



Reconstructing the first COVID-19 pandemic wave with minimal data in England

Siyu Chen^{a,*}, Jennifer A. Flegg^b, Katrina A. Lythgoe^{c,d}, Lisa J. White^{d,*}

^a Department of Public and Ecosystem Health, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

^b School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia

^c Big Data Institute, Nuffield Department of Medicine, University of Oxford, United Kingdom

^d Department of Biology, University of Oxford, United Kingdom

ARTICLE INFO

Keywords:

COVID-19
Exposure
Case detection rate
England
Serological data

ABSTRACT

Accurate measurement of exposure to SARS-CoV-2 in the population is crucial for understanding the dynamics of disease transmission and evaluating the impacts of interventions. However, it was particularly challenging to achieve this in the early phase of a pandemic because of the sparsity of epidemiological data. We previously developed an early pandemic diagnostic tool that linked minimum datasets: seroprevalence, mortality and infection testing data to estimate the true exposure in different regions of England and found levels of SARS-CoV-2 population exposure to be considerably higher than suggested by seroprevalence surveys. Here, we re-examine and evaluate the model in the context of reconstructing the first COVID-19 epidemic wave in England from three perspectives: validation against the Office for National Statistics (ONS) Coronavirus Infection Survey, relationship among model performance and data abundance and time-varying case detection ratios. We find that our model can recover the first, unobserved, epidemic wave of COVID-19 in England from March 2020 to June 2020 if two or three serological measurements are given as additional model inputs, while the second wave during winter of 2020 is validated by estimates from the ONS Coronavirus Infection Survey. Moreover, the model estimates that by the end of October in 2020 the UK government's official COVID-9 online dashboard reported COVID-19 cases only accounted for 9.1 % of cumulative exposure, dramatically varying across the two epidemic waves in England in 2020, 4.3 % vs 43.7 %.

1. Introduction

The COVID-19 pandemic inflicted devastating effects on global populations and economies (Aburto et al., 2022; Ozili and Arun, 2023) and is now still affecting countries in many different ways. Reviewing the challenges posed by the COVID-19 pandemic and evaluating previous responses is vitally important for future pandemic preparedness (Metcalf et al., 2020; Aguas et al., 2020; Pagel and Yates, 2022; Bollyky et al., 2022). Accurate estimation of exposure remains crucial for understanding the dynamics of disease transmission and assessing the impacts of interventions at the different stages of pandemic. However, this is particularly challenging in early phases since most of the characteristics of the pathogen are unknown and at the same time epidemiological data are sparse.

Confirmed COVID-19 cases were typically the first type of data to be collected and reported, mostly due to the syndrome surveillance systems

(Kennedy et al., 2022; Desjardins, 2020). However, these data tend to underestimate the true exposure in the population because of limited diagnostic capacity of, uncertainties about definition of cases, testing criteria, etc. Large-scale viral infection surveys in the community can help to solve the testing issue. For example, the UK Office for National Statistics (ONS) conducted a nation-wide COVID-19 viral testing survey, the Covid Infection Survey (CIS) (Pouwels et al., 2021) that successfully tracked the trajectories of COVID-19 infections in the community of UK starting in April of 2020. Because of its representative sampling across households in the general population this study is recognized to have the capacity to capture asymptomatic infections such as those missed by the symptomatic testing scheme in the early pandemic and could provide reliable estimates of prevalence over time (Office of National Statistics Coronavirus COVID-19 Infection Survey UK, 2023). However, sample collection only started in April 2020 so that reporting of daily incidence estimates only started in May, while the first death due to COVID-19

* Corresponding authors.

E-mail addresses: siyu.chen@cornell.edu (S. Chen), lisa.white@biology.ox.ac.uk (L.J. White).

<https://doi.org/10.1016/j.epidem.2025.100814>

Received 31 May 2024; Received in revised form 30 December 2024; Accepted 10 January 2025

Available online 14 January 2025

1755-4365/© 2025 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

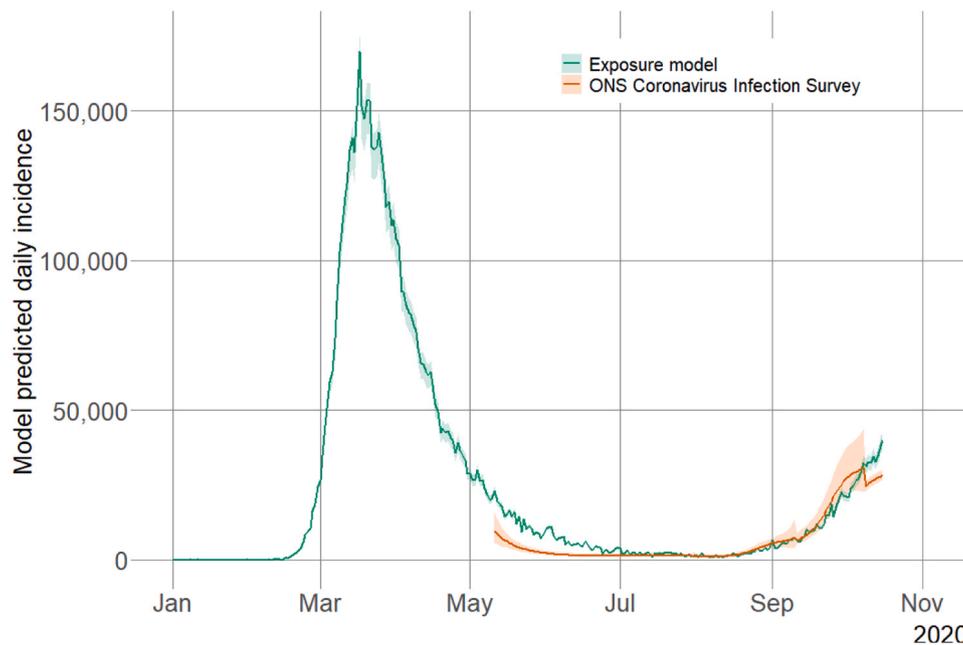


Fig. 1. Comparison of model-predicted daily incidence of SARS-CoV-2 in England. The green lines show the predictions of median daily incidence by our model (Chen et al., 2021) based on Eqs. (1) and (2) in the Materials and Methods section. The orange lines show the predictions of median daily incidence from the ONS Coronavirus Infection Survey while the orange shaded areas correspond to the 95 % CrI.

disease in the UK had been documented in February 2020 (GOV.UK, 2023). The transmission of COVID-19 in the community began months earlier than the survey, which could then not be able to recover the early epidemic curve.

Serologic studies measuring how many people have antibodies against the virus are a promising tool for pinning down the stage of the pandemic because of its ability to capture past infections regardless of clinical symptoms (Clapham et al., 2020). If the antibody elicited by the virus lasts for lifetime, representative sampling in a population followed by the antibody testing will provide robust estimates of exposure. However, cohort studies following individuals over time after they have had a known COVID-19 infection determined that antibodies are only measurable up to 6–9 months (Long et al., 2020; Ibarrondo et al., 2020; Wei et al., 2021), on average, varying across testing assays (Böger et al., 2021) and antigen types (Van Elslande et al., 2021). The immediate implication is that serological studies will inevitably under-estimate the number of people exposed, since some of these will have a low antibody count when the study is conducted and test negative. Linking multiple publicly available datasets, we previously proposed a method (Chen et al., 2021) for estimating the true level of exposure taking into account antibody decay. Here we further examine and evaluate the model in the context of reconstructing the first COVID-19 pandemic from three perspectives: validation against the ONS Infection Survey, relationship between model performance and data abundance and time-varying case detection ratios.

2. Result

2.1. Reconstruction of the early epidemic

In our previous paper (Chen et al., 2021), we presented a simple model to link together three key metrics for evaluating the progress of an epidemic, as applied to the context of SARS-CoV-2 in England: antibody seropositivity, infection incidence and number of deaths. We used these three metrics to estimate the antibody seroreversion rate and region-specific infection fatality ratios. In doing so, the cumulative number of infections in England were estimated, showing that cross-sectional seroprevalence data underestimated the true extent of

the SARS-CoV-2 epidemic in England in the early pandemic. Estimates for the IgG (spike) seroreversion rate and infection fatality ratio (IFR) is broadly consistent with other studies, which supports the validity of these findings. The modeling details are attached in Appendix A.

The model was set up based on the serendipitous observation on the COVID-19 infection timeline that seroconversion in individuals who survive occurs at approximately the same time as death for those who do not. Therefore, a simple ordinary differential equation (ODE) was formulated to model the rate of change in the number of seropositive individuals in different regions of England as increasing as new infections were generated, calculated by the daily number of deaths divided by the infection fatality ratio and that will decrease as antibody decays. The model predicted seropositive population were then fitted to observed seroprevalence using a Bayesian observation model.

2.2. Numerical convergence and efficiency assessment

Trace plots of samples versus the simulation index can be very useful in assessing convergence. Figure S1 shows the trace plot of posterior estimates for all parameters in the time-independent IFR model. The center of the chain appears to be around the mean values with very small fluctuations for each parameter (Table S2 and Figure S1). This indicates that the chain is likely to have reached the right distribution and is exploring the distribution by traversing to areas where its density is very low; the mixing is thus quite good. The summary results of other numerical diagnostic metrics (Table S2) show that \hat{R} of all posterior estimates is 1 and there are more than 100 samples per chain for ess-bulk and ess-tail. We thus conclude that the Markov Chain converged well for all parameters.

2.3. Validation from ONS infection survey

Comparing the incidence of SARS-CoV-2 in England estimated by our model with those inferred by the ONS Coronavirus Infection Survey (Fig. 1), we found that our model could reveal the first, unobserved, epidemic wave of COVID-19 in England from March 2020 to June 2020, with the second wave validated by the estimates from the ONS Infection Survey. Further, we found our model results were highly consistent with

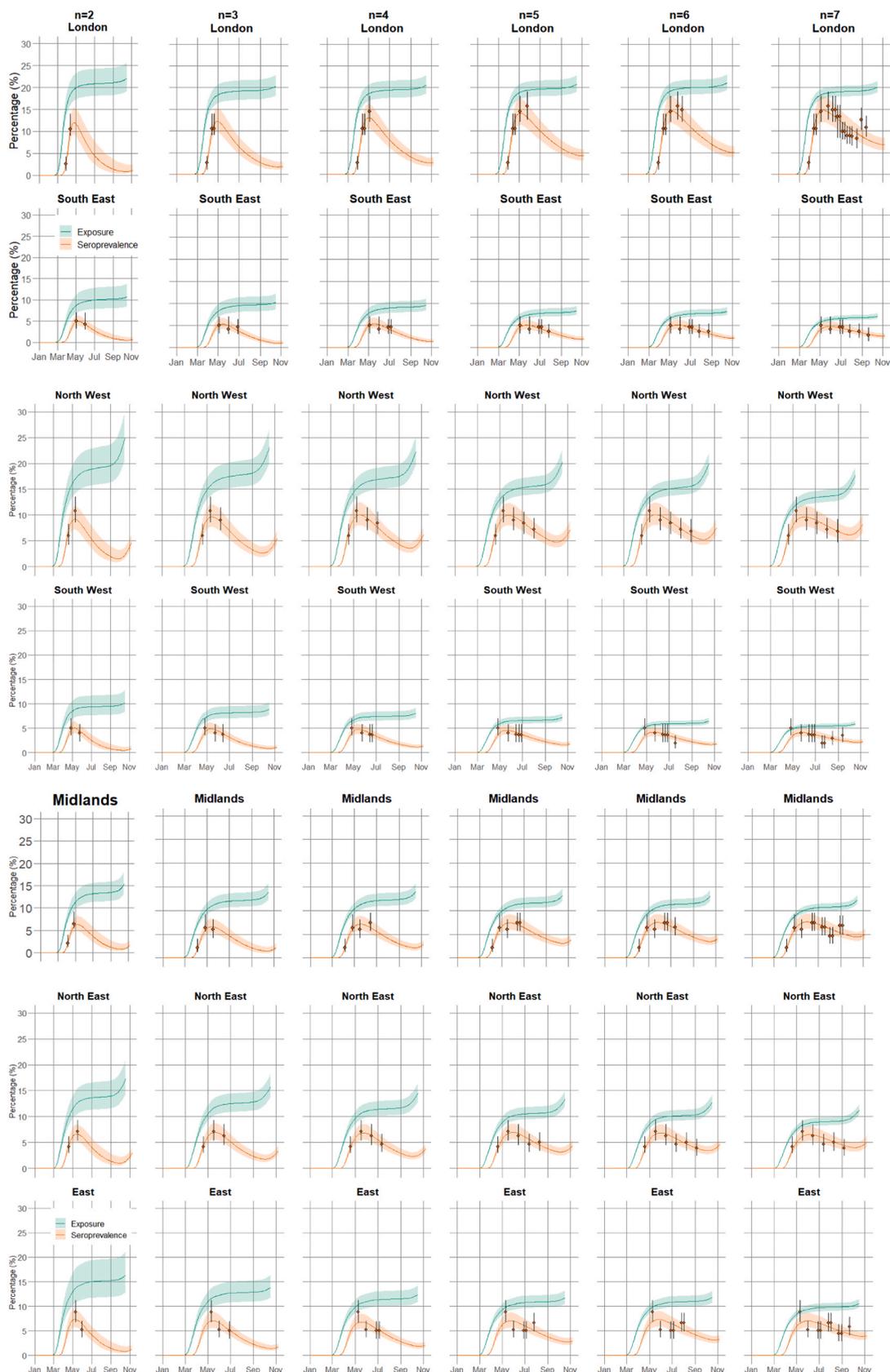


Fig. 2. Comparison of estimates of exposure in seven regions of England as more serological measurements are given as model inputs (left to right). The green and orange lines show the model predictions of median exposure and seroprevalence, respectively, while the shaded areas correspond to the 95 % CrI.

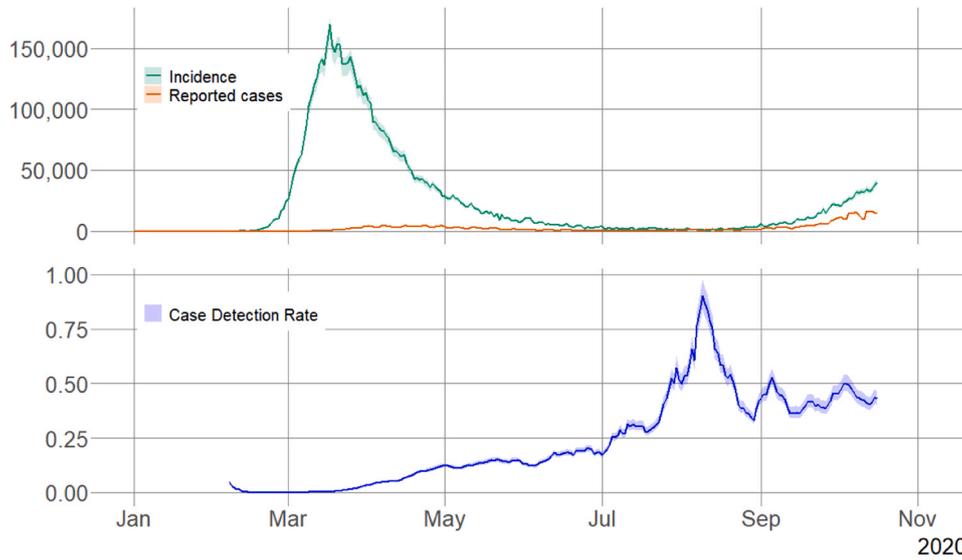


Fig. 3. Comparison between estimates of daily incidence with reported cases of SARS-CoV-2 in England and case detection ratios. Here, all serological measurements were used in the model fitting. In the top figure, the green lines show the predictions of median daily incidence by our model based on Eqs. (1) and (2) in the Materials and Methods section while the shaded areas correspond to the 95 % CrI. The red lines show the reported confirmed cases in England downloaded from GUV.UK dashboard. In the bottom figure, the blue lines show the estimates of median case reporting rate in England based on Eqs. (3) and (4) while the shaded areas correspond to the 95 % CrI.

those using SEIRS type compartmental models with time-varying force of infection (Knock et al., 2021; Russell et al., 2020).

2.4. Relationship between model performance and data abundance

We then examined the relationship between model performance and data abundance - how estimates of exposure from our model change as more serological data points are added into the fitting procedure one by one over time (Fig. 2). We found a highly robust pattern of exposure across different regions of England was estimated in general. Specifically, the model could only start estimating the quantities of interest: exposure and two parameters (infection fatality ratio and antibody decay rate) when at least two serological measurements from April to

June 2020 in each region were given as inputs. However, these estimates were already highly consistent with those when more serological measurements were added although the credible bands were wider. The wide credible bands suggested a greater uncertainty around the estimates when little information was available. When three serological measurements in each of region were included the estimates of exposure level became largely consistent with the results when all serological measurements were used. This might be attributed to the timing of these third serological measurements since then the seroprevalence in most regions started decreasing. With more and more serological measurements being added, the credible bands of estimates of exposure gradually narrowed.

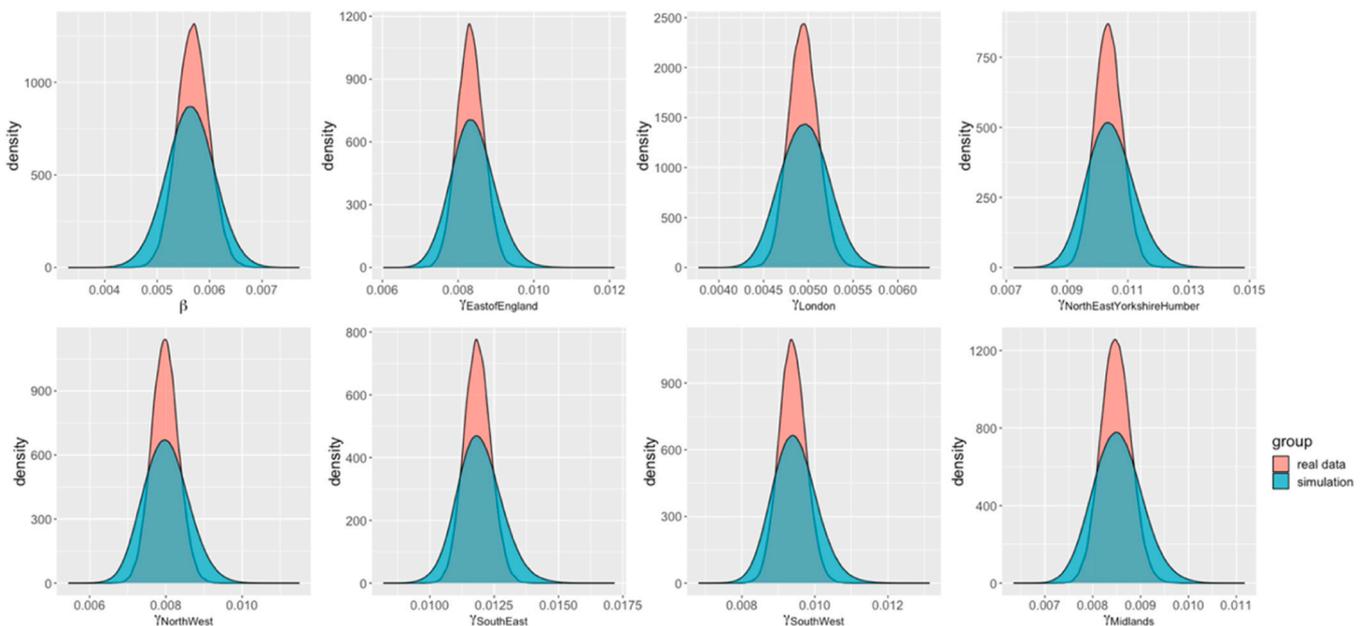


Fig. 4. Comparisons of posterior estimates generated through 1) fitting to real-world seroprevalence data colored in red and 2) fitting to simulated seroprevalence data for all parameters in time-independent IFR model colored in blue.

2.5. Time-varying case detection ratios

While comparing the reported cases with the incidence estimated by our model (Fig. 3), we found the UK government's official COVID-19 online dashboard (<https://coronavirus.data.gov.uk>) reported COVID-19 cases in England only accounted for 9.1 % (95 %CrI (8.7 %,9.8 %)) of cumulative exposure by the end of October 2020. Further, the relative size of two infection waves in England in 2020 estimated by our model, Spring wave from February to June and Autumn wave from September to November, were reversed compared those reported by the confirmed cases. The case detection rate relative to the total exposure was also dramatically different in these two-epidemic waves. By separating the two waves before and after the first of August 2020, we found during January 2020 to August 2020 the case detection rate was only 4.3 % (95 %CrI (4.1 %, 4.6 %)) which increased to 43.7 % (95 %CrI (40.7 %, 47.3 %)) during August 2020 to October 2020, highlighting the dominant effect of testing effort in shaping the case curve in the early stage of the pandemic. The testing issue, e.g. the limited capacity of tests and the symptom-based testing strategy posed a major challenge for understanding the early pandemic. Viral surveys in the general population can solve the sampling issue, but still have the problem of lack of sampling early on. Serological data from some convenient samples, e.g., blood donors can help to pin down the progress of the pandemic when antibody decay is teased out.

2.6. Numerical parameters identifiability assessment

After implementing the numerical simulation algorithm described in the Method, we found that all posterior estimates generated through fitting to simulated seroprevalence data for each parameter in the time-dependent IFR model have very close concentrations (median values) with the posterior estimates generated through fitting to real-world seroprevalence data (Fig. 4). In addition, the posterior estimates generated through fitting to simulated seroprevalence data generally have wider ranges, thus, wider credible intervals (Table S2).

3. Discussion

Accurate reconstruction of exposure time series is necessary to assess how policies influenced transmission over time, in particular when reporting is lagged, and multiple interventions may have been undertaken in succession. For example, (Chen et al., 2022) made use of the comparison of exposure between general population and pregnant women in New York City to conclude the effectiveness of shielding during pregnancy. Moreover, the prior exposure level in the population can be used to inform future intervention design, e.g., vaccination prioritization. For example, in the early stage of the COVID-19 vaccination campaign, when dose supply and administrative capacity were initially limited worldwide, a modeling study (Bubar et al., 2021) explored how uncertainty about previous exposure levels and about a vaccine's characteristics affects the prioritization strategies for reducing deaths and transmission. This model showed how the use of individual-level serological tests to redirect doses to seronegative individuals improved the marginal impact of each dose while potentially reducing existing inequities in COVID-19 impact.

Here, we evaluated a previously published dynamic model and demonstrated its ability to reconstruct the first epidemic wave before large-scale survey sampling by providing robust estimates of exposure over time. One key element of the model was fitting the model to serologic data generated from healthy adult blood donors supplied by the NHS Blood and Transplant (NHS BT collection) serum samples using the Euroimmun anti-spike IgG assay and reported in the Weekly National Influenza and COVID-19 Surveillance Report. This suggests that convenient samples, for example here serum samples from blood donors have the promise of providing primary information about epidemic progress in a short time frame especially during the emergence of a new

outbreak of a novel pathogen.

Because of its rigorous sampling design and robust estimation power the ONS Covid Infection Survey may be seen as the gold standard for estimating community prevalence. Our model does not take any results or estimates from the survey as inputs, so the comparison exercise that we conducted here between estimates of exposure from our model against the ONS Covid Infection Survey provides a real-world validation. Moreover, we showed that the modeling approach is a valuable early pandemic diagnostic tool and can clearly recover the first epidemic wave that the survey was unable to capture because of its late starting time. Using the inferred daily incidence, we explicitly demonstrated the variation of case detection ratios over two epidemic waves in England in 2020. This provides quantitative information for studying the association between the capacity, behavior and strategy of testing during the evolution of the epidemic and further supports the argument that confirmed cases largely underestimate the extent of disease transmission. Inspired by this idea, we inferred the daily incidence in later COVID-19 waves (October 2020 to March 2022) in Lower Tier Lower Authorities (LTLAs) of England based on a de-biased causal inference framework and assessed the unequal contributions of socioeconomic factors to the case detection ratios along the epidemics with the privilege of having access to the national COVID-19 mass testing datasets. We found that there were no notable persistent differences in case detection ratios according to deprivation level and throughout much of 2021, Black ethnic groups had higher rates of case detection; what underpins this relationship is unclear, but targeted community testing could have played a role (Bajaj et al., 2024).

Moreover, the simple structure of the model presented avoids the unnecessary complexity and structure-based uncertainty of a full dynamic model where compartmental exchanges simulating disease spread in different groups of the population, such as susceptible, exposed, infected and recovered, are developed. The exercise of studying the model performance against data abundance suggests the modeling results remain highly robust in data sparse setting that should be particularly relevant, for example, in Low- or Middle-Income Countries.

4. Materials and methods

4.1. Data sources

We used publicly available epidemiological data to conduct the analysis, as described below.

ONS estimated incidence

The Office for National Statistics (ONS) launched the Coronavirus (COVID-19) Infection Survey in England on 26 April 2020 to estimate how many people across England, Wales, Northern Ireland and Scotland would have tested positive for COVID-19 infection, regardless of whether they report experiencing symptoms that is one of the primary goals of the survey. The survey is based on a random sample of households to provide a nationally representative sample. Everyone aged 2 years and over in each household sample was asked to take a nose and throat swab for SARS-CoV-2 to be analysed using the reverse transcriptase polymerase chain reaction (RT-PCR). Every participant was swabbed once. They were then invited to have repeat tests every week for another four weeks and then monthly. More descriptions about the survey design can be found in (Coronavirus, 2020). Using a Bayesian multilevel generalized additive regression model to model the swab test result (positive or negative) as a function of age, sex, time, and region, the study estimated community prevalence of SARS-CoV-2 in England since April 2020 (Pouwels et al., 2021). Combining the estimates of community prevalence and estimates of duration of PCR testing positivity, the survey modeling team also published the estimates of daily incidence based on a deconvolution model (Coronavirus, 2020).

To conduct the comparison of estimates of incidence between our model and the ONS survey, we retrieved the SARS-CoV-2 daily incidence

Algorithm 1 numerical parameters identifiability check

- 1: Fit seromodel to real-world seroprevalence data
 - 2: Obtain marginal posterior samples for each epidemiological parameter, $\hat{\beta}_i^{real}, \hat{\gamma}_{1i}^{real}, \dots, \hat{\gamma}_{6i}^{real}$
 - 3: **for** $i = 1, 2, \dots, N$ **do**
 - 4: Generate simulated seroprevalence data from seromodel
 - 5: Fit seromodel to simulated seroprevalence data
 - 6: Obtain marginal posterior samples for each epidemiological parameter, $\hat{\beta}_i^{sim}, \hat{\gamma}_{1i}^{sim}, \dots, \hat{\gamma}_{7i}^{sim}$
 - 7: **end for**
 - 8: Compare $\hat{\beta}_i^{sim}, \hat{\gamma}_{1i}^{sim}, \dots, \hat{\gamma}_{7i}^{sim}$ with $\hat{\beta}_i^{real}, \hat{\gamma}_{1i}^{real}, \dots, \hat{\gamma}_{7i}^{real}$
-

Fig. 5. Algorithm of numerical checking for parameters identifiability.

in England in 2020 from the Office for National Statistics (ONS) ([Office of National Statistics Coronavirus COVID-19 Infection Survey UK, 2023](#)) on March 17, 2023 as shown in Fig. 1.

4.2. Model estimated exposure

Cumulative exposure to SARS-CoV-2 in seven regions of England estimated by the model that we published before were obtained from (Chen et al., 2021). Here, we firstly transformed and aggregated the cumulative exposure by region of England to daily incidence in England using Eq. (1) and Eq. (2).

4.3. 7-day average of reported COVID-19 cases in England

7-day average of reported COVID-19 daily cases in England in 2020 were retrieved from the UK government's official COVID-9 online dashboard ([GOV.UK, 2023](#)) on March 17, 2023 as shown in Fig. 3.

5. Method

5.1. Calculation of case detection ratios

We firstly calculated the incidence in England estimated by exposure model (Chen et al., 2021) by computing the difference of cumulative exposure in two successive days and adding together to the whole England as shown in Fig. 1 and Fig. 3:

$$I_i(t) = E_i(t+1) - E_i(t), \quad t = 1, 2, \dots, n, \quad i = 1, 2, \dots, 7 \quad (1)$$

$$I_{England}(t) = \sum_{i=1}^7 I_i(t) \quad (2)$$

Here, $E_i(t)$ is the daily exposure at region i estimated by exposure model (Chen et al., 2021), n is the total number of days from 1 January 2020 to 7 November 2020, $i = 1, \dots, 7$ represents London, Southwest, Southeast, Northeast, Northwest, East, Midland. $I_{England}(t)$ represents the daily incidence of England.

The 7-day average model predicted incidence can be calculated by

$$\bar{I}_{England}(t) = \frac{1}{7} \sum_{i=t-3}^{t+3} I_{England}(i), \quad t = 4, \quad 5, \dots, n-4 \quad (3)$$

Here, $t = 4$ refers to the fourth day of 2020, n is the end date of the comparison exercise, 7 November 2020.

The estimated reporting ratio as shown in Fig. 3 was calculated by

$$r(t) = \frac{\bar{I}_{England}(t)}{C(t)} \quad (4)$$

Here, C is the 7-day average reported cases in England from the UK

government's official COVID-9 online dashboard ([GOV.UK, 2023](#)).

While testing the relationship between model performance and data abundance in Fig. 2, we firstly obtained all the data and codes from paper (Chen et al., 2021) and rerun the model by adding the seroprevalence measurements one by one into the model.

5.2. Convergence and efficiency of Markov chain

We further assessed the convergence and efficiency performance for Markov Chains of the Bayesian estimations using multiple metrics, including trace plot, \hat{R} , bulk effective sample size, tail effective sample size.

The trace shows if the chain has not yet converged to its stationary distribution—that is, if it needs a longer burn-in period. A trace can also demonstrate whether the chain is mixing well. A chain might have reached stationarity if the distribution of points is not changing as the chain progresses. The aspects of stationarity that are most recognizable from a trace plot are a relatively constant mean and variance. A chain that mixes well traverses its posterior space rapidly, and it can jump from one remote region of the posterior to another in relatively few steps.

\hat{R} is a well-known index in the Bayesian convergence analysis literature that measures the between- and within- chain variances. It is a test that indicates a lack of convergence by comparing the variance between multiple chains to the variance within each chain. If the parameters are successfully exploring the full space for each chain, then $\hat{R} = 1$, since the between- chain and within- chain variance should be equal. If chains have not mixed well (i.e., the between- and within-chain estimates don't agree), \hat{R} is larger than 1. RStan reports \hat{R} which is the maximum of rank normalized split-R-hat and rank normalized folded-split-R-hat, which works for thick tailed distributions and is sensitive also to differences in scale.

MCMC samplers do not draw truly independent samples from the target distribution, which means that our samples are correlated. We can, however, measure the number of effectively independent samples we draw, which is called the effective sample size. It is a measure that combines information from the \hat{R} value with the autocorrelation estimates within the chains. ESS estimates come in two variants, ess-bulk and ess-tail. The rule of thumb for ess-bulk is for this value to be greater than 100 per chain on average. Since we ran four chains, we need ess-bulk to be greater than 400 for each parameter. The ess-tail is an estimate for effectively independent samples considering the more extreme values of the posterior. This is not the number of samples that landed in the tails of the posterior, but rather a measure of the number of effectively independent samples if we sampled the tails of the posterior. The rule of thumb for this value is also to be greater than 100 per chain on average.

5.3. Identifiability analyses

We assessed the parameters identifiability through numerical simulations (Fig. 5) where the estimation procedure was conducted twice, one on real-world seroprevalence data and another on simulated seroprevalence data. The rationale of this numerical simulation approach is that the model parameterized by the estimates from real-world data becomes a fully known ‘oracle’ model and its parameters can accordingly be seen as ‘oracle’ parameters. Any estimates generated from this ‘oracle’ model could then be checked through comparing with these oracle parameters. To test whether these two sets of parameter posteriors are similar enough, we performed a Kolmogorov–Smirnov test (KS test).

CRedit authorship contribution statement

Jennifer A Flegg: Writing – review & editing, Supervision, Investigation, Conceptualization. **Siyu Chen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lisa J White:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization. **Katrina A Lythgoe:** Writing – review & editing, Supervision, Data curation, Conceptualization.

Declaration of Competing Interest

The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for [*Epidemics*] and was not involved in the editorial review or the decision to publish this article.

Acknowledgments

The authors received no financial support for the research.

Author contributions

L.J.W., S.C., J.A.F. and K.A.L. conceived and designed the study. S.C. cleaned the data, S.C. and L.J.W. developed the methodology and conducted the formal analysis. S.C. and L.J.W. wrote the original manuscript. All authors reviewed and provided analytical input and approved the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.epidem.2025.100814](https://doi.org/10.1016/j.epidem.2025.100814).

Data availability

All codes and materials used in the analyses can be accessed at:

https://github.com/SiyuChenOxf/Exposure_ONS-modeling

References

- Aburto, J.M., et al., 2022. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *Int. J. Epidemiol.* 51 (1), 63–74.
- Aguas, R., et al., 2020. Modelling the COVID-19 pandemic in context: an international participatory approach. *BMJ Glob. Health* 5 (12), e003126.
- Bajaj, Sumali, et al., 2024. COVID-19 testing and reporting behaviours in England across different sociodemographic groups: a population-based study using testing data and data from community prevalence surveillance surveys. *Lancet Digit. Health* 6 (11), e778–e790.
- Böger, B., et al., 2021. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am. J. Infect. Control* 49 (1), 21–29.
- Bollyky, T.J., et al., 2022. Pandemic preparedness and COVID-19: an exploratory analysis of infection and fatality rates, and contextual factors associated with preparedness in 177 countries, from Jan 1, 2020, to Sept 30, 2021. *Lancet* 399 (10334), 1489–1512.
- Bubar, K.M., et al., 2021. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science* 371 (6532), 916–921.
- Chen, S., et al., 2021. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *PLOS Comput. Biol.* 17 (9), e1009436.
- Chen, S., et al., 2022. Estimating the effectiveness of shielding during pregnancy against SARS-CoV-2 in New York City during the first year of the COVID-19 pandemic. *Viruses* 14 (11), 2408.
- Clapham, H., et al., 2020. Seroepidemiologic study designs for determining SARS-COV-2 transmission and immunity. *Emerg. Infect. Dis.* 26 (9), 1978.
- Coronavirus (COVID-19) Infection Survey technical article: cumulative incidence of the number of people who have tested positive for COVID-19, UK: 22 April 2022. 2020.
- Desjardins, M.R., 2020. Syndromic surveillance of COVID-19 using crowdsourced data. *Lancet Reg. Health–West. Pac.* 4.
- GOV.UK. Coronavirus (COVID-19) in the UK 2023; Available from: (<https://coronavirus.data.gov.uk/details/deaths?areaType=nation&areaName=England>).
- Ibarrondo, F.J., et al., 2020. Rapid decay of anti-SARS-CoV-2 antibodies in persons with mild Covid-19. *N. Engl. J. Med.* 383 (11), 1085–1087.
- Kennedy, B., et al., 2022. App-based COVID-19 syndromic surveillance and prediction of hospital admissions in COVID symptom study Sweden. *Nat. Commun.* 13 (1), 2110.
- Knock, E.S., et al., 2021. Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Sci. Transl. Med.* 13 (602), eabg4262.
- Long, Q.-X., et al., 2020. Antibody responses to SARS-CoV-2 in COVID-19 patients: the perspective application of serological tests in clinical practice. *MedRxiv* 2020, 03.18.20038018.
- Metcalfe, C.J.E., Morris, D.H., Park, S.W., 2020. Mathematical models to guide pandemic response. *Science* 369 (6502), 368–369.
- Office of National Statistics Coronavirus (COVID-19) Infection Survey UK, 2023 (<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurvey/latest#strengths-and-limitations>). 2023.
- Ozili, P.K., Arun, T., 2023. Spillover of COVID-19: impact on the global economy. *Managing Inflation and Supply Chain Disruptions in the Global Economy*. IGI Global, pp. 41–61.
- Page, C., Yates, C.A., 2022. Role of mathematical modelling in future pandemic response policy. *bmj* 378.
- Pouwels, K.B., et al., 2021. Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS coronavirus infection survey. *Lancet Public Health* 6 (1), e30–e38.
- Russell, T.W., et al., 2020. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med.* 18 (1), 1–9.
- Van Elslande, J., et al., 2021. Estimated half-life of SARS-CoV-2 anti-spike antibodies more than double the half-life of anti-nucleocapsid antibodies in healthcare workers. *Clin. Infect. Dis.* 73 (12), 2366–2368.
- Wei, J., et al., 2021. Anti-spike antibody response to natural SARS-CoV-2 infection in the general population. *Nat. Commun.* 12 (1), 6250.