in patient-facing health-care role or resident-facing care home role); and lower evidence, comprising all other positive tests, which by definition were all in asymptomatic individuals not having patient-facing or resident-facing role with a single gene detected with a Ct value of 34 or higher.

Role of the funding source

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had access to all data reported in the study and accept responsibility for the decision to submit for publication.

Results

Between April 26 and Nov 1, 2020, results were available from 1 191 170 nose and throat swabs from 280 327 individuals. 5231 samples were positive overall, from 3923 individuals in 3056 households. The study is ongoing and many participants were only recruited in October, 2020, to achieve the target sample size; nevertheless, the median number of visits per individual was four (IQR three to five, maximum 13). Of participants enrolled sufficiently early to have multiple study visits before Nov 1, 2020, most had at least five study visits (appendix p 5).

Characteristics of participating individuals are shown in the appendix (pp 6–7). Representativeness of the sample was visualised by plotting proportions of the sample within each region and age and sex category and comparing with known distributions for individuals living in private households in England (appendix p 12). Small under-representations and over-representations of some groups (eg, individuals aged 2–11 years were slightly under-represented) were corrected for using dynamic multilevel regression and poststratification, with poststratification done every day. SARS-CoV-2 positivity rates dropped to consistently low levels during July and August, 2020, before increasing substantially at the end of August, 2020 (figure 1). When restricting the analysis to the most recent 7 weeks to make the flexible spline more responsive to recent changes, the increase in SARS-CoV-2 positivity rates seemed to start levelling off at the end of October, 2020, before the second national lockdown was implemented on Nov 5, 2020.

Observed patterns in SARS-CoV-2 positivity rates were similar between participants reporting symptoms and those not reporting symptoms, although in October, 2020, the positivity rate among those reporting symptoms started to increase less steeply (figure 2A). The modelled percentage of positive cases with reported symptoms around the test was lowest around mid-July, 2020 (32%), and highest around the beginning of October, 2020 (55%). The increase in SARS-CoV-2 positivity rates starting at the end of August, 2020, was almost entirely due to high-evidence positive tests, although the levels of moderate-evidence and (to a lesser extent) low-evidence positive tests started to increase slightly in September, 2020, as well (figure 2B). People might have become infected with lower viral loads and fewer symptoms during July and August, 2020, when

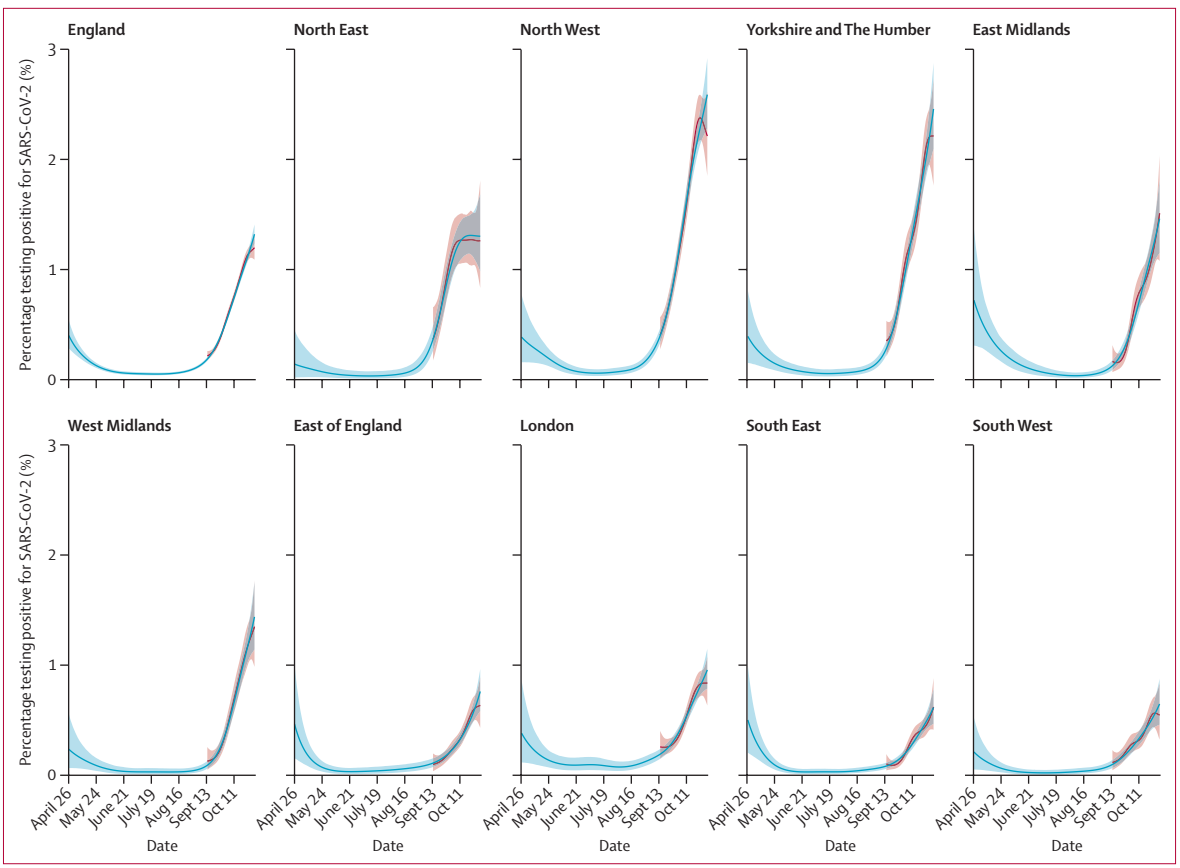

Figure 1: **Percentage of population living in private households testing positive for SARS-CoV-2 over time in England and the nine regions of England**
Shaded areas are 95% credible intervals. The blue curve is from a model fitted on data from the entire period (April 26 to Nov 1, 2020) whereas the red curve is from a model fitted on data from the 7 weeks up to Nov 1, 2020. SARS-CoV-2=severe acute respiratory syndrome coronavirus 2.

small increases were noted in low-evidence positive tests and few people reported symptoms when testing positive, but with higher viral loads in September, 2020, potentially leading to a higher proportion of cases with symptoms. A substantial proportion (varying between 68% in early July, 2020, to 45% in early October, 2020; figure 2A) of individuals who tested positive for SARS-CoV-2 did not report any symptoms on the day of the visit or at visits before or after the swab was taken.

SARS-CoV-2 positivity rates showed substantial regional differences, with increases in late August to October, 2020, largely occurring in regions in the north of England and (to a lesser extent) the Midlands regions (figure 1). The most important factor underlying the observed sharp increase in SARS-CoV-2 positivity was age, with earlier and greater increases apparent in younger adults (particularly those aged 17–24 years; figure 3), also shown by results from a model categorising age (appendix p 13). These data also show that near the end of October, 2020, the prevalence of SARS-CoV-2 positivity started to decrease in young adults. Importantly, clear diffusion of risk was seen from initial increases in younger age groups at lower risk of hospitalisation and death to older ages at higher risk.

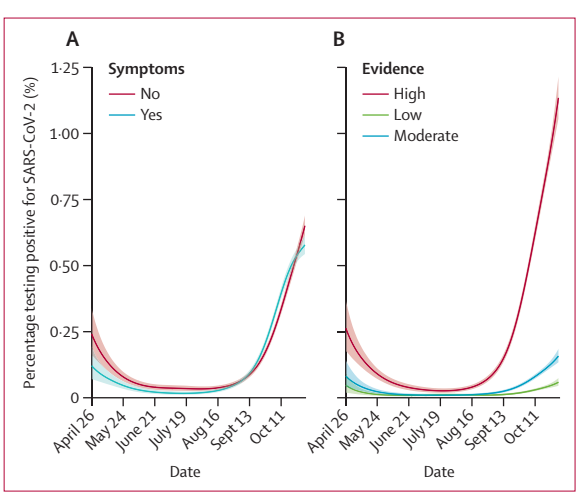

Figure 2: **Percentage of population living in private households testing positive for SARS-CoV-2**
Plots are with and without reporting symptoms (A) and stratified by high, moderate, and low evidence positivity (B). Shaded areas are 95% credible intervals.
SARS-CoV-2=severe acute respiratory syndrome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

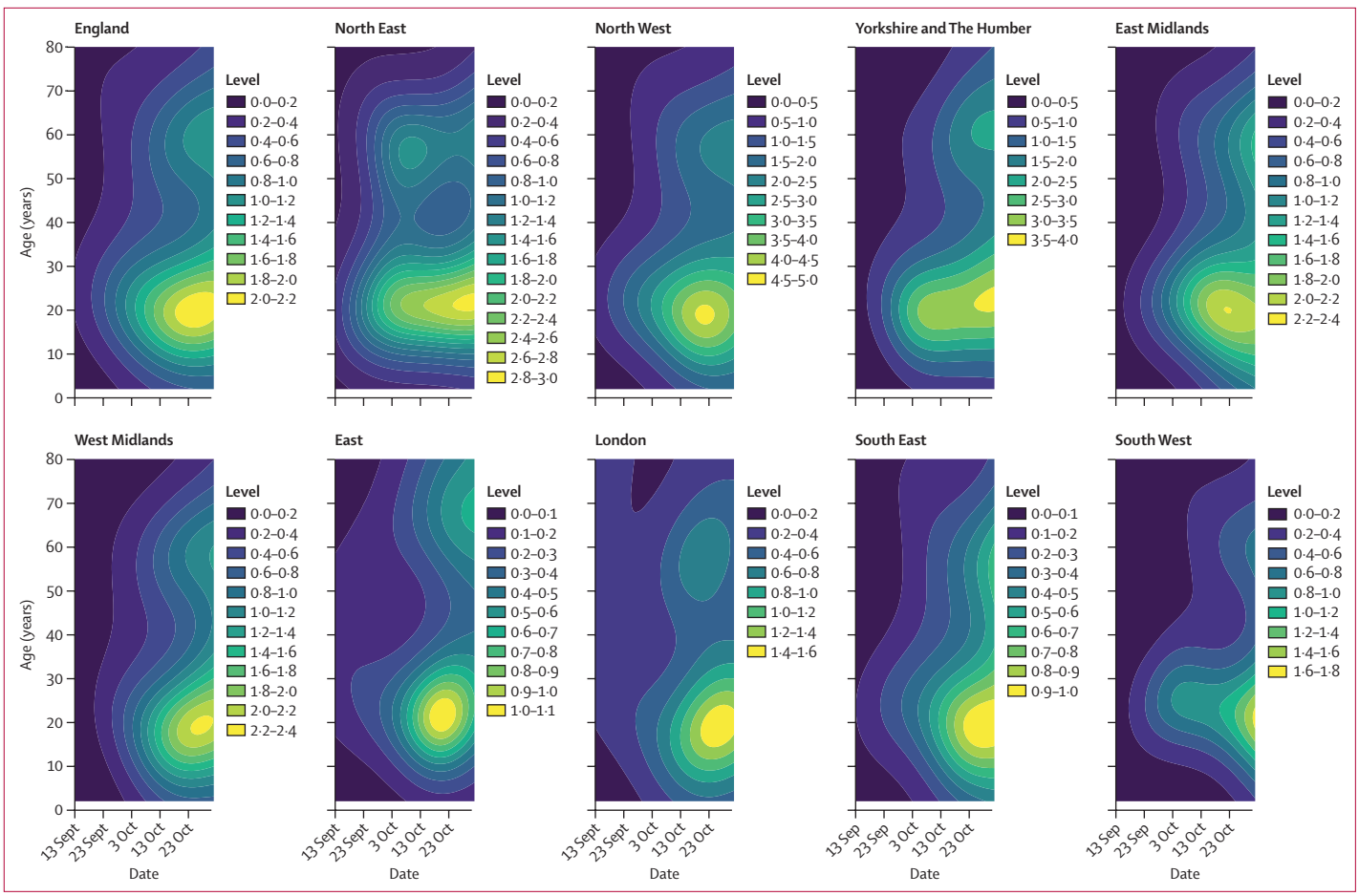

Figure 3: **Modelled estimates (posterior medians) of the distribution of positive SARS-CoV-2 tests by age over time**
Note that different scales are being used for each region. SARS-CoV-2=severe acute respiratory syndrome coronavirus 2.

Although working outside their home and in patient-facing health-care roles were clear risk factors during the first wave of high SARS-CoV-2 positivity that was captured by our study (April 26 to June 28, 2020), as was contact with hospitals (appendix p 8), there was no evidence that people working outside their home, working in patient-facing roles, or with hospital contact were driving initial increases after July and August, 2020 (appendix pp 14–16). Non-white ethnicity was also associated with greater SARS-CoV-2 positivity rates during the first wave but not the initial increases after July and August, 2020 (appendix pp 17–18). Although the probability of testing positive for SARS-CoV-2 increased in all age groups after July and August, 2020, the increase was especially pronounced in individuals younger than 25 years who shared a household with another person aged 17–24 years (appendix p 19).

Discussion

The findings of our longitudinal community survey show substantial changes over time in the percentage of people in private-residential households in England testing positive for SARS-CoV-2, with an initial decrease between April 26 and June 28, 2020, followed by consistently low levels during July and August, 2020, before substantial increases between the end of August and Nov 1, 2020. Our estimates have been regularly updated and shared with the UK Government and the Scientific Advisory Group for Emergencies Scientific Pandemic Influenza subgroup on Modelling to directly inform decisions about potential changes to the current alert level or relaxation of some restrictions. Notably, a substantial proportion of individuals who tested positive did not report any symptoms on the day of the visit or at visits before or after the swab was taken.

Bayesian dynamic multilevel generalised additive models are useful for monitoring the effect of different factors on SARS-CoV-2 positivity rates over time. In particular, these models show that the COVID-19 pandemic restarted in young people (particularly those aged 17–24 years), and that factors associated with an increased risk of testing positive for SARS-CoV-2 during April to June, 2020, such as working outside the home and having a job with direct patient contact, were not

important drivers of initial increases occurring since the end of August, 2020.

Although false-positive test results might be a concern when prevalence of SARS-CoV-2 is low, the low positivity for SARS-CoV-2 at the end of June, 2020 (0·05%), is also reassuring because 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 higher than 0·05%, even in the purely hypothetical situation that SARS-CoV-2 was not circulating in June, 2020.

According to findings of a systematic review of population-based prevalence surveys from 19 countries,[20] during the COVID-19 pandemic, two-thirds of studies (n=25 [68%]) reported only antibody testing, with many of those studies having a high risk of bias. The few PCR-based surveys (such as ours) were generally found to have a low risk of bias and, importantly, provide information about people currently infected and potentially able to transmit SARS-CoV-2.[20]

An important strength of our population-based study is that it can detect increases in the SARS-CoV-2 positivity rate potentially earlier and more systematically than can surveillance based on confirmed cases, hospital admissions, or deaths.[21,22] This advantage of our study will be most useful when new increases in SARS-CoV-2 positivity initially occur in a subgroup of the population at low risk of hospitalisation and death, but whose infections still contribute to transmission (including if asymptomatic),[23] which was the situation after July and August, 2020, with the increase in positive tests among young adults (particularly those aged 17–24 years). The sharp rise in cases in this age category that started in August, 2020, in the North West region of England subsequently resulted in increases in other age groups and preceded an upsurge in intensive care bed occupancy by patients with COVID-19 in larger cities in the North West region, such as Manchester, where 35% of beds were occupied by patients with COVID-19 on Oct 22, 2020.[24]

Interpretation of changes in SARS-CoV-2 incidence and positivity rates from tests that are taken for contact tracing or clinical cases is likely to be 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 multilevel regression and poststratification, thereby preventing difficulties with interpretation due to changes in testing practice.

A few other studies have 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 of SARS-CoV-2 among the general population between May 1 and June 1, 2020;[25] another cross-section from that study showed an increase in prevalence in September, 2020.[26] Among individuals who tested positive for SARS-CoV-2 in that study, the percentage reporting no symptoms varied between 50% and 81% in different cross-sections.[26] A study from Vo',[27] an Italian town with a population of 3275 individuals, reported the percentage of people who tested positive for SARS-CoV-2 who did not report any symptoms was 41·0–44·8%.[27] In a larger study from Iceland,[28] participants were recruited via an open invitation, which could bias the sample towards people with symptoms. 57% of individuals in that study who tested positive for SARS-CoV-2 reported having symptoms, although 29% of individuals testing negative also reported having symptoms.[28] A 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.[29] However, informing people that they were recently in close contact with a confirmed COVID-19 case can result in recall bias and overestimate the true prevalence of symptoms among a representative sample of infected people. Although we might have underestimated the true prevalence of symptoms among people with SARS-CoV-2 infection in the community, partly due to asking about current symptoms at visits up to July 23, 2020 (meaning that very transient symptoms only occurring between visits would have been missed), and symptoms in the past 7 days thereafter, our study adds to the growing evidence that a substantial proportion of SARS-CoV-2 in the community could be asymptomatic.[30–32]

For the **latest data from the UK Government on COVID-19** see https://coronavirus.data.gov.uk

Our survey shows that community supervised self-swabbing RT-PCR-based surveillance is achievable and practical. Community surveillance facilitates early detection of changes in the COVID-19 pandemic that are not driven by changes in testing, estimation of prevalence and incidence, evaluation of time-varying risk factors of testing positive for SARS-CoV-2, and changes in viral burden.[19] Our survey could serve as a model for other countries and future pandemics.

An important limitation of our study is that the number of people in the community who 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 effect on our results and, assuming the households we sampled are representative of households in general, our estimates should still reflect SARS-CoV-2 positivity rates in the target population as a whole.

Another limitation of our study is that, although we adjusted for potential non-representativeness in terms of age, sex, and region, there could be other factors for which we did not have detailed information about population distributions, which are also associated with testing positive for SARS-CoV-2. For example, people with non-white ethnicity were modestly under-represented in our survey, potentially underestimating the prevalence of SARS-CoV-2. Furthermore, associations (or, lack of associations) with testing positive for

SARS-CoV-2 could be due to residual confounding. We did forwards and backwards imputation for missing data, reflecting the relatively short timescales of the study.

A third limitation is that, in the absence of a true gold standard, we do not know the sensitivity and specificity of the PCR test, making assessment of the true prevalence of SARS-CoV-2 difficult. However, the true specificity is likely to be very close to 100%. These data cannot inform about test sensitivity without providing a very informative prior on the true prevalence.[33] Although self-swabbing was monitored by study workers and is used very widely, this procedure could still lead to underestimates of prevalence of SARS-CoV-2. However, use of self-swabbing should not affect trends over time.

For more on **accreditation** see https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme

In a rapidly evolving epidemic, during which ongoing surveillance is essential to guide public health response, Bayesian dynamic multilevel regression and poststratification is a powerful method to ensure population-representative estimates can be obtained. Specifically, our model showed that the percentage of individuals from the community in England testing positive for SARS-CoV-2 declined between April 26 and June 28, 2020, and remained stable for much of July and August, 2020, before increasing again from the end of August to October, 2020. Important risk factors for testing positive for SARS-CoV-2 varied substantially between the first and second waves of higher positivity rates, and a substantial proportion of SARS-CoV-2 infections were in individuals not reporting symptoms, indicating that continued monitoring for SARS-CoV-2 in the community will be important for managing the pandemic moving forwards.

Contributors

ASW, JF, JIB, JNN, IBe, ID, PB, KBP, and JVR designed and planned the study. KBP, TH, EP, K-DV, NB, IBo, HT, JL, AG, PJB, and ASW contributed to the statistical analysis. KBP drafted the report. All authors contributed to interpretation of data and revised the report. KBP and ASW are the guarantors of the study. All authors approved the final version of the report and agree to be accountable for all aspects of the work.

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 Airimitoaie, 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 B 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; and *Glasgow Lighthouse Laboratory*—Jodie Hay, Harper Vansteenhouse.

Declaration of interests

We declare no competing interests.

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 or visit the SRS website.

Acknowledgments

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 (NIHR) Health Protection Research Unit (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 the US National Institutes of Health and Office of Naval Research. ASW is also supported by the NIHR Oxford Biomedical Research Centre and by core support from the UK Medical Research Council (MRC) to the MRC Clinical Trials Unit (MC_UU_12023/22) and is an NIHR Senior Investigator. We acknowledge the support of the Huo Family Foundation. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, Department of Health, or PHE.

References

1 Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395:** 497–506.

2 WHO. Weekly epidemiological update. Nov 3, 2020. https://www.who.int/publications/m/item/weekly-epidemiological-update---3-november-2020 (accessed Nov 19, 2020).

3 Salje H, Tran Kiem C, Lefrancq N, et al. Estimating the burden of SARS-CoV-2 in France. *Science* 2020; **369:** 208–11.

4 Jarvis CI, Van Zandvoort K, Gimma A, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med* 2020; **18:** 124.

5 Davies NG, Kucharski AJ, Eggo RM, et al. 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; **5:** e375–85.

6 Guan D, Wang D, Hallegatte S, et al. Global supply-chain effects of COVID-19 control measures. *Nat Hum Behav* 2020; **4:** 577–87.

7 US Department of Labour. Unemployment insurance weekly claims. June 4, 2020. https://www.dol.gov/sites/dolgov/files/OPA/newsreleases/ui-claims/20201165.pdf (accessed Nov 19, 2020).

8 HM Government. COVID alert levels. May 11, 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/884352/slides_-_11_05_2020.pdf (accessed Nov 19, 2020).

9 Kojima N, Turner F, Slepnev V, et al. Self-collected oral fluid and nasal swab specimens demonstrate comparable sensitivity to clinician-collected nasopharyngeal swab specimens for the detection of SARS-CoV-2. *Clin Infect Dis* 2020; published online Oct 19. https://doi.org/10.1093/cid/ciaa1589.

10 Office for National Statistics. COVID-19 infection survey (pilot): methods and further information. Sept 21, 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation (accessed Nov 19, 2020).

11 Gelman A, Little TC. Poststratification into many categories using hierarchical logistic regression. *Surv Methodol* 1997; **23:** 127–35.

12 Gelman A, Lax J, Phillips J, Gabry J, Trangucci R. Using multilevel regression and poststratification to estimate dynamic public opinion. Aug 28, 2018. http://www.stat.columbia.edu/~gelman/research/unpublished/MRT(1).pdf (accessed Nov 19, 2020).

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:** 479–91.

14 Warshaw C, Rodden J. How should we measure district-level public opinion on individual issues? *J Polit* 2012; **74:** 203–19.

15 Kennedy L, Gelman A. Know your population and know your model: using model-based regression and poststratification to generalize findings beyond the observed sample. *arXiv* 2020; published online April 13. https://arxiv.org/abs/1906.11323v2 (preprint, version 2).

16 Si Y, Trangucci R, Gabry JS, Gelman A. Bayesian hierarchical weighting adjustment and survey inference. July 25, 2017. http://www.stat.columbia.edu/~gelman/research/unpublished/modelweighting.pdf (accessed Nov 19, 2020).

17 McCullagh P. Regression models for ordinal data. *J R Stat Soc B* 1980; **42:** 109–42.

18 Wood SN. Thin plate regression splines. *J R Stat Soc Series B Stat Methodol* 2003; **65:** 95–114.

19 Walker AS, Pritchard E, House T, et al. Viral load in community SARS-CoV-2 cases varies widely and temporally. *medRxiv* 2020; published online Oct 27. https://doi.org/10.1101/2020.10.25.20219048 (preprint, version 1).

20 Franceschi VB, Santos AS, Glaeser AB, et al. Population-based prevalence surveys during the COVID-19 pandemic: a systematic review. *medRxiv* 2020; published online Oct 22. https://doi.org/10.1101/2020.10.20.20216259 (preprint, version 1).

21 Office for National Statistics. Deaths registered weekly in England and Wales, provisional. Oct 30, 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending30october2020 (accessed Nov 19, 2020).

22 Public Health England. Weekly coronavirus disease 2019 (COVID-19) surveillance report: summary of COVID-19 surveillance systems. May 27, 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/888254/COVID19_Epidemiological_Summary_w22_Final.pdf (accessed Nov 19, 2020).

23 Chau NVV, Thanh Lam V, Thanh Dung N, et al. The natural history and transmission potential of asymptomatic severe acute respiratory syndrome coronavirus 2 infection. *Clin Infect Dis* 2020; published online June 4. https://doi.org/10.1093/cid/ciaa711.

24 Dobson C. The true intensive care unit figures in Greater Manchester's hospitals as our Nightingale is set to reopen. Oct 22, 2020. https://www.manchestereveningnews.co.uk/news/greater-manchester-news/true-intensive-care-unit-figures-19148243 (accessed Nov 19, 2020).

25 Riley S, Ainslie KEC, Eales O, et al. Community prevalence of SARS-CoV-2 virus in England during May 2020: REACT study. *medRxiv* 2020; published online July 11. https://doi.org/10.1101/2020.07.10.20150524 (preprint, version 1).

26 Riley S, Ainslie KEC, Eales O, 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; published online Oct 2. https://doi.org/10.1101/2020.09.30.20204727 (preprint, version 1).

27 Lavezzo E, Franchin E, Ciavarella C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* 2020; **584:** 425–29.

28 Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic population. *N Engl J Med* 2020; **382:** 2302–15.

29 Byambasuren O, Cardona M, Bell K, Clark J, McLaws M-L, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *J Assoc Med Microbiol Infect Dis Can* 2020; published online Oct 9. https://doi.org/10.3138/jammi-2020-0030.

30 Poletti P, Tirani M, Cereda D, et al. Probability of symptoms and critical disease after SARS-CoV-2 infection. *arRxiv* 2020; published online June 22. https://arxiv.org/abs/2006.08471v2 (preprint, version 2).

31 Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* 2020; **396:** 535–44.

32 Kasper MR, Geibe JR, Sears CL, et al. An outbreak of Covid-19 on an aircraft carrier. *N Engl J Med* 2020; published online Nov 11. https://doi.org/10.1056/NEJMoa2019375.

33 Lewis FI, Torgerson PR. 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:** 9.