Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys

Authors: Siyu Chen[1], Jennifer A Flegg[2], Lisa J White[1], Ricardo Aguas[3*]

Affiliations:

[1]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford.

[2]School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia.

[3]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

*Correspondence to: ricardo@tropmedres.ac

Abstract:

Accurate knowledge of levels of prior population exposure will inform preparedness plans for subsequent SARS-CoV-2 epidemic waves and vaccination strategies. Serological studies can be used to estimate levels of past exposure and thus position populations in their epidemic timeline. To circumvent biases introduced by decaying antibody titers over time, population exposure estimation methods should account for seroreversion. Here, we present a new method that combines multiple datasets (serology, mortality, and virus positivity ratios) to estimate seroreversion time and infection fatality ratios and simultaneously infer population exposure levels. The results indicate that the average time to seroreversion is five months, meaning that true exposure may be more than double the current seroprevalence levels reported for several regions of England.

Summary:

SARS-CoV-2 exposure in England could be double the reported seroprevalence, with implications for the impact of vaccines/other interventions.

Main Text:

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has spread across the globe with devastating effects on populations and economies (1). Levels and styles of reporting the epidemic progress in each country vary considerably (2). Cases are dramatically under-reported and case definitions have changed significantly over time; therefore, the scientific and public health communities turned to serological surveys as a means to position populations along their expected epidemic timeline (3, 4). Prospects that information gained from serological studies could resolve the true levels of population exposure to the virus, and thus provide valuable insights into COVID-19 lethality, were frustrated by apparent rapid declines in antibody levels following infection (5-7). Further research efforts have attempted to determine the correlates for protective immunity against disease and infection and have found that while antibody titers are poor indicators of sustained immunity, cellular immunity can play a determinant role in limiting susceptibility to further SARS-CoV-2 challenges in previously exposed individuals (8, 9). Unfortunately, performing T cell assays at scale is technically challenging and expensive, which justified the decision to conduct a series of serology surveys (some of which are still underway) in many locations globally to provide a better understanding of the extent of viral spread among populations (10).

In England, a nationwide survey sampling more than 100,000 adults was performed from 20 June to 13 July 2020. The results suggested that 13% and 6% of the population of London and England, respectively, had been exposed to SARS-CoV-2, giving an estimated overall infection fatality ratio (IFR) of 0.90% (11). Although corrections were made for the sensitivity and specificity of the test used to infer seroprevalence, declining antibody levels were not accounted for. This is a limitation of the approach, potentially resulting in underestimates of the true levels of population exposure (12) and an overestimate of the IFR.

We now have a much clearer picture of the time dynamics of humoral responses following SARS-CoV-2 exposure, with antibody titers remaining detectable for approximately 6 months (*13, 14*). Commonly used serological assays have a limit of antibody titer detection, below which a negative result is yielded. Hence, a negative result does not necessarily imply absence of antibodies, but rather that there is a dynamic process by which production of antigen targeted antibodies diminishes once infection has been resolved, resulting in decaying antibody titers over time. As antibody levels decrease beyond the limit of detection, seroreversion occurs. We define the seroreversion rate as the inverse of the average time taken following seroconversion for antibody levels to decline below the cut-off for testing seropositive. In a longitudinal follow-up study, antibodies remained detectable for at least 100 days (6). In another study (15), seroprevalence declined by 26% in approximately three months, which translates to an average time to seroreversion of around 200 days. However, this was not a cohort study, so newly admitted individuals could have seroconverted while others transitioned from positive to negative between rounds, leading to an overestimation of the time to seroreversion.

Intuitively, if serology were a true measure of past exposure, we would expect a continually increasing prevalence of seropositive individuals over time. However, data suggest this is not the case (16), with most regions in England showing a peak in seroprevalence at the end of May 2020. This suggests seroreversion plays a significant role in shaping the seroprevalence curve and that the time since the first epidemic peak will influence the extent to which subsequent seroprevalence measurements underestimate the underlying population attack size (proportion of the population exposed). We argue that the number of people infected during the course of the epidemic can be informed by data triangulation, i.e., by combining numbers of deceased and seropositive individuals over time. For this linkage to be meaningful, we need to carefully consider the typical SARS-CoV-2 infection and recovery timeline (Fig. 1).

Most individuals, once infected, experience an incubation period of approximately 4.8 days (95% confidence interval (CI): 4.5–5.8) (17), followed by the development of symptoms, including fever, dry cough, and fatigue, although some individuals will remain asymptomatic throughout. Symptomatic individuals may receive a diagnostic PCR test at any time after symptoms onset; the time lag between

symptoms onset and date of test varies by country and area, depending on local policies and testing capacity. Some individuals might, as their illness progresses, require hospitalization, oxygen therapy, or even intensive care, eventually either dying or recovering.

The day of symptoms onset, as the first manifestation of infection, is a critical point for identifying when specific events occur relative to each other along the infection timeline. The mean time from symptoms onset to death is estimated to be 17.8 days (95% credible interval (CrI): 16.9–19.2) and to hospital discharge 24.7 days (22.9–28.1) (18). The median seroconversion time for IgG (long-lasting antibodies thought to be indicators of prior exposure) is estimated to be 14 days post-symptom onset; the presence of antibodies is detectable in less than 40% of patients within 1 week of symptoms onset, rapidly increasing to 79.8% (IgG) at day 15 post-onset (19). We assume onset of symptoms occurs at day 5 post-infection and that it takes an average of 2 additional days for people to have a PCR test. Thus, we fix the time lag between exposure and seroconversion, δ_ϵ, at 21 days, the time lag between a PCR test and death, δ_P, as 14 days, and assume that seroconversion in individuals who survive occurs at approximately the same time as death for those who don't (Fig. 1).

Thus, we propose to use population level dynamics (changes in mortality and seroprevalence over time) to estimate three key quantities: the seroreversion rate, the IFR, and the total population exposure over time. We developed a Bayesian inference method to estimate said quantities, based on official epidemiological reports and a time series of serology data from blood donors in England, stratified by region (16) – see Supplementary Material for more details. This dataset informed the national COVID-19 serological surveillance and its data collection was synchronous with the "REACT" study (11). The two sero-surveys use different, but comparable, antibody diagnostic tests (20). While "REACT" used a lateral flow immunoassay (LFIA) test for IgG (11), the data presented here were generated using the Euroimmun® assay. The independent "REACT" study acts as a validation dataset, lending credence to the seroprevalence values used. For example, seroprevalence in London was reported by REACT to be 13.0% (12.3–13.6) for the period 20 June to 13 July 2020. In comparison, the London blood-donor time series indicated seroprevalence to be 13.3% (8.4–16) on 21 June 2020.

We developed a method that combines daily mortality data with seroprevalence in England, using a mechanistic mathematical model to infer the temporal trends of exposure and seroprevalence during the COVID-19 epidemic. We fit the mathematical model jointly to serological survey data from seven regions in England (London, North West, North East (North East and Yorkshire and the Humber regions), South East, South West, Midlands (East and West Midlands combined), and East of England) using a statistical observation model. For more details on the input data sources, mechanistic model and fitting procedure, see the Supplementary Material. We considered that mortality is perfectly reported and proceeded to use this anchoring variable to extrapolate the number of people infected 3-weeks prior. We achieved this by estimating region-specific IFRs (defined as γ_i), which we initially assumed to be time invariant, later relaxing this assumption. The identifiability of the IFR metric was guaranteed by using the serological data described above as a second source of information on exposure. From the moment of exposure, individuals seroconvert a fixed 21 days later and can then serorevert at a rate, β, that is estimated as a global parameter. We thus have both mortality and prevalence of seropositivity informing SARS-CoV-2 exposure over time.

Several other research groups have used mortality data to extrapolate exposure and as a result provide estimates for IFR. Some IFR estimates were published assuming serology cross-sectional prevalence to be a true reflection of population exposure, while others used infection numbers generated by mechanistic dynamic models fit to mortality data (21). Most recently, sophisticated statistical techniques have been used, which take into account the time lag between exposure and seroconversion when estimating the underlying population exposure from seroprevalence measurements (22), with one study also considering seroreversion (23). Our method is very much aligned with the latter but is applied at a subnational level while using a dataset that has been validated by an independent, largely synchronous study.

Results from the fixed IFR inference method show excellent agreement with serological data (Fig. 2). We found that, after seroconverting, infected individuals remain seropositive for about 161 days (95% CrI: 145-178) (Tables 1, S1 and Fig. S1). This relatively rapid (approximately five months)

seroreversion is similar to other estimates from experimental studies (13, 14). As a consequence of this rapid seroreversion, epidemic progression will result in an increasing gap between measured serology prevalence levels and cumulative population exposure to the virus. Ultimately, this may mean that more than twice as many people have been exposed to the virus relative to the number of people who are seropositive (Fig. 2), raising questions about the relevance of serological data for informing policy decisions moving forward. We also estimated age-independent IFRs for seven English regions (means ranging from 0.37% to 0.84% - Table 1) that are in very good agreement with other estimates for England (24). The estimated IFRs are noticeably lower for London, which can be explained by differences in population age structure across the seven regions considered here. London has a considerably younger population than other regions of England, which, associated with increasing severity of disease with increasing age (25, 26), results in a lower expected number of fatalities given a similar number of infections. This could be construed as a possible explanation for fluctuations in estimates of the country-wide IFR over time (24), as outbreaks occur intermittently across regions with different underlying IFRs. An alternative interpretation of IFR trends in England is that individuals who are more likely to die from infection (due to some underlying illness or other risk factors) will do so earlier. This means that as the epidemic progresses, a selection (through infection) for a decrease in average population frailty (a measure of death likelihood once infected) is taking place and, consequently, a reduction in the ratio of deaths to infections. To test this hypothesis, we constructed an alternative formulation of our modelling approach, whereby the IFR at a specific time is dependent on the stage of epidemic progression. It is extremely difficult to extrapolate the underlying risk of infection from reported case data due to the volatility in testing capacity. Hence, we propose that the optimal metric for epidemic progression is the cumulative test positivity ratio. In the absence of severe sampling biases, the test positivity ratio is a good indicator of changes in underlying population infection risk, as a larger proportion of people will test positive if infection prevalence increases. In fact, it is clear from Fig. S5 that the test positivity ratio is a much better indicator of exposure than the case fatality ratio (CFR) or the hospitalization fatality ratio (HFR), since mirrors the shape of the mortality incidence curve. For the time-varying IFR, we took the normalized cumulative test positivity

ratio time series and applied it as a scalar of the maximum IFR value estimated for each region – for more details see the Supplementary Materials.

Results from the time-varying IFR model indicate that the population of London might have undergone a significant frailty selection process during the first wave of the epidemic and now shows a significantly lower IFR compared with March/April 2020 (Fig. S4). Interestingly, no statistically significant time-dependence on IFR was inferred for any of the remaining regions (Figure S3), suggesting this phenomenon is dependent on age structure. We can eliminate exposure levels as the main driver of this process as there is no clear temporal signal for IFR for the only other region (North West) with a comparable force of selection (i.e., similar predicted exposure levels). It seems that in younger populations, with a lower subset of very frail individuals, this selection will be more pronounced. Overall, the estimates obtained with both models are very consistent, with the estimated credible intervals for the time-varying IFR model including the median estimated obtained for the fixed IFR model (Fig. S1, S3 and Tables 1, S3). Given the current polarization of opinion around COVID-19 natural immunity, we realize that our results are likely to be interpreted in one of two conflicting ways: (1) the rate of seroreversion is high, therefore achieving population (herd) immunity is unrealistic, or (2) exposure in more affected places such as London is much higher than previously thought, and population immunity has almost been reached, which explains the decrease in IFR over time. We would like to dispel both interpretations and stress that our results do not directly support either. Regarding (1), it is important to note that the rate of decline in neutralizing antibodies, reflective of the effective immunity of the individual, is not the same as the rate of decline in seroprevalence. Antibodies may visibly decline in individuals, yet remain above the detection threshold for antibody testing (*6*). Conversely, if the threshold antibody titer above which a person is considered immune is greater than the diagnostic test detection limit, individuals might test positive when in fact they are not effectively immune. The relationship between the presence and magnitude of antibodies (and therefore seropositive status) and protective immunity is still unclear, with antibodies that provide functional immunity only now being discovered (*13*). Furthermore, T cell mediated immunity is detectable in seronegative individuals and is associated with protection against disease (*8*). Therefore, the immunity

profile for COVID-19 goes beyond the presence of a detectable humoral response. We believe our methodology to estimate total exposure levels in England offers valuable insights and a solid evaluation metric to inform future health policies (including vaccination) that aim to disrupt transmission. With respect to (2), we must clarify that decreasing IFR trends can result solely from selection processes operating at the intersection of individual frailty and population age structure. Likewise, the lower IFR in London can be attributed to its relatively younger population when compared with populations in the other regions of England. In conclusion, a method that accounts for seroreversion using mortality data allows the total exposure to SARS-CoV-2 to be estimated from seroprevalence data. The associated estimate of time to seroreversion of 161 days (95% CrI: 145-178) lies within realistic limits derived from independent sources. The total exposure in regions of England estimated using this method is more than double the current seroprevalence measurements. Implications for the impact of vaccination and other future interventions depend on the, as yet uncharacterized, relationships between exposure to the virus, seroprevalence, and population immunity.

References

1. M. Nicola *et al.*, The socio-economic implications of the coronavirus and COVID-19 pandemic: a review. *International Journal of Surgery*, (2020).
2. A. Mahdi. (2020). Understanding the Global Impact of COVID-19 through Data Science https://covid19.eng.ox.ac.uk/
3. C. J. E. Metcalf, C. Viboud, D. J. Spiro, B. T. Grenfell, Using Serology with Models to Clarify the Trajectory of the SARS-CoV-2 Emerging Outbreak. *Trends in Immunology* **41**, 849-851 (2020).
4. J. Watson, A. Richter, J. Deeks, Testing for SARS-CoV-2 antibodies. *bmj* **370**, (2020).
5. F. J. Ibarrondo *et al.*, Rapid decay of anti–SARS-CoV-2 antibodies in persons with mild Covid-19. *New England Journal of Medicine* **383**, 1085-1087 (2020).
6. J. Seow *et al.*, Longitudinal evaluation and decline of antibody responses in SARS-CoV-2 infection. *MedRxiv*, (2020).
7. Q.-X. Long *et al.*, Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nature medicine*, 1-4 (2020).
8. T. Sekine *et al.*, Robust T Cell Immunity in Convalescent Individuals with Asymptomatic or Mild COVID-19. *Cell* **183**, 158-168.e114 (2020).
9. A. Grifoni et al., Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell 181, 1489-1501.e1415 (2020).

10. R. K. Arora *et al.*, SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *The Lancet. Infectious Diseases*, (2020).
11. H. Ward *et al.*, Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. *medRxiv*, (2020).
12. S. Burgess, M. J. Ponsford, D. Gill. (British Medical Journal Publishing Group, 2020).
13. T. J. Ripperger *et al.*, Orthogonal SARS-CoV-2 Serological Assays Enable Surveillance of Low Prevalence Communities and Reveal Durable Humoral Immunity. *Immunity*, (2020).
14. A. S. Iyer *et al.*, Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Science Immunology* **5**, eabe0367 (2020).
15. H. Ward *et al.*, Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults (2020).
16. P. H. England, "Weekly Coronavirus Disease 2019 (COVID-19) surveillance report (week 40)," (2020).
17. Q. Bi *et al.*, Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*, (2020).
18. R. Verity *et al.*, Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*, (2020).
19. J. Arehao *et al.*, Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clinical Infectious Diseases*, (2020).
20. T. Nicol *et al.*, Assessment of SARS-CoV-2 serological tests for the diagnosis of COVID-19 through the evaluation of three immunoassays: Two automated immunoassays (Euroimmun and Abbott) and one rapid lateral flow immunoassay (NG Biotech). *Journal of Clinical Virology* **129**, 104511 (2020).
21. G. Meyerowitz-Katz, L. Merone, A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* **101**, 138-148 (2020).
22. S. Stringhini *et al.*, Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet* **396**, 313-319 (2020).
23. N. F. Brazeau et al., "COVID-19 Infection Fatality Ratio: Estimates from Seroprevalence," (Imperial College London, London, 2020).
24. D. Howdon, J. Oke, C. Heneghan, "Estimating the infection fatality ratio in England," (The Centre for Evidence-Based Medicine, 2020).
25. F. Zhou et al., Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The lancet, (2020).
26. N. G. Davies et al., Age-dependent effects in the transmission and control of COVID-19 epidemics. Nature Medicine 26, (2020).
27. GOV.UK Coronavirus (COVID-19) in the UK. https://coronavirus.data.gov.uk/
28. P. H. England, "Weekly Coronavirus Disease 2019 (COVID-19) surveillance report (week 40)," (2020).
29. P. H. England, "National flu and COVID-19 surveillance report (week 45) " (2020). 20

30. O. f. N. Statistics, "Population age structure by single year of age and sex for local authorities, counties, regions and England as a whole, mid-2018 to mid-2043," (2020).
31. K. Mohd Hanafiah, J. Groeger, A. D. Flaxman, S. T. Wiersma, Global epidemiology of hepatitis C virus infection: new estimates of age - specific antibody to HCV seroprevalence. Hepatology 57, 1333-1342 (2013).
32. Stan Development Team, RStan: the R interface to Stan (2020).

Acknowledgments: We would like to thank Adam Bodley for scientific writing assistance (according to Good Publication Practice guidelines) and editorial support.

Funding: JF acknowledges receives funding from the Australian Research Council (DP200100747). LJW is funded by the Li Ka Shing Foundation, and the University of Oxford's COVID-19 Research Response Fund (BRD00230). RA is funded by the Bill and Melinda Gates Foundation (OPP1193472).

Author contributions: LJW and RA conceptualized the analysis; SC curated and formatted the data for analysis; SC and JAF developed the statistical model and performed the Bayesian inferences; LJW and RA wrote the first draft. SC, JAF, LJW and RA reviewed the results and edited the manuscript.

Competing interests: Authors declare no competing interests.

Data and materials availability: All data, code, and materials used in the analyses can be accessed at: https://github.com/SiyuChenOxf/COVID19SeroModel/tree/master. All parameter estimates and figures presented can be reproduced using the codes provided. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplementary Materials
Materials and Methods
Tables S1 to S3
Figures S1 to S5

Fig. 1. Progression of exposed individuals through the various clinical (below the timeline), and diagnostic (above the timeline) stages of infection and recovery. Stages marked in grey represent events that may happen, with a probability consistent with the darkness of the shade of grey.

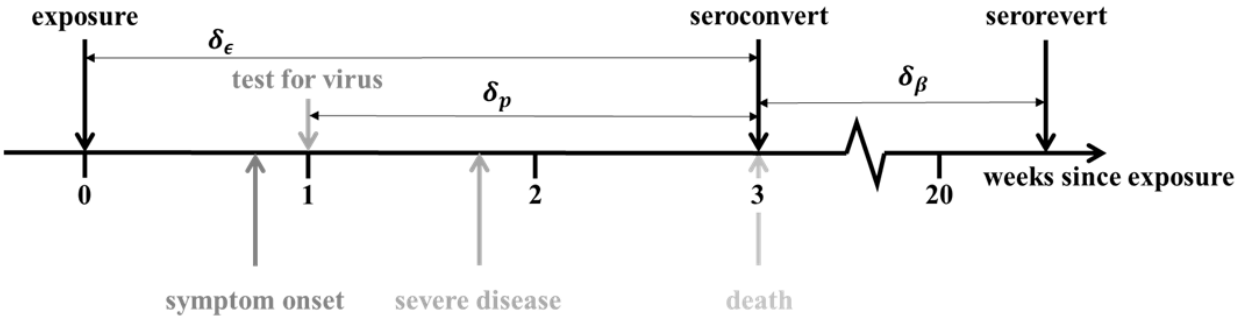

Fig. 2. Time course of the SARS-CoV-2 pandemic up to 7 November 2020 for seven regions in England. The solid orange circles and black error bars in each regional panel represent the observed seroprevalence data and their confidence interval, respectively, after adjusting for the sensitivity and specificity of the antibody test. The green and orange lines show the model predictions of median exposure and seroprevalence, respectively, while the shaded areas correspond to 95% CrI. The regional predicted exposure levels (expressed as the proportion of the population that has been infected) as of 17 October 2020 are shown on the map of England.

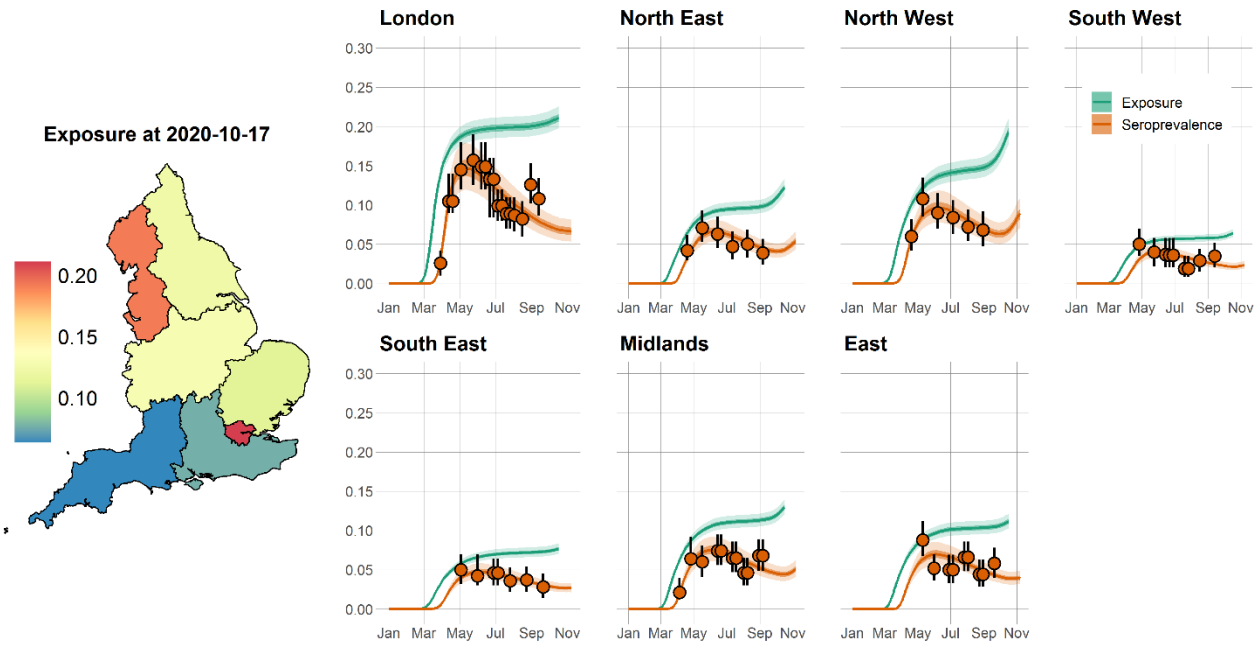

Table 1. Marginal median parameter estimates and 95% CrI for the constant IFR model. β is the rate of seroreversion and γ denotes the IFR. The estimated median time to seroreversion given by $1/\beta$ is 161 (95% CrI: 145-178) days.

Parameters	*Median (95% CrI)*
β	0.0062 (0.0056 - 0.0069)
γ_{London}	0.0037 (0.0035 - 0.0040)
$\gamma_{NorthWest}$	0.0064 (0.0059 - 0.0070)
$\gamma_{NortEast}$	0.0080 (0.0074 - 0.0088)
$\gamma_{SouthWest}$	0.0063 (0.0059 - 0.0068)
$\gamma_{SouthEast}$	0.0084 (0.0077 - 0.0092)
$\gamma_{Midlands}$	0.0068 (0.0064 - 0.0073)
γ_{East}	0.0070 (0.0065 - 0.0076)

Supplementary Materials for

Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys

Authors: Authors: Siyu Chen, Jennifer A Flegg, Lisa J White, Ricardo Aguas.

Correspondence to: ricardo@tropmedres.ac

This PDF file includes:

Materials and Methods
Tables S1 to S3
Figures S1 to S5

Materials and Methods

Data Sources

We used publicly available epidemiological data to infer the underlying exposure to SARS-CoV-2 over time, as described below:

Regional daily death

The observed daily mortality data for each of 7 English regions (London, North West, North East (contains both North East and Yorkshire and the Humber regions), South East, South West, Midlands (West and East Midlands combined) and East of England) from January 1st 2020 to November 11th 2020 relates to COVID-19 deaths within 28 days of a positive PCR test. This information was extracted from the UK government's official Covid-9 online dashboard (*27*) on Nov 9st 2020.

Regional adjusted seroprevalence

Region specific SARS-CoV-2 antibody seroprevalence measurements, adjusted for the sensitivity and specificity of the antibody test, were retrieved from Public Health England's week 40 National COVID-19 surveillance report (*28*).

Regional case positivity ratios

Weekly positivity ratios of laboratory confirmed COVID-19 cases were obtained from the week 40 and week 45 National COVID-19 surveillance reports (*28-29*). The first report contains Pillar 1 testing information spanning weeks 5 to 39 (2020), and Pillar 2 positivity ratios from week 19 up to week 39. The second report presents both Pillar 1 and Pillar 2 testing data from week 27 to week 44. We took the average of both Pillar 1 and Pillar 2 test positivity ratios where both data were available.

Regional population

Region specific population structures were obtained from the UK Office for National Statistics 2018 population survey (30).

Mechanistic model

We developed a mechanistic mathematical model that relates reported COVID-19 daily deaths to seropositive status by assuming all COVID-19 deaths are reported and estimating an infection fatality ratio that is congruent with the observed seroprevalence data. For each region, $i = 1, \dots, 7$ corresponding to London, North West, North East, South East, South West, Midlands and East of England respectively, we denote the infection fatality ratio at time t by $\alpha_i(t)$ and the number of daily deaths by $m_i(t)$. While we formulate the model in terms of a general, time-dependent, infection fatality ratio, we assume its default shape to be time invariant and later allow infection fatality ratio to vary with the stage of the epidemic.

Using the diagram in Fig. 1 as reference, and given a number of observed deaths at time t, $m_i(t)$, we can expect a number of infections $\frac{1}{\alpha_i(t)} m_i(t)$ to have occurred d_e days before. Of these infected individuals, $m_i(t)$ will eventually die, whilst the remaining $\frac{1-\alpha_i(t)}{\alpha_i(t)} m_i(t)$ will seroconvert from sero-negative to sero-positive. This assumes that seroconversion happens, on average, with the same delay from the moment of infection as death.

Assuming that seropositive individuals convert to seronegative (serorevert) at a rate β, the rate of change of the number of seropositive individuals in region i, $X_i(t)$, is given by:

$$\frac{dX_i}{dt} = \frac{1-\alpha_i(t)}{\alpha_i(t)} m_i(t) - \beta X_i \tag{1}$$

Solving Equation (1), subject to the initial condition $X_i(t_0) = 0$ where t is time since January 1st, 2020, gives:

$$X_i(t) = e^{-\beta t} \int_{t_0}^{t} e^{\beta w} \frac{(1-\alpha_i(w))}{\alpha_i(w))} m_i(w) dw \tag{2}$$

Discretizing Equation (2) with daily intervals $(\Delta w = 1)$ gives:

$$X_i(t) = e^{-\beta t} \sum_{w=t_0}^{t} \left[\frac{(1-\alpha_i(w))}{\alpha_i(w)} e^{\beta w} m_i(w) \right] \tag{3}$$

The model-predicted proportion of seropositive individuals in each population, $x_i(t)$, is calculated by dividing $X_i(t)$ (Equation (3)) by the respective region population size at time t, $P_i - \sum_{w=t_0}^{t} m_i(w)$, where P_i is the reported population in region i before the COVID-19 outbreak (*30*):

$$x_i(t) = e^{-\beta t} \left[P_i - \sum_{w=t_0}^{t} m_i(t) \right]^{-1} \sum_{w=t_0}^{t} \left[\frac{(1-\alpha_i(w))}{\alpha_i(w)} e^{\beta w} m_i(w) \right] \tag{4}$$

This is relatively straightforward when the serology data is already adjusted for test sensitivity and specificity as is the case. For unadjusted antibody test results, the proportion of the population that would test positive given the specificity (k_{sp}) and sensitivity (k_{se}) can be calculated as

$$z_i(t) = k_{se} x_i(t) + \left(1 - k_{sp}\right)\left(1 - x_i(t)\right).$$

As mentioned earlier, the method that we present in this paper allows for the infection fatality ratio, $\alpha_i(t)$, to be (a) constant or (b) vary over time with the stage of the epidemic:

(a) For constant infection fatality ratio, we have:

$$\alpha_i(t) = \gamma_i$$

(b) For time-varying infection fatality ratio, we first define the epidemic stage, $ES(t)$, as the normalized cumulative positivity ratio:

$$ES_i(t) = \frac{\sum_{t=t_0}^{t} y_i(t-\delta_p)}{\sum_{t=t_0}^{T} y_i(t-\delta_p)} \tag{5}$$

where $y_i(t)$ is the confirmed case positivity ratio at time t in the proportion of individuals testing positive for the virus, δ_p is the average time between testing positive and seroconversion (see Fig. 1) and T is the total number of days from t_0 until the last date of positivity data. In this work, we fixed $\delta_p = 7$ days (see Fig. 1 and main text). We assume that the infection fatality ratio is a linear function of the normalized cumulative positivity ratio as follows:

$$\alpha_i(t) = \gamma_i(1 - \eta_i)ES_i(t) \tag{6}$$

where $\eta_i \in [0,1]$ and $\gamma_i \in [0,1]$ are coefficients to be estimated. At the start of the epidemic when epidemic stage is 0 (see Equation (5)), then $\alpha_i(t) = \gamma_i$, whereas when epidemic stage is 1, $\alpha_i(t) = \gamma_i - \eta_i \times \gamma_i \leq \gamma_i$.

In Equation (5), $y_i(t)$ is taken from the regional weekly test positivity ratios (see Data section), converted to daily positivity ratios (taken to be the same over the week).

Once the model is parameterized, we can estimate the total proportion of the population that has been exposed, E_i, with the following formula:

$$E_i(t - \delta_\epsilon) = \left[P_i - \sum_{w=t_0}^{t} m(w)\right]^{-1} \sum_{w=t_0}^{t} \frac{1-\alpha_i(w)}{\alpha_i(w)} m_i(w) \tag{7}$$

where δ_ϵ is fixed to 21 days (see Fig. 1 and main text).

Observation model for statistical estimation of model parameters

We developed a hierarchical Bayesian model to estimate the model parameters θ, and present the posterior predictive distribution of the seroprevalence (Equation (4)) and exposure (Equation (7)) over time. Results are presented as the median of the posterior with the associated 95% credible intervals

(CrI). We assumed a negative binomial distribution (*31*) for the observed number of seropositive individuals in region i over time, $X_i^{obs}(t)$:

$$X_i^{obs}(t) = x_i^{obs}(t) \times \left(P_i - \sum_{w=t_0}^{t} m_i(w)\right) \tag{8}$$

where $x_i^{obs}(t)$ is the observed seroprevalence in region i over time. Then the observational model is specified for region i with observations at times $t_{i1}, t_{i2} \ldots, t_{in_i}$:

$$X_i^{obs}(t) \sim NB(X_i(t), \phi), \quad t = t_{i1}, t_{i2} \ldots, t_{in_i} \tag{9}$$

where $NB(X_i(t), \phi)$ is a negative binomial distribution, with mean $X_i(t)$ – given by equation (3)– and ϕ is an overdispersion parameter. We set ϕ to 100 to capture additional uncertainty in data points that would not be captured with a Poisson or binomial distribution (26). We assume uninformative beta priors for each of the parameters, according to the assumption made for how the infection fatality ratio is allowed to vary over time:

(a) For constant infection fatality ratio, we have $\theta = \left\{\{\gamma_i\}_{i=1}^{i=7}, \beta\right\}$ and take priors:

$$\gamma_i \sim Beta(1,1), \quad \beta \sim Beta(1,1) \tag{10}$$

(b) For time-varying infection fatality ratio, we have $\theta = \left\{\{\gamma_i\}_{i=1}^{i=7}, \{\eta_i\}_{i=1}^{i=7}, \beta\right\}$ and take priors:

$$\gamma_i \sim Beta(1,1), \quad \eta_i \sim Beta(1,1), \quad \beta \sim Beta(1,1) \tag{11}$$

We use Bayesian inference (Hamiltonian Monte Carlo algorithm) in RStan (32) to fit the model to seroprevalence data by running four chains of 20,000 iterations each (burn-in of 10,000). We use 2.5% and 97.5% percentiles from the resulting posterior distributions for 95% CrI for the parameters. The Gelman-Rubin diagnostics ($\hat{R}$) given in Tables S1 and S2 show values of 1, indicating that there is no evidence of non-convergence for either model formulation. Furthermore, the effective sample sizes

(n_{eff}) in Tables S1 and S2 are all more than 15,000, meaning that there are many samples in the posterior that can be considered independent draws.

Table S1. The effective sample size (n_{eff}) and the Gelman-Rubin diagnostic ($\hat{R}$) for the 8 model parameters in the default model (constant infection fatality ratio).

Parameter	n_{eff}	$\hat{R}$
β	16139	1
γ_{London}	22160	1
$\gamma_{NorthEast}$	34554	1
$\gamma_{SouthEast}$	29337	1
$\gamma_{NorthWest}$	34763	1
$\gamma_{SouthWest}$	27053	1
$\gamma_{Midlands}$	25317	1
γ_{East}	25620	1

Table S2. The effective sample size (n_{eff}) and the Gelman-Rubin diagnostic ($\hat{R}$) for the 15 model parameters in the time-varying infection fatality ratio model.

Parameter	n_{eff}	$\hat{R}$
β	29009	1
γ_{London}	30288	1
η_{London}	22488	1
$\gamma_{NorthEast}$	37606	1
$\eta_{NorthEast}$	35881	1
$\gamma_{SouthEast}$	24257	1
$\eta_{SouthEast}$	24320	1
$\gamma_{NorthWest}$	32036	1
$\eta_{NorthWest}$	24257	1
$\gamma_{SouthWest}$	31144	1
$\eta_{SouthWest}$	32177	1
$\gamma_{Midlands}$	30574	1
$\eta_{Midlands}$	26114	1
γ_{East}	21863	1
η_{East}	22204	1

Table S3. Marginal median parameter estimates and 95% CrI for the time-varying infection fatality ratio model.

Parameter	*Median (95%CrI)*
β	0.0068 (0.0061 - 0.0076)
γ_{London}	0.0044 (0.0038 - 0.0050)
η_{London}	0.45 (0.12 - 0.68)
$\gamma_{NorthEast}$	0.0080 (0.0073 - 0.0089)
$\eta_{NorthEast}$	0.062 (0.0023 - 0.27)
$\gamma_{NorthWest}$	0.0065 (0.0058 - 0.0076)
$\eta_{NorthWest}$	0.11 (0.0046 - 0.42)
$\gamma_{SouthWest}$	0.0062 (0.0057 - 0.0070)
$\eta_{SouthWest}$	0.049 (0.0018 - 0.23)
$\gamma_{SouthEast}$	0.0087 (0.0077 - 0.011)
$\eta_{SouthEast}$	0.12 (0.0052 - 0.43)
$\gamma_{Midlands}$	0.0071 (0.0064 - 0.0083)
$\eta_{Midlands}$	0.18 (0.010 - 0.46)
γ_{East}	0.0073 (0.0064 - 0.0093)
η_{East}	0.16 (0.0067 - 0.51)

Fig. S1. Marginal posterior distributions for parameters in the constant infection fatality ratio model. Vertical lines show median of distribution and grey shaded region shows 95% CrI.

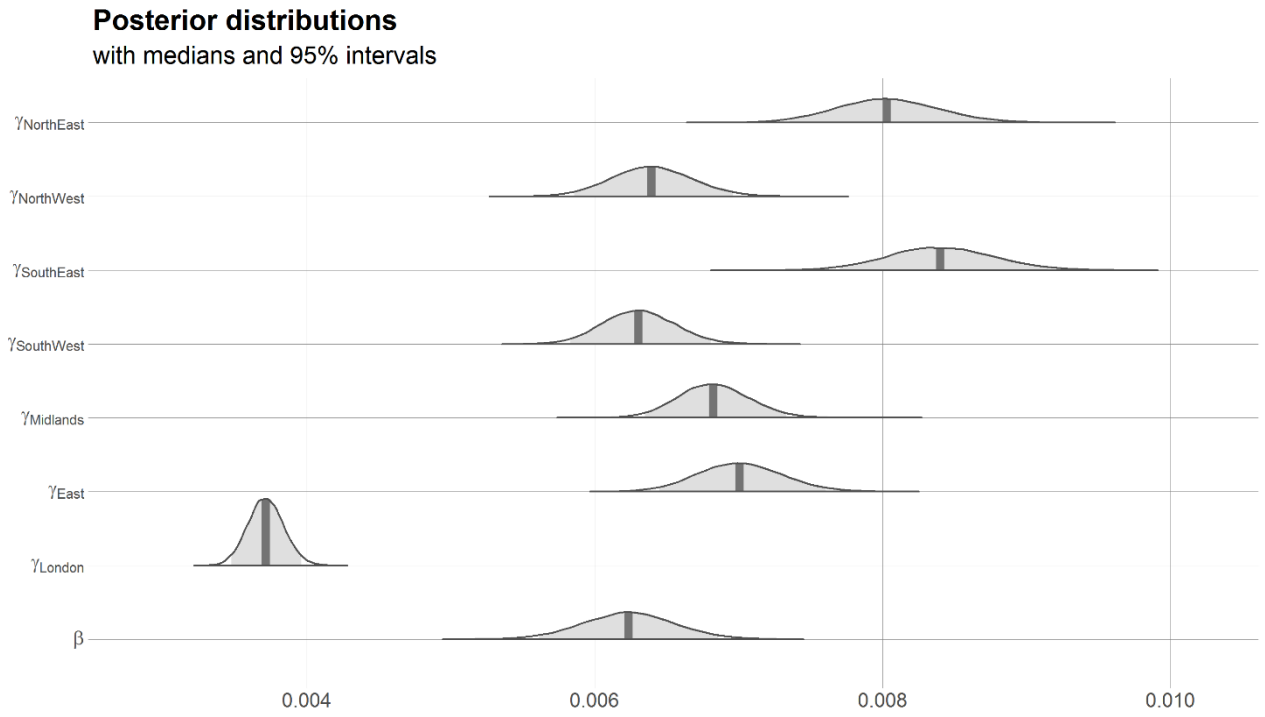

Fig. S2. Time course of the SARS-Cov-2 pandemic up to 7 November 2020 for seven regions in England. The orange solid circles and black error bars in each regional panel represent the observed seroprevalence data and its confidence interval after adjusting for the sensitivity and specificity of the antibody test. The green and orange lines show the median time-varying IFR model predictions for exposure, and seroprevalence, respectively, while the shaded areas correspond to 95% CrI. The regional median predicted exposure levels (expressed as the proportion of the population that has been infected) as of 17 October 2020 are shown on the map of England.

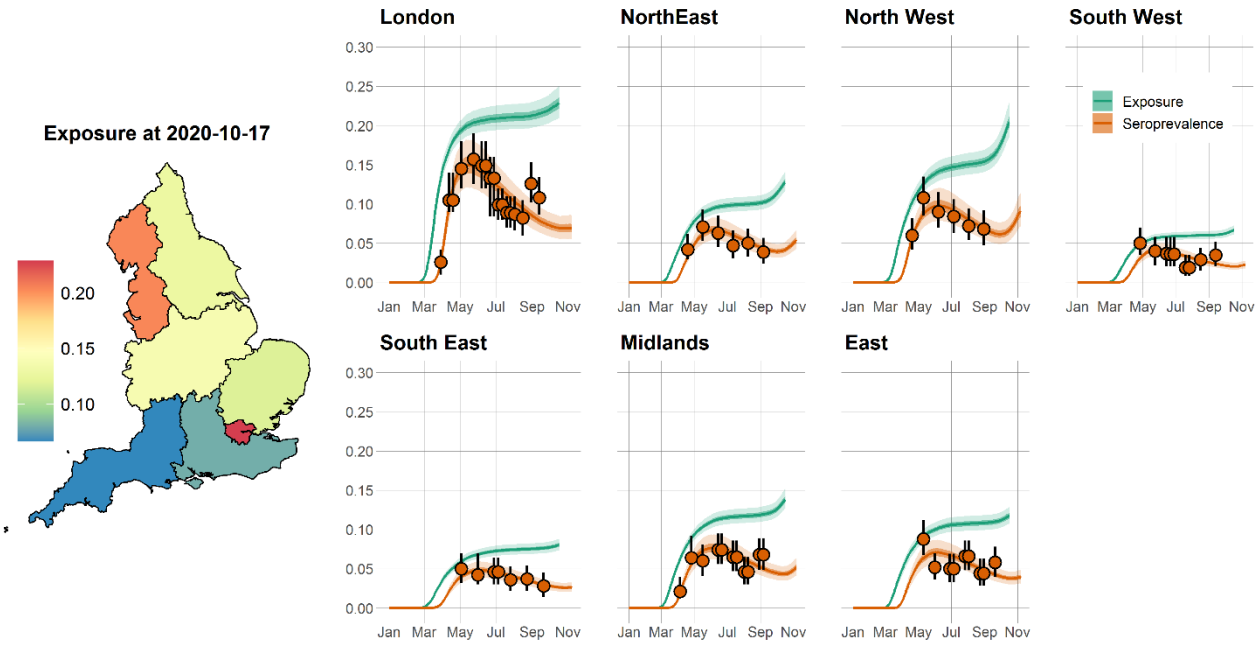

Fig. S3. Marginal posterior distributions for parameters in the time-varying infection fatality ratio model. Vertical lines show median of distribution and grey shaded region shows 95% CrI.

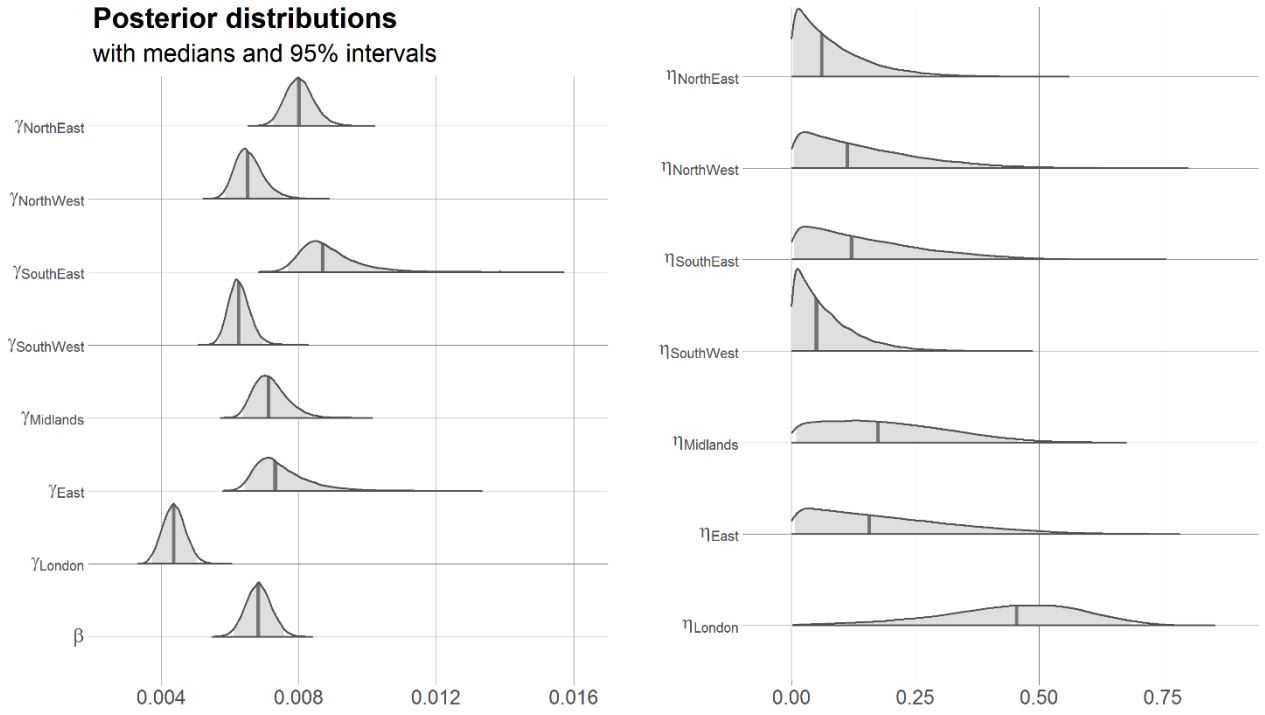

Fig. S4. Posterior predictive distribution of time-varying infection fatality ratio. Solid line shows median and shaded region 95% CrI.

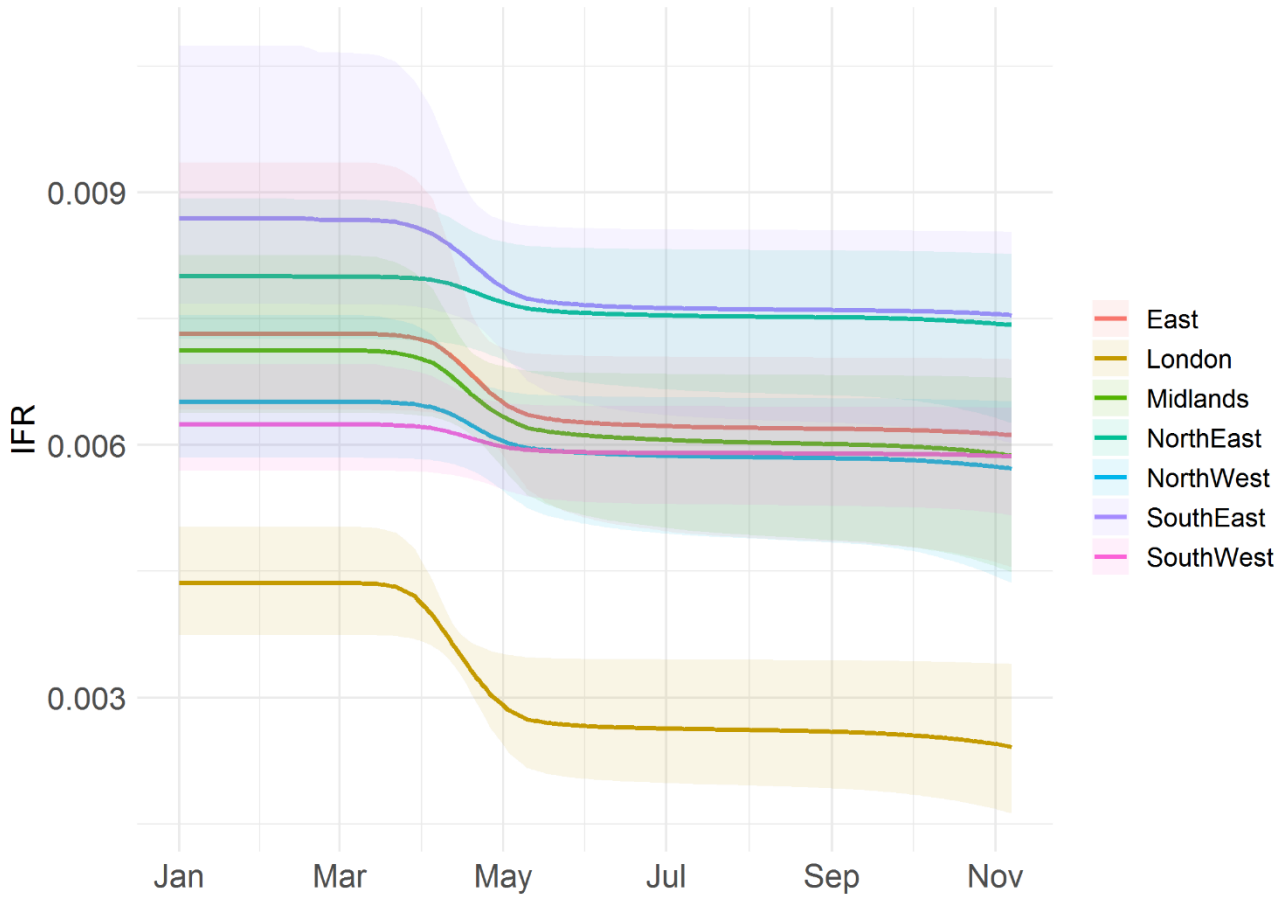
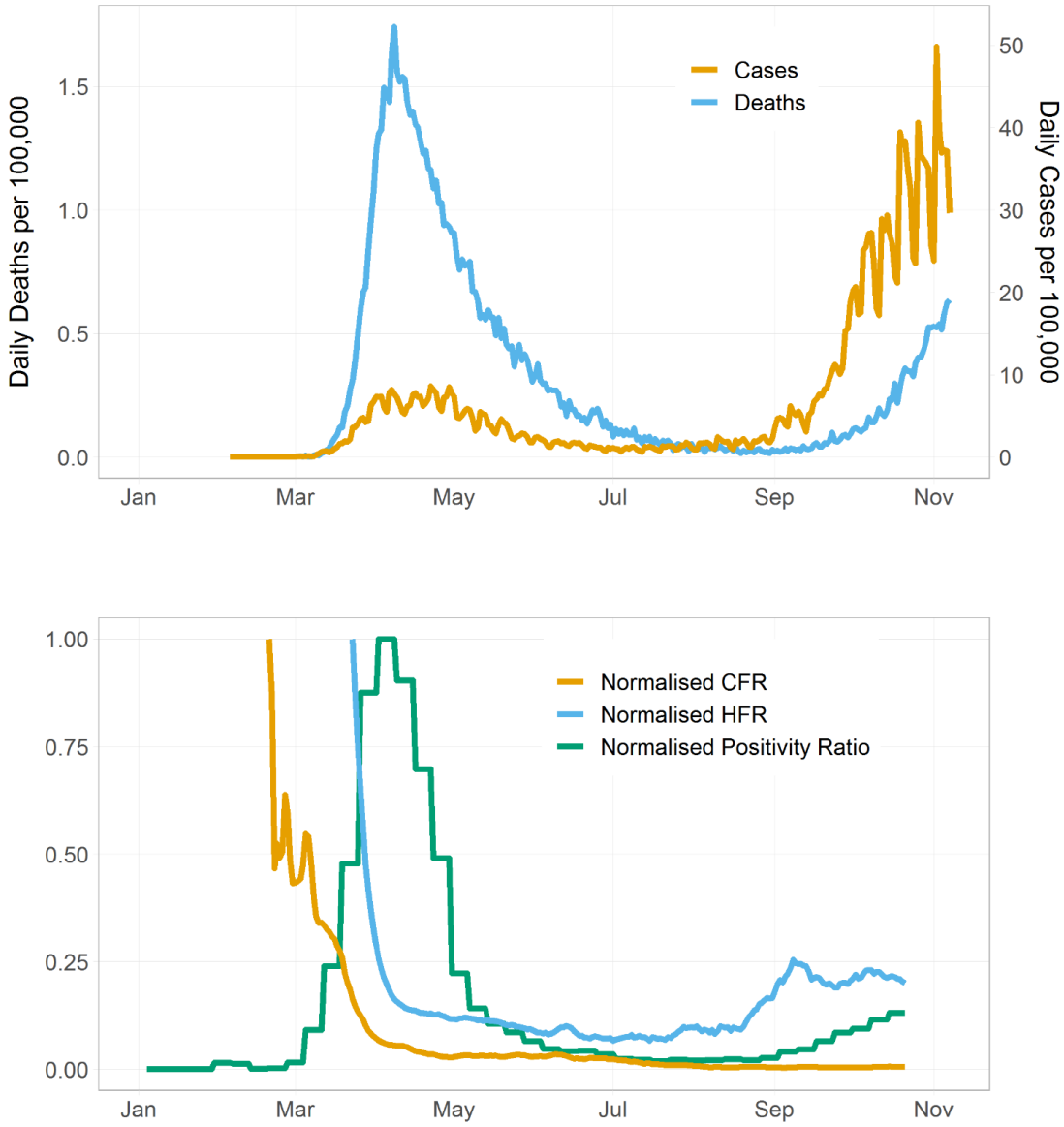
Figure S5. Relevant epidemiological metrics in England over the course of the pandemic. Top subplot shows COVID-19 cases and deaths (yellow and blue lines, respectively) per 100,000 population in England from Feb 5 to Nov 7, 2020. Bottom subplot shows the normalized case fatality rate, hospital fatality rate and virus positivity (yellow, blue, and green lines, respectively).