| 1 | Reconstructing the first COVID-19 pandemic wave with minimal data in the UK |
|----------|--|
| 2 | |
| 3 | Siyu Chen ^{1*} , Jennifer A Flegg ² , Katrina A Lythgoe ^{1,3} , Lisa J White ^{3*} |
| 4 | |
| 5 | ¹ Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield |
| 6 | Department of Medicine, University of Oxford, United Kingdom |
| 7 | ² School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia |
| 8 | ³ Department of Biology, University of Oxford, United Kingdom |
| 9 | |
| 10 | *Corresponding authors: <u>siyu.chen@ndm.ox.ac.uk</u> ; <u>lisa.white@biology.ox.ac.uk</u> |
| 11 | |
| 12 | Abstract |
| 13 | Accurate measurement of exposure to SARS-CoV-2 in the population is crucial for |
| 14 | understanding the dynamics of disease transmission and evaluating the impacts of |
| 15 | interventions. However, it is particularly challenging to achieve this in the early phase of a |
| 16 | pandemic because of the sparsity of epidemiological data. In our previous publication[1], we |
| 17 | developed an early pandemic diagnostic tool that can link minimum datasets: |
| 18 | seroprevalence, mortality and infection testing data to estimate the true exposure in |
| 19 | different regions of England and found levels of SARS-CoV-2 population exposure are |
| 20 | considerably higher than suggested by seroprevalence surveys. Here, we re-examined and |
| 21 | evaluated the model in the context of reconstructing the first COVID-19 epidemic wave in |
| 22 | England from three perspectives: validation from ONS Coronavirus Infection Survey, |
| 23 | relationship between model performance and data abundance and time-varying case |
| 24 | detection rate. We found that our model can recover the first but unobserved epidemic |
| 25 | wave of COVID-19 in England from March 2020 to June 2020 as long as two or three |
| 26 | serological measurements are given as model inputs additionally, with the second wave |
| 27 | during winter of 2020 validated by the estimates from ONS Coronavirus Infection Survey. |
| 28 | Moreover, the model estimated that by the end of October in 2020 the UK government's |
| 29 | official COVID-9 online dashboard reported COVID-19 cases only accounted for 9.1% (95%CrI |
| 30 | (8.7%,9.8%)) of cumulative exposure, dramatically varying across two epidemic waves in |
| 31 32 | England in 2020 (4.3% (95%CrI (4.1%, 4.6%)) vs 43.7% (95%CrI (40.7%, 47.3%))). NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice. |

33

34 Introduction

35 The COVID-19 pandemic has inflicted devastating effects on global populations and 36 economies [2, 3] and now still affecting countries in many different ways. Reviewing the 37 challenges posted by the COVID-19 pandemic and evaluating previous responses is vital important for future pandemic preparedness [4-7]. Accurate estimation of exposure 38 39 remains crucial for understanding the dynamics of disease transmission and assessing the 40 impacts of interventions along different stages of pandemic. However, this is particularly 41 challenging in the early phase since most of the characteristics of the pathogen are 42 unknown and at the same time epidemiological data are sparse. 43 Confirmed COVID-19 cases was typically the first type of data to be collected and reported 44 mostly due to the syndrome surveillance systems [8, 9]. However, it usually underestimates 45 the true exposure in the population because of the limited capacity of diagnoses, the 46 unsolid definition of cases, testing criteria and etc. Large-scale viral infection survey in the 47 community can help to solve the testing issue. For example, the UK Office for National Statistics (ONS) conducted a national wide COVID-19 viral testing survey, namely Covid 48 49 Infection Survey (CIS) [10] that successfully tracked the trajectories of COVID-19 infections in the community of UK since April of 2020. Because of its representative sampling across 50 51 households in the general population this study is recognised to have a strong power to 52 capture asymptomatic infections which might be missed out by symptomatic testing scheme 53 in the early pandemic and can provide reliable estimates of prevalence over time [11]. However, this study started collecting samples from April of 2020 and then reporting the 54 55 estimates of daily incidence from May of 2020 while the first death due to COVID-19 disease 56 in the UK was documented in February 2020 [12]. This implies that the transmission of 57 COVID-19 in the community began earlier than the survey, and the survey might not be able 58 to recover the early epidemic curve. 59 Serologic studies that measure how many people have antibodies against the virus are a

promising tool for pinning down the stage of the pandemic because of its ability of capturing past infections regardless of clinical symptoms [13]. 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've had a known COVID-19 infection were able to determine that antibodies

65 are only measurable up to 6–9 months [14-16], on average, varying across testing assay [17] 66 and antigen types [18]. The immediate implication is that serological studies will inevitably 67 under-estimate the number of people exposed, since some will have a lower antibody count 68 when the study is conducted and test negative. Linking multiple publicly available datasets, 69 we proposed a method that have been published previously [1] to estimate the true level of 70 exposure after considering the antibody decay. Here we further examined and evaluated 71 the model in the context of reconstructing the first COVID-19 pandemic from three perspectives: validation from ONS Infection Survey, relationship between model 72 73 performance and data abundance and time-varying case detection rate.

74

75 Result

76 **Reconstruction of the early epidemic**

77 In our previous paper, we presented a simple model to link together three key metrics for 78 evaluating the progress of an epidemic, applied to the context of SARS-CoV-2 in England: 79 antibody seropositivity, infection incidence and number of deaths. We use these three 80 metrics to estimate the antibody seroreversion rate and region-specific infection fatality 81 ratios. In doing so, the cumulative number of infections in England are estimated, showing 82 that cross-sectional seroprevalence data underestimate the true extent of the SARS-CoV-2 83 epidemic in England in the early pandemic. Estimates for the IgG (spike) seroreversion rate 84 and IFR are broadly consistent with other studies, which supports the validity of these 85 findings.

The model was set up based on the important observation about the COVID-19 infection 86 87 timeline that seroconversion in individuals who survive occurs at approximately the same 88 time as death for those who do not. Therefore, a simple ordinary differential equation (ODE) 89 was formulated to model the rate of change in the number of seropositive individuals in 90 different regions of England which will increase as new infections were generated that was 91 calculated by the daily number of deaths dividing by infection fatality ratio and will decrease 92 as antibody decay. The model predicted seropositive population were fitted to observed seroprevalence using a Bayesian observation model. 93 94 Validation from ONS Infection Survey

Comparing the incidence of SARS-CoV-2 in England estimated by our model with those
inferred by ONS Coronavirus Infection Survey (Figure 1), we found that our model could

97 reveal the first but unobserved epidemic wave of COVID-19 in England from March 2020 to

98 June 2020 additionally, with the second wave validated by the estimates from ONS Infection

- 99 Survey. Further, we found our model results were highly consistent with those using SEIRS
- 100 type compartmental models with time-varying force of infection [19, 20].
- 101

102 Relationship between model performance and data abundance

103 We then examined the relationship between model performance and data abundance - how 104 estimates of exposure from our model change with more serological data points being 105 added into the fitting procedure one by one over time (Figure 2). We found a highly robust 106 pattern of exposure across different regions of England was estimated in general. 107 Specifically, the model could only start estimating the interested quantities: exposure and 108 two parameters (infection fatality ratio and antibody decaying rate) when at least two 109 serological measurements from April to June 2020 in each region were given as inputs. 110 However, these estimates were already highly consistent with those when more serological 111 measurements were added although the credible bands were wider. The wide credible bands suggested a bigger uncertainty around the estimates when little information was 112 113 available. When three serological measurements in each of region were included the 114 estimates of exposure level became largely consistent at the results when all serological 115 measurements were used. This might be attributed to the timing of these third serological 116 measurements since then the seroprevalence in most regions started decreasing. With 117 more and more serological measurements being added, the credible bands of estimates of exposure gradually narrowed. 118

119

120 Time-varying case detection rate

121 While comparing the reported cases with the incidence estimated by our model (Figure 3),

122 we found the UK government's official COVID-9 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. If separating

129 the two waves from the first of August 2020, we found during January 2020 to August 2020 130 the case detection rate was only 4.3% (95%Crl (4.1%, 4.6%)) which increased to 43.7% 131 (95%Crl (40.7%, 47.3%)) during August 2020 to October 2020, highlighting the dominate 132 effect of testing effort in shaping the case curve in the early stage of a pandemic. The testing 133 issue, e.g. the limited capacity of tests and symptom-based testing strategy posted a big challenge for understanding the early pandemic. Viral surveys in the general population can 134 135 solve the sampling issue, but still have the problem of not sampling early on. Serological data even from some convenient samples, e.g., blood donors can help to pin down the 136 137 progress of the pandemic when antibody decay is teased out.

138

139 Discussion

140 Accurate reconstruction of exposure time series is necessary to assess how policies 141 influenced transmission over time, in particular when reporting is lagged, and multiple 142 interventions may have been undertaken in succession. For example, [21] made use of the 143 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 144 145 level in the population can be used to inform future intervention design, e.g., vaccination 146 prioritisation. For example, in the early stage of the COVID-19 vaccination campaign, when 147 dose supply and administrative capacity were initially limited worldwide, a modelling study [22] explored how uncertainty about previous exposure levels and about a vaccine's 148 149 characteristics affects the prioritization strategies for reducing deaths and transmission. This 150 model showed use of individual-level serological tests to redirect doses to seronegative 151 individuals improved the marginal impact of each dose while potentially reducing existing 152 inequities in COVID-19 impact.

153

Here, we evaluated a simple dynamic model that we published previously and
demonstrated its ability of reconstructing 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 model to serologic data that was 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

161 serum samples from blood donors have the promising power to provide primary

162 information of epidemic progress in a short timeframe especially during the emergency of a

- 163 new outbreak from a novel pathogen.
- 164

Because of the rigorous sampling design and robust estimation power ONS Covid Infection 165 Survey can almost be seen as a golden standard for estimating community prevalence. Our 166 167 model does not take any results or estimates from the survey as inputs, so the comparison 168 exercise that we conducted here between estimates of exposure from our model with ONS 169 Covid Infection Survey provides a real-world validation. Moreover, we showed the 170 modelling approach is a valuable early pandemic diagnostic tool and can clearly recover the 171 first epidemic wave that the survey was unable to capture because of late starting time. 172 Using the inferred daily incidence, we explicitly demonstrated the variation of case detection rates over two epidemic waves in England in 2020. It provides quantitative 173 174 information for studying the association between the capacity, behaviour, strategy of 175 testing with the epidemic evolution and further supported the argument that confirmed 176 cases largely underestimate the extent of disease transmission. 177

178 Moreover, the simple structure of the presented model avoids unnecessary complexity and 179 structure-based uncertainty in a full dynamic model where compartmental models 180 simulating the disease spread in different groups of population including susceptible, expose, infected and recovered are developed. The exercise of studying the model 181 182 performance against data abundance suggests the modelling results remain highly robust in 183 data sparse setting that is particularly important, for example, in Low- or Middle-Income 184 Country (LMIC).

185





Figure 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 [1] based on
Equation (1) and (2) in the Materials and Methods section. The orange lines show the
predictions of median daily incidence from ONS Coronavirus Infection Survey while the
orange shaded areas correspond to the 95% Crl.



- 197 **Figure 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
- 199 lines show the model predictions of median exposure and seroprevalence, respectively,
- while the shaded areas correspond to the 95% Crl.
- 201



202

203 Figure 3. Comparison between estimates of daily incidence with reported cases of SARS-204 CoV-2 in England and case detection rate. Here, all serological measurements were used in 205 the model fitting. In the top figure, the green lines show the predictions of median daily 206 incidence by our model based on Equation (1) and (2) in the Materials and methods section 207 while the shaded areas correspond to the 95% CrI. The red lines show the reported 208 confirmed cases in England downloaded from GUV.UK dashboard. In the bottom figure, the 209 blue lines show the estimates of median case reporting rate in England based on Equation 210 (3) and (4) while the shared areas correspond to the 95% Crl.

211

212 Materials and methods

213 Data sources

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

- 215 ONS estimated incidence
- 216 Office for National Statistics (ONS) launched Coronavirus (COVID-19) Infection Survey in
- 217 England on 26 April 2020 to estimate how many people across England, Wales, Northern

218 Ireland and Scotland would have tested positive for COVID-19 infection, regardless of

219 whether they report experiencing symptoms that is one of the primary goals of the survey.

- 220 The survey is based on a random sample of households to provide a nationally
- representative survey. Everyone aged 2 years and over in each household sample was asked
- to take a nose and throat swab for SARS-CoV-2 using reverse transcriptase polymerase chain
- reaction (RT-PCR). Every participant is swabbed once. they are then invited to have repeat
- tests every week for another four weeks and then monthly. More descriptions about the
- survey design can be found [23]. Using Bayesian multilevel generalised additive regression
- 226 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
- 228 2020 [10]. Combine the estimates of community prevalence and estimates of duration of
- 229 PCR testing positivity, the survey modelling team also published the estimates of daily
- 230 incidence based on a deconvolution model [23].
- 231 To conduct the comparison of estimates of incidence between our model and ONS survey,
- we retrieved the SARS-CoV-2 daily incidence in England in 2020 from the Office for National
- 233 Statistics (ONS) [11] on March 17, 2023 as shown in Figure 1.
- 234

235 Model estimated exposure

- 236 Cumulative exposure to SARS-CoV-2 in seven regions of England estimated by the model
- that we published before were obtained from [1]. Here, we firstly transformed and
- aggregated the cumulative exposure by region of England to daily incidence in England usingEquation (1) and Equation (2).
- 240

241 7-day average of reported COVID-19 cases in England

- 242 7-day average of reported COVID-19 daily cases in England in 2020 were retrieved from the
- 243 UK government's official COVID-9 online dashboard [12] on March 17, 2023 as shown in
- 244 Figure 3.
- 245

246 Method

- 247 We firstly calculated the incidence in England estimated by exposure model [1] by
- 248 computing the difference of cumulative exposure in two successive days and adding
- together to the whole England as shown in Figure 1 and Figure 3:

250

251
$$I_i(t) = E_i(t+1) - E_i(t), t = 1, 2, ..., n, i = 1, 2, ..., 7$$

252

253
$$I_{England}(t) = \sum_{i=1}^{7} I_i(t)$$

254

Here, $E_i(t)$ is the daily exposure at region *i* estimated by exposure model [1], *n* is the total

number of days from 1 January 2020 to 7 November 2020, i = 1, ... 7 represents London,

257 Southwest, Southeast, Northeast, Northwest, East, Midland. $I_{England}(t)$ represents the

258 daily incidence of England.

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

260
$$\bar{I}_{England}(t) = \frac{1}{7} \sum_{i=t-3}^{t+3} I_{England}(i), \qquad t = 4, 5, \dots, n-4$$

261

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

264 The estimated reporting ratio as shown in Figure 3 was calculated by

265

266
$$r(t) = \frac{\bar{I}_{England}(t)}{C(t)}$$

267

Equation (4)

Equation (3)

Equation (1)

Equation (2)

Here, *C* is the 7-day average reported cases in England from the UK government's official
COVID-9 online dashboard [12].

270 While testing the relationship between model performance and data abundance in Figure 2,

we firstly obtained all the data and codes from paper [1] and rerun the model by adding the

272 seroprevalence measurements one by one into the model.

273

274 **Acknowledgments:** The authors received no financial support for the research.

275

276 Author contributions:

277 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

278 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.
- 281
- 282 **Competing interests:** The authors have declared that no competing interests exist.
- 283
- 284 Data and materials availability: All codes and materials used in the analyses can be
- 285 accessed at: https://github.com/SiyuChenOxf/Exposure ONS-modelling. All parameter
- estimates and figures presented can be reproduced using the code provided. This work is
- 287 licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which
- 288 permits unrestricted use, distribution, and reproduction in any medium, provided the
- 289 original work is properly cited.
- 290

291 References

- 292
- Chen, S., et al., *Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys.* PLOS Computational Biology, 2021. **17**(9):
 p. e1009436.
- Aburto, J.M., et al., *Quantifying impacts of the COVID-19 pandemic through life- expectancy losses: a population-level study of 29 countries.* International journal of
 epidemiology, 2022. 51(1): p. 63-74.
- Ozili, P.K. and T. Arun, *Spillover of COVID-19: impact on the Global Economy*, in
 Managing Inflation and Supply Chain Disruptions in the Global Economy. 2023, IGI
 Global. p. 41-61.
- Metcalf, C.J.E., D.H. Morris, and S.W. Park, *Mathematical models to guide pandemic response.* Science, 2020. **369**(6502): p. 368-369.
- 3045.Aguas, R., et al., Modelling the COVID-19 pandemic in context: an international305participatory approach. BMJ global health, 2020. 5(12): p. e003126.
- Pagel, C. and C.A. Yates, *Role of mathematical modelling in future pandemic response policy.* bmj, 2022. **378**.
- 308 7. Bollyky, T.J., et al., Pandemic preparedness and COVID-19: an exploratory analysis of
 309 infection and fatality rates, and contextual factors associated with preparedness in
 310 177 countries, from Jan 1, 2020, to Sept 30, 2021. The Lancet, 2022. 399(10334): p.
 311 1489-1512.
- Kennedy, B., et al., *App-based COVID-19 syndromic surveillance and prediction of hospital admissions in COVID Symptom Study Sweden.* Nature Communications,
 2022. 13(1): p. 2110.
- Desjardins, M.R., Syndromic surveillance of COVID-19 using crowdsourced data. The
 Lancet Regional Health–Western Pacific, 2020. 4.

| 317 | 10. | Pouwels, K.B., et al., Community prevalence of SARS-CoV-2 in England from April to |
|-----|-----|---|
| 318 | | November, 2020: results from the ONS Coronavirus Infection Survey. The Lancet |
| 319 | 11 | Public Health, 2021. 6(1): p. e30-e38. |
| 320 | 11. | Office of National Statistics Coronavirus (COVID-19) Infection Survey OK,2023 |
| 321 | | <u>nttps://www.ons.gov.uk/peoplepopulationanacommunity/neathanasocialcare/cona</u> |
| 322 | | Itionsandalseases/bulletins/coronaviruscovia19in/ectionsurveypilot/latest#strengths- |
| 323 | 10 | ana-iimitations. 2023. |
| 324 | 12. | GOV.UK. <i>Coronavirus</i> (<i>COVID-19</i>) in the UK 2023; Available from: |
| 325 | | nttps://coronavirus.data.gov.uk/details/deaths?arealype=nation&arealyame=Engla |
| 326 | 4.2 | |
| 327 | 13. | Clapham, H., et al., Seroepidemiologic study designs for determining SARS-COV-2 |
| 328 | | transmission and immunity. Emerging Infectious Diseases, 2020. 26 (9): p. 1978. |
| 329 | 14. | Long, Qx., et al., Antibody responses to SARS-CoV-2 in COVID-19 patients: the |
| 330 | | perspective application of serological tests in clinical practice. MedRxiv, 2020: p. |
| 331 | | 2020.03. 18.20038018. |
| 332 | 15. | Ibarrondo, F.J., et al., Rapid decay of anti–SARS-CoV-2 antibodies in persons with |
| 333 | | <i>mild Covid-19.</i> New England Journal of Medicine, 2020. 383 (11): p. 1085-1087. |
| 334 | 16. | Wei, J., et al., Anti-spike antibody response to natural SARS-CoV-2 infection in the |
| 335 | | general population. Nature Communications, 2021. 12(1): p. 6250. |
| 336 | 17. | Böger, B., et al., Systematic review with meta-analysis of the accuracy of diagnostic |
| 337 | | tests for COVID-19. American journal of infection control, 2021. 49(1): p. 21-29. |
| 338 | 18. | Van Elslande, J., et al., Estimated half-life of SARS-CoV-2 anti-spike antibodies more |
| 339 | | than double the half-life of anti-nucleocapsid antibodies in healthcare workers. |
| 340 | | Clinical Infectious Diseases, 2021. 73 (12): p. 2366-2368. |
| 341 | 19. | Knock, E.S., et al., Key epidemiological drivers and impact of interventions in the 2020 |
| 342 | | SARS-CoV-2 epidemic in England. Science Translational Medicine, 2021. 13(602): p. |
| 343 | | eabg4262. |
| 344 | 20. | Russell, T.W., et al., Reconstructing the early global dynamics of under-ascertained |
| 345 | | COVID-19 cases and infections. BMC medicine, 2020. 18(1): p. 1-9. |
| 346 | 21. | Chen, S., et al., Estimating the effectiveness of shielding during pregnancy against |
| 347 | | SARS-CoV-2 in New York City during the first year of the COVID-19 pandemic. Viruses, |
| 348 | | 2022. 14 (11): p. 2408. |
| 349 | 22. | Bubar, K.M., et al., Model-informed COVID-19 vaccine prioritization strategies by age |
| 350 | | and serostatus. Science, 2021. 371 (6532): p. 916-921. |
| 351 | 23. | Coronavirus (COVID-19) Infection Survey technical article: Cumulative incidence of |
| 352 | | the number of people who have tested positive for COVID-19, UK: 22 April 2022. |
| 353 | | 2020. |
| 354 | | |
| | | |