

The Incidence and Prevalence of SARS-CoV-2 in the UK Population from the UKHSA Winter COVID Infection Study

Martyn Fyles^{1,*}, Jonathon Mellor¹, Robert Paton¹, Christopher E. Overton^{1,2}, Alexander M. Phillips^{1,3}, Alex Glaser¹, Thomas Ward¹

¹ *Infectious Disease Modelling Team, All Hazards Intelligence, Data Analytics and Surveillance, UK Health Security Agency*

² *Department for Mathematical Sciences, University of Liverpool*

³ *Department of Electrical Engineering and Electronics, University of Liverpool*

*Corresponding author: Martyn.Fyles@ukhsa.gov.uk

Abstract

The SARS-CoV-2 virus continues to cause substantial morbidity and mortality, particularly during the winter period. The Winter Covid Infection Study (WCIS) ran from the 14th of November 2023 to the 7th of March 2024, and enabled the UK Health Security Agency to publish fortnightly estimates of the incidence and prevalence of SARS-CoV-2. Testing was performed using Lateral Flow Devices (LFD), and a repeat testing design was used to estimate key epidemiological parameters of SARS-CoV-2. This facilitated the estimation of time-varying prevalence, incidence, and test sensitivity. A Bayesian multilevel regression and poststratification model was developed to produce representative and unbiased estimates. In England and Scotland, prevalence peaked at 4.54% (95% CI: 3.90 to 5.24), and incidence peaked at 498 (95% CrI: 429 to 585) new infections per 100,000 individuals per day. The average LFD test sensitivity in England and Scotland during the study was estimated to be 72.1% (95% CrI: 70.3, 74.0), though due to epidemic phase bias this varied from a minimum value of 68.6% (95% CrI: 66.4 to 70.7) to a maximum value of 77.2% (95% CrI: 75.3 to 79.2). The novel study design of WCIS addressed key survey design challenges faced by previous large-scale SARS-CoV-2 population prevalence studies. The study demonstrated the utility and cost-benefit of LFD tests in large community surveys of prevalence.

Key words:

SARS-CoV-2; prevalence; incidence; sensitivity; community; LFD; survey design

Introduction

Respiratory pathogens, such as influenza and SARS-CoV-2, contribute substantially to the burden of disease in the UK, particularly during the winter period [1]. Epidemic waves for respiratory pathogens can vary substantially in size from year to year [2], and consequently reliable surveillance of respiratory pathogens during the winter period is an essential tool in public health. Surveillance of SARS-CoV-2 is particularly important, as the virus continues to evolve within a shifting epidemiological and immunological landscape [3], and its seasonal dynamics are still uncertain due to its relatively recent emergence in 2019.

In the UK, the majority of respiratory pathogen surveillance is subject to a severity bias [2], e.g. reports of influenza-like illness, data from individuals who sought healthcare, and morbidity/mortality metrics. Older age groups are particularly affected by this severity bias, which can make it difficult to understand epidemic trends in the wider population. Community prevalence studies test randomly sampled individuals for infection with the pathogen of interest, which facilitates less biased surveillance of the epidemic in the general population. In England during the SARS-CoV-2 pandemic two large-scale community prevalence surveys, REACT (Real-time Assessment of Community Transmission) and CIS (Community Infection Survey) [4, 5, 6], were conducted. These studies provided decision makers with timely, reliable, and representative surveillance of the SARS-CoV-2 epidemic in the general population, which was of vital importance during a public health emergency.

From the 14th November 2023 to the 7th March 2024 the UK Health Security Agency (UKHSA), in partnership with the Office for National Statistics (ONS), ran the Winter Covid Infection Study (WCIS); a community prevalence study that published fortnightly estimates of the prevalence and incidence of SARS-CoV-2 in England and Scotland. Prevalence is defined as the proportion of the population currently infected with SARS-CoV-2, and incidence is defined as the rate of new SARS-CoV-2 infections per capita and per day. The study design of WCIS builds upon the work of the CIS, which was performed by ONS. Previous SARS-CoV-2 community prevalence studies used Polymerase Chain Reaction (PCR) tests, however WCIS moved to using Lateral Flow Device (LFD) tests. Additionally, WCIS introduced a repeat testing design for participants that test positive. This enabled the estimation of key epidemiological parameters [7] that are used to infer the prevalence and incidence.

In the primary analysis of REACT and CIS, PCR test sensitivity was not estimated or accounted for. As such, these studies reported SARS-CoV-2 PCR positivity, rather than SARS-CoV-2 prevalence. Given that the PCR test is highly sensitive [8], positivity and prevalence can be similar, however divergences may appear due to epidemic dynamics. The analysis of CIS used post-stratification to produce representative estimates, whereas REACT relied upon the ongoing recruitment of a representative sample for each study round. The model structure of REACT was therefore required to handle gaps between subsequent rounds. A study by Ward et. al. [9] combined the datasets of CIS and REACT, and further extended the methods of these studies to estimate the time-varying PCR test sensitivity, incidence rate, and infection mortality/hospitalisation risk. Inference of these additional quantities requires several epidemiological parameters, such as duration of positivity, to be estimated from repeat testing data. REACT and CIS were not designed with the goal of collecting repeat testing data, though in actuality CIS did collect a small amount of sparse repeat testing data that could be used to estimate the additional parameters. The sparsity of the data necessitated the use of complex model structures during analysis, however. To facilitate easier and more robust estimation of key epidemiological parameters, WCIS was designed to include a repeat testing design for all individuals who tested positive.

Previous SARS-CoV-2 community prevalence studies focused primarily on PCR tests. PCR tests are processed in a laboratory and are highly sensitive as a result, though the laboratory processing requirement makes PCR tests relatively expensive and necessitates a delay of up to several days to receive the result. Further, a community prevalence study requires sufficient laboratory capacity/infrastructure to be available throughout the study period, which is not guaranteed. In contrast, the LFD tests are self-contained units that can be performed at home, provide results in under 30 minutes, and have a substantially lower cost per test. LFD tests therefore have many advantages over PCR tests in the context of community prevalence surveys. The primary disadvantage of LFD tests is a reduced test sensitivity when compared to PCR tests [10], which makes it necessary to adjust for the test sensitivity of LFD tests accurately and robustly to obtain reliable estimates of the prevalence.

Several papers have demonstrated that average cycle threshold (Ct) values, a proxy for the average viral concentration in samples, varies depending on the recent epidemic dynamics [11] due to an effect known as epidemic phase bias. As LFD tests have been shown to have a reduced sensitivity to small viral concentrations [10], it is plausible that epidemic phase bias affects LFD test sensitivity more than it affects PCR test sensitivity. In order to produce reliable and unbiased estimates of the prevalence it is necessary to develop methods to account for the potentially time-varying nature of LFD test sensitivity.

This manuscript develops a Bayesian multilevel regression approach for estimating the SARS-CoV-2 incidence and prevalence over time. A convolution function approach is used, which obtains all quantities of interest as convolutions of the incidence time series. The resulting estimates of LFD test sensitivity are conditional upon the recent epidemic dynamics, and therefore the effects of epidemic phase bias on test sensitivity are accounted for in the model. A counterfactual analysis is performed to understand the error in prevalence had the effects of epidemic phase bias not been adjusted for. All model components, including parameter estimation models that use the WCIS repeat testing data from Overton et. al. [7], were implemented in a single Bayesian program allowing for full propagation of uncertainty throughout the model. Overall, the methodological improvements in the design and analysis of WCIS allowed LFD tests to be successfully utilised in a community prevalence studies to produce cost-effective and timely epidemic surveillance.

Methods

Epidemiological Data

Testing data were collected as part of the Winter Covid Infection Survey (WCIS). Every 4 weeks study participants were given an 8-day window, within which they were asked to perform an LFD test and complete a short questionnaire. If the participants LFD test was positive, they were asked to answer a follow-up questionnaire and participate in a repeat-testing regime, whereby they continue to take an LFD test every other day until testing negative for two consecutive tests. These data, including the initial test results, repeat tests and questionnaire results, were reported to ONS and then made available to UKHSA. Further details on the cohort, study design and data processing are provided in the supplementary materials. A breakdown of the sample size of each stratum is provided in Supplementary Table (1).

Demographic information about each individual surveyed is recorded alongside their testing results, including their age, sex, and location of residence in the UK. The age groups considered for analysis are as follows: 3 to 17 years, 18 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 to 74 years, and 75 years and over. Locations are reported at a combination of Scotland and the English

regions, giving ten total locations: East Midlands, East of England, London, North East, North West, Scotland, South East, South West, West Midlands, Yorkshire and the Humber.

Model structure

Modelling approach

To begin, we briefly outline the different datasets that are collected, the epidemiological parameters that need to be estimated, and the relationships between the key quantities of interest.

This study integrates epidemiological parameter models developed by Overton et. al. [7], which developed using the repeat testing data collected by WCIS. The definition of a currently infected individual is therefore inherited from Overton et. al., where an individual was defined to be “currently infected” if they have a non-zero probability of returning a positive LFD test result that is not a false positive. These individuals are referred to as being in the LFD-positive state. This definition is a natural consequence of the study being performed using only LFD tests, since there are no PCR tests to compare against.

During the study, participants were periodically asked to self-test which produced a daily time series of LFD outcomes across different strata to inform sample positivity. The repeat-testing data is used to provide estimates of the duration of LFD positivity in infected individuals, and the LFD test sensitivity profile over the course of an infection. The estimated duration of LFD positivity and the LFD test sensitivity profile are then used to obtain both the positivity and prevalence time series as convolutions of the incidence time series, with the positivity then being fitted to the data. This approach ensures the complex epidemiological relationships between incidence, prevalence, and positivity are fully accounted for. Study participants are provided with a 7-day window within which to perform their LFD test, however in reality participants were able to submit test results two days before the window opened, and one day after the window closed. Therefore, the effective window was 10 days in length, though submissions outside of the intended 7-day window were in the minority. Individuals who tested earlier in the window were observed to be more likely to test positive, and this window effect needed to be adjusted for when fitting to the observed positivity data.

The study population is stratified by 7 age groups, 10 locations, and sex, resulting in 140 different strata. Consequently, this can result in a small number of samples in each stratum each day. A multilevel regression approach is used to reduce the uncertainty in estimates by pooling information across strata. Poststratification is then performed to obtain representative estimates across subgroups by adjusting for the over/under representation of different strata within the sample.

The time-varying average test sensitivity in a population is derived from the estimated positivity and prevalence time series, rather than being supplied as an input to the calculation of prevalence from positivity [9]. This provides an estimate of the test sensitivity that accounts for epidemic phase bias [12]. Epidemic phase bias is where the epidemic phase, e.g. growth or decline, biases the distribution of times since infection in infected individuals. When the epidemic is growing rapidly, the time since infection is biased towards smaller values for example, as there are expected to be more new infections than the previous day. This alters the expected test sensitivity value, as viral load typically peaks a few days after infection and therefore test sensitivity peaks a few days after infection. To assess the importance of estimating a time-varying test sensitivity model, a counterfactual analysis is performed to estimate the error had a constant estimate of test sensitivity been used to estimate prevalence from positivity.

Convolution function approach

Any time series of an epidemic quantity, $Q(t)$, can be obtained as the convolution of the incidence time series $I(t')$ up until time t , and a function $f_Q(\tau)$ that describes the relationship between the incidence time series and the quantity of interest [13];

$$Q(t) = \int_0^{\infty} I(t - \tau) f_Q(\tau) d\tau.$$

The model therefore begins with the incidence time series and convolves the incidence with different functions that describe SARS-CoV-2 epidemiology to obtain a time series of expected positivity and prevalence.

Note that $Q(0)$ depends on $t - \tau < 0$, which implies that all epidemic quantity time series depends on values of the incidence time series that occurred prior to the start of the study. Since SARS-CoV-2 infections cannot remain positive indefinitely, for both the prevalence and positivity time series it is assumed is a maximum length of time that individuals can remain infected for, $W \in \mathbb{R}_+$. This implies that, the current prevalence and positivity are effectively only conditional upon the previous W days of incidence, i.e. for $\tau > W$ we have that $f_Q(\tau) = 0$. Therefore, the indefinite integral can be approximated as

$$Q(t) = \int_0^{\infty} I(t - \tau) f_Q(\tau) d\tau \approx \int_0^W I(t - \tau) f_Q(\tau) d\tau.$$

To compute $Q(t)$, the incidence time series is modelled on the interval $[-W, T]$, where T is the last time point in the study, and the positivity and prevalence time series are modelled on the interval $[0, T]$. The interval $[-W, 0)$ is referred to as the “warmup interval” and is weakly identified. Consequently, while the warmup interval is necessary for model fitting, it does not have a clear interpretation and is not presented in the results section of the paper. For this paper, a value of $W = 20$ was used, as the probability of an individual remaining LFD positive for more than 20 days is small [7].

Let $I_{a,l,s}(t) \in [0,1]$ denote the incidence time series for a given age, location, and sex combination, where incidence is defined as the proportion of that subgroup newly infected on day t . The prevalence time series, $\text{Prev}_{a,l,s}(t) \in [0,1]$, is defined as the proportion of that subpopulation that is in the LFD-positive state on day t . Let $f_{\text{prev}}^a(\tau)$ be the probability that an individual in age group a is in the LFD-positive state $\tau \in \mathbb{R}_+$ days after they were infected. The prevalence time series is then obtained by convolving $f_{\text{prev}}^a(\tau)$ with the incidence time series;

$$\text{Prev}_{a,l,s}(t) = \int_0^W I_{a,l,s}(t - \tau) f_{\text{prev}}^a(\tau) d\tau.$$

To calculate the expected sample positivity, it is necessary to account for prevalence, sensitivity, and specificity. The false positive probability and the true positive probability are analysed separately via

$$\text{Pos}_{a,l,s}(t) = \text{Pos}_{a,l,s}^{\text{TP}}(t) + \text{Pos}_{a,l,s}^{\text{FP}}(t),$$

where the superscript “TP” and “FP” represent true and false positives, respectively. The true positive probability is given by the probability that a randomly sampled member of the stratum is in the LFD-positive state and when tested returns a true positive result. The false positive probability is

the probability that an uninfected member of the population is tested and returns a false positive test result.

In [7], a model is developed to describe the probability of observing a true positive LFD test given that an individual is still LFD-positive τ days after they were infected, denoted by $f_{+|+}^a(\tau) \in [0,1]$. This is referred to as the test sensitivity profile of LFD-positive individuals. Individuals who were recently infected are likely to have high viral loads, and therefore the LFD test sensitivity is high for small values of τ . As τ increases, the test sensitivity decreases as individuals who remain LFD-positive approach the end of their infection, which implies a decreasing viral load.

The integrand for the prevalence convolution, $I_{a,l,s}(t - \tau)f_{\text{prev}}^a(\tau)$, provides the proportion of the population that were infected at time $t - \tau$ and were still LFD-positive τ days later at time t . For the positivity convolution, $I_{a,l,s}(t - \tau)f_{\text{prev}}^a(\tau)f_{+|+}^a(\tau)$ provides the probability that a randomly tested member of the subpopulation at time t was infected at time $t - \tau$, is still LFD-positive τ days and returns a true positive test result. Therefore, the true positive probability at time t is given by

$$\text{Pos}_{a,l,s}^{\text{TP}}(t) = \int_0^W I_{a,l,s}(t - \tau)f_{\text{prev}}^a(\tau)f_{+|+}^a(\tau)d\tau$$

Under the assumption that an uninfected individual returns a false positive test result independently at random with probability $1 - \gamma \in [0,1]$, where γ is the test specificity, the time-varying false positive probability is given by

$$\text{Pos}_{a,l,s}^{\text{FP}}(t) = (1 - \text{Prev}_{a,l,s}(t))(1 - \gamma)$$

Combining this together, the positivity time series is given by

$$\text{Pos}_{a,l,s}(t) = \int_0^W I_{a,l,s}(t - \tau)f_{\text{prev}}^a(\tau)f_{+|+}^a(\tau)d\tau + (1 - \text{Prev}_{a,l,s}(t))(1 - \gamma)$$

Given that $1 \approx (1 - \text{Prev}_{a,l,s}(t))$ and $(1 - \gamma) \approx 10^{-4}$, for computational efficiency the false positive rate is approximated as

$$\text{Pos}_{a,l,s}^{\text{FP}}(t) \approx (1 - \gamma).$$

Likelihood and model fitting

Additional effects are modelled to account for individuals' propensity to test at different points in their assigned testing window. Participants were observed to be more likely to test positive if they tested at the very start, or prior to their window opening. The window effects are assigned a weakly-informative prior of

$$\beta_w^{\text{window}} \sim \mathcal{N}(0, \sqrt{15}),$$

where $w \in \{-2, -1, \dots, 7\}$ is the number of days since the testing window opened. These testing effects are constrained so that the average effect on the prevalence time series is zero over the course of the testing window when weighted by the number of tests, n_i , taken on each window index, i :

$$\sum_{i=-2}^7 n_i \beta_i^{\text{window}} = 0.$$

Typically, this constraint would be implemented using a simplex, however for computational efficiency we implement this as a soft constraint using

$$\sum_{i=-2}^7 n_i \beta_i^{\text{window}} \sim \mathcal{N}(0, 10^{-3}).$$

The expected rate of positivity after accounting for window effects is then calculated as:

$$p_{a,l,s,w}(t) = \text{logit}^{-1} \left(\text{logit} \left(\text{Pos}_{a,l,s}(t) \right) + \beta_w \right).$$

The observed series of positive tests each day and in each of the subgroups, $y_{a,l,s,w}(t)$, is modelled using a binomial distribution with a probability of the predicted rate of positivity for each subgroup at each time point

$$y_{a,l,s,w}(t) \sim \text{Binomial} \left(N_{a,l,s,w}(t), p_{a,l,s,w}(t) \right).$$

The model generates posterior samples of the most credible incidence time series, prevalence time series, and epidemiological parameters that explain the observed data for each combination of age, location, and sex.

Multilevel regression

A Bayesian multilevel regression model is used to estimate incidence for each demographic stratum. A national level trend of log-odds incidence,

$$f^{\text{nat}}(t) \in \mathbb{R}^{T+W}$$

is combined with per-age, per-location, and per-sex deviations from the national-level trend, each of which are also allowed to vary over time;

$$f_l^{\text{loc}}(t) \in \mathbb{R}^{T+W} \quad \forall l \in L,$$

$$f_a^{\text{age}}(t) \in \mathbb{R}^{T+W} \quad \forall a \in A,$$

$$f^{\text{sex}}(t) \in \mathbb{R}^{T+W}.$$

Some values of the incidence time series and duration of positivity distributions will result in prevalence values that are greater than 1. These incidence time series would be impossible to observe, e.g. 75% of a subgroups population is infected two days in a row. However, during the MCMC initialisation and adaptation steps, this can result in many draws being rejected causing poor convergence and inefficient sampling. To prevent this, the incidence is restricted such that $I_{a,l,s}(t) \in [0, 0.05]$, implying that no more than 5% of any subgroup population is infected per day. SARS-CoV-2 has an average duration of LFD positivity of approximately 10 days [7], and therefore the upper limit of 5% incidence per day places an upper limit of approximately 50% prevalence – significantly higher than any observed community prevalence or positivity in the UK. Consequently, it is extremely unlikely that at any point since the emergence of SARS-CoV-2 the incidence has ever been above 5%.

Letting s take values of either -1 or 1 we have

$$I_{a,l,s}(t) = 0.05 \cdot \text{logit}^{-1} \left(f^{\text{nat}}(t) + f_l^{\text{loc}}(t) + f_a^{\text{age}}(t) + s f^{\text{sex}}(t) \right)$$

Each of the national, location, age and sex-level trends are assigned second-order random-walk (RW2) smoothing priors [14, 15] as follows:

$$\begin{aligned}
 f^{\text{nat}}(t) &\sim \text{RW2}(\mu^{\text{nat}}, \tau^{\text{nat}}), \\
 f_l^{\text{loc}}(t) &\sim \text{RW2}(\mu_l^{\text{loc}}, \tau^{\text{loc}}) \forall l \in L, \\
 f_a^{\text{age}}(t) &\sim \text{RW2}(\mu_a^{\text{age}}, \tau^{\text{age}}) \forall a \in A, \\
 f^{\text{sex}}(t) &\sim \text{RW2}(\mu^{\text{sex}}, \tau^{\text{sex}}).
 \end{aligned}$$

The $\tau \in \mathbb{R}_+$ parameters control the “smoothness” of the time series, such that a smaller value for τ indicates more smoothing, and each $\mu \in \mathbb{R}$ parameter indicates the starting value of the random walk. The following prior for smoothness was used,

$$\tau^{\text{nat}}, \tau^{\text{loc}}, \tau^{\text{age}}, \tau^{\text{sex}} \sim \text{Exponential}(100).$$

The intercepts μ for each location, age group and sex are given zero-mean priors with pooled variances across all locations, age-groups, and sexes:

$$\begin{aligned}
 \mu_l^{\text{loc}} &\sim \mathcal{N}(0, \sigma^{\text{loc}}), \\
 \mu_a^{\text{age}} &\sim \mathcal{N}(0, \sigma^{\text{age}}), \\
 \mu^{\text{sex}} &\sim \mathcal{N}(0, \sigma^{\text{sex}}), \\
 \sigma^{\text{loc}}, \sigma^{\text{age}}, \sigma^{\text{sex}} &\sim \text{Exponential}(100).
 \end{aligned}$$

For the national level trend intercept we assign the prior of

$$\mu^{\text{nat}} \sim \mathcal{N}(-4, 3).$$

The median of this prior gives an incidence rate of 90 new infections per day per 100,000 individuals, and the interval of (0.26, 4300) new infections per day per 100,000 individuals contains 95% of the prior probability mass. As such, the median value of this prior is on the right scale for the incidence rate, however there is substantial flexibility for the model to explore if necessary.

After model fitting, poststratification [16, 17] is performed to produce estimates of the incidence and prevalence that are representative of a given demographic strata, e.g. those aged between 3 to 17 years, by adjusting for the over/under representation of different stratum within a given strata. The size of each stratum is obtained using ONS household population estimate projections [18].

Infections per individual

The “size” of an epidemic wave can be contextualised by reporting the expected number of SARS-CoV-2 infections an individual will experience over the course of the study. Here, infection is used to refer to a successful SARS-CoV-2 transmission event. Let $\lambda_{a,l,s}(t) \in \mathbb{R}_+$ be the total number of infections that occurred between day 0 and day t of the study for a given stratum, given by

$$\lambda_{a,l,s}(t) = P_{a,l,s} \cdot \int_0^t I_{a,l,s}(\tau) d\tau$$

where $P_{a,l,s}$ is the stratum’s population size. The average number of infections that occurred per individual between day 0 and t of the study, $\alpha_{a,l,s}(t) \in \mathbb{R}_+$, is therefore given by

$$\alpha_{a,l,s}(t) = \frac{\lambda_{a,l,s}(t)}{P_{a,l,s}} = \int_0^t I_{a,l,s}(\tau) d\tau.$$

It is possible for individuals to be infected more than once, and consequently if the study had run for long enough the average number of infections per individual would eventually exceed 1. If the

average number of infections per individual was equal to 1, this would not imply that each member of the population had been infected once, as some individuals would have been infected more than once, while others may not have been infected at all. Due to the length of the study, it is unlikely that a substantial number of individuals were infected more than once.

Analysis of time-varying test sensitivity

The population average test sensitivity, $\text{Sens}_{a,l,s}(t)$, is defined as the probability of returning a positive test from a randomly selected LFD-positive member of a stratum at a given point in time. This differs from the test sensitivity profile of LFD-positive cases, which refers to the probability of returning a true positive test result given the time since infection of the LFD-positive individual. The population average test sensitivity is determined by the test sensitivity profile and the infectious age distribution, i.e. the distribution of times since infection in LFD positive individuals, of that population at that point in time.

For a given stratum, the population average test sensitivity can be expressed in terms of the prevalence and the expected true positive rate,

$$\text{Sens}_{a,l,s}(t) = \frac{\text{Pos}_{a,l,s}^{\text{TP}}(t)}{\text{Prev}_{a,l,s}(t)}.$$

Consequently, $\text{Sens}_{a,l,s}(t)$ is conditional upon the recent values of the incidence time series, weighted by the probability that infections several days ago remain positive and would return a false negative result. This results in a model of test sensitivity that is affected by epidemic phase bias, as it accounts for how the recent epidemic dynamics have affected the infectious age distribution and the epidemiological characteristics of SARS-CoV-2 infections within that population.

For LFD tests, the magnitude of the effect that epidemic phase bias has on the test sensitivity is unknown. Here, we develop an expression for the error incurred had a constant value of test sensitivity been used, as this would not account for epidemic phase bias. Let $\overline{\text{Sens}}_{a,s,l}$ be the average of value of $\text{Sens}_{a,l,s}(t)$. The estimate of prevalence obtained using the average value of sensitivity is given by

$$\widehat{\text{Prev}}_{a,l,s}(t) = \frac{\text{Pos}_{a,l,s}(t)}{\overline{\text{Sens}}_{a,s,l}}.$$

The relative error in the prevalence at a given point in time from using a constant test sensitivity estimate is then calculated as

$$\text{Error}_{a,l,s}(t) = \frac{\text{Prev}_{a,l,s}(t) - \widehat{\text{Prev}}_{a,l,s}(t)}{\text{Prev}_{a,l,s}(t)} = 1 - \frac{\text{sens}_{a,l,s}(t)}{\overline{\text{sens}}_{a,l,s}}.$$

As the time-varying sensitivity is conditional on the recent epidemic dynamics, the error in the prevalence estimate is highly related to the exponential growth rate of the epidemic for each stratum, $r_{a,l,s}(t) \in \mathbb{R}^T$. Under the assumptions of an exponential growth model, we have that

$$I_{a,s,l}(t) = I_{a,s,l}(t-1) \cdot e^{r_{a,l,s}(t)},$$

which allows the daily exponential growth rate to be estimated as

$$r_{a,l,s}(t) = \log(I_{a,l,s}(t)) - \log(I_{a,l,s}(t-1)).$$

Implementation

The Bayesian model is implemented using Stan [19], a probabilistic programming language for Bayesian inference. The parameter estimation sub-models from Overton et. al. [7], and the methodology described here are implemented in a single Stan program. This was technically challenging to perform, and the resulting Stan program is complex, however this approach ensures that uncertainty is properly propagated throughout the model. Model fitting was performed using Hamiltonian MCMC, with 8 chains using 500 warmup iterations and 500 sampling draws each. Model fitting took approximately 12 hours to complete using 8 CPU cores.

Results

Prevalence

The SARS-CoV-2 prevalence in England and Scotland between the 14th November 2023 and the 7th March 2024 is provided in Figure (1). Rapid growth in SARS-CoV-2 occurred at the beginning of December 2023 before reaching a peak prevalence of 4.54% (95% CI: 3.90 to 5.24) on the 22nd December 2023, after which prevalence began to rapidly decline. Between the 18th January 2024 and the 1st February 2024, the prevalence plateaued at approximately 2% before continuing to decline until the end of the study. A minimum prevalence of 0.718% (95% CI: 0.543 to 0.945) occurred on the 1st March 2024.

Age-stratified estimates of SARS-CoV-2 prevalence are provided in Figure (2). The shape of the epidemic trajectory was broadly consistent across age groups, though it varied in magnitude. Prevalence was consistently highest in those aged between 35 to 44 years, with those aged 18 to 34 years a close second. For those aged 45 and above, the prevalence decreased as the group age increased. Supplementary Table (2) provides estimates of when the prevalence peaked, and the corresponding peak prevalence value.

Location-stratified estimates of SARS-CoV-2 prevalence are provided in Figure (3), with the corresponding peak prevalence dates and values provided in Supplementary Table (3). There was limited variation in the prevalence between different regions. The highest peak prevalence of 5.37% (95% CI: 4.60 to 6.27) occurred in London, and the North East and Scotland had the lowest peak prevalence of 3.92% (95% CI: 3.17 to 4.70) and 3.92% (95% CI: 3.28 to 4.60) respectively.

Sex-stratified prevalence is provided in Supplementary Figure (1). Overall, there were negligible difference in prevalence between the sexes. Prevalence peaked at 4.59% (95% CrI: 3.95% to 5.32%) in females, and at 4.47% (95% CrI: 3.83% to 5.18%) in males.

Supplementary Figures (2-4) plot the raw sample positivity against the poststratified estimates of positivity and prevalence. Differences between the sample positivity and the poststratified positivity are due to the reweighting to adjust for over/under representation, and the correction for the window effects. The posterior estimates of the window effects are provided in Supplementary Figure (5) and highlights how individuals who tested earlier in the window were more likely to test positive.

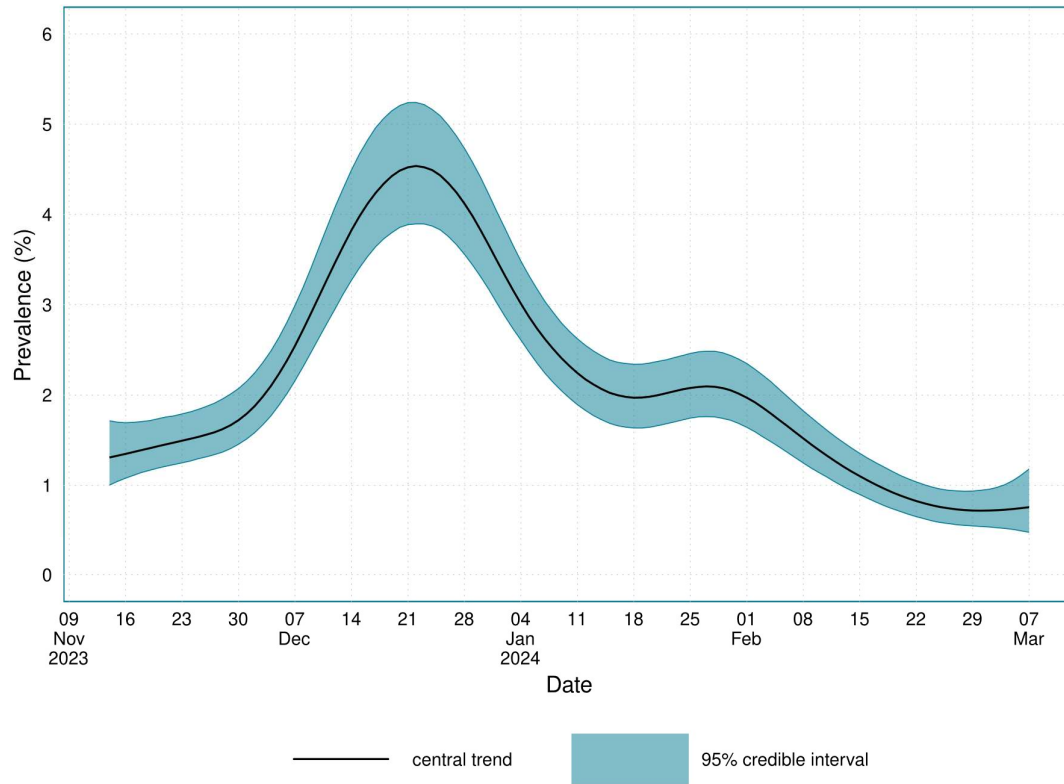


Figure 1: The prevalence of SARS-CoV-2 in England and Scotland between the 14th November 2023 and the 7th March 2024.

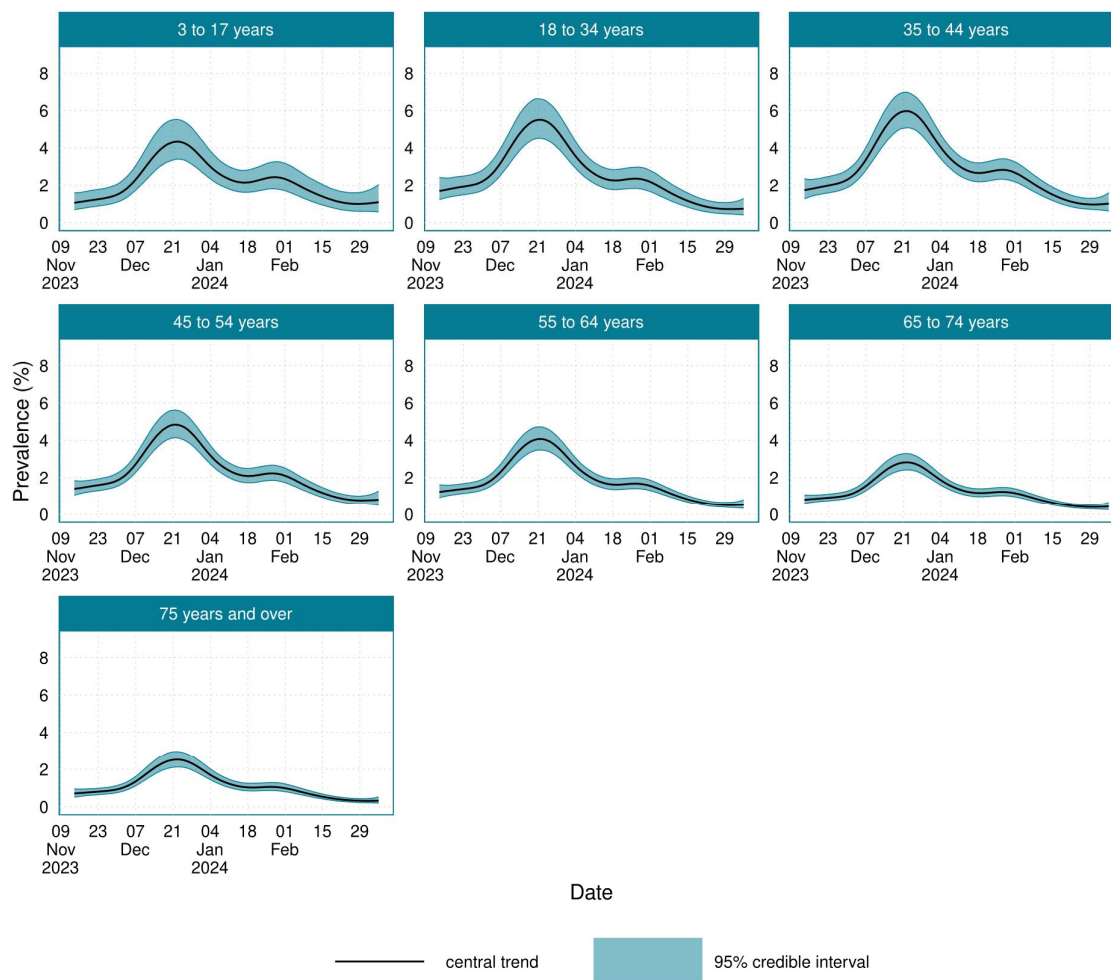


Figure 2: The age-stratified prevalence of SARS-CoV-2 in England and Scotland between the 14th November 2023 and the 7th March 2024.

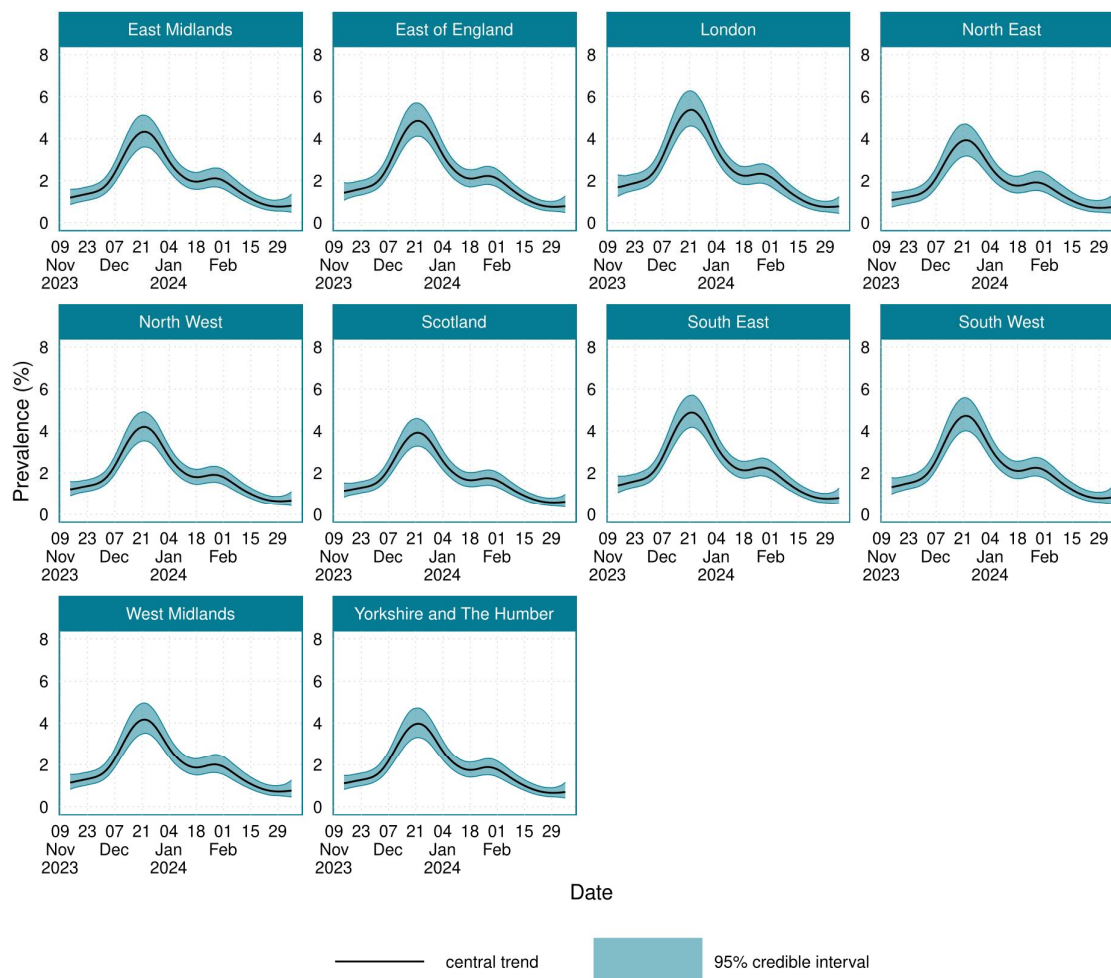


Figure 3: The location-stratified prevalence of SARS-CoV-2 in England and Scotland between the 14th November 2023 and the 7th March 2024.

Incidence

The estimated incidence for England and Scotland is presented in Figure (4). In England and Scotland, incidence peaked on the 17th December 2023 at 498 (95% CrI: 429 to 585) new infections per 100,000 individuals per day. The results suggest that after the main epidemic peak had occurred, the incidence rate increased slightly around mid-January 2024, which would explain the plateau in prevalence that was observed during this period.

The age-stratified incidence is plotted in Figure (5), and the corresponding timing and values of peak incidence are provided in Supplementary Table (4). In general, throughout the study period incidence was highest in the 35 to 44 years age group, with a peak incidence of 643 (95% CI: 540 to 763) new infections per 100,000 people per day, and lowest in those aged 75 years and over, with a peak incidence of 249 (95% CI: 206 to 299) new infections per 100,000 people per day. Around mid-January 2024, after the primary epidemic peak, an increase in the incidence rate is observed in those aged under 54 years. The location-stratified estimates of incidence are provided in Figure (6), and the timings and values of minimum and maximum incidence are provided in Supplementary Table (5). There was limited variation in incidence across different locations. The sex-stratified incidence is plotted in Supplementary Figure (3), again showing negligible differences between the sexes.

Incidence peaked in females at 503 (95% CrI: 428 to 593) new infections per 100,000 individuals per day, and in males peaked at 492 (95% CrI: 418 to 579).

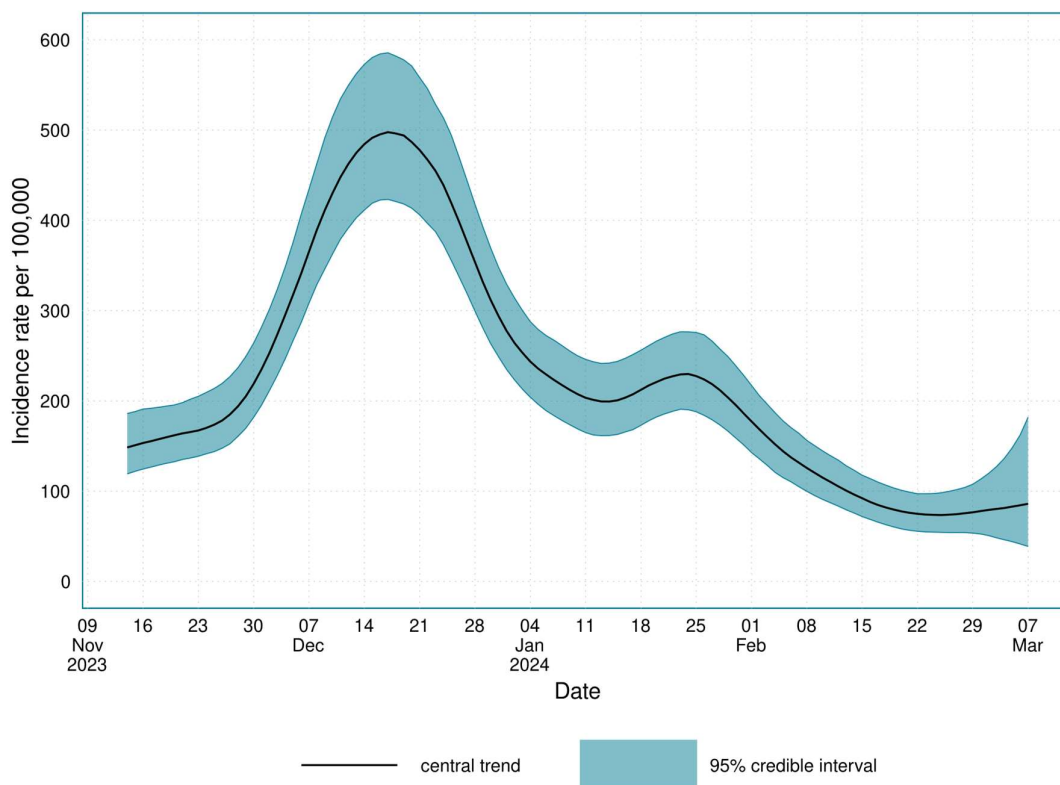


Figure 4: The incidence of SARS-CoV-2 infections in England and Scotland from the 14th November 2023 to the 7th March 2024.

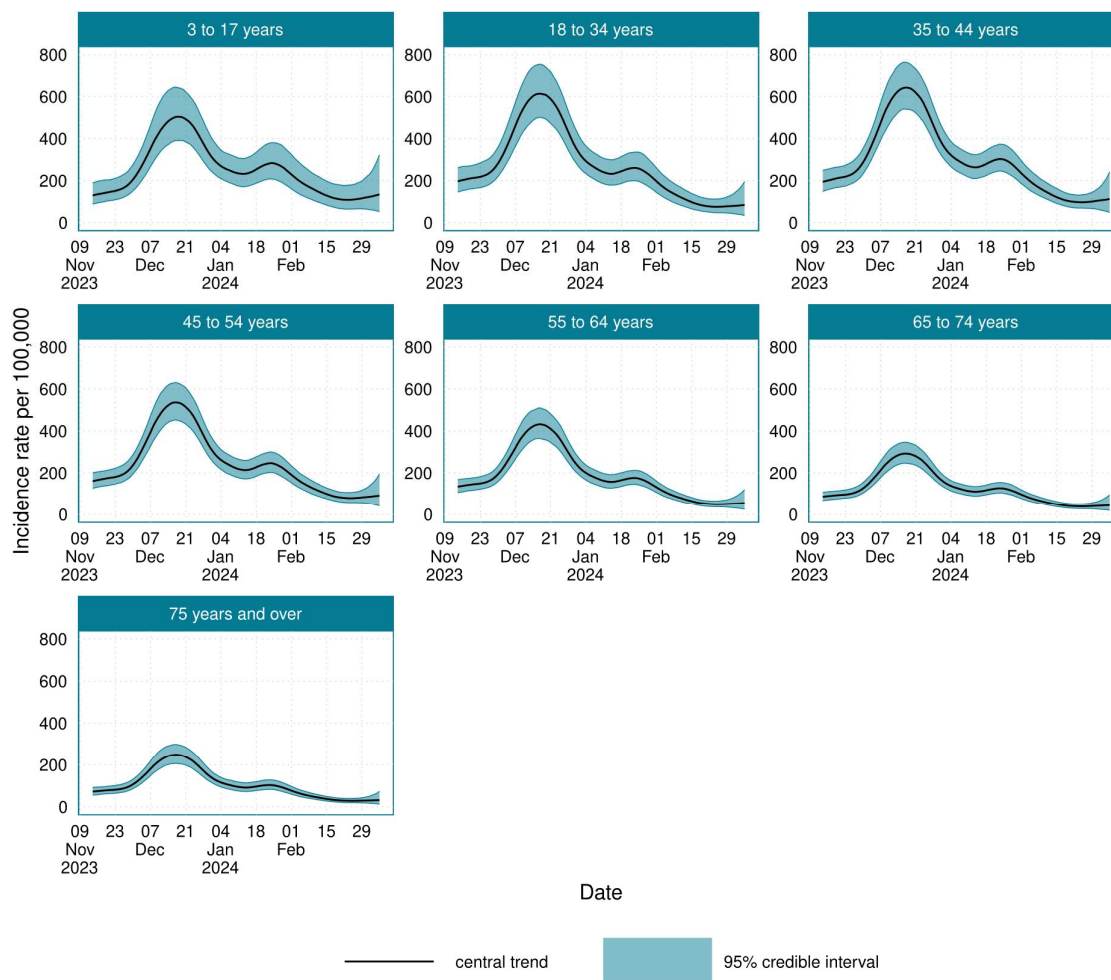


Figure 5: The age-stratified incidence of SARS-CoV-2 infections in England and Scotland from the 14th November 2023 to the 7th March 2024.

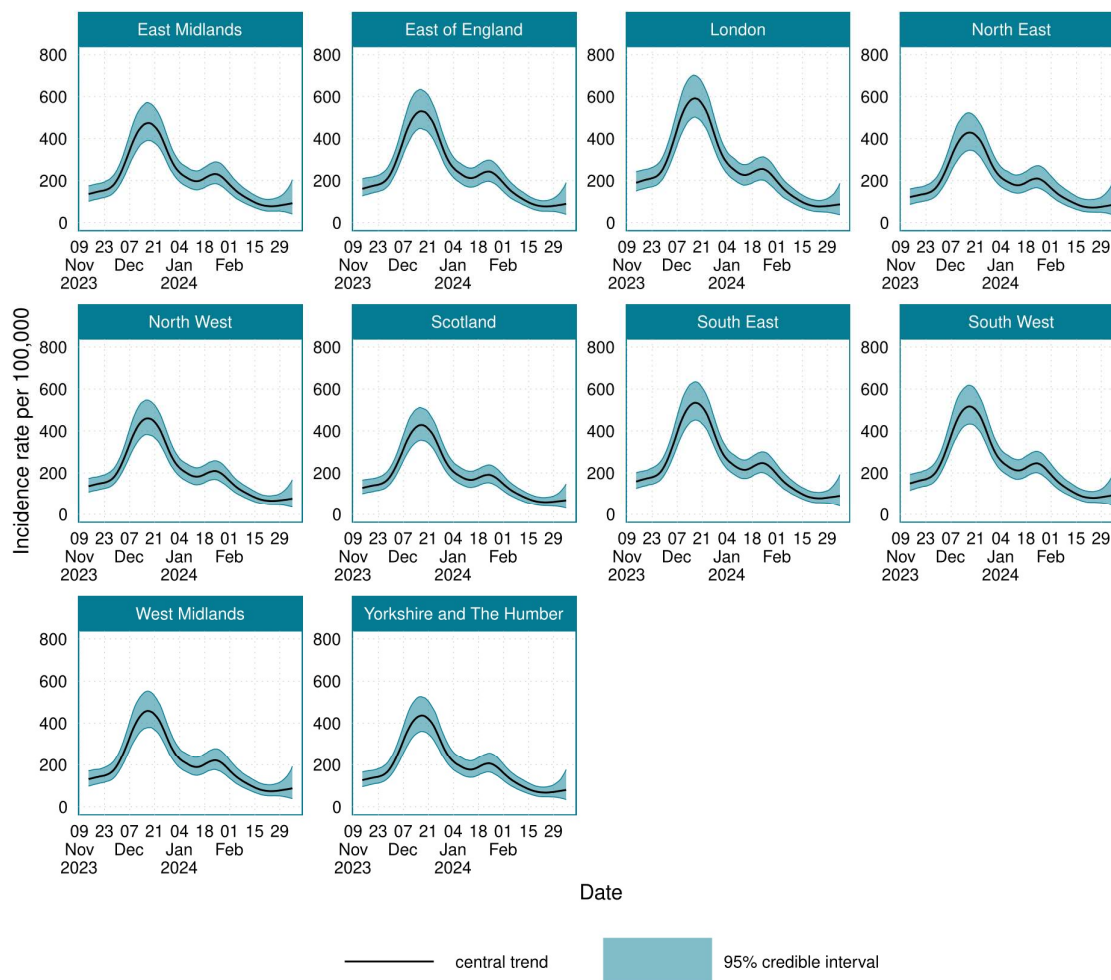


Figure 6: The location-stratified incidence of SARS-CoV-2 infections in England and Scotland from the 14th November 2023 to the 7th March 2024.

Average number of infections per individual

The average number of infections per individual that occurred over the whole the study period is presented in Figure (7). For England and Scotland, during the study period there was on average 0.258 (95% CI: 0.224 to 0.298) infections per individual. The average number of infections per individual was highest in those aged 35 to 44 years at 0.338 (95% CI: 0.288 to 0.395) infections per individual, and lowest in those aged 75 years and over with 0.122 (95% CI: 0.103 to 0.143) infections per individual. The location with the most infections per individual was London with 0.301 (95% CI: 0.258 to 0.35) infections per individual, and the location with the least infections per individual was Scotland with 0.219 (95% CI: 0.184 to 0.259) infections per individual. There was little difference in the average number of infections per individual between the sexes. The time-varying average number of infections per individual over the course of the study is plotted in Supplementary Figures (4-7).

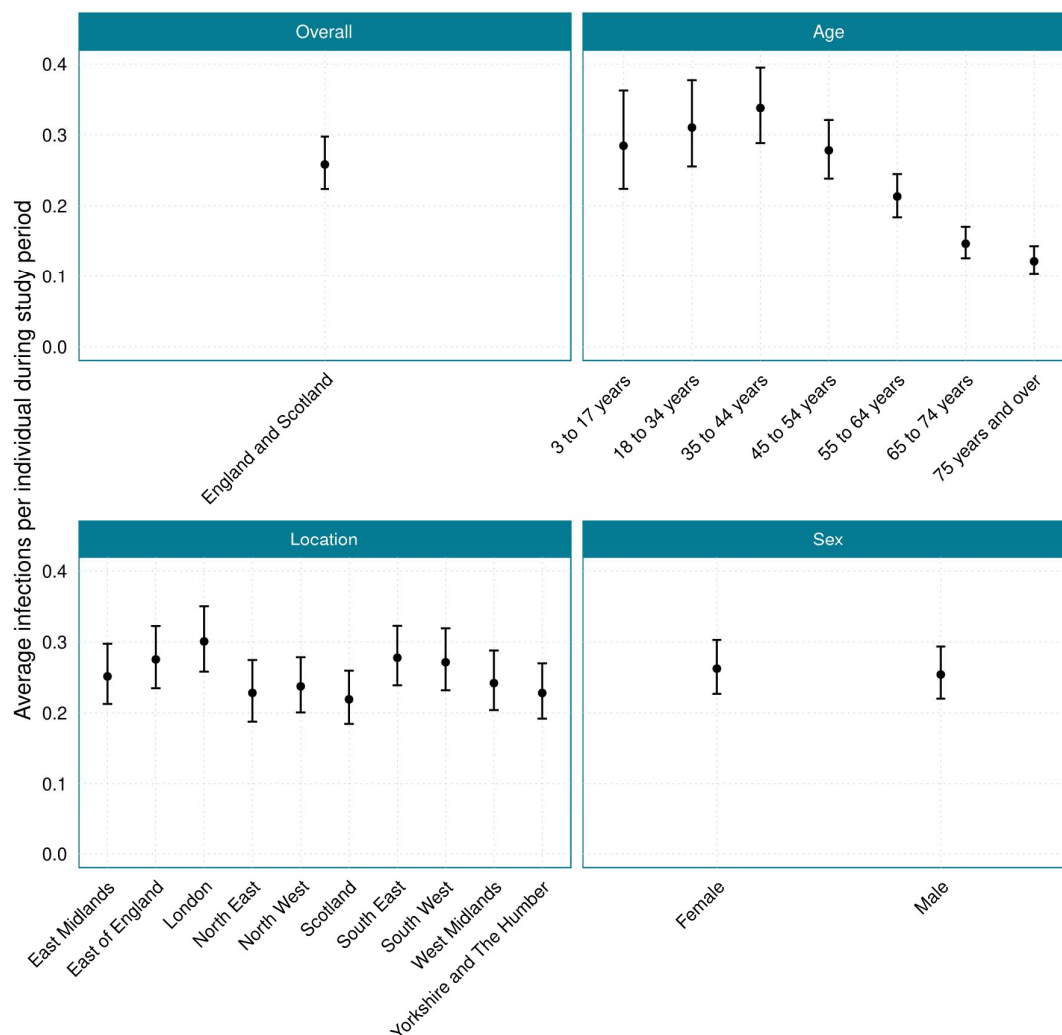


Figure 7: The average number of infections that occurred per individual during the study period, 14th November 2023 to 7th March 2024, for England and Scotland and when stratified by age, location, and sex.

Sensitivity

The population average test sensitivity for England and Scotland combined is provided in Figure (8), alongside the prevalence growth rate for England and Scotland. During the study period, the average test sensitivity in England and Scotland was 72.1% (95% CrI: 70.3, 74.0), however, due to epidemic phase bias the average test sensitivity in England and Scotland varied substantially over the course of the study. In England and Scotland, the maximum value of the population average test sensitivity was 77.2% (95% CrI: 75.3 to 79.2), which occurred on the 9th December 2023, and the minimum population average test sensitivity value was 68.6% (95% CrI: 66.4 to 70.7), which occurred on the 4th January 2024. The population average test sensitivity was strongly correlated with epidemic growth rate, and the maximum test sensitivity value occurred when the epidemic growth rate in England and Scotland was approximately 0.06 day⁻¹, and the minimum test sensitivity occurred when the epidemic growth rate in England and Scotland was approximately -0.05 day⁻¹.

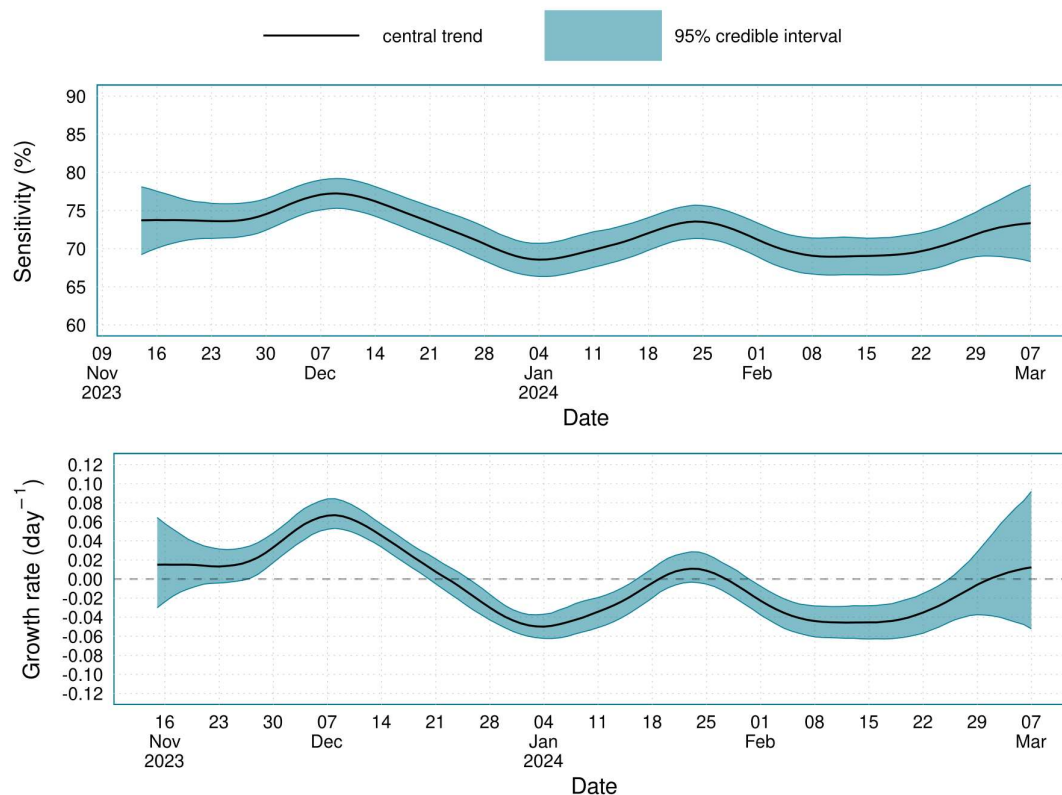


Figure 8: Estimates of average LFD sensitivity and the incidence exponential growth rate for England and Scotland from the 14th November 2023 to the 7th March 2024.

The age-stratified population average test sensitivity value is shown in Figure (9), and in a summary of the average, maximum and minimum population average test sensitivity is provided in Table (6). The average test sensitivity for those aged 3 to 17 years was substantially lower than the age groups at 62.0% (95% CI: 55.8 to 67.9). The average test sensitivity was broadly consistent across those aged 18-74 years, with those aged 75 years and over having the highest average test sensitivity at 79.2% (95% CI: 76.9 to 81.3). The location-stratified average test sensitivity over time is provided in Supplementary Figure (1), with regions following the overall trend in England Scotland closely. The average, minimum and maximum test sensitivity for each location is provided in Supplementary Table (6).

To understand the importance of accounting for time-varying LFD test sensitivity when modelling prevalence, the expected error incurred from not adjusting for time-varying test sensitivity is provided in Figure (10). While the actual variations in the test sensitivity are relatively minor, as the reciprocal sensitivity is used to obtain prevalence from positivity these small differences could result in potentially large errors. In those aged 3 to 17 years, where the average test sensitivity was lowest, the error incurred from using a constant estimate of test sensitivity exceeded 7% during the study. In those aged 75 years and over, where average test sensitivity is highest, the relative error in prevalence incurred from a constant estimate of test sensitivity occasionally exceeded 4%.



Figure 9: The age-stratified LFD test sensitivity in England and Scotland from the 14th November 2023 to the 7th March 2024.

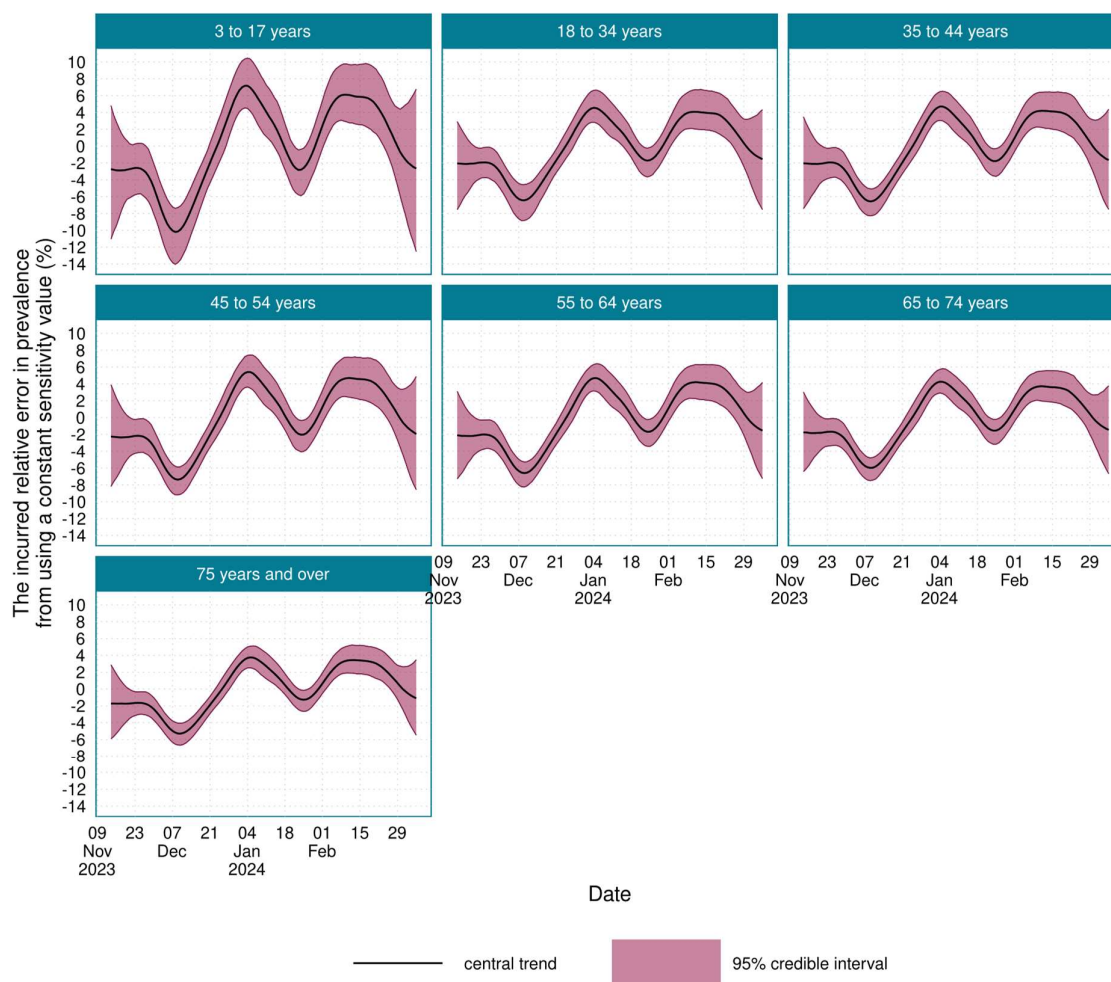


Figure 10: The relative error in the prevalence estimate, $Error_{a,s,l}(t)$, that is incurred from using a constant value of LFD test sensitivity, rather than the time-varying value of LFD test sensitivity that was estimated by the model.

Discussion/conclusion

This study demonstrated that LFD tests, with the appropriate study design and statistical adjustment, can be used to develop robust estimates of incidence and prevalence. This results in a community prevalence study design that is far more cost effective than earlier population wide PCR based studies, such as CIS and REACT, that were rapidly stood up in response to the public health emergency caused by SARS-CoV-2. As such, the design of this study is more suited to the current context where SARS-CoV-2 is no longer a public health emergency, though it continues to cause substantial morbidity and mortality.

Between the 14th November 2023 and the 7th March 2024 SARS-CoV-2 prevalence reached a peak value of 4.54% (95% CrI: 3.90% to 5.24%) in England and Scotland on 22nd December 2023. After the

primary epidemic wave, there is some evidence of a short period of positive epidemic growth during mid-January 2024. This roughly coincides with the return to work and school after the winter holiday period in the UK. The variant JN.1 grew rapidly in the UK during the study period [20], which may have driven some of the epidemic wave. However, it is challenging to disaggregate the relative roles of variant pressure compared to expected increases in incidence due to seasonal mixing patterns.

There was limited spatial variation in the incidence, prevalence, and sensitivity. Though incidence/prevalence was slightly higher in London and slightly lower in the North East of England and Scotland. Previous research has demonstrated that there was substantial SARS-CoV-2 spatial heterogeneity early in the SARS-CoV-2 pandemic, however this heterogeneity later disappeared, likely due to changes in mixing behaviours and shifting immunological dynamics [21]. Differences in incidence/prevalence between the sexes were negligible. Prevalence and incidence varied across different age groups, with the highest incidence/prevalence occurring in young adults, and the lowest incidence/prevalence occurring in those aged 75 years or over. During the study period, an individual aged 35 to 44 years would have experienced on average 0.338 (95% CI: 0.288 to 0.395) infections, whereas an individual aged 75 years or over would have experienced on average 0.122 (95% CI: 0.103 to 0.143) infections. This likely results from the higher contact rates in younger age groups leading to higher levels of transmission [22].

The model structure used a convolution approach to obtain the various quantities of interest as convolutions of the incidence. This enabled estimating a time-varying test sensitivity value that accounts for the recent epidemic dynamics and could therefore adjust for the altered LFD test sensitivity during periods of rapid epidemic growth or decline. During the study period, the average test sensitivity was 72.1% (95% CI: 70.3 to 74.0), with variation between a maximum value of 77.2% (95% CI: 75.3 to 79.2) and a minimum value of 68.6% (95% CI: 66.4 to 70.7). The average test sensitivity was lowest in 3 to 17 years age group. This is likely due to a difference in the viral load dynamics in those aged between 3 to 17 years. The average LFD test sensitivity values are consistent with the results from Eyre et. al. [10].

The results demonstrated it is important to account for time-varying test sensitivity in LFD-based SARS-CoV-2 prevalence studies, as using a constant test sensitivity value would have incurred an absolute relative error in the estimated prevalence of up to 7% in the youngest age groups during this study. If time-varying test sensitivity is not accounted for, then prevalence would be over-estimated during periods of rapid epidemic growth, and under-estimated during period of rapid epidemic decline. This is concerning, as periods of rapid epidemic growth/decline is when accurate surveillance is particularly important.

The exponential growth rate of the 23/24 winter SARS-CoV-2 wave peaked at around 0.06 day^{-1} , however substantially higher growth rates were observed during the early pandemic [23, 24]. Adjusting for the effect of epidemic phase bias on LFD test sensitivity is therefore an important consideration for any future prevalence studies that use LFD tests, particularly if the epidemic wave is growing rapidly. LFD-based prevalence studies for other pathogens may be more impacted by epidemic phase bias in the test sensitivity due to the different viral load dynamics of those pathogens [25].

Multilevel regression and poststratification (MRP) was used to produce representative estimates. This was particularly important given the study sample population was strongly biased towards older age groups, likely a consequence of study participants not being compensated unlike the previous CIS study. Study designs where participants are volunteers allows for an increased sample size due to reduced costs per participant, when compared to compensated individuals, and therefore may be

preferable if the non-representativeness of the sample can be fully accounted for. Despite the under-representation of some age groups, it was possible to produce estimates that parsimoniously explained the observed data for each stratum of the study population and adjusted for over/under representation in the sample.

Limitations

As part of the study design, individuals were provided with a 10-day window to report their test result within, however individuals who tested earlier in the window were observed to be more likely to return a positive result. Therefore, it was necessary to adjust for this to remove unintended daily effects induced by the overlapping test windows schedule. Likely, this was symptomatic individuals wanting to use their LFD test and waiting until the testing window was open. While this unintended behaviour was corrected for in the model, there remains open questions regarding the ideal length of the testing window. A shorter testing window may reduce study participation from individuals who do not have currently have symptoms and therefore cause participation bias, whereas a longer testing window allowed for modelling the testing window phenomenon.

Given the limited spatial variation, future SARS-CoV-2 community prevalence studies may consider reducing the amount of sampling required by focussing on obtaining a representative age-stratified sample, rather than a geographically stratified sample. Study participants were originally provided with 4 LFD tests as part of the main survey, and an additional 10 LFD tests for use in repeat testing. The repeat testing regime therefore added substantial cost to the study, however, it was necessary to accurately estimate SARS-CoV-2 prevalence and incidence while adjusting for time-varying LFD test sensitivity. While participation in the main study was consistently high, uptake and completion of the repeat testing regime was lower, and this area of the study design may benefit from further incentivisation, given that a large number of tests were provided for use in repeat testing.

Conclusion

The novel study design of WCIS addressed several key challenges faced in previous SARS-CoV-2 prevalence studies and successfully demonstrated the use of LFD tests in respiratory pathogen community surveillance. The use of LFD tests, rather than PCR tests, substantially reduced the cost of performing the study, making community prevalence studies more feasible in the current context, as the SARS-CoV-2 epidemic is no longer a public health emergency [26]. However, the use of LFD tests introduces several challenges to community prevalence studies, such as needing to accurately account for time-varying test sensitivity. To overcome these challenges, a novel study design that facilitated efficient parameter estimation and a modelling approach that accounted for the complex relationships between the different quantities of interest was developed. Future community prevalence studies should explore the use of multiplex LFD tests [25], which simultaneously test for a panel of different pathogens using a single swab. A community prevalence study design using the methods and design of WCIS with multiplex LFD tests would provide cost-effective and accurate surveillance of many different pathogens.

Ethics Approval

The study received ethical approval from the National Statistician's Data Ethics Advisory Committee.

Transparency Declaration

The lead author affirms that the manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Conflict of Interest

The authors have declared that no competing interests exist. There was no financial support for this work, completed as part of the author's employment.

Patient and Public Participation

Patient and public participation was not deemed necessary for this research beyond the routine engagement of survey participants.

Acknowledgements

The authors would like to thank the UKHSA Surveillance and Immunity team and ONS WCIS analysis team for their contributions.

Data Availability Statement

UKHSA operates a robust governance process for applying to access protected data that considers:

- the benefits and risks of how the data will be used
- compliance with policy, regulatory and ethical obligations
- data minimisation
- how the confidentiality, integrity, and availability will be maintained
- retention, archival, and disposal requirements
- best practice for protecting data, including the application of 'privacy by design and by default', emerging privacy conserving technologies and contractual controls

Access to protected data is always strictly controlled using legally binding data sharing contracts.

UKHSA welcomes data applications from organisations looking to use protected data for public health purposes.

To request an application pack or discuss a request for UKHSA data you would like to submit, contact DataAccess@ukhsa.gov.uk.

Bibliography

- [1] UK Health Security Agency, "COVID-19," 04 09 2024. [Online]. Available: <https://ukhsa-dashboard.data.gov.uk/topics/covid-19>.
- [2] UK Health Security Agency, "Surveillance of Influenza and other seasonal respiratory viruses in the UK, winter 2023 to 2024," 6 September 2024. [Online]. Available: <https://www.gov.uk/government/statistics/surveillance-of-influenza-and-other-seasonal-respiratory-viruses-in-the-uk-winter-2023-to-2024/surveillance-of-influenza-and-other-seasonal-respiratory-viruses-in-the-uk-winter-2023-to-2024#vaccination>.
- [3] D. Planas, I. Staropoli, V. Michel, F. Lemoine, F. Donati, M. Prot, F. Porrot, F. Guivel-Benhassine, B. Jeyarajah, A. Brisebarre, O. Dehan, L. Avon, W. H. Bolland, M. Hubert, J. Buchrieser, T. Vanhoucke, P. Rosenbaum, D. Veyer, H. Péré, B. Lina, S. Trouillet-Assant, L. Hocqueloux, T. Prazuck, E. Simon-Loriere and O. Schwartz, "Distinct evolution of SARS-CoV-2 Omicron XBB and BA.2.86/JN.1 lineages combining increased fitness and antibody evasion," *Nature Communications*, vol. 15, no. 2254, 2024.

- [4] P. Elliot, M. Whitaker, D. Tang, O. Eales, N. Steyn, B. Bodinier, H. Wang, J. Elliott, C. Atchinson, D. Ashby, W. Barclay, G. Taylor, A. Darzi, G. S. Cooke, H. Ward, C. A. Donnelly, S. Riley and M. Chadeau-Hyam, "Design and Implementation of a National SARS-CoV-2 Monitoring Program in England: REACT-1 Study," *American Journal of Public Health*, vol. 113, no. 5, pp. 545-5544, 2023.
- [5] H. Ward, C. Atchinson, M. Whitaker, B. Davies, D. Ashby, A. Darzi, M. Chadeau-Hyam, S. Riley, C. A. Donnelly, W. Barclay, G. S. Cooke and P. Elliott, "Design and Implementation of a National Program to Monitor the Prevalence of SARS-CoV-2 IgG Antibodies in England Using Self-Testing: The REACT-2 Study," *American Journal of Public Health*, vol. 113, no. 11, pp. 1201-1209, 2023.
- [6] B. P. Koen, T. House, E. Pritchard, J. V. Robotham, P. J. Birrell, A. Gelman, K.-D. Vihta, N. Bowers, I. Boreham, H. Thomas, J. Lewis, I. Bell, J. I. Bell, J. N. Newton, J. Farrar, I. Diamond, P. Benton, A. S. Walker and COVID-19 Infection Survey Team, "Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus Infection Survey," *Lancet Public Health*, vol. 6, no. 6, pp. e30-e38, 2020.
- [7] C. E. Overton, M. Fyles, J. Mellor, R. S. Paton, A. M. Phillips, A. Glaser, A. Charlett and T. Ward, "Epidemiological Parameters of SARS-CoV-2 in the UK during the 2023/2024 Winter: A Cohort Study," *medRxiv*, 2024.
- [8] A. Wolf, J. Hulmes and S. Hopkins, "Lateral flow device specificity in phase 4 (post-marketing) surveillance," 10 March 2021. [Online]. Available: <https://www.gov.uk/government/publications/lateral-flow-device-specificity-in-phase-4-post-marketing-surveillance>.
- [9] T. Ward, M. Fyles, A. Glaser, R. S. Paton, W. Ferguson and C. E. Overton, "The real-time infection hospitalisation and fatality risk across the COVID-19 pandemic in England," *Nature Communications*, 2024.
- [10] D. W. Eyre, M. Futschik, S. Tunkel, J. Wei, J. Cole-Hamilton and R. Saquib, "Performance of antigen lateral flow devices in the UK during the alpha, delta, and omicron waves of the SARS-CoV-2 pandemic: a diagnostic and observational study," *The Lancet Infectious Diseases*, pp. 922-932, 2023.
- [11] J. A. Hay, L. Kennedy-Schaffer, S. Kanjilal, N. J. Lennon, B. S. Gabriel, M. Lipsitch and M. J. Mina, "Estimating epidemiologic dynamics from cross-sectional viral load distributions," *Science*, vol. 373, 2021.
- [12] M. Makhoul, F. Abou-Hijleh, S. Seedat, G. R. Mumtaz, H. Chemaitelly, H. Ayoub and L. J. Abu-Raddad, "Analyzing inherent biases in SARS-CoV-2 PCR and serological epidemiologic metrics," *BMC Infectious Disease*, 2022.
- [13] O. Eales and S. Riley, "Differences between the true reproduction number and the apparent reproduction number of an epidemic time series," *Epidemics*, vol. 46, 2024.
- [14] F. Lindgren and H. Rue, "On the Second-Order Random Walk Model for Irregular Locations," *Scandinavian Journal of Statistics*, vol. 35, pp. 691-700, 2008.
- [15] H. Rue and L. Held, *Gaussian Markov Random Fields*, 2005.
- [16] D. K. Park, A. Gelman and J. Bafumi, "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls," *Political Analysis*, 2017.
- [17] R. J. A. Little, "Post-Stratification: A Modeler's Perspective," *Journal of the American Statistical Association*, 1993.
- [18] Office for National Statistics, "Population projections," 2024. [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections>.
- [19] J. Gabry, R. Češnovar and A. Johnson, *cmdstanr: R Interface to 'CmdStan'*, 2023.

- [20] UK Health Security Agency, "SARS-CoV-2 genome sequence prevalence and growth rate update: 21 February 2024," 21 February 2024. [Online]. Available: <https://www.gov.uk/government/publications/sars-cov-2-genome-sequence-prevalence-and-growth-rate/sars-cov-2-genome-sequence-prevalence-and-growth-rate-update-21-february-2024>.
- [21] T. Ward, M. Mitzi, A. Gelman, B. Carpenter, W. Ferguson, C. Overton and M. Fyles, "Bayesian spatial modelling of localised SARS-CoV-2 transmission through mobility networks across England," *PLOS Computational Biology*, 2023.
- [22] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. Scalia Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, J. W. Edmunds and M. Rosinska, "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases," *PLOS Medicine*, 2008.
- [23] L. Pellis, F. Scarabel, H. B. Stage, C. E. Overton, L. H. K. Chappell, E. Fearon, E. Bennett, K. A. Lythgoe, T. A. House, I. Hall and University of Manchester COVID-19 Modelling Group, "Challenges in control of COVID-19: short doubling time and long delay to effect of interventions," *Philosophical Transaction of the Royal Society B*, 2021.
- [24] UK Health Security Agency, "The R value and growth rate," 23 December 2022. [Online]. Available: <https://www.gov.uk/guidance/the-r-value-and-growth-rate>.
- [25] M. Fyles, C. E. Overton, T. Ward, E. Bennett, T. Fowler and I. Hall, "Modelling multiplex testing for outbreak control," *Journal of Infection*, 2024.
- [26] World Health Organisation, "Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic," 14 10 2024. [Online]. Available: [https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic).
- [27] Office for National Statistics, "Coronavirus (COVID-19) Infection Survey, UK: 24 March 2023," 24 March 2023. [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/24march2023>.
- [28] Office for National Statistics, "Winter Coronavirus (COVID-19) Infection Study Quality and Methods Information," 24 June 2024. [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/wintercoronaviruscovid19infectionstudyqmi>. [Accessed 29 July 2024].
- [29] Office for National Statistics, "Winter Coronavirus (COVID-19) Infection Study," [Online]. Available: <https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/wintercoronaviruscovid19infectionstudy>. [Accessed 29 July 2024].