

11 **Abstract**

12 Accurate knowledge of accurate levels of prior population exposure has critical ramifications for
13 preparedness plans of subsequent SARS-CoV-2 epidemic waves and vaccine prioritization strategies.
14 Serological studies can be used to estimate levels of past exposure and thus position populations in
15 their epidemic timeline. To circumvent biases introduced by decaying antibody titers over time,
16 population exposure estimation methods should account for seroreversion, to reflect that changes in
17 seroprevalence measures over time are the net effect of increases due to recent transmission and
18 decreases due to antibody waning. Here, we present a new method that combines multiple datasets
19 (serology, mortality, and virus positivity ratios) to estimate seroreversion time and infection fatality
20 ratios and simultaneously infer population exposure levels. The results indicate that the average time
21 to seroreversion is six months, and that true exposure may be more than double the current
22 seroprevalence levels reported for several regions of England.

23

24 **Key words:** SARS-CoV-2; serology; antibody decline; seroreversion; exposure; infection fatality
25 ratio

26 **Introduction**

27 The COVID-19 pandemic has inflicted devastating effects on global populations and economies¹.
28 Levels and styles of reporting epidemic progress vary considerably across countries², with cases
29 consistently being under-reported and case definitions changing significantly over time. Therefore,
30 the scientific and public health communities turned to serological surveys as a means to position
31 populations along their expected epidemic timeline, and thus provide valuable insights into
32 COVID-19 lethality^{3,4}. Those prospects were frustrated by apparent rapid declines in antibody levels
33 following infection⁵⁻⁷. Population wide antibody prevalence measurements can significantly
34 underestimate the amount of underlying population immunity with obvious implications for
35 intervention strategy design and vaccine impact measurement.

36 Continued research efforts to determine the correlates for protective immunity against disease and
37 infection have found that while antibody titers are poor indicators of sustained immunity, cellular
38 immunity can play a determinant role in limiting susceptibility to further SARS-CoV-2 challenges in
39 previously exposed individuals^{8,9}. Unfortunately, performing T cell assays at scale is technically
40 challenging and expensive, which justified the decision to conduct a series of serology surveys (some
41 of which are still underway) in many locations globally to provide a better understanding of the
42 extent of viral spread among populations¹⁰.

43 In England, a nationwide survey sampling more than 100,000 adults was performed from 20 June to
44 13 July 2020. The results suggested that 13% and 6% of the population of London and England,
45 respectively, had been exposed to SARS-CoV-2, giving an estimated overall infection fatality ratio
46 (IFR) of 0.90%¹¹. Although corrections were made for the sensitivity and specificity of the test used

47 to infer seroprevalence, declining antibody levels were not accounted for. This is a limitation of the
48 approach, potentially resulting in underestimates of the true levels of population exposure¹² and an
49 overestimate of the IFR.

50

51 We now have a much clearer picture of the time dynamics of humoral responses following
52 SARS-CoV-2 exposure, with antibody titers remaining detectable for approximately 6 months^{13,14}.

53 Commonly used serological assays have a limit of antibody titer detection, below which a negative
54 result is yielded. Hence, a negative result does not necessarily imply absence of antibodies, but rather
55 that there is a dynamic process by which production of antigen targeted antibodies diminishes once
56 infection has been resolved, resulting in decaying antibody titers over time. As antibody levels
57 decrease beyond the limit of detection, seroreversion occurs.

58

59 We define the seroreversion rate as the inverse of the average time taken following seroconversion
60 for antibody levels to decline below the cut-off for testing seropositive. In a longitudinal follow-up
61 study, antibodies remained detectable for at least 100 days⁶. In another study¹⁵, seroprevalence
62 declined by 26% in approximately three months, which translates to an average time to seroreversion
63 of around 200 days. However, this was not a cohort study, so newly admitted individuals could have
64 seroconverted while others transitioned from positive to negative between rounds, leading to an
65 overestimation of the time to seroreversion.

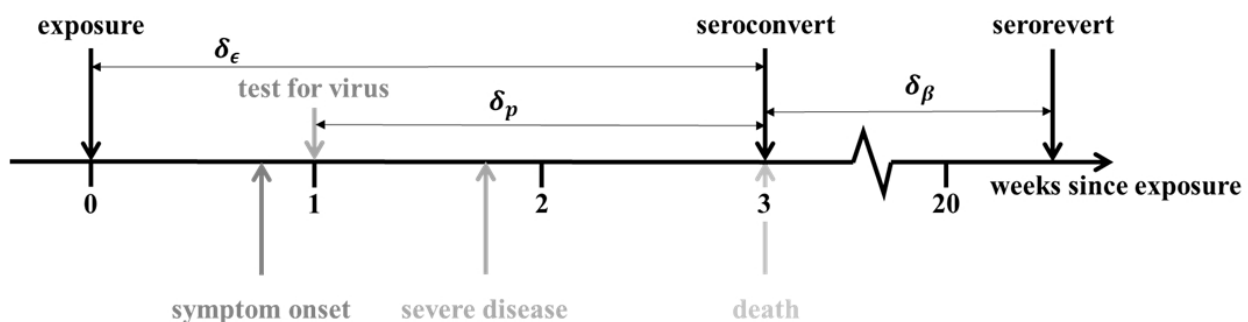
66

67 Intuitively, if serology were a true measure of past exposure, we would expect a continually
68 increasing prevalence of seropositive individuals over time. However, data suggest this is not the
69 case¹⁶, with most regions in England showing a peak in seroprevalence at the end of May 2020. This
70 suggests seroreversion plays a significant role in shaping the seroprevalence curve and that the time
71 since the first epidemic peak will influence the extent to which subsequent seroprevalence
72 measurements underestimate the underlying population attack size (proportion of the population
73 exposed). We argue that the number of people infected during the course of the epidemic can be
74 informed by data triangulation, i.e., by combining numbers of deceased and seropositive individuals
75 over time. For this linkage to be meaningful, we need to carefully consider the typical SARS-CoV-2
76 infection and recovery timeline (Figure 1).

77

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

81



82

83

84 Most individuals, once infected, experience an incubation period of approximately 4.8 days (95%
85 confidence interval (CI): 4.5–5.8)¹⁷, followed by the development of symptoms, including fever, dry

86 cough, and fatigue, although some individuals will remain asymptomatic throughout. Symptomatic
87 individuals may receive a diagnostic PCR test at any time after symptoms onset; the time lag
88 between symptoms onset and date of test varies by country and area, depending on local policies and
89 testing capacity. Some individuals might, as their illness progresses, require hospitalization, oxygen
90 therapy, or even intensive care, eventually either dying or recovering.

91 The day of symptoms onset, as the first manifestation of infection, is a critical point for identifying
92 when specific events occur relative to each other along the infection timeline. The mean time from
93 symptoms onset to death is estimated to be 17.8 days (95% credible interval (CrI): 16.9–19.2) and to
94 hospital discharge 24.7 days (22.9–28.1)¹⁸. The median seroconversion time for IgG (long-lasting
95 antibodies thought to be indicators of prior exposure) is estimated to be 14 days post-symptom onset;
96 the presence of antibodies is detectable in less than 40% of patients within 1 week of symptoms onset,
97 rapidly increasing to 79.8% (IgG) at day 15 post-onset¹⁹. We assume onset of symptoms occurs at
98 day 5 post-infection and that it takes an average of 2 additional days for people to have a PCR test.
99 Thus, we fix the time lag between exposure and seroconversion, δ_e , at 21 days, the time lag between
100 a PCR test and death, δ_p , as 14 days, and assume that seroconversion in individuals who survive
101 occurs at approximately the same time as death for those who don't (Figure 1).

102

103 Thus, we propose to use population level dynamics (changes in mortality and seroprevalence over
104 time) to estimate three key quantities: the seroreversion rate, the IFR, and the total population
105 exposure over time. We developed a Bayesian inference method to estimate said quantities, based on
106 official epidemiological reports and a time series of serology data from blood donors in England,

107 stratified by region¹⁶ – see Materials and Methods for more details. This dataset informed the
108 national COVID-19 serological surveillance and its data collection was synchronous with the
109 “REACT” study¹¹. The two sero-surveys use different, but comparable, antibody diagnostic tests²⁰.
110 While “REACT” used a lateral flow immunoassay (LFIA) test for IgG¹¹, the data presented here
111 were generated using the Euroimmun® assay. The independent “REACT” study acts as a validation
112 dataset, lending credence to the seroprevalence values used. For example, seroprevalence in London
113 was reported by REACT to be 13.0% (12.3–13.6) for the period 20 June to 13 July 2020. In
114 comparison, the London blood-donor time series indicated seroprevalence to be 13.3% (8.4–16) on
115 21 June 2020.

116 We developed a method that combines daily mortality data with seroprevalence in England, using a
117 mechanistic mathematical model to infer the temporal trends of exposure and seroprevalence during
118 the COVID-19 epidemic. We fit the mathematical model jointly to serological survey data from
119 seven regions in England (London, North West, North East (North East and Yorkshire and the
120 Humber regions), South East, South West, Midlands (East and West Midlands combined), and East
121 of England) using a statistical observation model. For more details on the input data sources,
122 mechanistic model and fitting procedure, see the Materials and Methods section. We considered that
123 mortality is perfectly reported and proceeded to use this anchoring variable to extrapolate the number
124 of people infected 3-weeks prior. We achieved this by estimating region-specific IFRs (defined as
125 γ_i), which we initially assumed to be time invariant, later relaxing this assumption. The
126 identifiability of the IFR metric was guaranteed by using the serological data described above as a
127 second source of information on exposure. From the moment of exposure, individuals seroconvert a

128 fixed 21 days later and can then serorevert at a rate, β , that is estimated as a global parameter. We
129 thus have both mortality and prevalence of seropositivity informing SARS-CoV-2 exposure over
130 time.

131

132 Several other research groups have used mortality data to extrapolate exposure and as a result
133 provide estimates for IFR. Some IFR estimates were published assuming serology cross-sectional
134 prevalence to be a true reflection of population exposure, while others used infection numbers
135 generated by mechanistic dynamic models fit to mortality data²¹. Most recently, sophisticated
136 statistical techniques have been used, which take into account the time lag between exposure and
137 seroconversion when estimating the underlying population exposure from seroprevalence
138 measurements²², with one study also considering seroreversion²³. Our method is very much aligned
139 with the latter but is applied at a subnational level while using a dataset that has been validated by an
140 independent, largely synchronous study.

141

142 **Results**

143 Results from the fixed IFR inference method show excellent agreement with serological data (Figure
144 2). We found that, after seroconverting, infected individuals remain seropositive for about 176 days
145 on average (95% CrI: 159-197) (Table 1, Table S1, and Figure 2—figure supplement 1).

146

147 **Table 1. Marginal median parameter estimates and 95% CrI for the constant IFR model.** β is
148 the rate of seroreversion and γ denotes the IFR. The estimated median time to seroreversion given
149 by $1/\beta$ is 176 (95% CrI: 159-197 days).

150

<i>Parameters</i>	<i>Median (95% CrI)</i>
β	0.0057 (0.0051 - 0.0063)
γ_{London}	0.0049 (0.0046 - 0.0063)
$\gamma_{NorthWest}$	0.0080 (0.0073 - 0.0087)
$\gamma_{NorthEast}$	0.0103 (0.0095 - 0.0112)
$\gamma_{SouthWest}$	0.0094 (0.0087 - 0.0101)
$\gamma_{SouthEast}$	0.0118 (0.0109 - 0.0129)
$\gamma_{Midlands}$	0.0085 (0.0079 - 0.0091)
γ_{East}	0.0083 (0.0077 - 0.0090)

151

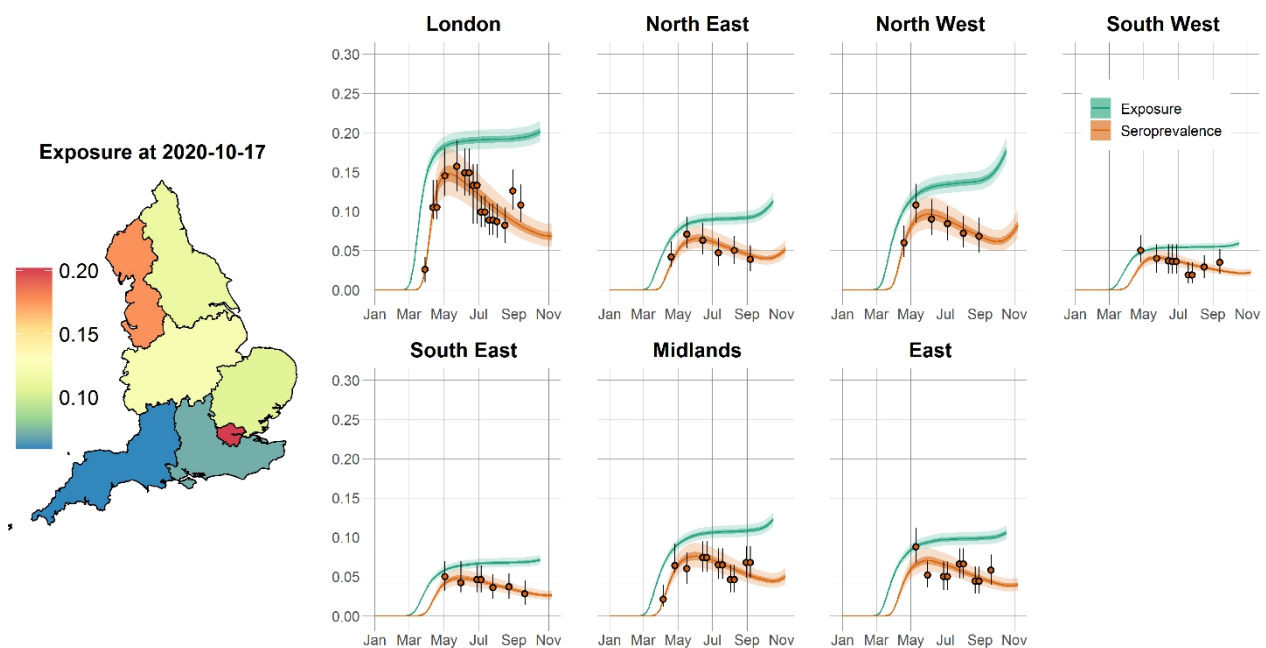
152 This relatively rapid (approximately six months) seroreversion is similar to other estimates from
 153 experimental studies^{13,14}, and might explain the reported 83% protection against reinfection within 6
 154 months of disease in UK patients²⁴. As a consequence of this rapid seroreversion, epidemic
 155 progression will result in an increasing gap between measured serology prevalence levels and
 156 cumulative population exposure to the virus. Ultimately, this may mean that more than twice as
 157 many people have been exposed to the virus relative to the number of people who are seropositive
 158 (Figure 2), raising questions about the relevance of serological data for informing policy decisions
 159 moving forward. We also estimated age-independent IFRs for seven English regions (means ranging
 160 from 0.49% to 1.18% - Table 1) that are in very good agreement with other estimates for England²⁵.

161

162

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

170



171

172

173 The estimated IFRs are noticeably lower for London, which can be explained by differences in
 174 population age structure across the seven regions considered here. London has a considerably
 175 younger population than other regions of England, which, associated with increasing severity of
 176 disease with increasing age^{26,27}, results in a lower expected number of fatalities given a similar
 177 number of infections. This could be construed as a possible explanation for fluctuations in estimates
 178 of the country-wide IFR over time²⁵, as outbreaks occur intermittently across regions with different
 179 underlying IFRs. An alternative interpretation of IFR trends in England is that individuals who are

180 more likely to die from infection (due to some underlying illness or other risk factors) will do so
181 earlier. This means that as the epidemic progresses, a selection (through infection) for a decrease in
182 average population frailty (a measure of death likelihood once infected) is taking place and,
183 consequently, a reduction in the ratio of deaths to infections. To test this hypothesis, we constructed
184 an alternative formulation of our modelling approach, whereby the IFR at a specific time is
185 dependent on the stage of epidemic progression – Figure 3. It is extremely difficult to extrapolate the
186 underlying risk of infection from reported case data due to the volatility in testing capacity. Hence,
187 we propose that the optimal metric for epidemic progression is the cumulative test positivity ratio. In
188 the absence of severe sampling biases, the test positivity ratio is a good indicator of changes in
189 underlying population infection risk, as a larger proportion of people will test positive if infection
190 prevalence increases. In fact, it is clear from (Figure 3 – figure supplement 3 that the test positivity
191 ratio is a much better indicator of exposure than the case fatality ratio (CFR) or the hospitalization
192 fatality ratio (HFR), since it mirrors the shape of the mortality incidence curve. For the time-varying
193 IFR, we took the normalized cumulative test positivity ratio time series and applied it as a scalar of
194 the maximum IFR value estimated for each region – for more details can be found in the Materials
195 and Methods section.

196

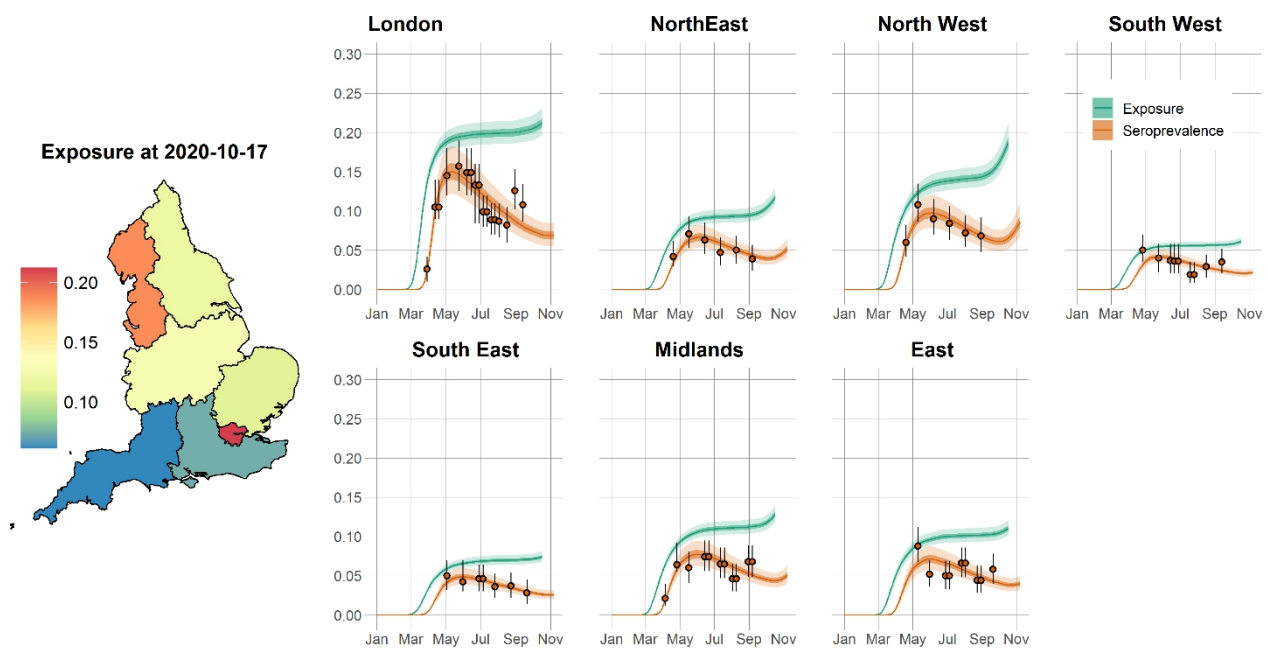
197 Results from the time-varying IFR model indicate that the population of London might have
198 undergone a significant frailty selection process during the first wave of the epidemic and now shows
199 a significantly lower IFR compared with March/April 2020 (Figure 3 – figure supplement 2).
200 Interestingly, no statistically significant time-dependence on IFR was inferred for any of the

201 remaining regions (Figure 3 – figure supplement 1), suggesting this phenomenon is dependent on age
202 structure.

203

204 **Figure 3. Time course of the SARS-Cov-2 pandemic up to 7 November 2020 for seven regions**
205 **in England for time-varying IFR model.** The orange solid circles and black error bars in each
206 regional panel represent the observed seroprevalence data and its confidence interval after adjusting
207 for the sensitivity and specificity of the antibody test. The green and orange lines show the median
208 time-varying IFR model predictions for exposure, and seroprevalence, respectively, while the shaded
209 areas correspond to 95% CrI. The regional median predicted exposure levels (expressed as the
210 proportion of the population that has been infected) as of 17 October 2020 are shown on the map of
211 England.

212



213

214

215 We can eliminate exposure levels as the main driver of this process as there is no clear temporal
216 signal for IFR for the only other region (North West) with a comparable force of selection (i.e.,

217 similar predicted exposure levels). It seems that in younger populations, with a lower subset of very
218 frail individuals, this selection will be more pronounced. Overall, the estimates obtained with both
219 models are very consistent, with the estimated credible intervals for the time varying IFR model
220 including the median estimated obtained for the fixed IFR model (Figure 3 – figure supplement 1 and
221 Table S3).

222

223 **Discussion**

224 Given the current polarization of opinion around COVID-19 natural immunity, we realize that our
225 results are likely to be interpreted in one of two conflicting ways: (1) the rate of seroreversion is high,
226 therefore achieving population (herd) immunity is unrealistic, or (2) exposure in more affected places
227 such as London is much higher than previously thought, and population immunity has almost been
228 reached, which explains the decrease in IFR over time. We would like to dispel both interpretations
229 and stress that our results do not directly support either. Regarding (1), it is important to note that the
230 rate of decline in neutralizing antibodies, reflective of the effective immunity of the individual, is not
231 the same as the rate of decline in seroprevalence. Antibodies may visibly decline in individuals yet
232 remain above the detection threshold for antibody testing⁶. Conversely, if the threshold antibody titer
233 above which a person is considered immune is greater than the diagnostic test detection limit,
234 individuals might test positive when in fact they are not effectively immune. The relationship
235 between the presence and magnitude of antibodies (and therefore seropositive status) and protective
236 immunity is still unclear, with antibodies that provide functional immunity only now being
237 discovered¹³. Furthermore, T cell mediated immunity is detectable in seronegative individuals and is

238 associated with protection against disease⁸. Therefore, the immunity profile for COVID-19 goes
239 beyond the presence of a detectable humoral response. We believe our methodology to estimate total
240 exposure levels in England offers valuable insights and a solid evaluation metric to inform future
241 health policies (including vaccination) that aim to disrupt transmission. With respect to (2), we must
242 clarify that decreasing IFR trends can result solely from selection processes operating at the
243 intersection of individual frailty and population age structure. Likewise, the lower IFR in London can
244 be attributed to its relatively younger population when compared with populations in the other
245 regions of England. In conclusion, a method that accounts for seroreversion using mortality data
246 allows the total exposure to SARS-CoV-2 to be estimated from seroprevalence data. The associated
247 estimate of time to seroreversion of 176 days (95% CrI: 159-197) lies within realistic limits derived
248 from independent sources. The total exposure in regions of England estimated using this method is
249 more than double the last seroprevalence measurements. Implications for the impact of vaccination
250 and other future interventions depend on the, as yet uncharacterized, relationships between exposure
251 to the virus, seroprevalence, and population immunity. To assess vaccination population impact one
252 can consider the population at risk to be those individuals who are seronegative, those with no past
253 exposure (confirmed or predicted), or those with no T cell reactivity. In this manuscript, we offer an
254 extra dimension to the evidence base for immediate decision-making, as well as anticipating future
255 information from the immunological research community about the relationship between
256 SARS-CoV-2 exposure and immunity.

257

258 **Materials and Methods**

259 **Data Sources**

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

262

263 *Regional daily death*

264 The observed daily mortality data for each of 7 English regions (London, North West, North East
265 (contains both North East and Yorkshire and the Humber regions), South East, South West, Midlands
266 (West and East Midlands combined) and East of England) from January 1st 2020 to November 11th
267 2020 relates to daily deaths with COVID-19 on the death certificate by date of death. This
268 information was extracted from the UK government's official Covid-9 online dashboard³² on March
269 8, 2021.

270

271 *Regional adjusted seroprevalence*

272 Region specific SARS-CoV-2 antibody seroprevalence measurements, adjusted for the sensitivity
273 and specificity of the antibody test, were retrieved from Public Health England's National
274 COVID-19 surveillance report¹⁶.

275

276 *Regional case positivity ratios*

277 Weekly positivity ratios of laboratory confirmed COVID-19 cases for each of 7 English regions were
278 obtained from the week 40 and week 45 (2020) Public Health England's National COVID-19
279 surveillance reports^{16,33}. The first report contains Pillar 1 testing information spanning weeks 5 (2020)

280 to 39 (2020), and Pillar 2 positivity ratios from week 19 up to week 39 (2020). The second report
281 presents both Pillar 1 and Pillar 2 testing data from week 27 to week 44 (2020). We took the average
282 of both Pillar 1 and Pillar 2 test positivity ratios where both data were available.

283

284 *Regional population*

285 Region specific population structures were obtained from the UK Office for National Statistics 2018
286 population survey³⁴.

287

288 **Mechanistic model**

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

297 Using the diagram in Figure 1 as reference, and given a number of observed deaths at time t , $m_i(t)$,
298 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
299 individuals, $m_i(t)$ will eventually die, whilst the remaining $\frac{1-\alpha_i(t)}{\alpha_i(t)}m_i(t)$ will seroconvert from

300 sero-negative to sero-positive. This assumes that seroconversion happens, on average, with the same
301 delay from the moment of infection as death.

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

$$304 \quad \frac{dX_i}{dt} = \frac{1-\alpha_i(t)}{\alpha_i(t)} m_i(t) - \beta X_i \quad (1)$$

305 Solving Equation (1), subject to the initial condition $X_i(t_0) = 0$ where t is time since January 1st,
306 2020, gives:

$$307 \quad X_i(t) = e^{-\beta t} \int_{t_0}^t e^{\beta w} \frac{(1-\alpha_i(w))}{\alpha_i(w)} m_i(w) dw \quad (2)$$

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

$$309 \quad X_i(t) = e^{-\beta t} \sum_{w=t_0}^t \left[\frac{(1-\alpha_i(w))}{\alpha_i(w)} e^{\beta w} m_i(w) \right] \quad (3)$$

310 The model-predicted proportion of seropositive individuals in each population, $x_i(t)$, is calculated
311 by dividing $X_i(t)$ (Equation (3)) by the respective region population size at time t , $P_i -$
312 $\sum_{w=t_0}^t m_i(w)$, where P_i is the reported population in region i before the COVID-19 outbreak³⁴:

$$313 \quad x_i(t) = e^{-\beta t} \left[P_i - \sum_{w=t_0}^t m_i(w) \right]^{-1} \sum_{w=t_0}^t \left[\frac{(1-\alpha_i(w))}{\alpha_i(w)} e^{\beta w} m_i(w) \right] \quad (4)$$

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

$$317 \quad z_i(t) = k_{se} x_i(t) + (1 - k_{sp})(1 - x_i(t)).$$

318 As mentioned earlier, the method that we present in this paper allows for the infection fatality ratio,
319 $\alpha_i(t)$, to be (a) constant or (b) vary over time with the stage of the epidemic:

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

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

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

$$323 \quad ES_i(t) = \frac{\sum_{t=t_0}^t y_i(t-\delta_p)}{\sum_{t=t_0}^T y_i(t-\delta_p)} \quad (5)$$

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

$$330 \quad \alpha_i(t) = \gamma_i(1 - \eta_i)ES_i(t) \quad (6)$$

331 where $\eta_i \in [0,1]$ and $\gamma_i \in [0,1]$ are coefficients to be estimated. At the start of the epidemic
332 when epidemic stage is 0 (see Equation (5)), then $\alpha_i(t) = \gamma_i$, whereas when epidemic stage is 1,
333 $\alpha_i(t) = \gamma_i - \eta_i \times \gamma_i \leq \gamma_i$.

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

336

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

$$339 \quad E_i(t - \delta_\epsilon) = \left[P_i - \sum_{w=t_0}^t m(w) \right]^{-1} \sum_{w=t_0}^t \frac{1-\alpha_i(w)}{\alpha_i(w)} m_i(w) \quad (7)$$

340 where δ_ϵ is fixed to 21 days (Figure 1).

341

342 **Observation model for statistical estimation of model parameters**

343 We developed a hierarchical Bayesian model to estimate the model parameters θ , and present the
344 posterior predictive distribution of the seroprevalence (Equation (4)) and exposure (Equation (7))
345 over time. Results are presented as the median of the posterior with the associated 95% credible
346 intervals (CrI). We assumed a negative binomial distribution³⁵ for the observed number of
347 seropositive individuals in region i over time, $X_i^{obs}(t)$:

$$348 \quad X_i^{obs}(t) = x_i^{obs}(t) \times \left(P_i - \sum_{w=t_0}^t m_i(w) \right) \quad (8)$$

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

$$351 \quad X_i^{obs}(t) \sim NB(X_i(t), \phi), \quad t = t_{i1}, t_{i2}, \dots, t_{in_i} \quad (9)$$

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

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

$$358 \quad \gamma_i \sim Beta(1,1), \quad \beta \sim Beta(1,1) \quad (10)$$

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

$$360 \quad \gamma_i \sim Beta(1,1), \quad \eta_i \sim Beta(1,1), \quad \beta \sim Beta(1,1) \quad (11)$$

361 We use Bayesian inference (Hamiltonian Monte Carlo algorithm) in RStan³⁶ to fit the model to
362 seroprevalence data by running four chains of 20,000 iterations each (burn-in of 10,000). We use
363 2.5% and 97.5% percentiles from the resulting posterior distributions for 95% CrI for the parameters.
364 The Gelman-Rubin diagnostics (\hat{R}) given in Table S1 and Table S
365 2 show values of 1, indicating that there is no evidence of non-convergence for either model
366 formulation. Furthermore, the effective sample sizes (n_{eff}) in Table S1 and Table S2 are all more
367 than 10,000, meaning that there are many samples in the posterior that can be considered
368 independent draws.

369

370 **Acknowledgments:** We would like to thank Adam Bodley for scientific writing assistance
371 (according to Good Publication Practice guidelines) and editorial support. JF acknowledges funding
372 from the Australian Research Council (DP200100747). LJW is funded by the Li Ka Shing
373 Foundation, and the University of Oxford's COVID-19 Research Response Fund (BRD00230). RA
374 is funded by the Bill and Melinda Gates Foundation (OPP1193472).

375

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

380

381 **Competing interests:** Authors declare no competing interests.

382

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

388

389 **References**

- 390 1. Nicola, Maria, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis,
391 Maliha Agha, and Riaz Agha. "The socio-economic implications of the coronavirus and
392 COVID-19 pandemic: a review." *International journal of surgery* (2020). DOI:
393 <https://doi.org/10.1016/j.ijssu.2020.04.018>
- 394 2. Mahdi, A., Blaszczyk, P., Dlotko, P., Salvi, D., Chan, T.S.T., Harvey, J., Gurnari, D., Wu, Y.,
395 Farhat, A., Hellmer, N. and Zarebski, A.E., 2020. OxCOVID19 Database: a multimodal data
396 repository for better understanding the global impact of COVID-19. medRxiv.
- 397 3. Metcalf, C.J.E., Viboud, C., Spiro, D.J. and Grenfell, B.T., 2020. Using Serology with Models to
398 Clarify the Trajectory of the SARS-CoV-2 Emerging Outbreak. *Trends in Immunology*, 41(10),
399 pp.849-851. DOI: <https://doi.org/10.1016/j.it.2020.06.011>
- 400 4. Watson, J., Richter, A. and Deeks, J., 2020. Testing for SARS-CoV-2 antibodies. *bmj*, 370. DOI:
401 <https://doi.org/10.1136/bmj.m3325>

- 402 5. Ibarondo, F.J., Fulcher, J.A., Goodman-Meza, D., Elliott, J., Hofmann, C., Hausner, M.A.,
403 Ferbas, K.G., Tobin, N.H., Aldrovandi, G.M. and Yang, O.O., 2020. Rapid decay of
404 anti-SARS-CoV-2 antibodies in persons with mild Covid-19. *New England Journal of Medicine*,
405 383(11), pp.1085-1087. DOI: [10.1056/NEJMc2025179](https://doi.org/10.1056/NEJMc2025179)
- 406 6. Seow, J., Graham, C., Merrick, B., Acors, S., Steel, K.J., Hemmings, O., O'Byrne, A., Kouphou,
407 N., Pickering, S., Galao, R. and Betancor, G., 2020. Longitudinal evaluation and decline of
408 antibody responses in SARS-CoV-2 infection. *MedRxiv*. Preprint at
409 <https://medrxiv.org/content/10.1101/2020.07.09.20148429v1.full>.
- 410 7. Long, Q.X., Liu, B.Z., Deng, H.J., Wu, G.C., Deng, K., Chen, Y.K., Liao, P., Qiu, J.F., Lin, Y.,
411 Cai, X.F. and Wang, D.Q., 2020. Antibody responses to SARS-CoV-2 in patients with
412 COVID-19. *Nature medicine*, 26(6), pp.845-848. DOI:
413 <https://doi.org/10.1038/s41591-020-0897-1>
- 414 8. Sekine, T., Perez-Potti, A., Rivera-Ballesteros, O., Strålin, K., Gorin, J.B., Olsson, A.,
415 Llewellyn-Lacey, S., Kamal, H., Bogdanovic, G., Muschiol, S. and Wullimann, D.J., 2020.
416 Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell*,
417 183(1), pp.158-168. DOI: <https://doi.org/10.1016/j.cell.2020.08.017>
- 418 9. Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings,
419 S.A., Sutherland, A., Premkumar, L., Jadi, R.S. and Marrama, D., 2020. Targets of T cell
420 responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed
421 individuals. *Cell*, 181(7), pp.1489-1501. DOI: <https://doi.org/10.1016/j.cell.2020.05.015>

- 422 10. Arora, R.K., Joseph, A., Van Wyk, J., Rocco, S., Atmaja, A., May, E., Yan, T., Bobrovitz, N.,
423 Chevrier, J., Cheng, M.P. and Williamson, T., 2020. SeroTracker: a global SARS-CoV-2
424 seroprevalence dashboard. *The Lancet. Infectious Diseases*. DOI:
425 [10.1016/S1473-3099\(20\)30631-9](https://doi.org/10.1016/S1473-3099(20)30631-9)
- 426 11. Ward, H., Atchison, C.J., Whitaker, M., Ainslie, K.E., Elliot, J., Okell, L.C., Redd, R., Ashby,
427 D., Donnelly, C.A., Barclay, W. and Darzi, A., 2020. Antibody prevalence for SARS-CoV-2 in
428 England following first peak of the pandemic: REACT2 study in 100,000 adults. *MedRxiv*.
429 Preprint at [https:// medrxiv.org/content/10.1101/2020.08.12.20173690v1](https://medrxiv.org/content/10.1101/2020.08.12.20173690v1) (2020).
- 430 12. Burgess, S., Ponsford, M.J. and Gill, D., 2020. Are we underestimating seroprevalence of
431 SARS-CoV-2? DOI: <https://doi.org/10.1136/bmj.m3364>
- 432 13. Ripperger, T.J., Uhrlaub, J.L., Watanabe, M., Wong, R., Castaneda, Y., Pizzato, H.A.,
433 Thompson, M.R., Bradshaw, C., Weinkauf, C.C., Bime, C. and Erickson, H.L., 2020.
434 Orthogonal SARS-CoV-2 serological assays enable surveillance of low-prevalence communities
435 and reveal durable humoral immunity. *Immunity*, 53(5), pp.925-933. DOI:
436 <https://doi.org/10.1016/j.immuni.2020.10.004>
- 437 14. Iyer, A.S., Jones, F.K., Nodoushani, A., Kelly, M., Becker, M., Slater, D., Mills, R., Teng, E.,
438 Kamruzzaman, M., Garcia-Beltran, W.F. and Astudillo, M., 2020. Persistence and decay of
439 human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in
440 COVID-19 patients. *Science immunology*, 5(52). DOI: [10.1126/sciimmunol.abe0367](https://doi.org/10.1126/sciimmunol.abe0367)
- 441 15. Ward, H., Cooke, G., Atchison, C.J., Whitaker, M., Elliott, J., Moshe, M., Brown, J.C., Flower,
442 B., Daunt, A., Ainslie, K.E. and Ashby, D., 2020. Declining prevalence of antibody positivity to

- 443 SARS-CoV-2: a community study of 365,000 adults. MedRxiv. Preprint at
444 <https://www.medrxiv.org/content/10.1101/2020.10.26.20219725v1>.
- 445 16. Public Health England, " National COVID-19 surveillance reports" (2020).
446 <https://www.gov.uk/government/publications/national-covid-19-surveillance-reports>
- 447 17. Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T.
448 and Gao, W., 2020. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their
449 close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*,
450 20(8), pp.911-919.DOI: [https://doi.org/10.1016/S1473-3099\(20\)30287-5](https://doi.org/10.1016/S1473-3099(20)30287-5)
- 451 18. Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg,
452 G., Thompson, H., Walker, P.G., Fu, H. and Dighe, A., 2020. Estimates of the severity of
453 coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*, 20(6),
454 pp.669-677.DOI: [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7)
- 455 19. Zhao, J., Yuan, Q., Wang, H., Liu, W., Liao, X., Su, Y., Wang, X., Yuan, J., Li, T., Li, J. and
456 Qian, S., 2020. Antibody responses to SARS-CoV-2 in patients with novel coronavirus disease
457 2019. *Clinical infectious diseases*, 71(16), pp.2027-2034. DOI:
458 <https://doi.org/10.1093/cid/ciaa344>
- 459 20. Nicol, T., Lefevre, C., Serri, O., Pivert, A., Joubaud, F., Dubée, V., Kouatchet, A., Ducancelle,
460 A., Lunel-Fabiani, F. and Le Guillou-Guillemette, H., 2020. Assessment of SARS-CoV-2
461 serological tests for the diagnosis of COVID-19 through the evaluation of three immunoassays:
462 Two automated immunoassays (Euroimmun and Abbott) and one rapid lateral flow

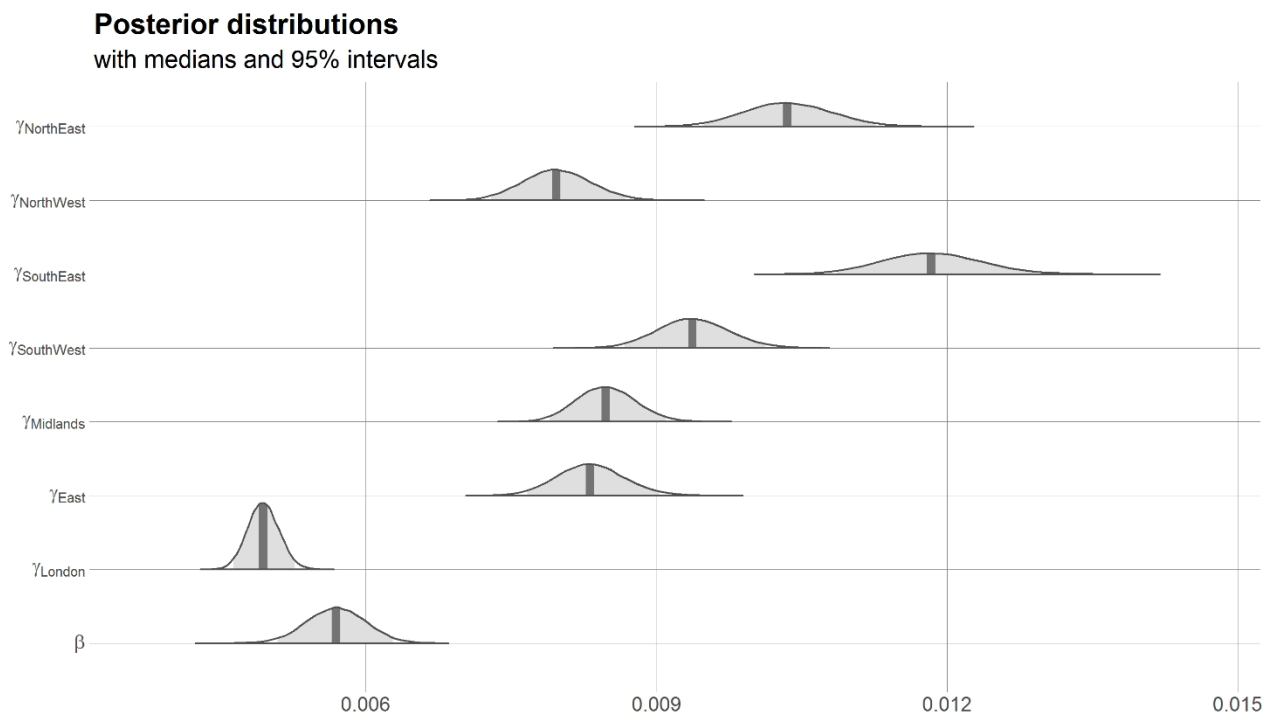
- 463 immunoassay (NG Biotech). *Journal of Clinical Virology*, 129, p.104511.DOI:
464 <https://doi.org/10.1016/j.jcv.2020.104511>
- 465 21. Meyerowitz-Katz, G. and Merone, L., 2020. A systematic review and meta-analysis of published
466 research data on COVID-19 infection-fatality rates. *International Journal of Infectious Diseases*.
467 DOI: <https://doi.org/10.1016/j.ijid.2020.09.1464>
- 468 22. Stringhini, S., Wisniak, A., Piumatti, G., Azman, A.S., Lauer, S.A., Baysson, H., De Ridder, D.,
469 Petrovic, D., Schrempft, S., Marcus, K. and Yerly, S., 2020. Seroprevalence of
470 anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based
471 study. *The Lancet*, 396(10247), pp.313-319. DOI:
472 [https://doi.org/10.1016/S0140-6736\(20\)31304-0](https://doi.org/10.1016/S0140-6736(20)31304-0)
- 473 23. Brazeau, N., Verity, R., Jenks, S., Fu, H., Whittaker, C., Winskill, P., Dorigatti, I., Walker, P.,
474 Riley, S., Schnekenberg, R.P. and Heltgebaum, H., 2020. Report 34: COVID-19 infection
475 fatality ratio: estimates from seroprevalence. DOI: <https://doi.org/10.25561/83545>
- 476 24. Hall, V. *et al.* Do antibody positive healthcare workers have lower SARS-CoV-2 infection rates
477 than antibody negative healthcare workers? Large multi-centre prospective cohort study (the
478 SIREN study), England: June to November 2020. Preprint at
479 <https://www.medrxiv.org/content/10.1101/2021.01.13.21249642v1.full-text> (2021).
- 480 25. Howdon, D., Oke, J. & Heneghan, C. Estimating the infection fatality ratio in England. *The*
481 *Centre for Evidence-Based Medicine*.
482 <https://www.cebm.net/covid-19/estimating-the-infection-fatality-ratio-in-england/> (2020)

- 483 26. Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X. and
484 Guan, L., 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19
485 in Wuhan, China: a retrospective cohort study. *The lancet*, 395(10229), pp.1054-1062. DOI:
486 [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)
- 487 27. Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M. and Eggo, R.M., 2020. Age-dependent
488 effects in the transmission and control of COVID-19 epidemics. *Nature medicine*, 26(8),
489 pp.1205-1211. DOI: <https://doi.org/10.1038/s41591-020-0962-9>
- 490 28. Salje, H., Kiem, C.T., Lefrancq, N., Courtejoie, N., Bosetti, P., Paireau, J., Andronico, A., Hozé,
491 N., Richet, J., Dubost, C.L. and Le Strat, Y., 2020. Estimating the burden of SARS-CoV-2 in
492 France. *Science*, 369(6500), pp.208-211. DOI: [10.1126/science.abc3517](https://doi.org/10.1126/science.abc3517)
- 493 29. Office for National Statistics. 2019 Datasets: Population estimates - local authority based by
494 single year of age. Available from:
495 <https://www.nomisweb.co.uk/query/construct/submit.asp?forward=yes&menuopt=201&subcom>
496 [p=](https://www.nomisweb.co.uk/query/construct/submit.asp?forward=yes&menuopt=201&subcomp=)
- 497 30. e Silva, L.V., de Andrade Abi, M.D.P., Dos Santos, A.M.T.B., de Mattos Teixeira, C.A., Gomes,
498 V.H.M., Cardoso, E.H.S., da Silva, M.S., Vijaykumar, N.L., Carvalho, S.V. and Frances, C.R.L.,
499 2020. COVID-19 mortality underreporting in Brazil: analysis of data from government internet
500 portals. *Journal of Medical Internet Research*, 22(8), p.e21413. DOI: [10.2196/21413](https://doi.org/10.2196/21413)
- 501 31. Chatterjee, P., 2020. Is India missing COVID-19 deaths?. *The Lancet*, 396(10252), p.657. DOI:
502 [https://doi.org/10.1016/S0140-6736\(20\)31857-2](https://doi.org/10.1016/S0140-6736(20)31857-2)

- 503 32. The official UK Government website for data and insights on Coronavirus (COVID-19). Public
504 Health England and NHSX. <https://coronavirus.data.gov.uk/> (2020).
- 505 33. National flu and COVID-19 surveillance report (week 45). Public Health England.
506 <https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports> (2020).
- 507 34. Population age structure by single year of age and sex for local authorities, counties, regions and
508 England as a whole, mid-2018 to mid-2043. Office for National Statistics.
509 <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/ukpopulationpyramidinteractive/2020-01-08> (2020).
- 511 35. Mohd Hanafiah, K., Groeger, J., Flaxman, A.D. and Wiersma, S.T., 2013. Global epidemiology
512 of hepatitis C virus infection: new estimates of age - specific antibody to HCV seroprevalence.
513 *Hepatology*, 57(4), pp.1333-1342. DOI: <https://doi.org/10.1002/hep.26141>
- 514 36. Stan Development Team, RStan: The R interface to Stan. <https://mc-stan.org/> (2020).
515

516 **Figure 2 — figure supplement 1. Marginal posterior distributions for parameters in the**
517 **constant infection fatality ratio model.** Vertical lines show median of distribution and grey shaded
518 region shows 95% CrI.

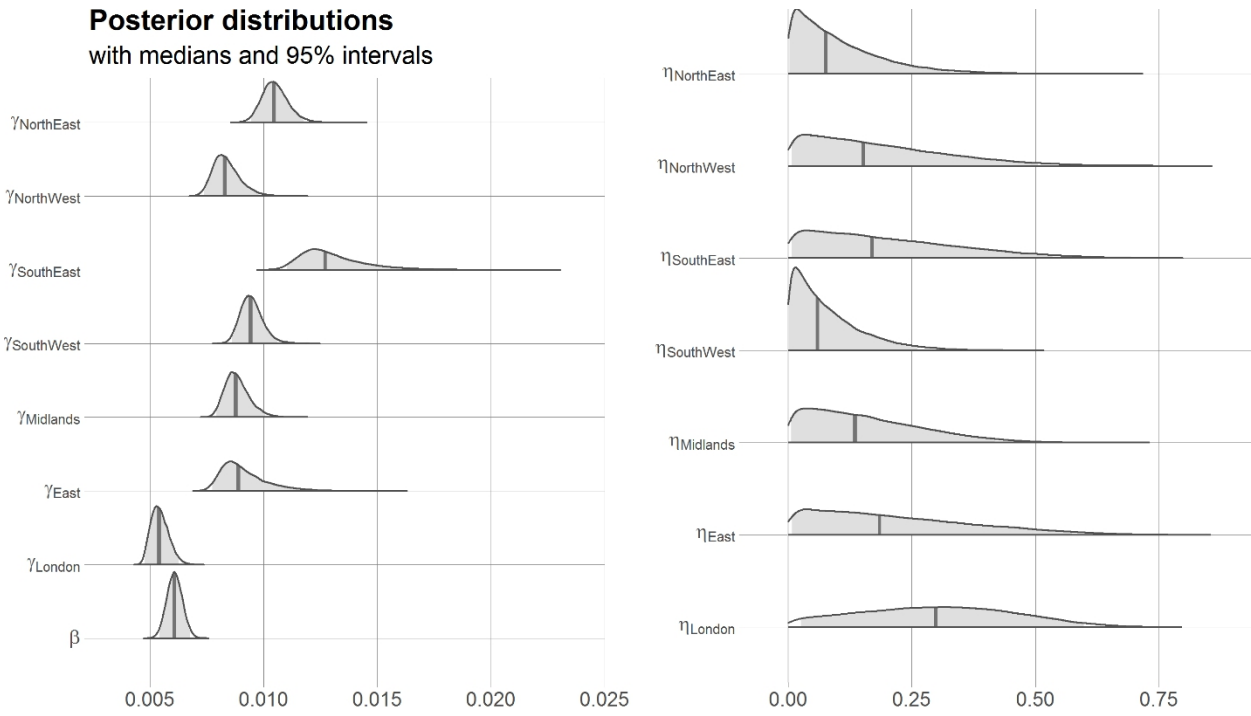
519



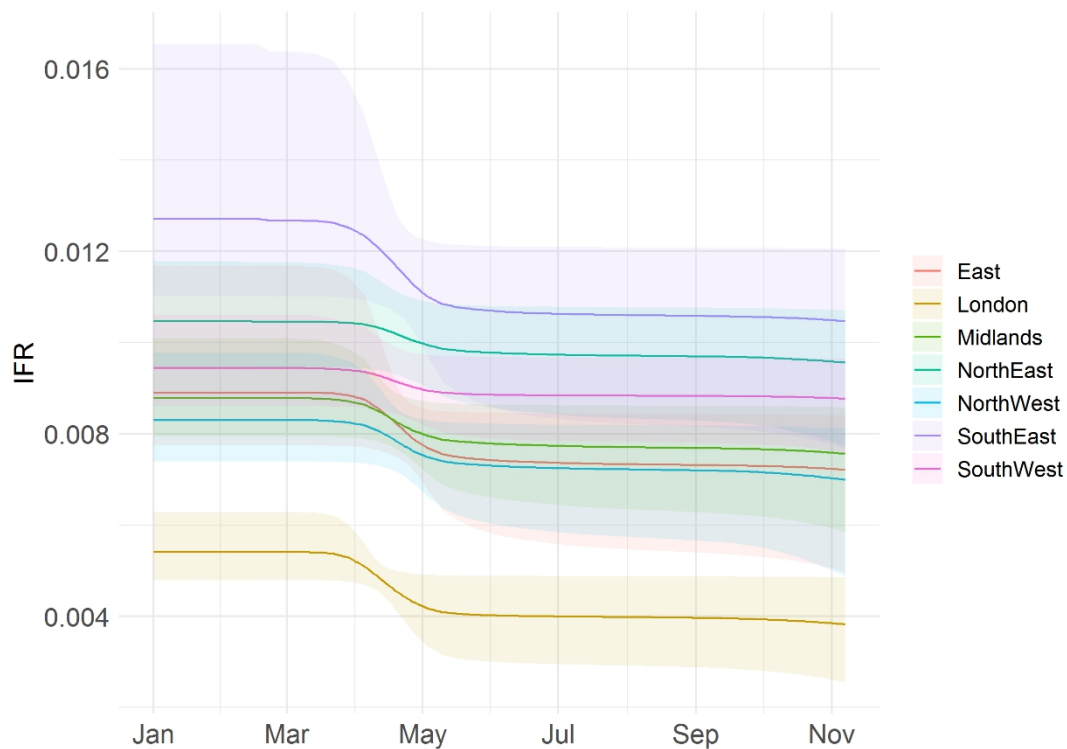
520

521 **Figure 3 — figure supplement 1. Marginal posterior distributions for parameters in the**
522 **time-varying infection fatality ratio model.** Vertical lines show median of distribution and grey
523 shaded region shows 95% CrI.

524

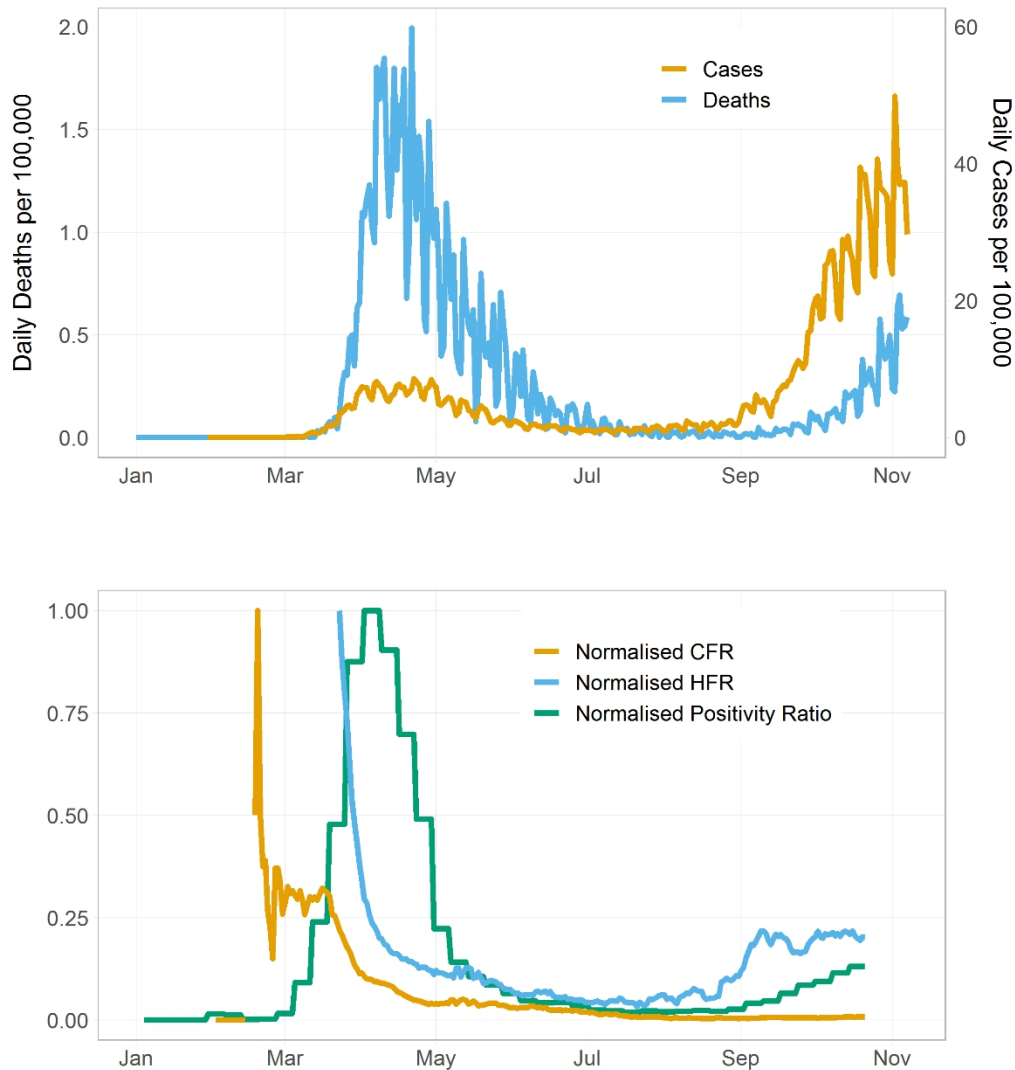


526 **Figure 3 – figure supplement 2. Posterior predictive distribution of time-varying infection**
527 **fatality ratio.** Solid line shows median and shaded region 95% CrI.



528

529 **Figure 3 —figure supplement 3. Relevant epidemiological metrics in England over the course**
530 **of the pandemic.** Top subplot shows COVID-19 cases and deaths (yellow and blue lines,
531 respectively) per 100,000 population in England from Feb 5 2020 to Nov 7, 2020. Bottom subplot
532 shows the normalized case fatality rate, hospital fatality rate and virus positivity (yellow, blue, and
533 green lines, respectively). We assumed fixed time lags of $\delta_p = 14$ days between PCR testing and
534 death and $\delta_h = 12$ days between PCR testing and hospitalization.



535

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

538

<i>Parameter</i>	n_{eff}	\hat{R}
β	12410	1
γ_{London}	18054	1
$\gamma_{NorthEast}$	29625	1
$\gamma_{SouthEast}$	23952	1
$\gamma_{NorthWest}$	28611	1
$\gamma_{SouthWest}$	22006	1
$\gamma_{Midlands}$	21334	1
γ_{East}	20992	1

539

540 **Table S2.** The effective sample size (n_{eff}) and the Gelman-Rubin diagnostic (\hat{R}) for the 15 model
 541 parameters in the time-varying infection fatality ratio model.
 542

<i>Parameter</i>	n_{eff}	\hat{R}
β	35229	1
γ_{London}	32184	1
η_{London}	26680	1
$\gamma_{NorthEast}$	38159	1
$\eta_{NorthEast}$	33166	1
$\gamma_{SouthEast}$	24036	1
$\eta_{SouthEast}$	24430	1
$\gamma_{NorthWest}$	32887	1
$\eta_{NorthWest}$	29965	1
$\gamma_{SouthWest}$	24430	1
$\eta_{SouthWest}$	31047	1
$\gamma_{Midlands}$	31943	1
$\eta_{Midlands}$	27558	1
γ_{East}	24703	1
η_{East}	25506	1

543

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

546

<i>Parameter</i>	<i>Median (95%CrI)</i>
β	0.0061 (0.0054 - 0.0068)
γ_{London}	0.0054 (0.0048- 0.0063)
η_{London}	0.30 (0.26 - 0.60)
$\gamma_{NorthEast}$	0.011 (0.0095 - 0.012)
$\eta_{NorthEast}$	0.077 (0.0029 - 0.33)
$\gamma_{NorthWest}$	0.0083 (0.0074 - 0.0098)
$\eta_{NorthWest}$	0.15 (0.0069 - 0.51)
$\gamma_{SouthWest}$	0.0094 (0.0086 - 0.011)
$\eta_{SouthWest}$	0.060 (0.0021- 0.26)
$\gamma_{SouthEast}$	0.0013 (0.011 - 0.017)
$\eta_{SouthEast}$	0.17 (0.0070 - 0.53)
$\gamma_{Midlands}$	0.0088 (0.0079- 0.010)
$\eta_{Midlands}$	0.14 (0.0060 - 0.42)
γ_{East}	0.0089 (0.0077 - 0.012)
η_{East}	0.18 (0.0080 - 0.57)

548