

1 **Herd immunity thresholds for SARS-CoV-2 estimated from** 2 **unfolding epidemics**

3 Ricardo Aguas^{1†}, Rodrigo M. Corder^{2†}, Jessica G. King³, Guilherme Gonçalves⁴,
4 Marcelo U. Ferreira², M. Gabriela M. Gomes^{5,6*}

5 ¹ *Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine,*
6 *University of Oxford, Oxford, United Kingdom.*

7 ² *Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil.*

8 ³ *Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United*
9 *Kingdom.*

10 ⁴ *Unidade Multidisciplinar de Investigação Biomédica, Instituto de Ciências*
11 *Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal.*

12 ⁵ *Department of Mathematics and Statistics, University of Strathclyde, Glasgow,*
13 *United Kingdom.*

14 ⁶ *Centro de Matemática da Universidade do Porto, Porto, Portugal.*

15

16 † These authors contributed equally to this work.

17 * Correspondence and requests for materials should be addressed to M.G.M.G. (email:
18 gabriela.gomes@strath.ac.uk).

19 **As severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spreads, the**
20 **susceptible subpopulation declines causing the rate at which new infections occur**
21 **to slow down. Variation in individual susceptibility or exposure to infection**
22 **exacerbates this effect. Individuals that are more susceptible or more exposed**
23 **tend to be infected and removed from the susceptible subpopulation earlier. This**
24 **selective depletion of susceptibles intensifies the deceleration in incidence.**
25 **Eventually, susceptible numbers become low enough to prevent epidemic growth**
26 **or, in other words, the herd immunity threshold is reached. Here we fit**
27 **epidemiological models with inbuilt distributions of susceptibility or exposure to**
28 **SARS-CoV-2 outbreaks to estimate basic reproduction numbers (R_0) alongside**
29 **coefficients of individual variation (CV) and the effects of containment strategies.**
30 **Herd immunity thresholds are then calculated as $1 - (1/R_0)^{1/(1+CV^2)}$ or $1 -$**
31 **$(1/R_0)^{1/(1+2CV^2)}$, depending on whether variation is on susceptibility or**
32 **exposure. Our inferences result in herd immunity thresholds around 10-20%,**
33 **considerably lower than the minimum coverage needed to interrupt transmission**
34 **by random vaccination, which for R_0 higher than 2.5 is estimated above 60%.**
35 **We emphasize that the classical formula, $1 - 1/R_0$, remains applicable to**
36 **describe herd immunity thresholds for random vaccination, but not for**
37 **immunity induced by infection which is naturally selective. These findings have**
38 **profound consequences for the governance of the current pandemic given that**
39 **some populations may be close to achieving herd immunity despite being under**
40 **more or less strict social distancing measures.**

41 Scientists throughout the world have engaged with governments, health agencies, and
42 with each other, to address the ongoing pandemic of coronavirus disease (COVID-
43 19). Mathematical models have been central to important decisions concerning

44 contact tracing, quarantine, and social distancing, to mitigate or suppress the initial
45 pandemic spread¹. Successful suppression, however, may leave populations at risk to
46 resurgent waves due to insufficient acquisition of immunity. Models have thus also
47 addressed longer term SARS-CoV-2 transmission scenarios and the requirements for
48 continued adequate response². This is especially timely as countries apply, relax and
49 reapply lockdown measures with varying levels of success in tackling national
50 outbreaks.

51 Here we demonstrate that individual variation in susceptibility or exposure
52 (connectivity) accelerates the acquisition of immunity in populations. More
53 susceptible and more connected individuals have a higher propensity to be infected
54 and thus are likely to become immune earlier. Due to this selective immunization by
55 natural infection, heterogeneous populations require less infections to cross their herd
56 immunity threshold (HIT) than suggested by models that do not fully account for
57 variation. We integrate continuous distributions of susceptibility or connectivity in
58 otherwise basic epidemic models for COVID-19 which account for realistic
59 intervention effects and show that as coefficients of variation (CV) increase from 0 to
60 5, HIT declines from over 60%^{3,4} to less than 10%. We then fit these models to series
61 of daily new cases to estimate CV alongside basic reproduction numbers (R_0) and
62 derive the corresponding HITs.

63 **Effects of individual variation on SARS-CoV-2 transmission**

64 SARS-CoV-2 is transmitted primarily by respiratory droplets and modelled as a
65 susceptible-exposed-infectious-recovered (SEIR) process.

66 *Variation in susceptibility to infection*

67 Individual variation in susceptibility is integrated as a continuously distributed factor
68 that multiplies the force of infection upon individuals⁵ as

$$69 \quad \dot{S}(x) = -\lambda(x)xS(x), \quad (1)$$

$$70 \quad \dot{E}(x) = \lambda(x)x[S(x) + \sigma R(x)] - \delta E(x), \quad (2)$$

$$71 \quad \dot{I}(x) = \delta E(x) - \gamma I(x), \quad (3)$$

$$72 \quad \dot{R}(x) = (1 - \phi)\gamma I(x) - \sigma\lambda(x)xR(x), \quad (4)$$

73 where $S(x)$ is the number of individuals with susceptibility x , $E(x)$ and $I(x)$ are the
74 numbers of individuals who originally had susceptibility x and became exposed and
75 infectious, while $R(x)$ counts those who have recovered and have their susceptibility
76 reduced to a reinfection factor σ due to acquired immunity. δ is the rate of
77 progression from exposed to infectious, γ is the rate of recovery or death, ϕ is the
78 proportion of individuals who die as a result of infection and $\lambda(x) =$
79 $(\beta/N) \int [\rho E(y) + I(y)] dy$ is the average force of infection upon susceptible
80 individuals in a population of size N and transmission coefficient β . Standardizing so
81 that susceptibility distributions have mean $\int xg(x) dx = 1$, given a probability
82 density function $g(x)$, the basic reproduction number is

$$83 \quad R_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right), \quad (5)$$

84 where ρ is a factor measuring the infectivity of individuals in compartment E in
85 relation to those in I . The coefficient of variation in individual susceptibility $CV =$
86 $\sqrt{\int (x - 1)^2 g(x) dx}$ is explored as a parameter. Non-pharmaceutical interventions
87 (NPIs) designed to control transmission typically reduce β and hence R_0 . We denote
88 the resulting controlled reproduction number by R_c . The effective reproduction

89 number R_{eff} is another useful indicator obtained by multiplying R_c by the
90 susceptibility of the population, in this case written as $R_{eff}(t) =$
91 $R_c(t) \int xS(x, t) dx/N(t)$ to emphasize its time dependence.

92 Top panels of Figure 1 depict model trajectories fitted to suppressed epidemics
93 (orange) in 2 European countries (Belgium and England) assuming gamma distributed
94 susceptibility and no reinfection ($\sigma = 0$). We estimate: R_0 rounding 5 (Belgium) and
95 2.9 (England); individual susceptibility CV reaching 3.9 (Belgium) and 1.5 (England);
96 and overall intervention efficacy at maximum (typically during lockdown) being 60%
97 (Belgium) and 54% (England). Other estimated parameters are the day when NPIs
98 begin to affect transmission, after which we assume a linear intensification from
99 baseline over 21 days, remaining at maximum intensity for T_{max} days and linearly
100 lifting back to baseline over a period of T_{lift} days (both T_{max} and T_{lift} are estimated).

101 Denoting by $d(t)$ the proportional reduction in average risk of infection due to
102 interventions, in this case we obtain $R_c(t) = [1 - d(t)]R_0$ which is depicted for each
103 country, alongside $R_{eff}(t)$, underneath the respective epidemic trajectories. To assess
104 the potential for case numbers to overshoot if NPIs had not been applied, we rerun the
105 model with $d(t) = 0$ and obtain the unmitigated epidemics (black). Further details
106 are described in Methods.

107 *Variation in connectivity*

108 In a directly transmitted infectious disease, such as COVID-19, variation in exposure
109 to infection is primarily governed by patterns of connectivity among individuals. We
110 incorporate this in the system (Equations 1-4) assuming that individuals mix at
111 random (but see Methods for more general formulations that enable other mixing
112 patterns). Under random mixing and heterogeneous connectivity⁶, the force of

113 infection is written as $\lambda(x) = (\beta/N)(\int y[\rho E(y) + I(y)] dy / \int yg(y) dy)$, the basic
114 reproduction number is

$$115 \quad R_0 = (1 + CV^2)\beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right), \quad (6)$$

116 $R_c(t)$ is as above and $R_{eff}(t)$ is derived by a more general expression given in
117 Methods. The results from this basic variable connectivity model are shown in
118 Extended Data Figure 1). To allow for the possibility that social distancing (d) may
119 change not only the scale but also the shape of connectivity distributions, we consider
120 an extended model where connectivity is reformulated as $(1 - d)[1 +$
121 $(1 - d)(x - 1)]$ (Extended Data Figure 2). This does not change the way the model
122 is written but special care is needed in analysis and interpretation to account for the
123 dynamic contact patterns. The basic reproduction number, in particular, depends
124 explicitly on a CV which is now dependent on social distancing.

125 Applying this model to the same epidemics as before we obtain the bottom panels of
126 Figure 1. The estimated epidemiological parameters are: R_0 rounding 8.9 (Belgium)
127 and 3.7 (England); individual connectivity CV reaching 3.9 (Belgium) and 1.6
128 (England); and intervention efficacy during lockdown being 45% (Belgium) and 37%
129 (England). The reported CVs correspond to baseline contact patterns.

130 Comparing the two models, variation in connectivity systematically leads to higher R_0
131 estimates. The effect attributed to NPIs is lower in England than in Belgium. In both
132 countries, contacts appear to reach minimal levels around lockdown in April (R_c
133 plots), beginning to re-intensify thereafter but remaining below about half of pre-
134 lockdown levels by the end of the data series used for model fittings (solid portions of
135 the R_c line), consistently with the CoMix contact survey⁷. To illustrate how the

136 epidemic may unfold beyond the data used in this study we simulate a hypothetical
137 scenario whereby contacts continue to reactivate linearly until eventually reaching the
138 pre-pandemic baseline. This leads to low epidemic activity over a period that
139 coincides with the Summer months, followed by a resurgence as contacts intensify. This
140 resurgence is likely to be aggravated by seasonality as the Winter approaches,
141 although seasonality has not been included in our models. The percentage of the
142 population required to be immune to curb the epidemic and prevent future waves
143 when interventions are lifted appears remarkably conserved across models: 10 vs 11%
144 (Belgium); and 27 vs 25% (England).

145 We then fit to the same data the model constrained by $CV = 0$, i.e. assuming no
146 relevant heterogeneity in susceptibility or exposure to infection. The results are shown
147 in Figure 2. In this case epidemic potential is considerably larger and hence the effect
148 attributed to NPIs to fit the data must also be larger. The percentage of the population
149 required to be immune to prevent epidemic growth when interventions are lifted rises
150 to 73% (Belgium) and 63% (England).

151 **Herd immunity thresholds**

152 Individual variation in risk of acquiring infection is under selection by the force of
153 infection, whether individual differences are due to biological susceptibility,
154 exposure, or both. The most susceptible or exposed individuals are selectively
155 removed from the susceptible pool as they become infected and eventually recover
156 (some die), resulting in decelerated epidemic growth and accelerated induction of
157 immunity in the population. In essence, the *herd immunity threshold* defines the
158 percentage of the population that needs to be immune to reverse epidemic growth and
159 prevent future waves. When individual susceptibility or connectivity is gamma-

160 distributed and mixing is random, HIT curves can be derived analytically⁸ from the
161 model systems (Equations 1-4, with the respective forces of infections). In the case of
162 variation in susceptibility to infection we obtain

$$163 \quad HIT = 1 - \left[\frac{1 - \sigma R_0}{(1 - \sigma)R_0} \right]^{\frac{1}{1+CV^2}}, \quad (7)$$

164 while variable connectivity results in

$$165 \quad HIT = 1 - \left[\frac{1 - \sigma R_0}{(1 - \sigma)R_0} \right]^{\frac{1}{1+2CV^2}}. \quad (8)$$

166 In more complex cases (such as the variable connectivity with dynamic CV or
167 assortative mixing) HIT curves can be approximated numerically. Figure 3 shows the
168 expected downward trends in HIT and the sizes of the respective unmitigated
169 epidemics for SARS-CoV-2 without reinfection ($\sigma = 0$) as the coefficients of
170 variation are increased (here we adopt gamma distributions; for robustness of the
171 trends to other distributions see Gomes et al⁹). Values of R_0 and CV estimated for our
172 study countries are overlaid to mark the respective HIT and final epidemic sizes.
173 While herd immunity is expected to require 60-80% of a homogeneous population to
174 have been infected, at the cost of infecting almost the entire population if left
175 unmitigated, given an R_0 between 2.5 and 5, these percentages drop to the range 10-
176 20% when CV is roughly between 1.5 and 4.

177 When acquired immunity is not 100% effective ($\sigma > 0$) HITs are relatively higher
178 (Extended Data Figure 3). However, there is an upper bound for how much it is
179 reasonable to increase σ before the system enters a qualitatively different regime.
180 Above $\sigma = 1/R_0$ – the *reinfection threshold*^{10,11} – infection becomes stably endemic

181 and the HIT concept no longer applies. Respiratory viruses are typically associated
182 with epidemic dynamics below the reinfection threshold, characterized by seasonal
183 epidemics intertwined with periods of low detection.

184 Individual variation in exposure, in contrast with susceptibility, accrues from complex
185 patterns of human behaviour which have been simplified in our basic model. To
186 explore the scope of our results we generalised the models by relaxing some key
187 assumptions. First, we allowed connectivity distributions to change in shape (not only
188 scale) when subject to social distancing (Figure 1, bottom panels). Second, we enable
189 mixing to be assortative in the sense that individuals contact predominantly with those
190 of similar connectivity (Methods). Formally, an individual with connectivity x , rather
191 than being exposed uniformly to individuals of all connectivities y , has contact
192 preferences described by a normal distribution on the difference $y - x$. We find this
193 modification to have negligible effect on HIT (Extended Data Figure 4).

194 **Herd immunity thresholds and seroprevalence at sub-national levels**

195 As countries conduct immunological surveys to assess the extent of exposure to
196 SARS-CoV-2 in populations it is of practical importance to understand how HIT may
197 vary across regions. We have redesigned our analyses to address this question. Series
198 of daily new cases in 4 European countries (Belgium, England, Portugal and Spain)
199 were stratified by region. Fitting the models simultaneously to the multiple series
200 enabled the estimation of local parameters (R_0 and CV) while the effects of NPIs
201 were estimated at country level. Extended Data Figures 5-8 show how the modelled
202 epidemics fit the regional data and include an additional metric to describe the
203 cumulative infected percentage. These model projections are comparable to data from
204 seroprevalence studies such as in Spain¹². In addition to their practical utility these

205 results begin to unpack some of the variation in HIT within countries: Belgium (9.4-
206 11%), England (16-26%), Portugal (7.1-9.9%) and Spain (7.5-21%). The sub-national
207 stratification has also enabled the application of the models to countries where the
208 epidemic was sufficiently asynchronous across regions to compromise the ability of
209 the models to fit the aggregate data (Portugal and Spain).

210 **Discussion**

211 The concept of *herd immunity* was developed in the context of vaccination
212 programs^{13,14}. Defining the percentage of the population that must be immune to
213 cause infection incidences to decline, HITs constitute useful targets for vaccination
214 coverage. In idealized scenarios of vaccines delivered at random and individuals
215 mixing at random, HITs are given by a simple formula ($1 - 1/R_0$) which, in the case
216 of SARS-CoV-2, suggests that 60-80% of randomly chosen subjects of the population
217 would need be immunized to halt spread considering estimates of R_0 between 2.5 and
218 5. This formula does not apply to infection-induced immunity because natural
219 infection does not occur at random. Individuals who are more susceptible or more
220 exposed are more prone to be infected and become immune, providing greater
221 community protection than random vaccination¹⁵. In our model, the HIT declines
222 sharply when coefficients of variation increase from 0 to 2 and remains below roughly
223 20% for more variable populations. The magnitude of the decline depends on what
224 property is heterogeneous and how it is distributed among individuals, but the
225 downward trend is robust as long as susceptibility or exposure to infection are
226 variable (Figure 3 and Extended Data Figures 4) and acquired immunity is efficacious
227 enough to keep transmission below the reinfection threshold (Extended Data Figure
228 3).

229 Several candidate vaccines against SARS-CoV-2 are showing promising safety and
230 immunogenicity in early-phase clinical trials^{16,17}, although it is not yet known how
231 this will translate into effective protection. We note that the reinfection threshold^{10,11}
232 informs not only the requirements on naturally acquired immunity but, similarly, it
233 sets a target for how efficacious a vaccine needs to be in order to effectively interrupt
234 transmission. Specifically, given an estimated value of R_0 we should aim for a
235 vaccine efficacy of $1 - 1/R_0$ (60% or 80% if R_0 is 2.5 or 5, respectively), which
236 seems to be materialising according to preliminary results from phase 3 trials.

237 Heterogeneity in the transmission of respiratory infections has traditionally focused
238 on variation in exposure summarized into age-structured contact matrices. Besides
239 overlooking differences in susceptibility given exposure, the aggregation of
240 individuals into age groups reduces coefficients of variation. We calculated CV for
241 the landmark POLYMOD matrices^{18,19} and obtained values between 0.3 and 0.5.
242 Recent studies of COVID-19 integrated contact matrices with age-specific
243 susceptibility to infection (structured in three levels)²⁰ or with social activity (three
244 levels also)²¹ which, again, resulted in coefficients of variation less than unity. We
245 show that models with coefficients of variation of this magnitude would appear to
246 differ only moderately from homogeneous approximations when compared with our
247 estimates, which are consistently above 1 in England and above 2 in Belgium,
248 Portugal and Spain. In contrast with reductionist procedures that aim to reconstruct
249 variation from correlate markers left on individuals (such as antibody or reactive T
250 cells for susceptibility, or contact frequencies for exposure), we have embarked on a
251 holistic approach designed to infer the whole extent of individual variation from the
252 imprint it leaves on epidemic trajectories. Our estimates are therefore expected to be
253 higher and should ultimately be confronted with more direct measurements as these

254 become available. Adam et al²² conducted a contact tracing study in Hong Kong and
255 estimated a coefficient of variation of 2.5 for the number of secondary infections
256 caused by individuals, attributing 80% of transmission to 20% of cases. This
257 statistical dispersion has been interpreted as reflecting a common pattern of contact
258 heterogeneity which has been corroborated by studies that specifically measure
259 mobility²³. According to our inferences, 20% of individuals may be responsible for
260 47-94% infections depending on model and country. In parallel, there is accumulating
261 evidence of individual variation in the immune system's ability to control SARS-
262 CoV-2 infection following exposure^{24,25}. While our inferences serve their purpose of
263 improving accuracy in model predictions, diverse studies such as these are necessary
264 for developing interventions targeting individuals who may be at higher risk of being
265 infected and propagating infection in the community.

266 Country-level estimates of R_0 reported here (Figures 1, 2) are in the range 3-5 when
267 individual variation in susceptibility is factored and 4-9 when accounting for variation
268 in connectivity. The homogeneous version of our models would have estimated R_0
269 around 2.7 and 3.7, in line with other studies²⁶. Estimates for England suggest lower
270 baseline R_0 and lower CV in comparison with Belgium. The net effect is a slightly
271 higher HIT in England which nevertheless we estimate around 25-27% depending on
272 characteristics of the individual variation. NPIs reveal less impact when heterogeneity
273 is contemplated (37-60%), appearing to inflate and agree with Flaxman et al²⁶ when
274 homogeneity assumptions are made (66-78%), although this does not affect the HIT
275 which relates to pre-pandemic societies.

276 More informative than reading these numbers, however, is to look at simulated
277 projections of how cases may unfold over future months (Figures 1, 2). In both

278 countries, when individual variation is considered, we foresee HIT being achieved
279 and the COVID-19 epidemic being mostly resolved by the end of 2020. Under the
280 homogeneous approximation, however, epidemic potential appears to justify
281 considerations of second lockdowns to flatten epidemic curves. Determining the level
282 of confidence on either scenario is highly relevant to policy decisions and hence of
283 central importance to public health. According to the Akaike information criterion
284 (Extended Data Table 5) data favours the models that account for individual variation
285 and a more natural resolution for the pandemic. Models that account for individual
286 variation in susceptibility score particularly well and amongst those that account for
287 variation in exposure those that allow this variation to reduce with social distancing
288 perform better.

289 Looking back, we conclude that NPIs had a crucial role in halting the growth of the
290 initial wave between February and April irrespective of individual variation. Although
291 the most extreme lockdown strategies may not be sustainable for longer than a month
292 or two, they proved effective at preventing overshoot, keeping cases within health
293 system capacities, and may have done so without impairing the development of herd
294 immunity.

295

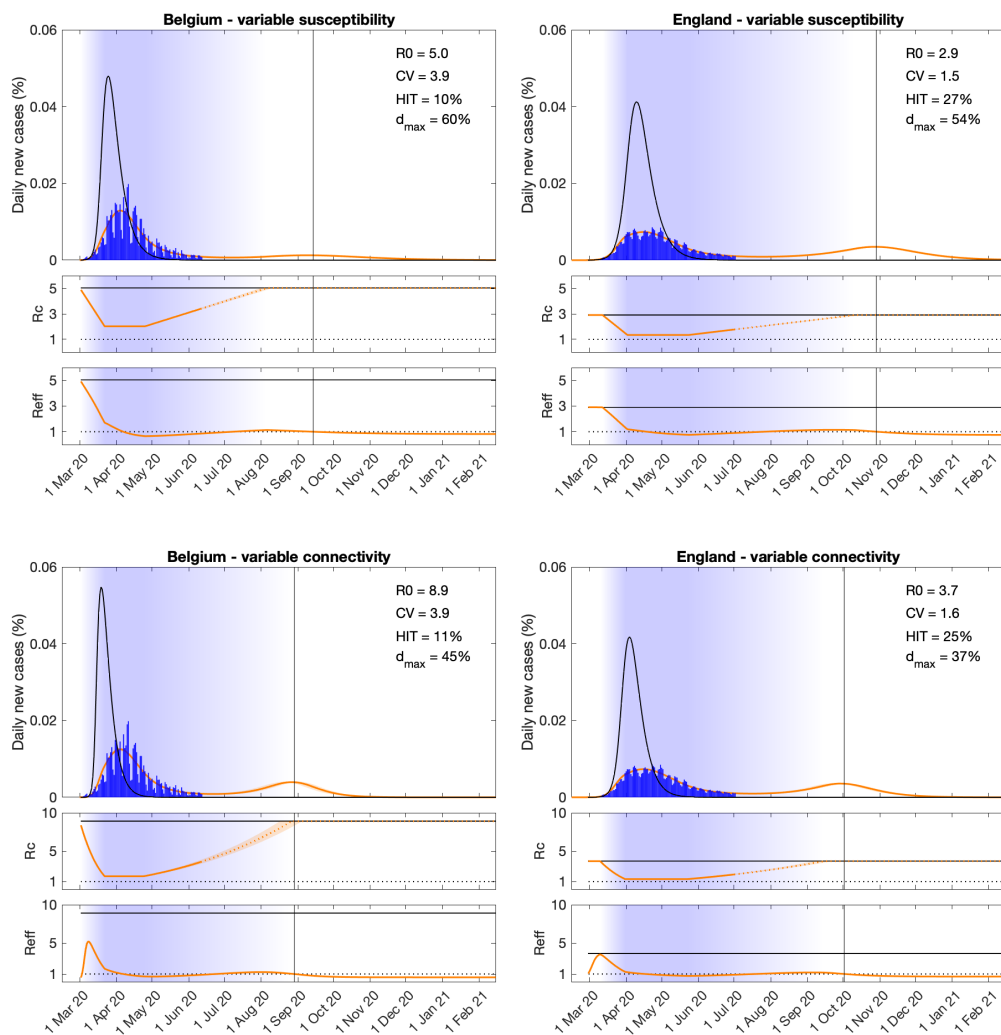
- 296 1. Ferguson, N. M., et al. Impact of non-pharmaceutical interventions (NPIs) to reduce
297 COVID-19 mortality and healthcare demand (Imperial College COVID-19 Response
298 Team, 2020). 10.25561/77482.
- 299 2. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the
300 transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**,
301 860-868 (2020).

- 302 3. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic
303 and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a
304 modelling study. *Lancet* **395**, 689-697 (2020).
- 305 4. Kwok, K. O., Lai, F., Wei, W. I., Wong, S. Y. S. & Tang, J. Herd immunity – estimating
306 the level required to halt the COVID-19 epidemics in affected countries. *J. Infect.* **80**,
307 e32-e33 (2020).
- 308 5. Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. On the definition and computation
309 of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous
310 populations. *J. Math. Biol.* **28**, 365-382 (1990).
- 311 6. Pastor-Satorras, R. & Vespignani, A. Epidemic dynamics and endemic states in
312 complex networks. *Phys. Rev. E* **63**, 066117 (2001).
- 313 7. Jarvis, C. I., et al. Quantifying the impact of physical distance measures on the
314 transmission of COVID-19 in the UK. *BMC Medicine* **18**, 206 (2020).
- 315 8. Montalbán, A., Corder, R. M. & Gomes, M. G. M. Herd immunity under
316 individual variation and reinfection. *arXiv* 200800098v2. (2020).
- 317 9. Gomes, M. G. M., et al. Individual variation in susceptibility or exposure to SARS-CoV-
318 2 lowers the herd immunity threshold. *medRxiv* 10.1101/2020.04.27.20081893 (2020).
- 319 10. Gomes, M. G. M., White, L. J. & Medley, G. F. Infection, reinfection, and vaccination
320 under suboptimal immune protection: Epidemiological perspectives. *J. Theor. Biol.* **228**,
321 539-549 (2004).
- 322 11. Gomes, M. G. M., Gjini, E., Lopes, J. S., Souto-Maior, C. & Rebelo, C. A theoretical
323 framework to identify invariant thresholds in infectious disease epidemiology. *J. Theor.*
324 *Biol.* **395**, 97-102 (2016).
- 325 12. Pollán, M., et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide,
326 population-based seroepidemiological study. *Lancet* 10.1016/s0140-6736(20)31483-5
327 (2020).

- 328 13. Gonçalves, G. Herd immunity: recent uses in vaccine assessment. *Expert Rev. Vaccines*
329 7, 1493-1506 (2008).
- 330 14. Fine, P., Eames, K. & Heymann, D. L. “Herd immunity”: a rough guide, *Clin. Infect.*
331 *Dis.* **52**, 911-916 (2011).
- 332 15. Ferrari, M. J., Bansal, S., Meyers, L. A. & Bjornstad, O. N. Network frailty and the
333 geometry of herd immunity. *Proc. R. Soc. B* **273**, 2743-2748 (2006).
- 334 16. Folegatti, P. M., et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine
335 against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised
336 controlled trial. *Lancet* 10.1016/S0140-6736(20)31604-4 (2020).
- 337 17. Zhu, F.-C., et al. Immunogenicity and safety of a recombinant adenovirus type-5-
338 vectored COVID-19 vaccine in healthy adults aged 18 years or older: a randomised,
339 double-blind, placebo-controlled, phase 2 trial. *Lancet* 10.1016/S0140-6736(20)31611-1
340 (2020).
- 341 18. Mossong, J., et al. Social contacts and mixing patterns relevant to the spread of
342 infectious diseases. *PLOS Med.* **5**, e74 (2008).
- 343 19. Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using
344 contact surveys and demographic data. *PLOS Comput. Biol.* **13**, e1005697 (2017).
- 345 20. Zhang, J., et al. Changes in contact patterns shape the dynamics of the COVID-19
346 outbreak in China. *Science* **368**, 1481-1486 (2020).
- 347 21. Britton, T., Ball, F. & Trapman, P. A mathematical model reveals the influence of
348 population heterogeneity on herd immunity to SARS-CoV-2. *Science*
349 10.1126/science.abc6810 (2020).
- 350 22. Adam, D., et al. Clustering and superspreading potential of severe acute respiratory
351 syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. 10.21203/rs.3.rs-
352 29548/v1
- 353 23. Eubank, S., et al. Modelling disease outbreaks in realistic urban social networks. *Nature*
354 **429**, 180-184 (2004).

- 355 24. Grifoni, A., et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans
356 with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489-1501.e15 (2020).
- 357 25. Le Bert, N., et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and
358 SARS, and uninfected controls. *Nature* 10.1038/s41586-020-2550-z (2020).
- 359 26. Flaxman, S., et al. Estimating the effects of non-pharmaceutical interventions on
360 COVID-19 in Europe. *Nature* 10.1038/s41586-020-2405-7 (2020).
- 361

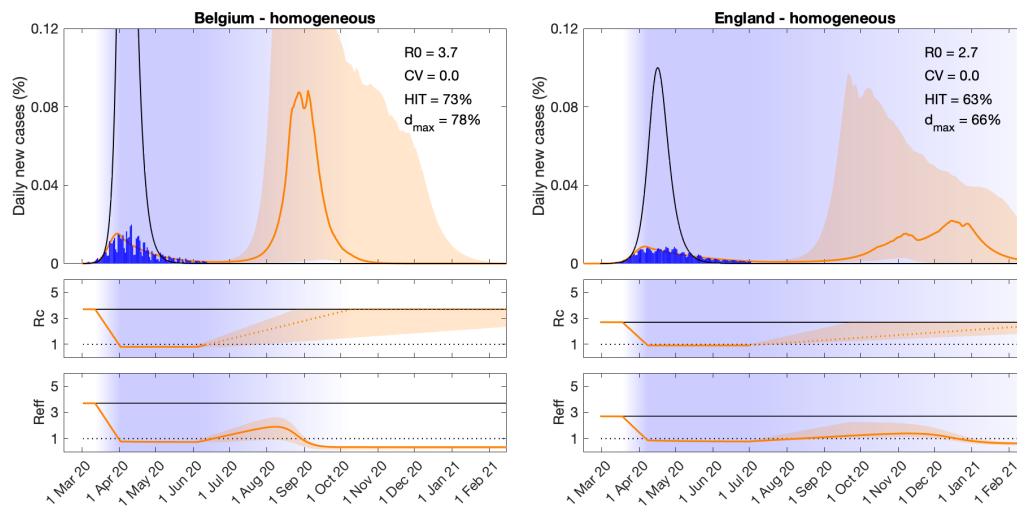
362



363

364 **Fig. 1| SARS-CoV-2 transmission with individual variation.** Variation in
 365 susceptibility (top panels); variation in connectivity (CV reducing in proportion to
 366 social distancing) (bottom panels). Susceptibility or connectivity factors implemented
 367 as gamma distributions. Suppressed wave and subsequent dynamics in Belgium and
 368 England (orange). Estimated epidemic in the absence of interventions revealing
 369 overshoot (black). Blue bars are daily new cases. Controlled (R_c) and effective (R_{eff})
 370 reproduction numbers are displayed on shallow panels underneath the main plots.
 371 Blue shades represent social distancing (intensity reflected in R_c trends and shade
 372 density). R_c values in the dotted portion of the orange lines do not interfere with the
 373 fittings and are only used to illustrate how the epidemic may unfold beyond the data
 374 analysed here. Consensus parameter values (Methods): $\delta = 1/4$ per day; $\gamma = 1/4$ per
 375 day; and $\rho = 0.5$. Fraction of infected individuals identified as positive (reporting
 376 fraction): 0.06 (Belgium); 0.024 (England). Basic reproduction number, coefficients
 377 of variation and social distancing parameters estimated by Bayesian inference as
 378 described in Methods (estimates in Extended Data Tables 1 and 3). Curves represent
 379 mean model predictions from 10^4 posterior samples. Orange shades represent 95%
 380 credible intervals. Vertical lines represent the expected time when herd immunity
 381 threshold will be achieved.
 382

383

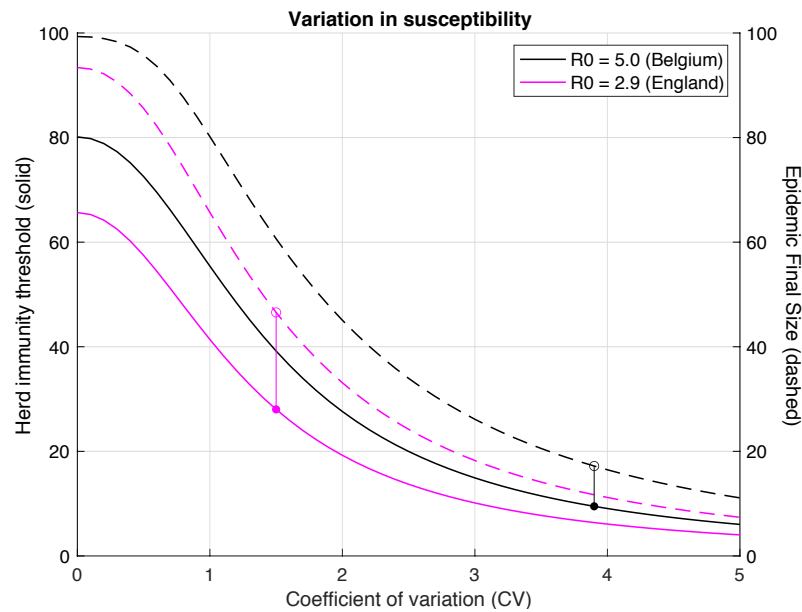


384

385 **Fig. 2| SARS-CoV-2 transmission in homogeneous populations.** Suppressed wave
 386 and subsequent dynamics in Belgium and England (orange). Estimated epidemic in
 387 the absence of interventions revealing overshoot (black). Blue bars are daily new
 388 cases. Controlled (R_c) and effective (R_{eff}) reproduction numbers are displayed on
 389 shallow panels underneath the main plots. Blue shades represent social distancing
 390 (intensity reflected in R_c trends and shade density). R_c values in the dotted portion of
 391 the orange lines do not interfere with the fittings and are only used to illustrate how
 392 the epidemic may unfold beyond the data analysed here. Consensus parameter values
 393 (Methods): $\delta = 1/4$ per day; $\gamma = 1/4$ per day; and $\rho = 0.5$. Fraction of infected
 394 individuals identified as positive (reporting fraction): 0.06 (Belgium); 0.024
 395 (England); 0.09 (Portugal); 0.06 (Spain). Basic reproduction number, coefficients of
 396 variation and social distancing parameters estimated by Bayesian inference as
 397 described in Methods (estimates in Extended Data Table 4). Curves represent mean
 398 model predictions from 10^4 posterior samples. Orange shades represent 95% credible
 399 intervals.

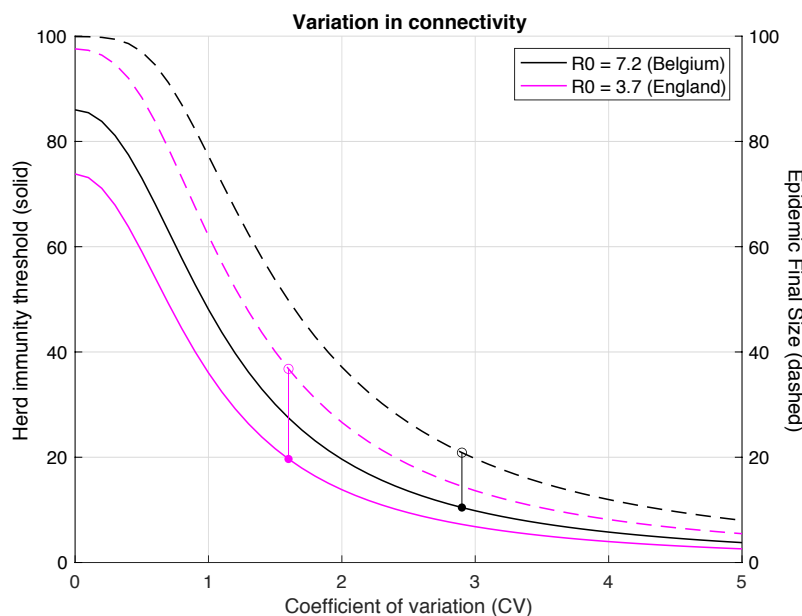
400

401



402

403



404

405 **Fig. 3| Herd immunity threshold with gamma-distributed susceptibility or**
 406 **exposure to infection.** Curves generated with the SEIR model (Equation 1-4)
 407 assuming values of R_0 estimated for the study countries (Extended Data Tables 1 and
 408 2) assuming gamma-distributed: susceptibility (top); connectivity (bottom) with
 409 constant CV. Herd immunity thresholds (solid curves) are calculated according to the
 410 formula $1 - (1/R_0)^{1/(1+CV^2)}$ for heterogeneous susceptibility and $1 -$
 411 $(1/R_0)^{1/(1+2CV^2)}$ for heterogeneous connectivity. Final sizes of the corresponding
 412 unmitigated epidemics are also shown (dashed).

413 **METHODS**

414 **Model structure and underlying assumptions**

415 The model presented here is a differential equation SEIR model, where susceptible
416 individuals become exposed at a rate that depends on their susceptibility, the number
417 of potentially infectious contacts they engage in, and the total number of infectious
418 people in the population per time unit. Upon exposure, individuals enter an
419 asymptomatic incubation phase, during which they slowly become infectious²⁷⁻³⁰.
420 Thus, infectivity of exposed individuals is made to be 1/2 of that of infectious ones
421 ($\rho = 0.5$). After a few days, individuals develop symptoms – on average 4 days after
422 the exposure to the virus ($\delta = 1/4$) – and thus become fully infectious³¹⁻³³. They
423 recover, i.e., they are no longer infectious 4 days after that ($\gamma = 1/4$), on average³⁴.

424 *Efficacy of acquired immunity*

425 We conducted the core of our analysis under the assumption that no reinfection occurs
426 after recovery due to acquired immunity ($\sigma = 0$). To analyse the sensitivity of these
427 results to leakage in immune response ($\sigma > 0$) we calculated herd immunity
428 thresholds (HIT) as a function of coefficients of variation (CV) for different values of
429 σ . The results displayed in Extended Data Figure 3 confirm the expectation that as the
430 efficacy of acquired immunity decreases (σ increases) larger percentages of the
431 population are infected before herd immunity is reached. Less intuitive is that there is
432 an upper bound for how much it is reasonable to increase σ before the system enters a
433 qualitatively different regime – the reinfection threshold^{10,11} ($\sigma = 1/R_0$) – above
434 which infection becomes stably endemic and the notion of herd immunity threshold

435 no longer applies. Respiratory viruses are typically associated with epidemics
436 dynamics below the reinfection threshold.

437 *Effective reproduction number*

438 The effective reproduction number (R_{eff} , also denoted by R_e or R_t by other authors)
439 is a time-dependent quantity which we calculate as the incidence of new infections
440 divided by the total number of active infections (affected by ρ for individuals in E)
441 multiplied by the average duration of infection (also affected by ρ for individuals in
442 E)

$$443 \quad R_{eff}(t) = \frac{\int \lambda(x, t)x[S(x, t) + \sigma R(x, t)] dx}{\int \rho E(x, t) + I(x, t) dx} \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right). \quad (9)$$

444 *Assortative mixing*

445 In the main text we assumed random mixing among individuals, but human
446 connectivity patterns are assortative due societal structures and human behaviours. To
447 explore the sensitivity of our results to deviations from random mixing, we develop
448 an extended formalism that allows individuals to connect preferentially with those
449 with similar connectivity, formally $\lambda(x) =$
450 $(\beta/N)(\int y h(y - x)[\rho E(y) + I(y)] dy / \int yg(y) dy)$, where $h(y - x)$ is a normal
451 distribution on the difference between connectivity factors (Extended Data Figure 2).

452 *Non-pharmaceutical interventions*

453 We implemented non-pharmaceutical interventions (NPI) as a gradual decrease in
454 viral transmissibility in the population and thus a lowering of the controlled and
455 effective reproduction numbers (R_c and R_{eff}). Once containment measures are put in

456 place in each country, we postulate it takes 21 days until the maximum effectiveness
457 of social distancing measures is reached, approximately linearly. In the simulations
458 presented throughout we have held this condition (maximum “lockdown” efficacy)
459 for T_{max} days, where T_{max} is estimated for each study country in countrywide
460 analyses and fixed at 30 in regional analyses. Eventually, lockdowns were lifted and
461 social distancing measures progressively relaxed. This relaxation is implemented
462 linearly over a period of T_{lift} days, which is also estimated. We note however that the
463 estimation of epidemiological parameters is not affected by the shape assumed for the
464 relaxation of interventions beyond the date of the last data point (June 11th in Belgium
465 and July 1st in England). In particular, assuming that social distancing measures
466 remained constant from these dates onwards, for example, would make no difference
467 to the estimated parameters.

468 **Bayesian Inference**

469 The model laid out above is amenable to theoretical exploration as presented in the
470 main manuscript and provides a perfect framework for inference. Fundamentally, to
471 be able to reproduce the inception of any epidemic, we would need to estimate when
472 local transmission started to occur (t_0), and the pace at which individuals infected
473 each other in the very early stages of the epidemic (R_0). All countries, to different
474 extents and at different timepoints of the epidemic, enforced some combination of
475 social distancing measures. To fully understand the interplay between herd immunity
476 and the impact of NPIs, we then set out to estimate the time at which social distancing
477 measures started to have an impact on daily incidence (t_0^d), what their maximum
478 efficacy (d_{max}) was and how long it lasted (T_{max}), a parameter used to determine the
479 rate at which contacts resume as restrictions are lifted (T_{lift} , which more specifically

480 represents the time it would take to reach pre-pandemic contact intensity should the
481 lifting of restricting have linear effect), the basic reproduction number (R_0) and what
482 the underlying variance in heterogeneity is for susceptibility or exposure to infection.

483 In order to preserve identifiability, we made two simplifying assumptions: (i) the
484 fraction of infectious individuals reported as COVID-19 cases (reporting fraction) is
485 constant throughout the study period and is comparable between countries
486 proportionally to the number of tests performed per person; (ii) local transmission
487 starts (t_0) when countries/regions report 1 case per 5 million population in one day.
488 To calculate the reporting rates, we used the Spanish national serological survey¹² as a
489 reference and divided the total number of reported cases up to May 11th by the
490 estimated number of people that had been exposed to the virus. This gives us a
491 reporting rate for Spain around 6%. Unfortunately, there are no other national
492 serological surveys that could inform the proportion of the population infected in
493 other countries, so we had to extrapolate the reporting rate for those. Assuming the
494 reporting rate is highly dependent on the testing effort employed in each country,
495 reflected in the number of tests per individual, we estimate the reporting rate by
496 scaling the reporting rate recorded in Spain according to the ratio of PCR tests per
497 person in other countries relative to the Spanish reference of 0.9 tests per thousand
498 people (<https://ourworldindata.org/coronavirus-testing>). This produced estimated case
499 reporting rates (ratio of reported cases to infections) of 9% for Portugal, 6% for
500 Belgium (and Spain) and 2.4% for England.

501 Whist national case and mortality data is easily available for most countries, more
502 spatially resolute data is difficult to find in the public domain. Thus, we restricted our
503 analysis to countries for which disaggregated regional case data was easily available.

504 We collected the data at two time points. First, we compiled all available data from
505 the day the countries started reporting COVID-19 cases to the initial collection date
506 (May 20th) and later collated available data from May 21st to July 10th.

507 Parameter estimation was performed with the software MATLAB, using PESTO
508 (Parameter ESTimation Toolbox)³⁵, and assuming the reported case data can be
509 accurately described by a Poisson process. We first fixed the beginning of local
510 transmission (parameter t_0) in each data series as the day in which reported cases
511 surpassed 1 in 5 million individuals. Next, we optimized the model for the set of
512 parameters $\theta = \{R_0, CV, t_0^d, d_{max}, T_{max}, T_{lift}\}$ by maximizing the logarithm of the
513 likelihood (LL) (Equation 10) of observing the daily reported number of cases in each
514 country $D = \{(k, \tilde{y}_k)\}_{k=0}^n$:

$$LL(\theta|D) = - \sum_{k=1}^n y(k, \theta) + \left(\sum_{k=1}^n \tilde{y}(k) \ln(y(k, \theta)) \right) - \ln \left(\prod_{k=1}^n \tilde{y}(k)! \right) \quad (10)$$

515 in which $y(k, \theta)$ is the simulated model output number of COVID-19 cases at day k
516 (with respect to t_0), and n is the total number of days included in the analysis for each
517 country.

518 To ensure that the estimated maximum is a global maximum, we performed 50 multi-
519 starts optimizations, and selected the combination of parameters resulting in the
520 maximal Loglikelihood as a starting point for 10^4 Markov Chain Monte-Carlo
521 iterations. From the resulting posterior distributions, we extract the median estimates
522 for each parameter and the respective 95% credible intervals for the set of parameters.
523 We used uniformly distributed priors with ranges $\{1-9, 0.0025-8, 1-60, 0-0.7, 1-90, 60-$
524 $1000\}$.

525 When fitting the model to disaggregated data, we follow the procedure outlined above
526 but fix $T_{max} = 30$ days, $T_{lift} = 120$ days and estimate region-specific R_0 and CV ,
527 with common t_0^d and d_{max} .

528 The countrywide fitting procedure was applied to 2 countries (Belgium and England)
529 and repeated for each of the 4 model variants considered here (homogeneous,
530 heterogeneous susceptibility, heterogeneous connectivity with constant CV , and
531 heterogeneous connectivity with CV reducing in proportion to social distancing). Our
532 study included two more countries (Portugal and Spain), where countrywide analyses
533 could not be performed due to the epidemic being geographically desynchronised. We
534 implemented stratified analyses on all 4 countries. In the fitting procedures using sub-
535 national data, we assumed regions had the same start date for interventions that
536 mitigate transmission (t_0^d), and that these measures produced the same maximum
537 impact on transmission (d_{max}) everywhere. Thus, the only region-specific parameters
538 to be estimated are $\{R_{0_i}, CV_i\}$. Parameter estimates obtained from each of the model
539 variants are displayed in Extended Data Table 1 (heterogeneity in susceptibility),
540 Extended Data Table 2 (heterogeneity in connectivity with constant CV), Extended
541 Data Table 3 (heterogeneity in connectivity with dynamic CV) and Extended Data
542 Table 4 (homogeneous model), are comparable to those obtained in other studies^{7,26,37-}
543 ⁴⁰. Finally, we apply the Akaike information criterion (AIC) for each estimation
544 procedure to inform on the quality of each model's fit to the datasets of reported cases
545 (Extended Data Table 5). In all cases, heterogeneous models are preferred over the
546 homogeneous approximation. The three heterogeneous models are roughly equally
547 well supported by the data used in this study. Further research should complement
548 this with discriminatory data types and hybrid models to enable the integration of
549 different forms of individual variation.

550 **Data and code availability**

551 Datasets are publicly available at the respective national ministry of health websites

552 (41-45). Core models implemented in MATLAB available from:

553 <https://github.com/mgmgomes1/covid>

554

555 27. Wei, W. E., et al. Presymptomatic Transmission of SARS-CoV-2 — Singapore, January

556 23–March 16, 2020. *MMWR Morb Mortal Wkly Rep* [Internet]. 2020 Apr 10 [cited

557 2020 May 4];69(14):411–5. Available from:

558 http://www.cdc.gov/mmwr/volumes/69/wr/mm6914e1.htm?s_cid=mm6914e1_w

559 28. To, K. K. W., et al. Temporal profiles of viral load in posterior oropharyngeal saliva

560 samples and serum antibody responses during infection by SARS-CoV-2: an

561 observational cohort study. *Lancet Infect. Dis.* **20**, 565-74 (2020).

562 29. Arons, M. M., et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a

563 Skilled Nursing Facility. *N. Engl. J. Med.* **382**, 2081-2090 (2020).

564 30. He, X., et al. Temporal dynamics in viral shedding and transmissibility of COVID-19.

565 *Nat. Med.* **26**, 672-675 (2020).

566 31. Lauer, S. A., et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19)

567 From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern.*

568 *Med.* **172**, 577-582 (2020).

569 32. Li, Q., et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-

570 infected pneumonia. *N. Engl. J. Med.* **382**, 1199-1207 (2020).

571 33. Zhang, J., et al. Evolving epidemiology and transmission dynamics of coronavirus

572 disease 2019 outside Hubei province, China: A descriptive and modelling study. *Lancet*

573 *Infect. Dis.* **20**, 793-802 (2020).

574 34. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus

575 (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284-6 (2020).

- 576 35. Stapor, P., et al. PESTO: Parameter ESTimation TOolbox. *Bioinformatics* **34**, 705-707
577 (2018).
- 578 36. Prem, K., et al. The effect of control strategies to reduce social mixing on outcomes of
579 COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261-
580 e270.
- 581 37. Tian, H., et al. An investigation of transmission control measures during the first 50 days
582 of the COVID-19 epidemic in China. *Science* **368**, 638-642.
- 583 38. Kucharski, A. J., et al. Effectiveness of isolation, testing, contact tracing and physical
584 distancing on reducing transmission of SARS-CoV-2 in different settings: a
585 mathematical modelling study. *Lancet Infect. Dis.* 10.1016/s1473-3099(20)30457-6
586 (2020).
- 587 39. Salje, H., et al. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208-211
588 (2020).
- 589 40. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boelle, P.-Y. & Colizza, V. Expected
590 impact of lockdown in Île-de-France and possible exit strategies. *medRxiv*
591 10.1101/2020.04.13.20063933 (2020).
- 592 41. [https://ourworldindata.org/coronavirus-testing#source-information-country-by-](https://ourworldindata.org/coronavirus-testing#source-information-country-by-country)
593 [country](https://ourworldindata.org/coronavirus-testing#source-information-country-by-country). Accessed on July 10th 2020.
- 594 42. <https://cnecovid.isciii.es/covid19>.
- 595 43. <https://covid-19.sciensano.be/nl/covid-19-epidemiologische-situatie>.
- 596 44. <https://covid19.min-saude.pt/ponto-de-situacao-atual-em-portugal>.
- 597 45. <https://coronavirus.data.gov.uk>.

598

599 **Acknowledgements**

600 We thank Jan Hasenauer and Antonio Montalbán for helpful discussions concerning
601 statistical inference and mathematics, respectively. R.M.C. and M.U.F. receive

602 scholarships from the Conselho Nacional de Desenvolvimento Científico e
603 Tecnológico (CNPq), Brazil.

604

605 **Author contributions**

606 M.G.M.G. conceived the study. R.A. and R.M.C. and M.G.M.G. performed the
607 analyses. All authors interpreted the data and wrote the paper.

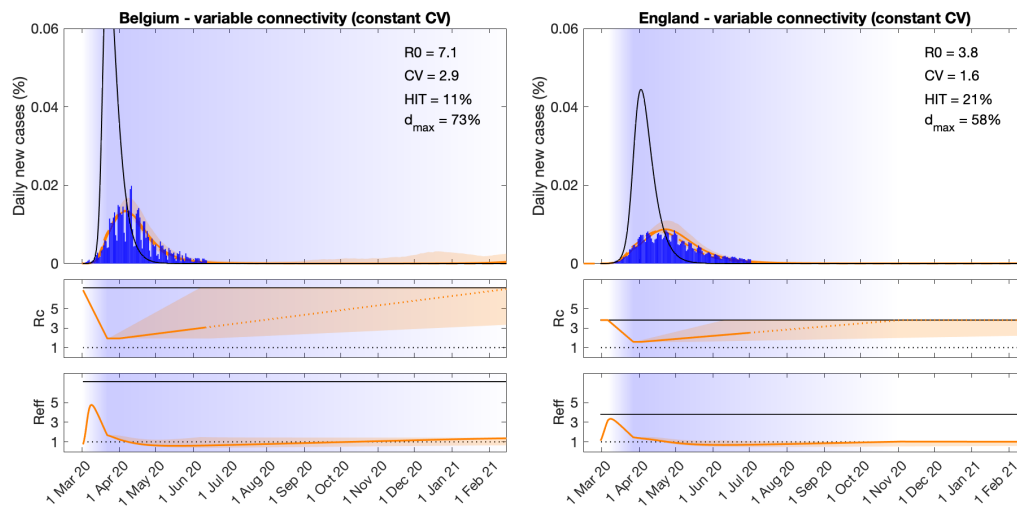
608

609 **Competing interests**

610 The authors declare no competing interests.

611

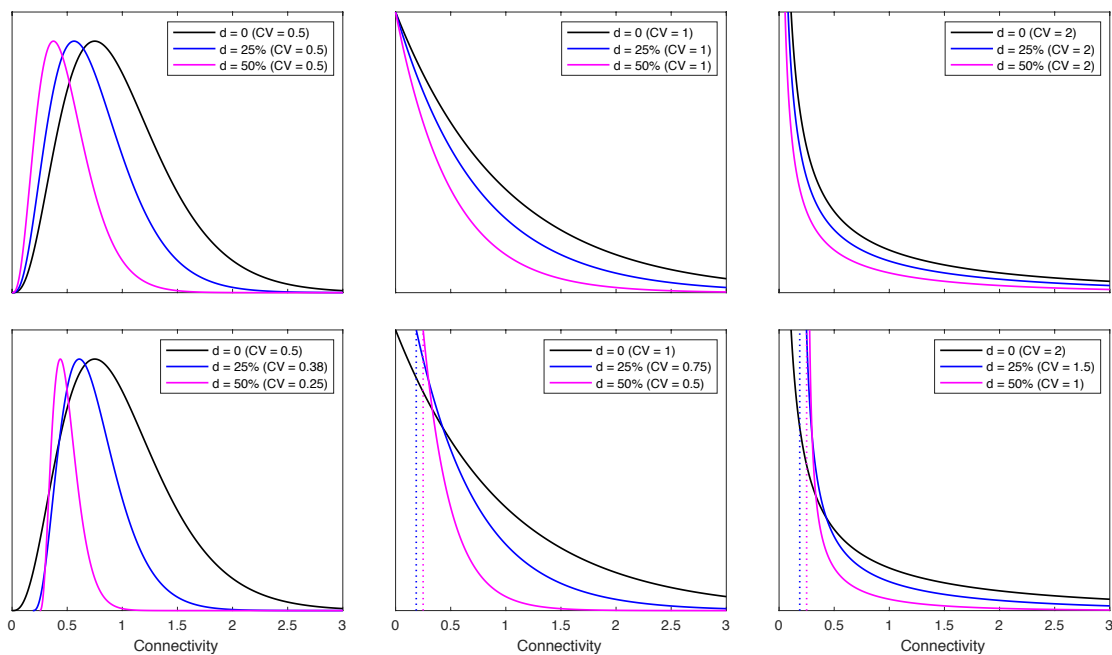
612



613

614 **Extended Data Fig. 1 | SARS-CoV-2 transmission with individual variation in**
 615 **connectivity.** Variation in connectivity (constant CV). Connectivity factors
 616 implemented as gamma distributions. Suppressed wave and subsequent dynamics in
 617 Belgium and England (orange). Estimated epidemic in the absence of interventions
 618 revealing overshoot (black). Blue bars are daily new cases. Controlled (R_c) and
 619 effective (R_{eff}) reproduction numbers are displayed on shallow panels underneath the
 620 main plots. Blue shades represent social distancing (intensity reflected in R_c trends
 621 and shade density). R_c values in the dotted portion of the orange lines do not interfere
 622 with the fittings and are only used to illustrate how the epidemic may unfold beyond
 623 the data analysed here. Consensus parameter values (Methods): $\delta = 1/4$ per day; $\gamma =$
 624 $1/4$ per day; and $\rho = 0.5$. Fraction of infected individuals identified as positive
 625 (reporting fraction): 0.06 (Belgium); 0.024 (England). Basic reproduction number,
 626 coefficients of variation and social distancing parameters estimated by Bayesian
 627 inference as described in Methods (estimates in Extended Data Table 2). Curves
 628 represent mean model predictions from 10^4 posterior samples. Orange shades
 629 represent 95% credible intervals. Vertical lines represent the expected time when herd
 630 immunity threshold will be achieved.
 631

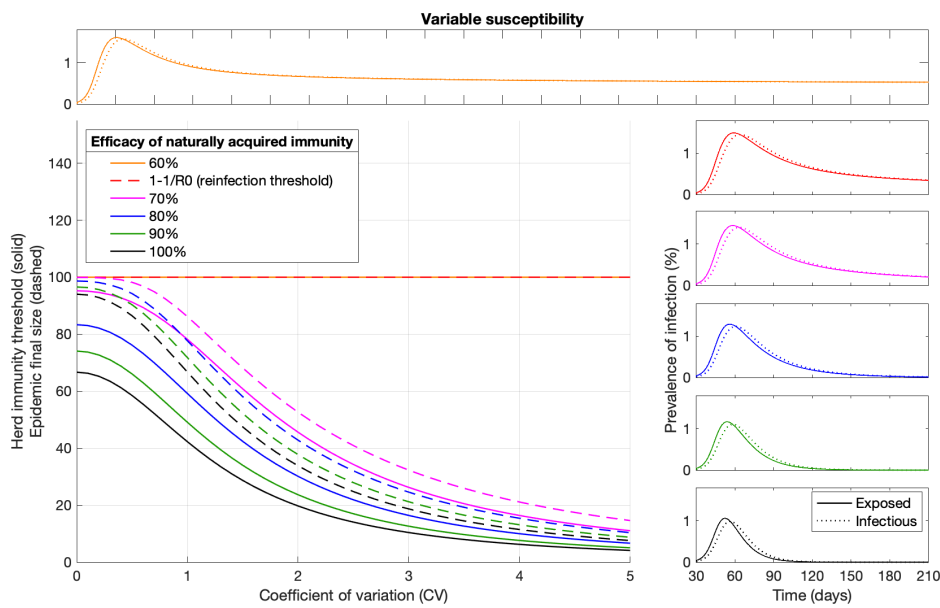
632



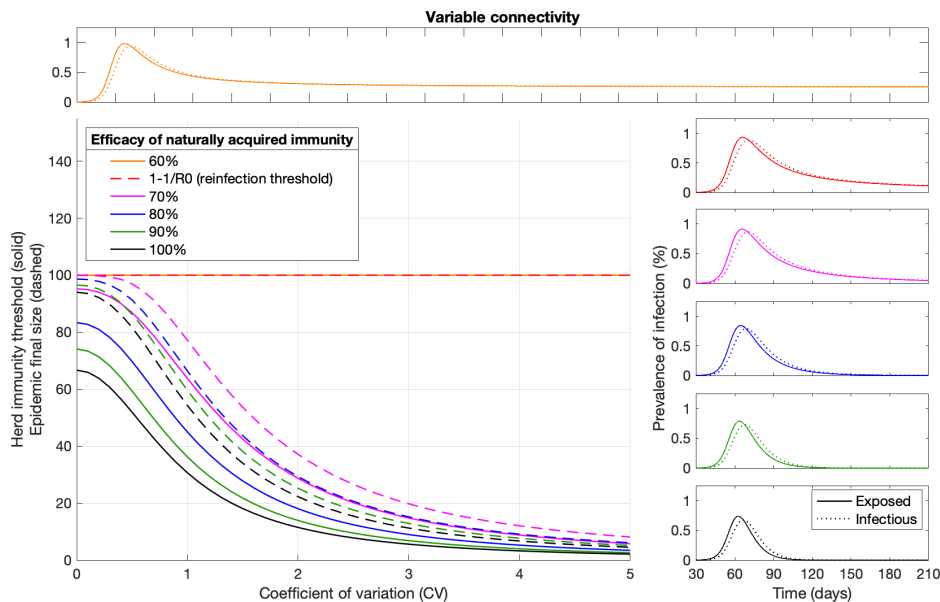
633

634 **Extended Data Fig. 2| Connectivity distributions with reducing coefficient of**
 635 **variation in proportion to social distancing.** Individual variation in connectivity is
 636 originally implemented as a gamma distribution of mean 1 parameterised by the
 637 coefficient of variation (CV) (black). Social distancing is initially implemented as a
 638 reduction in connectivity by the same factor to every individual, from x to $(1 - d)x$
 639 (top panels). A more general formulation where CV may reduce with social
 640 distancing is implemented by modifying x to $(1 - d)[1 + (1 - d)(x - 1)]$ (bottom
 641 panels).
 642

643



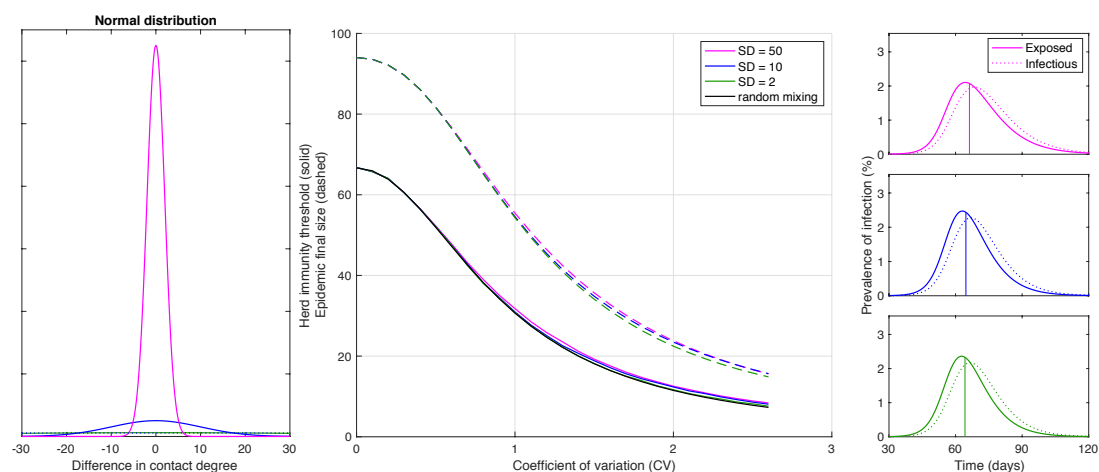
644



645

646 **Extended Data Fig. 3| Herd immunity threshold and epidemic final size with**
 647 **reinfection.** Curves in the main panels generated with the SEIR model (Equation 1-4)
 648 assuming $R_0 = 3$ and gamma-distributed susceptibility (top) or connectivity (bottom)
 649 with constant CV. Efficacy of acquired immunity is captured by a reinfection
 650 parameter σ , potentially ranging between $\sigma = 0$ (100% efficacy) and $\sigma = 1$ (0
 651 efficacy). This illustration depicts final sizes of unmitigated epidemics and associated
 652 HIT curves for 6 values of σ : $\sigma = 0$ (black); $\sigma = 0.1$ (green); $\sigma = 0.2$ (blue); $\sigma = 0.3$
 653 (magenta); $\sigma = 1/3$ (red); and $\sigma = 0.4$ (orange);. Above $\sigma = 1/R_0$ (reinfection
 654 threshold (Gomes et al 2004; 2016)) the infection becomes stably endemic and there
 655 is no herd immunity threshold. Representative epidemics of the regime $\sigma \leq 1/R_0$ are
 656 shown on the right while the regime $\sigma > 1/R_0$ is illustrated on top. All depicted
 657 dynamics are based on the rightmost CVs represented on the main panel.
 658

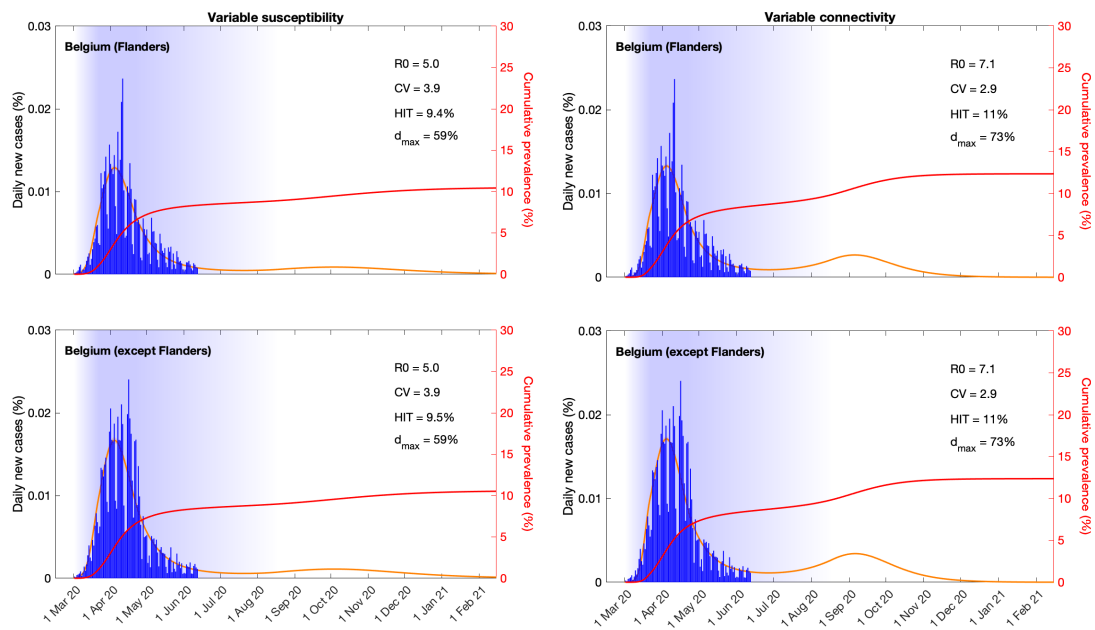
659



660

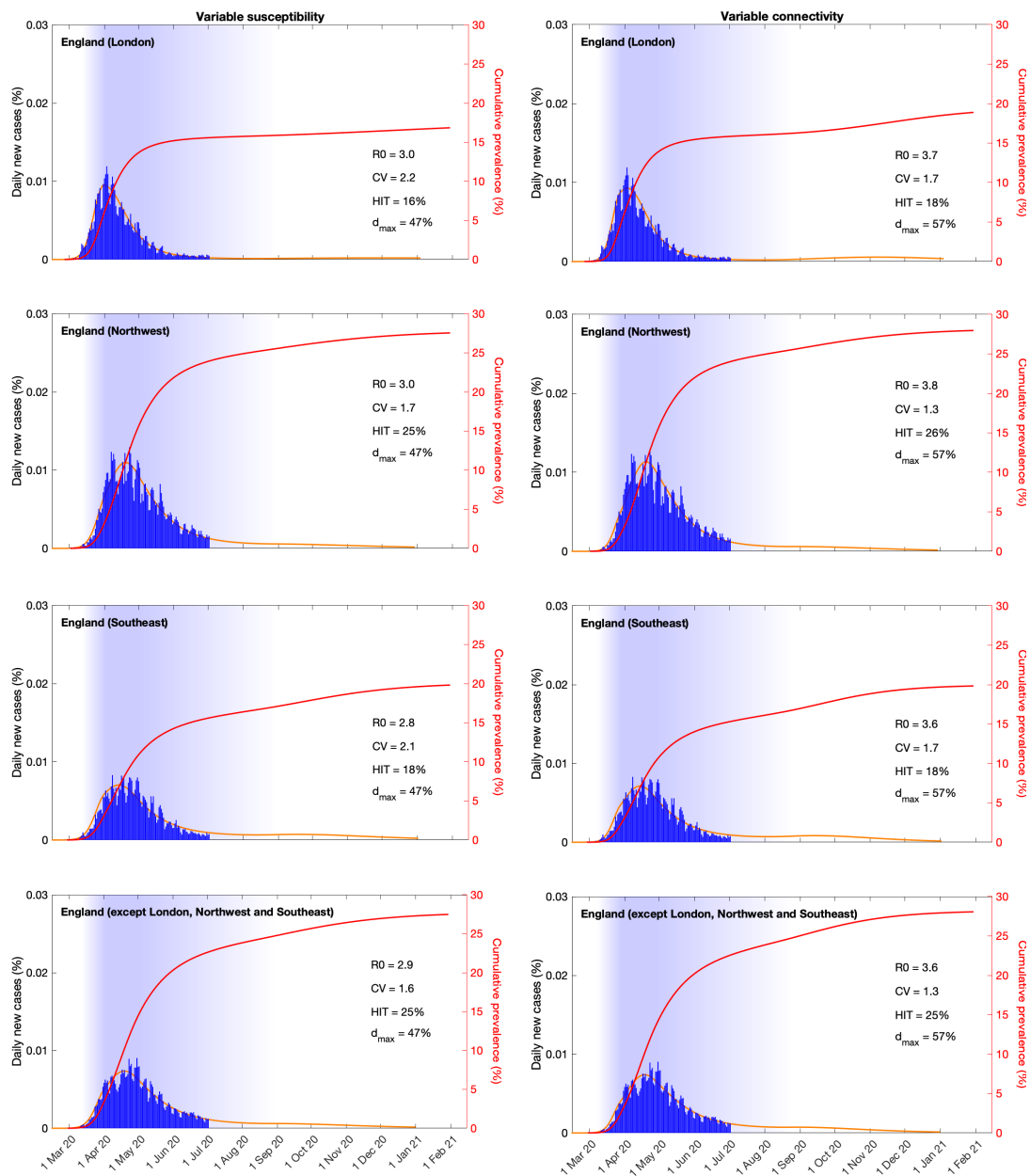
661 **Extended Data Fig. 4| Herd immunity threshold and epidemic final size with**
 662 **gamma-distributed exposure to infection and assortative mixing.** Curves in
 663 central panel generated with the SEIR model (Equation 1-4) assuming $R_0 = 3$ and
 664 gamma-distributed connectivity. Assortative mixing is implemented by imposing a
 665 normal distribution for contact preferences such that individuals contact preferentially
 666 with those with the similar contact degree (left). This illustration used normal
 667 distributions with standard deviation $SD = 50$ (green); $SD = 10$ (blue); and $SD = 2$
 668 (magenta). More assortative mixing leads to more skewed epidemics. Herd immunity
 669 thresholds were calculated numerically as the percentage of the population no longer
 670 susceptible when new outbreaks are effectively prevented (approximately when the
 671 exposed fraction crosses the peak in the absence of mitigation). Final sizes of the
 672 corresponding unmitigated epidemics are also shown. Representative epidemics are
 673 depicted on the right based on the rightmost CVs represented on the main panel (with
 674 vertical lines marking the point when herd immunity is achieved).
 675

676



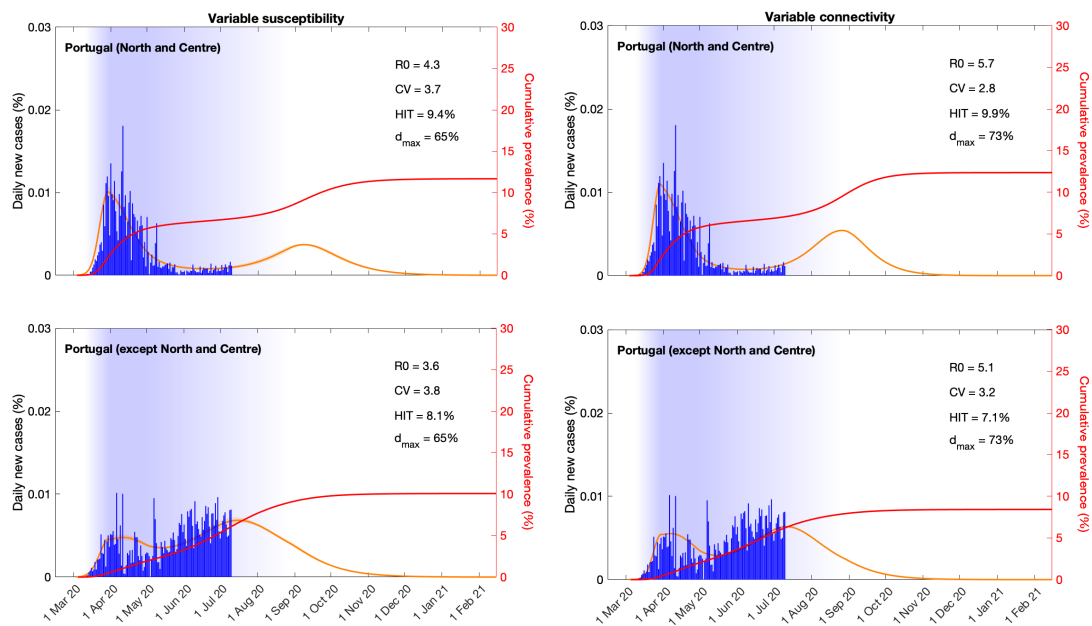
677 **Extended Data Fig. 5 | SARS-CoV-2 transmission at subnational levels in Belgium.**
 678 Suppressed wave and subsequent dynamics in Flanders and the rest of Belgium, with
 679 individual variation in susceptibility (left) or exposure (right). Blue bars are daily new
 680 cases. Shades represent social distancing (intensity reflected in shade density).
 681 Susceptibility or exposure factors implemented as gamma distributions. Consensus
 682 parameter values (Methods): $\delta = 1/4$ per day; $\gamma = 1/4$ per day; and $\rho = 0.5$.
 683 Fraction of infected individuals identified as positive (reporting fraction): 0.06. Basic
 684 reproduction number, coefficients of variation and social distancing parameters
 685 estimated by Bayesian inference as described in Methods (estimates in Extended Data
 686 Table 1 and 2). Curves represent mean model predictions from 10^4 posterior samples.
 687 Orange shades represent 95% credible intervals. Red curves represent cumulative
 688 infected percentages.
 689

690



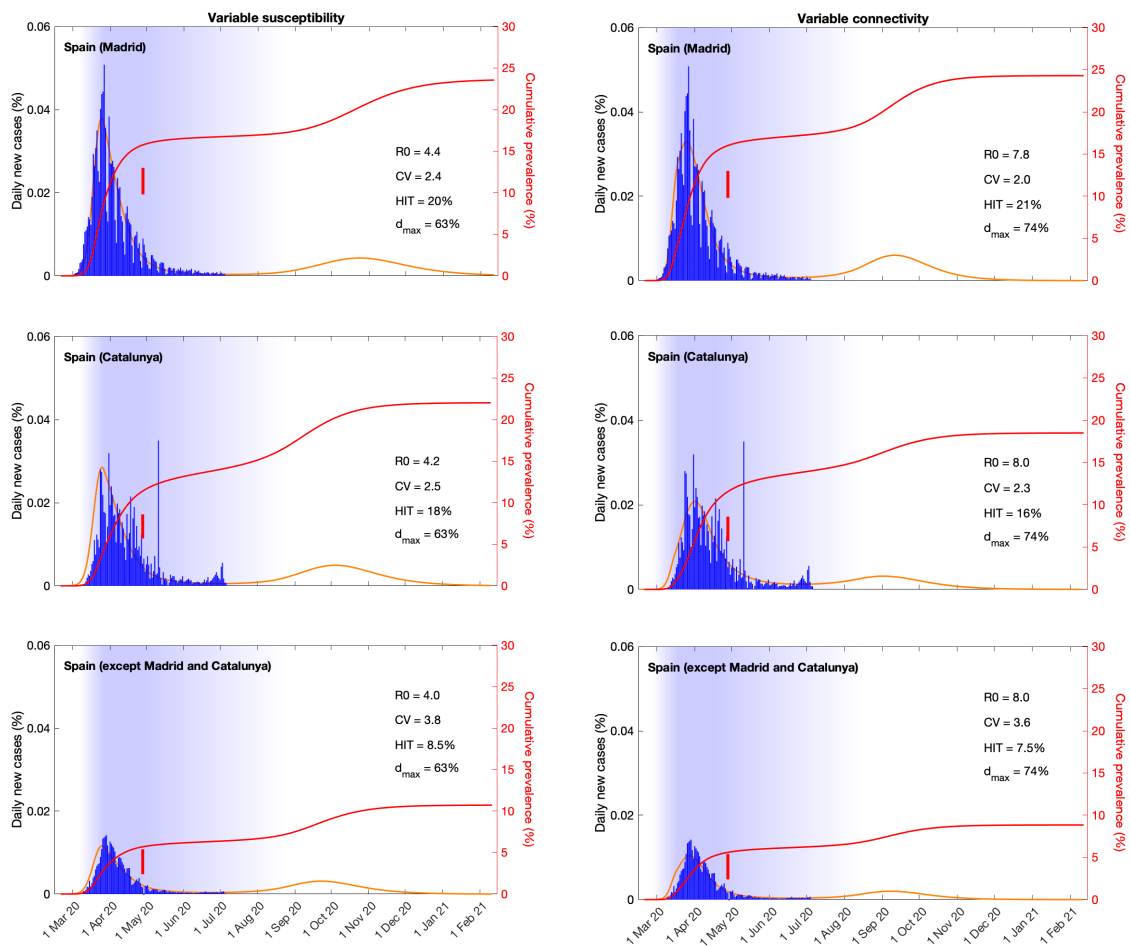
691 **Extended Data Fig. 6 | SARS-CoV-2 transmission at subnational levels in England.**
 692 Suppressed wave and subsequent dynamics in London, Northwest, Southeast and the
 693 rest of England, with individual variation in susceptibility (left) or exposure (right).
 694 Blue bars are daily new cases. Shades represent social distancing (intensity reflected
 695 in shade density). Susceptibility or exposure factors implemented as gamma
 696 distributions. Consensus parameter values (Methods): $\delta = 1/4$ per day; $\gamma = 1/4$ per
 697 day; and $\rho = 0.5$. Fraction of infected individuals identified as positive (reporting
 698 fraction): 0.024. Basic reproduction number, coefficients of variation and social
 699 distancing parameters estimated by Bayesian inference as described in Methods
 700 (estimates in Extended Data Table 1 and 2). Curves represent mean model predictions
 701 from 10^4 posterior samples. Orange shades represent 95% credible intervals. Red
 702 curves represent cumulative infected percentages.
 703

704



705 **Extended Data Fig. 7 | SARS-CoV-2 transmission at subnational levels in Portugal.**
 706 Suppressed wave and subsequent dynamics in the North and Centre regions versus the
 707 rest of Portugal, with individual variation in susceptibility (left) or exposure (right).
 708 Blue bars are daily new cases. Shades represent social distancing (intensity reflected
 709 in shade density). Susceptibility or exposure factors implemented as gamma
 710 distributions. Consensus parameter values (Methods): $\delta = 1/4$ per day; $\gamma = 1/4$ per
 711 day; and $\rho = 0.5$. Fraction of infected individuals identified as positive (reporting
 712 fraction): 0.09. Basic reproduction number, coefficients of variation and social
 713 distancing parameters estimated by Bayesian inference as described in Methods
 714 (estimates in Extended Data Table 1 and 2). Curves represent mean model predictions
 715 from 10^4 posterior samples. Orange shades represent 95% credible intervals. Red
 716 curves represent cumulative infected percentages.
 717

718



719 **Extended Data Fig. 8 | SARS-CoV-2 transmission at subnational levels in Spain.**
 720 Suppressed wave and subsequent dynamics in Madrid, Catalunya and the rest of
 721 Spain, with individual variation in susceptibility (left) or exposure (right). Blue bars
 722 are daily new cases. Shades represent social distancing (intensity reflected in shade
 723 density). Susceptibility or exposure factors implemented as gamma distributions.
 724 Consensus parameter values (Methods): $\delta = 1/4$ per day; $\gamma = 1/4$ per day; and $\rho =$
 725 0.5 . Fraction of infected individuals identified as positive (reporting fraction): 0.06 .
 726 Basic reproduction number, coefficients of variation and social distancing parameters
 727 estimated by Bayesian inference as described in Methods (estimates in Extended Data
 728 Table 1 and 2). Curves represent mean model predictions from 10^4 posterior samples.
 729 Orange shades represent 95% credible intervals. Red curves represent cumulative
 730 infected percentages and vertical red segments mark seroprevalences (95% CI)
 731 according to a recent study¹².
 732

733 **Extended Data Table 1| Estimated parameters for heterogeneous susceptibility**
 734 **model.** Estimates generated from model fit to the national datasets are in the grey
 735 shaded rows. The remaining rows provide the region-specific estimates. Best
 736 parameter estimates are presented as a bold median bounded by the lower and upper
 737 ends for the 95% credible interval. Model runs are initiated on the day (t_0) when
 738 reported cases surpassed 1 in 5 million individuals: Belgium (day 1); England (day
 739 29); Portugal (day 3); Spain (day 8).
 740

Country/Region	R_0			CV			$1 - d_{max}$			t_0^d		
Belgium	5.01	5.03	5.05	3.84	3.87	3.89	0.40	0.40	0.40	1.00	1.02	1.08
Flanders	4.96	5.00	5.02	3.89	3.91	3.93	0.41	0.41	0.41	1.00	1.02	1.15
Rest	4.97	5.01	5.03	3.87	3.89	3.91						
Portugal												
North/Centre	4.27	4.32	4.36	3.69	3.72	3.74	0.34	0.35	0.35	7.51	7.73	8.00
Rest	3.54	3.58	3.61	3.72	3.76	3.79						
Spain												
Madrid	4.38	4.39	4.39	2.37	2.37	2.38	0.37	0.37	0.37	16.40	16.41	16.44
Catalunya	4.20	4.21	4.21	2.49	2.50	2.50						
Rest	3.96	3.97	3.97	3.80	3.81	3.82						
England	2.90	2.91	2.92	1.54	1.54	1.55	0.46	0.46	0.46	42.03	42.19	42.41
London	2.95	2.96	2.96	2.24	2.24	2.24	0.52	0.53	0.53	41.35	41.51	41.64
NorthWest	3.03	3.04	3.05	1.66	1.67	1.68						
SouthEast	2.80	2.81	2.82	2.07	2.07	2.07						
Rest	2.88	2.88	2.89	1.64	1.64	1.65						

741
 742
 743

744 **Extended Data Table 2| Estimated parameters for heterogeneous connectivity**
 745 **model (constant CV).** Estimates generated from model fit to the national datasets are
 746 in the grey shaded rows. The remaining rows provide the region-specific estimates.
 747 Best parameter estimates are presented as a bold median bounded by the lower and
 748 upper ends for the 95% credible interval. Model runs are initiated on the day (t_0)
 749 when reported cases surpassed 1 in 5 million individuals: Belgium (day 1); England
 750 (day 29); Portugal (day 3); Spain (day 8).
 751

Country/Region	R_0			CV			$1 - d_{max}$			t_0^d		
Belgium	7.13	7.15	7.16	2.87	2.87	2.87	0.27	0.27	0.28	1.00	1.00	1.02
Flanders	7.09	7.11	7.13	2.86	2.87	2.89	0.27	0.27	0.28	1.00	1.01	1.03
Rest	7.11	7.13	7.15	2.86	2.87	2.89						
Portugal												
North/Centre	5.06	5.08	5.09	3.24	3.24	3.24	0.25	0.25	0.25	7.21	7.22	7.24
Rest	5.68	5.69	5.69	2.79	2.81	2.83						
Spain												
Madrid	7.81	7.83	8.82	1.98	1.99	2.06	0.24	0.26	0.26	5.38	7.02	7.06
Catalunya	8.00	8.02	9.00	2.33	2.34	2.43						
Rest	7.97	7.99	8.96	3.58	3.59	3.72						
England	3.81	3.82	3.84	1.55	1.55	1.55	0.42	0.42	0.42	36.38	37.49	36.54
London	3.70	3.70	3.71	1.69	1.69	1.70	0.43	0.43	0.43	37.50	37.52	37.55
NorthWest	3.83	3.83	3.84	1.32	1.32	1.32						
SouthEast	3.58	3.59	3.59	1.66	1.67	1.68						
Rest	3.60	3.60	3.61	1.30	1.31	1.31						

752
 753
 754
 755

756 **Extended Data Table 3| Estimated parameters for heterogeneous connectivity**
 757 **model (dynamic CV).** Estimates generated from model fit to the national datasets are
 758 in the grey shaded rows. The remaining rows provide the region-specific estimates.
 759 Best parameter estimates are presented as a bold median bounded by the lower and
 760 upper ends for the 95% credible interval. Model runs are initiated on the day (t_0)
 761 when reported cases surpassed 1 in 5 million individuals: Belgium (day 1); England
 762 (day 29); Portugal (day 3); Spain (day 8).
 763

Country/Region	R_0			CV			$1 - d_{max}$			t_0^d		
Belgium	8.88	8.91	8.94	3.89	3.92	3.95	0.55	0.55	0.55	1.00	1.00	1.00
Flanders	8.78	8.82	8.86	3.98	4.00	4.02	0.56	0.56	0.56	1.00	1.01	1.04
Rest	8.82	8.86	8.89	3.98	4.00	4.02						
Portugal												
North/Centre	6.65	6.72	6.80	3.75	3.78	3.81	0.56	0.57	0.57	5.84	6.02	6.19
Rest	5.98	6.05	6.13	4.09	4.14	4.19						
Spain												
Madrid	6.19	6.20	6.21	2.43	2.43	2.44	0.57	0.57	0.57	13.80	13.81	13.83
Catalunya	6.30	6.32	6.33	2.61	2.62	2.62						
Rest	6.34	6.35	6.36	3.80	3.81	3.82						
England	3.67	3.67	3.68	1.56	1.57	1.57	0.63	0.63	0.63	39.81	39.82	39.83
London	3.78	3.79	3.80	1.99	2.00	2.01	0.67	0.67	0.67	39.06	39.20	39.28
NorthWest	3.91	3.92	3.94	1.64	1.65	1.66						
SouthEast	3.67	3.69	3.70	1.89	1.89	1.90						
Rest	3.64	3.65	3.66	1.58	1.58	1.59						

764
 765
 766

767 **Extended Data Table 4| Estimated parameters for the homogenous model.**
 768 Estimates generated from model fit to the national datasets are in the grey shaded
 769 rows. The remaining rows provide the region-specific estimates. Best parameter
 770 estimates are presented as a bold median bounded by the lower and upper ends for the
 771 95% credible interval. Model runs are initiated on the day (t_0) when reported cases
 772 surpassed 1 in 5 million individuals: Belgium (day 1); England (day 29); Portugal
 773 (day 3); Spain (day 8).
 774

Country/Region	R_0			$1 - d_{max}$			t_0^d		
Belgium	3.683	3.698	3.712	0.218	0.219	0.220	11.948	12.054	12.173
Flanders	3.235	3.239	3.308	0.208	0.212	0.213	16.324	17.039	17.064
Rest	3.235	3.238	3.307						
Portugal									
North/Centre	3.542	3.578	3.608	0.343	0.345	0.348	7.514	7.725	7.999
Rest	4.274	4.321	4.361						
Spain									
Madrid	4.113	4.116	4.120	0.111	0.111	0.112	20.000	20.000	20.000
Catalunya	4.208	4.214	4.218						
Rest	3.735	3.751	3.752						
England	2.690	2.703	2.708	0.359	0.360	0.360	54.577	54.578	54.578
London	2.307	2.308	2.310						
NorthWest	2.602	2.603	2.604						
SouthEast	2.368	2.370	2.371						
Rest	2.502	2.503	2.504						

775
 776
 777

778 **Extended Data Table 5| Model selection criteria.** Displays the maximum
 779 Loglikelihood obtained for each combination of model and data partitioning for each
 780 country, as well as the Akaike information criterion. Models are labelled by a sort
 781 name as follows: homog (homogenous); hetsus (heterogeneity in susceptibility);
 782 hetcon (heterogeneity in connectivity with constant CV); hetdyn (heterogeneity in
 783 connectivity with dynamic CV).

784

<i>Country</i>	<i>Model</i>	<i>LL</i>	<i>AIC</i>
<i>Countrywide Data</i>			
<i>England</i>	homog	1.6765E+06	-3.3529E+06
	hetsus	1.6793E+06	-3.3585E+06
	hetcon	1.6785E+06	-3.3571E+06
	hetdyn	1.6790E+06	-3.3579E+06
<i>Belgium</i>	homog	3.3204E+05	-6.6407E+05
	hetsus	3.3325E+05	-6.6649E+05
	hetcon	3.3269E+05	-6.6537E+05
	hetdyn	3.3278E+05	-6.6554E+05
<i>Regional Data</i>			
<i>Portugal</i>	homog	3.36E+05	-6.73E+05
	hetsus	3.55E+05	-7.11E+05
	hetcon	3.55E+05	-7.10E+05
	hetdyn	3.55E+05	-7.11E+05
<i>Spain</i>	homog	1.87E+06	-3.75E+06
	hetsus	1.96E+06	-3.91E+06
	hetcon	1.96E+06	-3.91E+06
	hetdyn	1.47E+06	-2.95E+06
<i>England</i>	homog	2.59E+06	-5.18E+06
	hetsus	2.63E+06	-5.25E+06
	hetcon	2.62E+06	-5.25E+06
	hetdyn	2.62E+06	-5.25E+06
<i>Belgium</i>	homog	3.74E+05	-7.48E+05
	hetsus	3.78E+05	-7.56E+05
	hetcon	3.78E+05	-7.55E+05
	hetdyn	3.78E+05	-7.55E+05

785