

1 **Title:** The impact of health inequity on regional variation of COVID-19 transmission in
2 England

3 **Authors:** Thomas Rawson^{1*}, Wes Hinsley¹, Raphael Sonabend¹, Elizaveta Semenova²,
4 Anne Cori¹, Prof. Neil M Ferguson^{1,3}

5 **Affiliations:**

6 1. MRC Centre for Global Infectious Disease Analysis, Jameel Institute, School of Public
7 Health, Imperial College London, London, UK

8 2. Department of Computer Science, University of Oxford, Oxford, UK

9 3. National Institute for Health Research Health Protection Research Unit in Modelling
10 Methodology, Imperial College London, Public Health England, London School of Hygiene &
11 Tropical Medicine, London, UK

12 [*t.rawson@imperial.ac.uk](mailto:t.rawson@imperial.ac.uk)

13

14

15 **Abstract**

16 Considerable spatial heterogeneity has been observed in COVID-19 transmission across
17 administrative regions of England throughout the pandemic. This study investigates what
18 drives these differences. We constructed a probabilistic case count model for 306
19 administrative regions of England across 95 weeks, fit using a Bayesian evidence synthesis
20 framework. We include the mechanistic impact of acquired immunity, of spatial exportation of
21 cases, and 16 spatially-varying socio-economic, socio-demographic, health, and mobility
22 variables. Model comparison assesses the relative contributions of these respective
23 mechanisms. We find that regionally-varying and time-varying differences in week-to-week
24 transmission were definitively associated with differences in: time spent at home, variant-of-
25 concern proportion, and adult social care funding. However, model comparison
26 demonstrates that the mechanistic impact of these terms was of negligible impact compared
27 to the role of spatial exportation between regions. While these results confirm the impact of
28 some, but not all, measures of regional inequity in England, our work corroborates the
29 finding that observed differences in regional disease transmission during the pandemic were
30 predominantly driven by underlying epidemiological factors rather than the demography and
31 health inequity between regions.

32 **Author Summary**

33 During the COVID-19 pandemic, different geographic areas of England saw different
34 patterns in the number of confirmed cases over time. This study investigated whether
35 demographic differences between these areas (such as the amount of deprivation, the age
36 and ethnicity of the populations, or differences in where people spent their time) were linked
37 to these differences in disease transmission. We also considered whether this was
38 associated with the number of cases in neighbouring areas as well. Using a mathematical
39 model fit to multiple data streams, we discovered that a statistically significant link between
40 some demographic variables (time spent at home, COVID-19 variant, and the amount of
41 adult social care funding) and week-to-week transmission exists, but this relationship is very
42 small, and the influence of cases in neighbouring areas was far more impactful in explaining
43 differences in transmission between areas over time.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

44 Introduction

45 During the COVID-19 pandemic, measures of deprivation have been identified as impacting
46 health outcomes, with more deprived regions reporting higher COVID-19 attributed mortality,
47 both in England¹ and globally². Less well-understood is the impact these measures have on
48 disease incidence – confirmed cases. Descriptive studies early in the pandemic identified
49 that English regions with a higher Index of Multiple Deprivation (IMD) reported more cases of
50 COVID-19 than those with lower IMD scores^{3,4} during the first wave of disease incidence in
51 2020. However, such patterns do not persist throughout the entire epidemic, and for some
52 periods the opposite trend can now be observed (Figure 1B).

53 England is divided into Lower Tier Local Authorities (LTLA) – areas of social service
54 provisioning (see Supplementary Material 1.1). For these different areas, data is available on
55 the socio-demographic makeup – the average age, ethnic population proportions, population
56 density; on socio-economic metrics – median earnings, employment, education; and
57 epidemiological data throughout the pandemic – daily new cases, variant proportions,
58 COVID-19 support funding allocated, and mobility data recording time spent at different
59 locations. These variables vary greatly across LTLAs, and similarly disease incidence and
60 rates of infection have varied across LTLAs during the pandemic⁵. Figure 1 demonstrates
61 spatial variation in two covariates of interest, and mean per capita weekly incidence of
62 COVID-19 stratified by these variables (see Supplementary Material 1.6 for plots of other
63 covariates). Some variables change weekly (community mobility, variant proportion), others
64 change annually (funding allocation, income), while others are fixed by LTLA for the entire
65 duration.

66 Recognising the variation in disease incidence across administrative regions, the UK
67 government briefly implemented a tiered lockdown system on October 14th 2020⁶, where
68 more stringent rules on social mixing were applied to areas of the country with a greater
69 incidence of COVID-19. This system was retired the following month for a second nationwide
70 stay-at-home order. It remains unclear as to whether these observed epidemiological
71 differences can be explained solely by spatial drivers of disease spread, or whether the
72 intrinsic factors associated with each LTLA influenced the epidemic trajectory in each
73 respective LTLA. Hypothetically, for example, populations in wealthier LTLAs may have
74 been more able to work from home, or may have had more access to space to self-isolate in.
75 LTLAs with a higher proportion of elderly residents may have been more susceptible to
76 infection, or may have seen less social mixing. LTLAs that received more COVID-19 support
77 funding per head, may have subsequently achieved better disease suppression.

78 Here, we model the number of weekly pillar 2 (general population testing) PCR-confirmed
79 COVID-19 cases in 306 English LTLAs, for 95 weeks, from the week beginning May 10th
80 2020 to the week beginning February 27th 2022. We assume the number of weekly cases in
81 an LTLA is determined by the previous week's number of cases, plus a proportion of
82 imported infections from adjacent LTLAs controlled by a model parameter. Additional model
83 parameters then control the relative influence of 16 socio-economic and -demographic
84 variables, and time- and LTLA- varying terms on the observed increases and decreases in
85 cases. Model parameters are fit to English COVID-19 surveillance data by LTLA through a
86 Bayesian evidence synthesis framework.

87 Real-time modelling studies provided valuable insights and projections into key
88 epidemiological parameters throughout the pandemic⁷, through regular reports integrating
89 the latest epidemiological data. In this study we investigate how the composition of a
90 population, and population-level covariates, contributes to week-to-week transmission

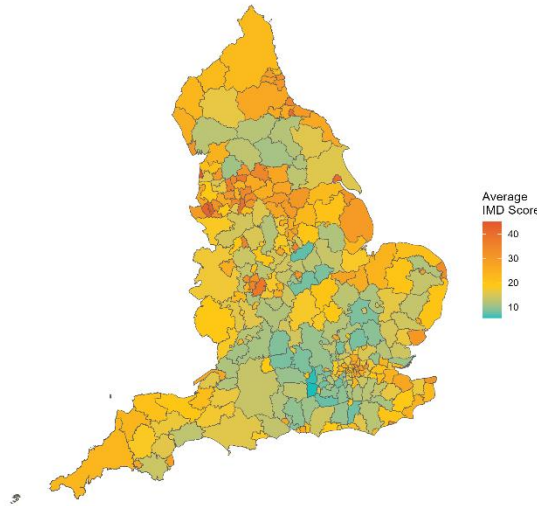
91 potential. Identification of such contributions would inform whether real-time-modelling efforts
92 could be improved in the future by integrating such socio-economic and -demographic data.

93

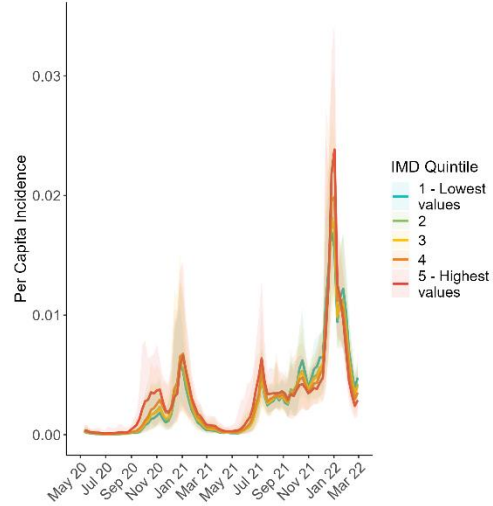
94

Variation in socio-demographic factors by LTLA and the respective stratification of per capita COVID-19 incidence

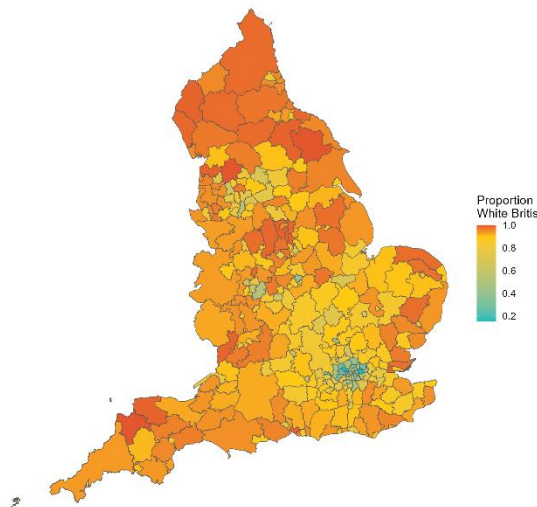
A Average Index of Multiple Deprivation (IMD) by LTLA



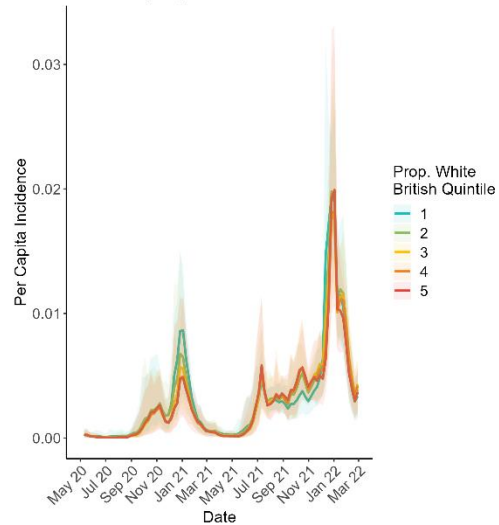
B Mean per capita weekly incidence of COVID-19 Cases in UK LTLAs by IMD Quintile



C White British Proportion of LTLA Population



D Mean per capita weekly incidence of COVID-19 Cases in UK LTLAs by Proportion White British Quintile



95

96 *Figure 1: Variation in socio-demographic factors by LTLA, and the respective differences in average per capita*
97 *incidence of COVID-19 cases when stratified by these socio-demographic factors. (A/C) Plots depict how IMD*
98 *and White British population proportion vary across the 306 English LTLAs we consider. (B/D) The 306*
99 *considered LTLAs are partitioned into quintiles (blue being the lowest values quintile and red the highest values*
100 *quintile) based on their (B) IMD scores and (D) White British population proportion respectively. Lines display the*
101 *mean per capita weekly incidence of COVID-19 across all LTLAs in each quintile. Shaded regions depict the 95%*
102 *quantiles. Quintile binning in plots B/D is for illustrative purposes - model fitting is performed to the continuous*
103 *measures presented in plots A/C. Boundary source: Office for National Statistics licensed under the Open*
104 *Government Licence v.3.0²⁹. Contains OS data © Crown copyright and database right [2024].*

105

106

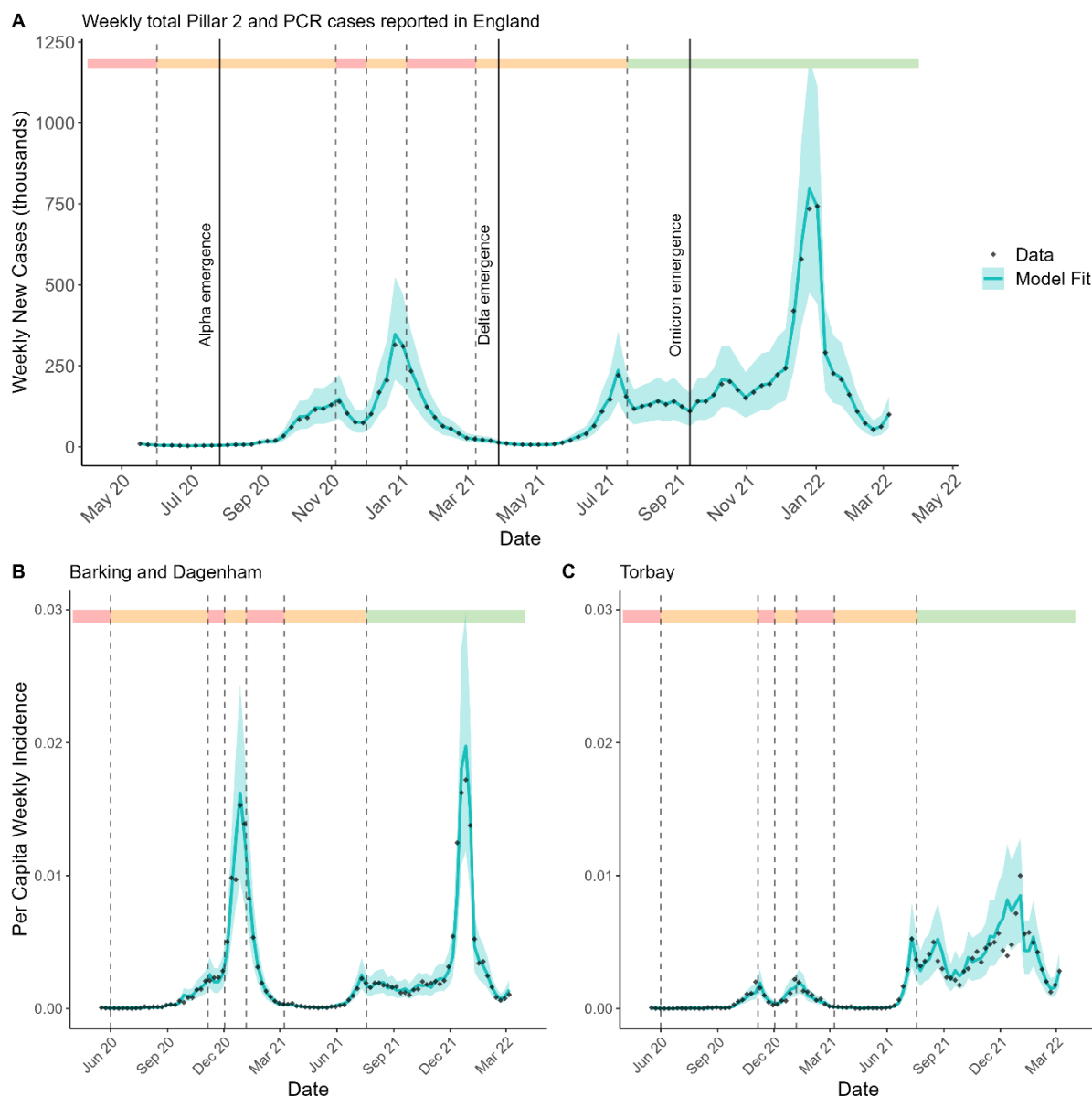
107

108 Results

109 Model fit

110 The model successfully captured the variation in LTLA-specific epidemic trajectories. Figure
111 2A sums the model fit across all 306 LTLAs, while plots 2B and 2C show the model fit, for
112 example, to the two LTLAs with the greatest variation in their epidemic trajectory as
113 assessed via dynamic time warping (DTW) distance⁸ (a metric for analysing similarity in time
114 series data). All LTLAs see broadly three principle epidemic waves, initiated by the
115 emergence of the Alpha, Delta, and Omicron variants respectively.

116



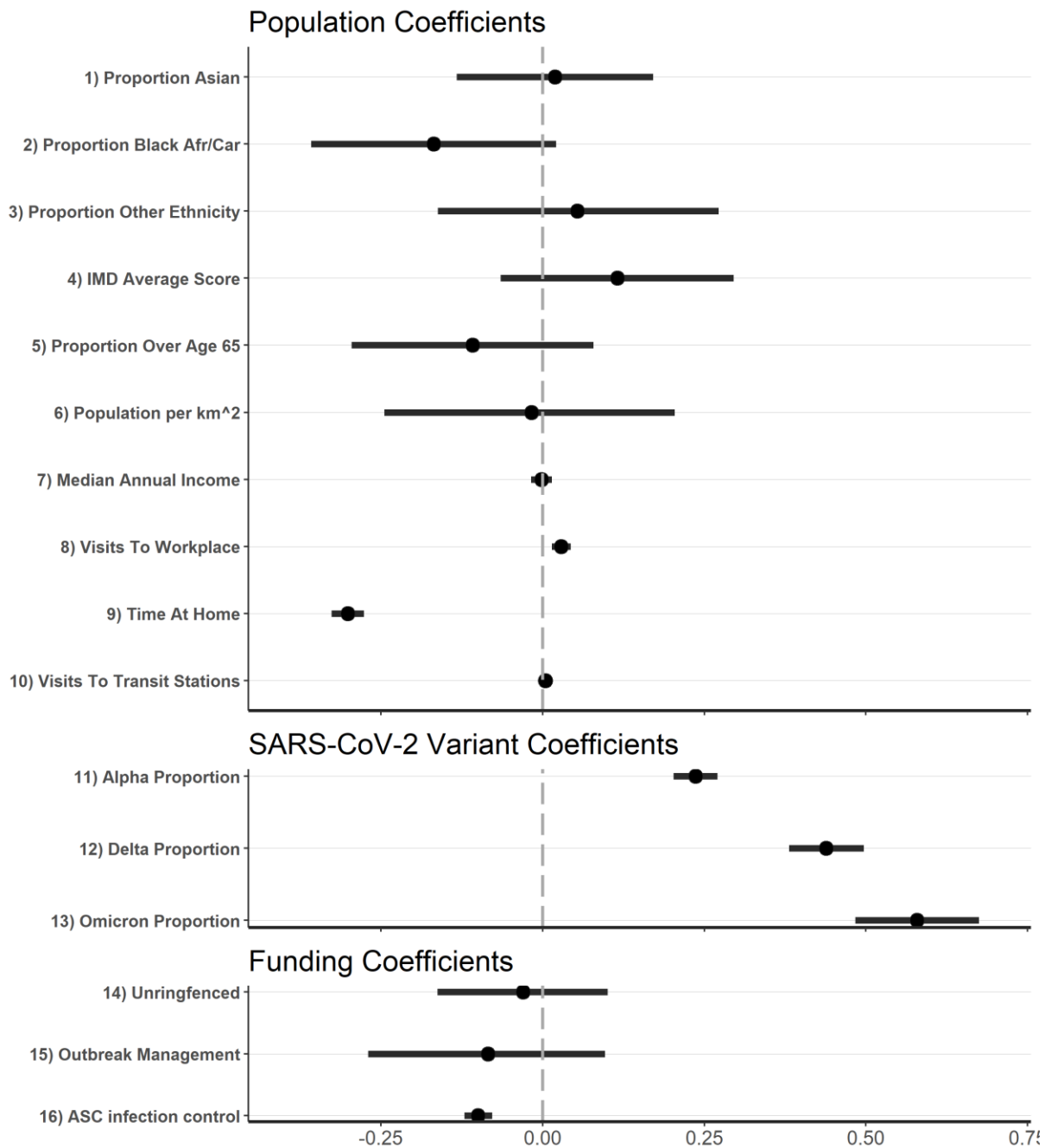
117

118 *Figure 2: Model fit to data, at national level and two LTLAs as examples. (A) The number of weekly new COVID-*
119 *19 cases in England, summing the model fit across all 306 LTLAs. (B/C) The per capita weekly incidence of*
120 *COVID-19 in two specific LTLAs with greatly different epidemic trajectories. Black dots show data and the blue*
121 *line represents mean model fit. Shaded blue regions depicts the 95% credible interval (CrI). Dashed lines*
122 *represent significant changes in nationwide non-pharmaceutical interventions imposed⁹. As shaded in the top*

123 *bar, red areas depict times of full “stay-at-home” orders, orange depicts partial restrictions on social mixing, green*
124 *depicts no barriers to social mixing.*

125 The effective reproduction number, $R_{i,t}^{eff}$, is an epidemiological parameter dictating the
126 number of secondary infections caused by a primary infection in LTLA i at week t . Hence,
127 when $R_{i,t}^{eff} > 1$, cases are observed to increase in LTLA i at week t . Likewise, when $R_{i,t}^{eff} <$
128 1 , cases are observed to decrease. Our model assumes that $R_{i,t}^{eff}$ is made up of three
129 principle elements. First, a time-varying random walk term, z_t , observed across all LTLAs,
130 capturing the impact of multiple time-varying factors such as changes to non-pharmaceutical
131 interventions (NPIs), vaccination uptake, school closures, and national holidays. Second, an
132 LTLA-varying error term, θ_i , to capture any unexplained intrinsic differences between the
133 reproduction number across LTLAs. Third, a term capturing the impact of 16 covariates of
134 interest – data compiled from multiple sources (see Section 1 of the Supplementary Material)
135 capturing: the population ethnicity proportions, the index of multiple deprivation (IMD)
136 scores, the population age proportions, the population densities, the median annual
137 incomes, the time spent at certain locations, the proportion of new COVID-19 variants, and
138 the amount of COVID-19 funding allocated, across all 306 LTLAs considered. The impact of
139 these variables is captured in the term $x_{i,t} \beta$, where $x_{i,t}$ represents the 16 covariates
140 introduced above for LTLA i at week t , and β is a model parameter of coefficients controlling
141 the relative contributions of each of the 16 covariates.

142 Figure 3 shows the mean and 95% CrI of the posterior distributions for these covariate
143 coefficients (parameter β above). Model covariates ($x_{i,t}$) were standardised to have mean 0
144 and standard deviation 1 before model fitting to enable comparison of relative covariate
145 coefficients.



146

147 *Figure 3: Posterior estimates of all covariate coefficients (parameter β). These values are separated into three*
 148 *categories: those capturing population effects, those capturing variants-of-concern, and those capturing funding*
 149 *allocations. Black dots represent the mean estimate, black lines the 95% CrI. The dashed grey line marks 0. A*
 150 *positive value indicates that the effective reproduction number increases with higher values of the associated*
 151 *covariate, a negative value indicates that the effective reproduction number decreases with higher values of the*
 152 *associated covariate.*

153

154 Unsurprisingly, the time spent at home was the strongest covariate effect (outside of COVID-
 155 19 variant) in determining changes to transmission. LTLAs and weeks where populations
 156 spent more time in residential areas saw reduced effective reproduction numbers. Similarly,
 157 LTLA-weeks with more visits to (non-home) workplaces saw increased reproduction
 158 numbers. Additionally, the LTLAs with greater allocations of Adult Social Care (ASC)
 159 infection control funding per head saw reduced reproduction numbers.

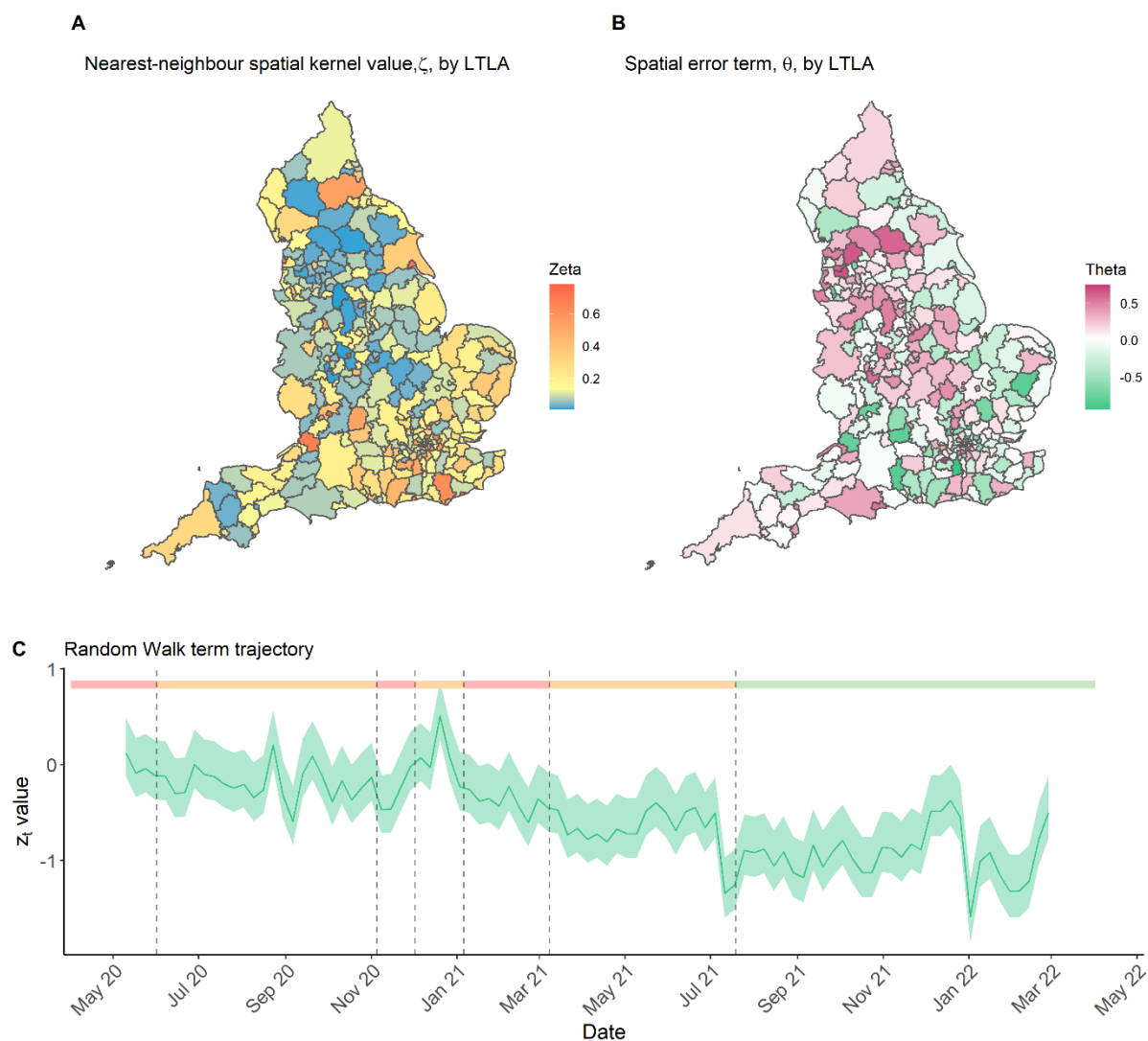
160 Each new variant was associated with a sequential increase in the reproduction number, in
161 alignment with similar studies into the transmission potential of each variant¹⁰.

162 Our analyses suggested there was no statistically significant impact on the effective
163 reproduction number by population ethnicity proportions, IMD, population proportion over the
164 age of 65, population density, median annual income, visits to transit stations, un-ringfenced
165 funding or Contain Outbreak Management Fund (COMF) funding allocated – all these
166 coefficients' 95% CrI overlaps 0 in Figure 3.

167 Our model also assumes that the number of weekly cases in an LTLA is not just driven by
168 the previous weekly number of cases in that LTLA, but that some new infections can be
169 triggered by infections in adjacent LTLAs. This is a process known as spatial exportation,
170 whereby a primary case in one LTLA may visit a neighbouring LTLA, and subsequently
171 cause a secondary infection outside of their home boundaries. We assume that the
172 proportion of spatial importations varies by LTLA, as some areas, like city centres, may
173 attract more visitors than other, more rural, LTLAs. The model parameter ζ_i is defined as the
174 proportion of all weekly cases in adjacent LTLAs that contribute secondary infections each
175 week in LTLA i .

176 Figure 4 shows the impact of all other model variables that contribute to the effective
177 reproduction number – the proportion of case importations from neighbouring LTLAs, ζ_i ,
178 (median value 0.134, interquartile range (IQR) 0.069 – 0.247) the LTLA-varying error terms
179 θ_i (median value 0.028, IQR -0.159 – 0.222), and the time-varying random walk trajectory,
180 z_t (median value -0.489, IQR -0.871 - -0.243). Figure 4A shows that the majority of LTLAs
181 imported only a small proportion of cases from neighbouring LTLAs - LTLAs shaded in blue
182 saw less than 20% of the total cases in LTLAs they share a border with causing onwards
183 infections within their own boundaries. Denser populated LTLAs like city centres have higher
184 ζ_i values in general, (areas shaded red in Figure 4A) demonstrating the increased
185 transmission risk caused by individuals travelling into population centres from more rural
186 LTLAs. The greatest contribution of the three terms comprising our $R_{i,t}^{eff}$ is thus the random
187 walk term, followed by the covariate impacts and LTLA-specific error term θ_i .

188



189

190 *Figure 4: Model variables contributing to the LTLA-varying and week-varying effective reproduction number. (A)*
 191 *Estimates of the LTLA-varying parameter ζ_i , denoting what proportion of the cases in adjacent LTLAs cause*
 192 *secondary cases in LTLA i the following week. (B) The remaining spatial error term, θ_i , capturing underlying*
 193 *differences in LTLA reproduction numbers not explained by the sixteen considered covariates $(x_{i,t-1} \beta)$. (C) The*
 194 *random walk term applied to all LTLAs capturing a baseline time-varying change to the reproduction number.*
 195 *Solid line shows the mean estimate and the shaded region the 95% CrI. Dashed lines and shaded bar at the top*
 196 *of the plot again mark areas of full (red), partial (orange) and no (green) stay-at-home orders. Boundary source:*
 197 *Office for National Statistics licensed under the Open Government Licence v.3.0²⁹. Contains OS data © Crown*
 198 *copyright and database right [2024].*

199 Figure 4B shows the spatial variation in the reproduction number that is not explained by the
 200 16 covariates $(x_{i,t-1})$. The apparent clustering of higher and lower θ_i values suggests this
 201 pattern may not simply be noise, but that there could be spatial variables yet to be identified
 202 that may also be influencing the transmission of COVID-19 across LTLAs in England.

203 Figure 4C shows the random walk trajectory over time. There are multiple time-varying
 204 aspects that will have influenced the epidemic in England. We specifically include COVID-19
 205 variant proportions, and our mobility covariates have been shown to capture the influence of
 206 NPIs over time¹¹. Figure 4C demonstrates a gradual decrease in the background effective
 207 reproduction number for all LTLAs, in line with the uptake in vaccination and changes in
 208 behaviour.

209

210

211 *Model Comparison*

212 To assess the relative contribution of each of these modelled mechanisms towards
 213 improving model fit, the model is fit multiple times with some mechanisms removed.

214 Alongside the main analysis (MA) presented above, where all model terms are included,
 215 three additional sensitivity analyses are presented: (SA1) we fit the model without any
 216 covariates included – i.e. β is hard-coded to 0, (SA2) we fit the model with no spatial
 217 exportation included (but covariates are still included) – i.e. ζ_i is hard-coded to 0, (SA3) we
 218 fit the model with no spatial exportation or covariates included – i.e. both β and ζ_i are fixed
 219 at 0. θ_i and z_t remain in all three sensitivities.

220 Model comparison is performed using the expected log pointwise predictive density (elpd)
 221 metric under a “leave future out” (LFO) cross-validation scheme, detailed in section 4.2 of
 222 the Supplementary Material. elpd (LFO) assesses the relative goodness of fit and predictive
 223 performance of different model versions by evaluating each model's ability to predict held-
 224 out future sections of the time series. Higher elpd (LFO) values (lower magnitude) indicate
 225 better model performance.

226 The elpd (LFO) values for the main analysis (MA) and three sensitivity analyses (SA1-3) are
 227 shown in Table 1 below. A greater elpd (LFO) value suggests a better model fit.

Model Formulation	elpd (LFO) value (point estimate / standard error)
(MA) Spatial importation included, covariates included.	-81,006 (SE 505)
(SA1) Spatial importation included, covariates excluded.	-81,124 (SE 502)
(SA2) Spatial importation excluded, covariates included.	-83,945 (SE 667)
(SA3) Spatial importation excluded, covariates excluded.	-84,018 (SE 666)

228 *Table 1: Model comparison via the elpd leave-future-out cross-validation measure. A greater value suggests a*
 229 *better model fit. “Spatial importation excluded” indicates that ζ_i is fixed to 0, and “covariates excluded” indicates*
 230 *that β is fixed to 0. When included these are both fit model parameters.*

231 The greatest elpd value, and hence the best performing model, is the main analysis (MA),
 232 containing both a nearest-neighbour spatial importation mechanism, and population / variant
 233 / funding covariates. This is to be expected as the sensitivities are nested models of the
 234 main analysis, and the elpd does not directly penalise increased model complexity.
 235 However, the improvement offered by including the model covariates (MA vs. SA1, and, SA2
 236 vs. SA3) is insignificant once the standard errors in the elpd estimates are considered, and
 237 therefore not worth the complexity trade-off of their inclusion. The improvement offered by
 238 the inclusion of spatial importation mechanisms however (an estimated elpd increase of
 239 2,939, MA vs. SA2) are significant, and support the inclusion of spatial importation as
 240 important to explaining variation in disease incidence between LTLAs.

241 Other model formulations are included as supplementary results for comparison in section 4
 242 of the Supplementary Material, including alternate spatial kernels, alternate data sources,
 243 univariate models, and consideration of different reporting assumptions.

244 **Discussion**

245 Our study explored the informative potential of multiple spatially varying health inequity,
 246 socio-demographic, and socio-economic factors on week-to-week transmission potential
 247 within a population. We investigated how these variables related to the observed differences
 248 in COVID-19 week-to-week transmission across 306 administrative regions of England over

249 a period of 95 weeks. In conclusion, the majority of these variables were not found to be
250 significantly associated with COVID-19 transmission; however, we did detect a significant
251 association for two population variables – the time spent at home, and the number of visits to
252 workplaces, and one funding variable – the amount of ASC infection control funding
253 allocated per head to an LTLA.

254 Starting with ethnicity, Black and South Asian populations have been shown to have
255 increased COVID-19 mortality risk¹². In their global systematic review of the impact of
256 ethnicity on COVID-19 health outcomes, Irizar et al. (2023)¹³ report mixed results when
257 comparing the risk of infection for Asian and “other” ethnicity populations with White majority
258 populations, in accordance with our results in Figure 3. However, they show a far more
259 conclusive increased relative risk in Black populations compared to white majority
260 populations. Our results suggest a mild decrease in week-to-week transmission potential for
261 LTLAs with a higher proportion of Black African / Caribbean residents on average (though
262 still statistically insignificant). One possible explanation is that put forward by Harris &
263 Brunsdon (2021)¹⁴, who show that the distribution of COVID-19 cases by ethnicity changes
264 over time in England, with Black populations reporting far higher relative incidence during the
265 peak of the first wave, before then changing to capturing the minority of cases proportionally.
266 Similarly, Mathur et al. (2021)¹⁵ report a lower risk of infection in Black populations compared
267 to White populations during the second wave. They demonstrate this is likely due to the
268 heterogeneity in spatial incidence over time. Our study both directly factors for spatial
269 heterogeneity in incidence and considers a longer time period than these studies
270 investigating this association, potentially explaining our finding. Section 4.3 of the
271 Supplementary Material shows that the mean coefficient magnitude for the population Black
272 African / Caribbean proportion is further reduced in the absence of all other covariates,
273 suggesting that some degree of covariate correlation is also influencing the estimated
274 importance of the covariate.

275 Crucially; however, care must be taken when comparing our results to those of community-
276 targeted infection risk studies. Population prevalence studies, such as the REal-time
277 Assessment of Community Transmission (REACT) studies, and those conducted via the
278 OpenSAFELY platform, directly investigate how COVID-19 prevalence differed by
279 demographic indicators like those considered in this study. Ward et al. (2021)¹⁶ identified a
280 three-fold increase in testing antibody-positive within Black populations compared to White
281 populations (reducing to two-fold when adjusted for confounding factors such as age, sex,
282 IMD quintile, household size). Mathur et al. (2021)¹⁵ identified a similar risk for the period
283 February 1st – August 3rd 2020, though this increased risk is not identified for the “second
284 wave” of September 1st – December 31st 2020. Such results should not be directly compared
285 to the findings of this study, which investigates a fundamentally different result – we do not
286 consider denominator populations, or individual-level infection results; rather, we consider
287 how the composition of a population contributes to week-to-week transmission potential.

288 Our results also show an inconclusive impact of IMD on transmission, though lean towards
289 higher reproduction numbers seen in more deprived areas. In their systematic review of
290 socioeconomic COVID-19 impacts, Benita et al. (2022)¹⁷ list only nine UK-specific studies,
291 and report a global trend of mixed and inconclusive findings as to the impact of poverty
292 metrics on COVID-19 infection. The trend we have shown in Figure 1B, of differences in
293 case incidence by deprivation quantiles seeming pronounced in some time periods, before
294 reversing in others is seen in multiple other countries^{18,19}.

295 The strong negative effect of the “time spent at home” variable is unsurprising given its
296 inherent epidemiological importance, and its direct impact on disease cases has been

297 demonstrated for multiple countries²⁰. While this variable predominantly changes temporally
298 in relation to NPI measures, strong variation is also observed across LTLAs at any given
299 time point (see Figure S15 of Supplementary Material). While it is possible that the impact of
300 some other covariates is captured within the “time spent at home” variable, i.e. LTLAs with
301 higher incomes also see more time spent at home on average, additional sensitivity
302 analyses exploring the removal of this variable in the Supplementary Material (section 4.3)
303 shows that results are broadly unchanged by their inclusion.

304 Of the three COVID-19 funding pools provided during the pandemic that we consider, only
305 the ASC infection control fund proved significant. COMF funds are provided for activities
306 such as targeted testing of hard-to-reach bodies, additional contact tracing, community
307 support, communication materials, as well as enforcement and compliance expenditure²¹.
308 The ASC infection control fund meanwhile was specifically for use in preventing onward
309 transmission in care home settings. As with many European countries, care homes in
310 England were hit particularly hard during the first wave of the epidemic in England²²,
311 motivating this specific fund. Our results show that this specific targeting of the most
312 vulnerable populations was effective in reducing transmission and is the first study to our
313 knowledge to investigate the associated impact of these funding provisions. All funding
314 allocations were informed by the specific health needs and population demographics of
315 respective LTLAs, meaning that some covariates such as IMD and age distribution will have
316 some degree of correlation.

317 To investigate the time-specific impact of these covariates, we conducted a sensitivity
318 analysis whereby the model was fit to three distinct subsections of the overall time series
319 (section 4.5 of the Supplementary Material). Our results are unchanged across specific time
320 periods, save for the significance of the ASC infection control fund disappearing for August
321 8th 2021 onwards, as would be expected, as this was when NPIs had been lifted.

322 Section 1.7 of the Supplementary Material presents the degree of autocorrelation present
323 amongst variables. This unavoidable aspect of the dataset is a limitation of the study;
324 however, we address this through multiple supplementary sensitivity analyses including
325 univariate model formulations. All fundamental results presented in this study are maintained
326 under these sensitivities, save for the impact of IMD, which does achieve statistical
327 significance upon exclusion of the other fifteen covariates.

328 We have not included vaccination directly, as the vaccination rollout was itself influenced by
329 the epidemic trajectory, with greater dose uptake encouraged in response to novel variants²³.
330 As such, since our model does not mechanistically include the effect of vaccination, nor the
331 impact of waning effectiveness, the random walk term will capture both the uptake and
332 impact of vaccination, but also unique temporal aspects such as public holidays, sporting
333 events, seasonal patterns, and others. We see in general a reduction in the reproduction
334 number over time, in line with the vaccine rollout, but also note increases in December likely
335 aligned with the Christmas holidays, and other adjustments such as an increase in June/July
336 2021 around the time of the UEFA European Football Championship²⁴, followed by a drop
337 after the event.

338 A caveat of this study is the heterogeneity within each LTLA for some covariates of interest.
339 LTLAs are areas of geographic administration and service provisioning, and as such differ in
340 population sizes. While many LTLAs are close to the median LTLA population size of
341 142,622 people (IQR 104,869 – 237,616, see section 1.8 of the Supplementary Material),
342 some outliers are considerably different, the largest being Birmingham with a population of
343 1,140,525. Heterogeneity in covariates, such as IMD, within these larger LTLAs can be
344 observed at the Lower layer Super Output Areas (LSOAs) scale, of 32,844 regions in

345 England, however COVID-19 case data at this scale is too sparse to model. Thus, the LTLA-
346 scale considered demonstrates a trade-off between demographic detail, data availability,
347 and modelling feasibility.

348 We also note a modelling assumption made whereby reported “first episodes” of disease
349 incidence in an LTLA contributed to fully immunising protection against onwards infection.
350 While protection against repeat infection was strong for the majority of the time period we
351 considered²⁵, this was likely to wane more against the Omicron variant. We explored this
352 modelling assumption through multiple sensitivity analyses where the model was fit to
353 different time periods, and where waning of acquired immunity was assumed, which are
354 presented in sections 4.5 and 4.6 of the Supplementary Material. Our model results were
355 unchanged in these analyses.

356 The overarching question motivating this study was whether population health and
357 demographic variables held informative potential such that their inclusion might improve real-
358 time modelling efforts that currently do not incorporate such data streams. Only a minority of
359 covariates were found to be impactful, and the improvement to model fit they offer is
360 insignificant. However, substantive improvements are offered by including mechanisms of
361 spatial spread. Detailed study of lineage exports by Kraemer et al. (2021)²⁶ have previously
362 demonstrated how human travel alone was able to explain the spatial heterogeneity
363 observed during the emergence of the Alpha variant in the UK, further supporting our
364 findings.

365 While real time modelling efforts are often limited by computational power and thus are
366 limited in what level of spatial disaggregation can be allowed for, we have effectively
367 demonstrated that mechanisms of case exportation are a worthwhile inclusion for improving
368 model fit, and that the benefits of incorporating broader socio-demographic data are unlikely
369 to be worth the time needed to gather and incorporate the relevant and up-to-date data.

370

371 **Methods**

372

373 *Study population and data*

374 Confirmed COVID-19 cases data were taken from the UK Health Security Agency (UKHSA)
375 national line list, collected by the Department of Health and Social Care as part of
376 surveillance activities and shared with us. Only pillar 2 cases (swab testing of the wider
377 population, not setting-specific) confirmed via PCR were used to account for changes in the
378 availability of lateral flow devices (LFDs). Cases were then aggregated by week (beginning
379 Monday) and LTLA. S-gene target failure (SGTF) data for each case was similarly obtained
380 from the UKHSA line list to identify the proportion of COVID-19 variants each week. The
381 cumulative number of first episodes by LTLA is obtained from the national data dashboard.
382 Population data on ethnicity, age, population density, income, was taken from Office for
383 National Statistics (ONS) reports. IMD data is taken from the Ministry of Housing,
384 Communities & Local Government (MHCLG) report on English Indices of Deprivation 2019
385 (IoD2019). Data on time spent at locations is taken from Google community mobility reports.
386 Data on COVID-19 funding allocations was taken from the associated Department for
387 Levelling Up, Housing and Communities reports.

388 In 2021 England was split into 309 LTLAs. Following the format used for the COVID-19
389 cases data release, we combine the LTLAs of Cornwall and Isles of Scilly; City of London

390 and Hackney. The Isle of Wight LTLA is removed. Thus, this study reports on 306 English
391 LTLAs total.

392 Detailed descriptions of all covariates are provided in section 1 of the Supplementary
393 Material.

394

395 *Epidemiological model and fitting*

396 Using Bayesian evidence synthesis inference we fit a probabilistic model to data $Y_{i,t}$, the
397 number of weekly pillar 2 PCR-confirmed COVID-19 cases in LTLA i at week t , via a
398 negative binomial distribution of the form

$$399 \quad Y_{i,t} \sim \text{NegBinom}(\mu_{i,t}, \phi)$$

400 for mean $\mu_{i,t}$ and overdispersion parameter ϕ . The mean takes the form

$$401 \quad \mu_{i,t} = (\lambda S_{i,t-1} \left(Y_{i,t-1} + \zeta_i \sum_{j \in \Omega_i} Y_{j,t-1} \right) \exp(x_{i,t-1} \beta + z_{t-1} + \theta_i))$$

402 where $S_{i,t-1}$ is an estimate of the proportion of the population of LTLA i that has no acquired
403 immunity in week $t - 1$, calculated as $1 - (\text{total number of recorded "first episodes" in LTLA } i$
404 $\text{by week } t / \text{LTLA } i \text{ population})$, and λ is a scaling factor parameter, between 0 and 1, scaling
405 the acquired-immunity lag term to account for the impact of under-reporting of first episodes,
406 incomplete protection of acquired immunity, and other nationwide scaling effects. Ω_i is the
407 set of all LTLAs that share a boundary with LTLA i , and ζ_i is a model parameter between 0
408 and 1 denoting the proportion of cases in neighbouring LTLAs which will cause secondary
409 cases in LTLA i . This represents a "nearest neighbours" spatial kernel formulation. $x_{i,t-1}$ is a
410 vector of the sixteen covariates considered in this study for LTLA i , at week $t - 1$. β is the
411 vector of coefficients capturing the relative impact of each covariate. z_{t-1} represents the $(t -$
412 $1)^{\text{th}}$ step in a Gaussian random walk process, and θ_i is an LTLA-specific error-term.

413 Heuristically, the left-hand side of the expression represents the number of cases
414 contributing to the next week's number of cases, and the right-hand side may be considered
415 an estimate of the time-varying reproduction number.

416 We model 95 weeks in total, from the week beginning May 10th 2020 to the week beginning
417 February 27th 2022, as case testing rates become inconsistent outside of this window.

418 Analyses were conducted in R version 4.1.1. The model was run in Stan via the `rstan`²⁷
419 package. All associated code is available in our online repository
420 (<https://github.com/thomrawson/Rawson-spatial-covid>). See section 5 of the Supplementary
421 Material for full details of package versions.

422 Further methodological detail is provided in sections 2 and 3 of the Supplementary Material.

423

424 *Sensitivity Analyses*

425 Model comparison is performed via the expected log pointwise predictive density (elpd)
426 score under a leave-future-out cross-validation process, detailed in section 4.2 of the
427 Supplementary Material.

428 As a supplementary result we also test the impact of using different data streams. While
429 SGTF is considered a highly accurate indicator for discerning between variants of concern
430 (VOC)²⁸, for completeness the Supplementary Material also presents analyses where variant
431 proportion is instead confirmed by whole genome sequencing (WGS). We also conduct a
432 sensitivity analysis where case data is expanded to include both pillar 1 and pillar 2 cases
433 and includes LFD cases. In both sensitivities, the results remain unchanged from the
434 inclusion of these data.

435 Sensitivity analyses exploring acquired-immunity assumptions are presented in section 4.6
436 of the Supplementary Material.

437 Sensitivity analyses exploring differences in reporting by ethnicity and IMD are presented in
438 section 4.7 of the Supplementary Material.

439 Other model formulations are included for comparison in section 4 of the Supplementary
440 Material, including alternate spatial kernels and univariate models.

441

442

443 **Inclusion and Ethics**

444 Ethics permission was sought for the study via Imperial College London's (London, UK)
445 standard ethical review processes and was approved by the College's Research
446 Governance and Integrity Team (ICREC reference 21IC6945). Patient consent was not
447 required as the research team accessed fully anonymised data only, which were collected as
448 part of routine public health surveillance activities by the UK Government.

449 The funders of this study had no role in the study design, data collection, data analysis, data
450 interpretation, or writing of the report.

451 **Data sharing agreement**

452 All code, scripts, and data used to create the results in this paper are available at:
453 <https://github.com/thomrawson/Rawson-spatial-covid>

454

455 **Contributors**

456 TR and NMF conceived the study. WH sourced and processed population data. TR and
457 NMF analysed the data. TR, RS, and ES contributed to the code base. TR created the
458 figures and wrote the original draft. All authors contributed to interpretation, investigation,
459 and writing (review and editing).

460 **Declaration of interests**

461 All authors declare no competing interests.

462

463 **Acknowledgments**

464 We thank all colleagues at the UK Health Security Agency (UKHSA, formerly Public Health
465 England) and front-line health professionals who have not only driven and continue to drive
466 the daily response to the COVID-19 epidemic in England but also provided the necessary
467 data to inform this study. This work would not have been possible without the dedication and

468 expertise of said colleagues and professionals. The use of pillar-2 PCR testing data and the
469 variant and mutation data was made possible thanks to UKHSA colleagues.

470 For the purpose of open access, the author has applied a 'Creative Commons Attribution'
471 (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

472 References

- 473
474
- 475 1 Kontopantelis E, Mamas MA, Webb RT, *et al.* Excess deaths from COVID-19 and
476 other causes by region, neighbourhood deprivation level and place of death during the
477 first 30 weeks of the pandemic in England and Wales: A retrospective registry study.
478 *Lancet Reg Heal - Eur* 2021; **7**: 100144.
- 479 2 McGowan VJ, Bambra C. COVID-19 mortality and deprivation: pandemic, syndemic,
480 and endemic health inequalities. *Lancet Public Health* 2022; **7**: e966–75.
- 481 3 Morrissey K, Spooner F, Salter J, Shaddick G. Area level deprivation and monthly
482 COVID-19 cases: The impact of government policy in England. *Soc Sci Med* 2021;
483 **289**: 114413.
- 484 4 Prats-Uribe A, Paredes R, Prieto-Alhambra D, Prats Uribe A. Ethnicity, comorbidity,
485 socioeconomic status, and their associations with COVID-19 infection in England: a
486 cohort analysis of UK Biobank data. *medRxiv* 2020; : 2020.05.06.20092676.
- 487 5 Elson R, Davies TM, Lake IR, *et al.* The spatio-temporal distribution of COVID-19
488 infection in England between January and June 2020. *Epidemiol Infect* 2021; **149**: e73.
- 489 6 Scott E. Covid-19 local alert levels: Three-tier system for England. House Lords Libr.
490 2020. [https://lordslibrary.parliament.uk/covid-19-local-alert-levels-three-tier-system-](https://lordslibrary.parliament.uk/covid-19-local-alert-levels-three-tier-system-for-england/)
491 [for-england/](https://lordslibrary.parliament.uk/covid-19-local-alert-levels-three-tier-system-for-england/) (accessed Oct 10, 2023).
- 492 7 Scientific Pandemic Influenza Group on Modelling Operational sub-group (SPI-M-O).
493 SPI-M-O: Summary of further modelling of easing restrictions – Roadmap Step 4 on
494 19 July 2021, 7 July 2021 - GOV.UK. Gov.uk. 2021.
495 [https://www.gov.uk/government/publications/spi-m-o-summary-of-further-modelling-](https://www.gov.uk/government/publications/spi-m-o-summary-of-further-modelling-of-easing-restrictions-roadmap-step-4-on-19-july-2021-7-july-2021)
496 [of-easing-restrictions-roadmap-step-4-on-19-july-2021-7-july-2021](https://www.gov.uk/government/publications/spi-m-o-summary-of-further-modelling-of-easing-restrictions-roadmap-step-4-on-19-july-2021-7-july-2021) (accessed Oct 10,
497 2023).
- 498 8 Giorgino T. Computing and visualizing dynamic time warping alignments in R: The
499 dtw package. *J Stat Softw* 2009; **31**: 1–24.
- 500 9 GOV.UK. Coronavirus: A history of English lockdown laws. House Commons Libr.
501 2023. <https://commonslibrary.parliament.uk/research-briefings/cbp-9068/> (accessed
502 Oct 10, 2023).
- 503 10 Perez-Guzman PN, Knock E, Imai N, *et al.* Epidemiological drivers of transmissibility
504 and severity of SARS-CoV-2 in England. *Nat Commun* 2023; **14**.
505 DOI:10.1038/s41467-023-39661-5.
- 506 11 Tully MA, McMaw L, Adlakha D, *et al.* The effect of different COVID-19 public
507 health restrictions on mobility : A systematic review. *PLoS One* 2021; **16**:
508 12::e0260919

- 509 12 Yates T, Summerfield A, Razieh C, *et al.* A population-based cohort study of obesity,
510 ethnicity and COVID-19 mortality in 12.6 million adults in England. *Nat Commun*
511 2022; **13**: 1–9.
- 512 13 Irizar P, Pan D, Kapadia D, *et al.* Ethnic inequalities in COVID-19 infection,
513 hospitalisation, intensive care admission, and death: a global systematic review and
514 meta-analysis of over 200 million study participants. *eClinicalMedicine* 2023; **57**:
515 101877.
- 516 14 Harris R, Brunson C. Measuring the exposure of Black, Asian and other ethnic
517 groups to COVID-infected neighbourhoods in English towns and cities. *Appl Spat*
518 *Anal Policy* 2021; : 621–46.
- 519 15 Mathur R, Rentsch CT, Morton CE, *et al.* Ethnic differences in SARS-CoV-2 infection
520 and COVID-19-related hospitalisation, intensive care unit admission, and death in 17
521 million adults in England: an observational cohort study using the OpenSAFELY
522 platform. *Lancet* 2021; **397**: 1711–24.
- 523 16 Ward H, Atchison C, Whitaker M, *et al.* SARS-CoV-2 antibody prevalence in England
524 following the first peak of the pandemic. *Nat Commun* 2021; **12**: 1–8.
- 525 17 Benita F, Rebollar-Ruelas L, Gaytan-Alfaro ED. What have we learned about
526 socioeconomic inequalities in the spread of COVID-19? A systematic review.
527 *Sustainable Cities and Society* 2022, **86**:104158.
- 528 18 Rohleder S, Costa DD, Bozorgmehr PK. Area-level socioeconomic deprivation, non-
529 national residency, and Covid-19 incidence: A longitudinal spatiotemporal analysis in
530 Germany. *eClinicalMedicine* 2022; **49**: 101485.
- 531 19 Landier J, Bassez L, Bendiane MK, *et al.* Social deprivation and SARS-CoV-2 testing:
532 a population-based analysis in a highly contrasted southern France region. *Front*
533 *Public Heal* 2023; **11**: 1–11.
- 534 20 Ilin C, Phan SA, Tai XH, Mehra S, Hsiang S, Blumenstock JE. Public mobility data
535 enables COVID - 19 forecasting and management at local and global scales. *Sci Rep*
536 2021; : 1–11.
- 537 21 Department of Health & Social Care. Contain Outbreak Management Fund: guidance –
538 financial year 2021 to 2022. Gov.uk. 2022.
539 [https://www.gov.uk/government/publications/contain-outbreak-management-fund-](https://www.gov.uk/government/publications/contain-outbreak-management-fund-2021-to-2022/contain-outbreak-management-fund-guidance-financial-year-2021-to-2022)
540 [2021-to-2022/contain-outbreak-management-fund-guidance-financial-year-2021-to-](https://www.gov.uk/government/publications/contain-outbreak-management-fund-2021-to-2022/contain-outbreak-management-fund-guidance-financial-year-2021-to-2022)
541 [2022](https://www.gov.uk/government/publications/contain-outbreak-management-fund-2021-to-2022/contain-outbreak-management-fund-guidance-financial-year-2021-to-2022) (accessed Oct 9, 2023).
- 542 22 Knock ES, Whittles LK, Lees JA, *et al.* Key epidemiological drivers and impact of
543 interventions in the 2020 SARS-CoV-2 epidemic in England. *Sci Transl Med* 2021; :
544 eabg4262.
- 545 23 Imai N, Rawson T, Knock ES, *et al.* Quantifying the effect of delaying the second

- 546 COVID-19 vaccine dose in England: a mathematical modelling study. *Lancet Public*
547 *Heal* 2023; **8**: e174–83.
- 548 24 Smith JAE, Hopkins S, Turner C, *et al.* Public health impact of mass sporting and
549 cultural events in a rising COVID-19 prevalence in England. *Epidemiol Infect* 2022;
550 **150**: 1–9.
- 551 25 Kojima, N., N. K. Shrestha, and J. D. Klausner. A systematic review of the protective
552 effect of prior SARS-CoV-2 infection on repeat infection. *Evaluation & the Health*
553 *Professions* 2021; **44**(4): 327-332.
- 554 26 Kraemer MUG, Hill V, Ruis C, *et al.* Spatiotemporal invasion dynamics of SARS-
555 CoV-2 lineage B.1.1.7 emergence. 2021; **895**: 889–95.
- 556 27 Stan Development Team. RStan: The R interface to Stan. 2023. <https://mc-stan.org/>
557 (accessed Oct 10, 2023).
- 558 28 McMillen T, Jani K, Robilotti E V., Kamboj M, Babady NE. The spike gene target
559 failure (SGTF) genomic signature is highly accurate for the identification of Alpha and
560 Omicron SARS-CoV-2 variants. *Sci Rep* 2022; **12**: 1–8.
- 561 29 Office for National Statistics (ONS), released 18 February 2022, ONS website,
562 Dataset, “Local Authority Districts (December 2021) Boundaries GB BFC” URL:
563 [https://geoportal.statistics.gov.uk/datasets/ons::local-authority-districts-december-](https://geoportal.statistics.gov.uk/datasets/ons::local-authority-districts-december-2021-boundaries-gb-bfc/about)
564 [2021-boundaries-gb-bfc/about](https://geoportal.statistics.gov.uk/datasets/ons::local-authority-districts-december-2021-boundaries-gb-bfc/about) [accessed: April 5th 2024]
- 565