

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

3 Ricardo Aguas¹, Jessica G. King², Guilherme Gonçalves³, Marcelo U. Ferreira⁴, M.
4 Gabriela M. Gomes^{5*}

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

7 ² *Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United*
8 *Kingdom.*

9 ³ *Unidade Multidisciplinar de Investigação Biomédica, Instituto de Ciências*
10 *Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal.*

11 ⁴ *Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil.*

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

14

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

17 **Variation in individual susceptibility or exposure to infection accelerates the rate**
18 **at which populations acquire immunity by natural infection. Individuals that are**
19 **more susceptible or more exposed tend to be infected earlier and hence**
20 **selectively removed from the susceptible pool, decelerating the incidence of new**
21 **infections. Eventually, susceptible numbers become low enough to prevent**
22 **epidemic growth or, in other words, the herd immunity threshold is reached.**
23 **Here we fit epidemiological models, with inbuilt distributions of susceptibility or**
24 **exposure, to SARS-CoV-2 epidemics in Spain and Portugal, to estimate basic**
25 **reproduction numbers (R_0) alongside coefficients of individual variation (CV)**
26 **and the effects of control measures. Herd immunity thresholds are then**
27 **calculated as $1 - (1/R_0)^{1/(1+CV^2)}$ or $1 - (1/R_0)^{1/(1+2CV^2)}$, depending on**
28 **whether variation is in susceptibility or exposure. Our inferences result in lower**
29 **herd immunity thresholds than what would be expected if population immunity**
30 **was to be induced by random infection or vaccination, $1 - 1/R_0$.**

31

32 **Introduction**

33 Selection acting on unmeasured individual variation is a well-known source of bias in
34 the analysis of populations. It has been shown to affect measured rates of mortality
35 (Keyfitz and Littman; Vaupel et al 1979; Vaupel and Yashin 1985), the survival of
36 endangered species (Kendall and Fox 2002; Jenouvrier et al 2018), the scope of
37 neutral theories of biodiversity and molecular evolution (Steiner and Tuljapurkar
38 2012, Gomes et al 2019), the risk of diseases whether non-communicable (Aalen et al
39 2015; Stensrud and Valberg 2017) or infections (Anderson et al 1986; Dwyer et al
40 1997; Smith et al 2005; Bellan et al 2015; Gomes et al 2019; Corder et al 2020), and
41 the efficacy of interventions such as vaccines (Halloran et al 1996; O'Hagan et al;
42 Gomes et al 2014; Gomes et al 2016; Langwig et al 2017) or symbionts (Pessoa et al
43 2016; King et al 2018). Building on this knowledge, we addressed how selection on
44 individual variation might affect the course of the coronavirus disease (COVID-19)
45 pandemic (Gomes et al 2020).

46 COVID-19 is an infectious respiratory disease caused by a virus (severe acute
47 respiratory syndrome coronavirus 2 [SARS-CoV-2]), which was first identified in
48 China in late 2019 and has since spread worldwide leading to considerable human
49 suffering and social disruption. European and American continents have been the
50 most affected, with 0.16% and 0.20% of the respective total populations having died
51 as of the 15 July 2021 (WHO 2021). Here we analyse series of daily deaths attributed
52 to COVID-19 in Spain and Portugal (Iberian Peninsula) to study how individual
53 variation in susceptibility and exposure to a respiratory virus affects its epidemic
54 trajectory. Besides adding to the compendium of neglected effects of selection in
55 population dynamics we hope to stimulate a new approach to study epidemic

56 dynamics.

57 The main idea developed here is that individual variation in susceptibility or exposure
58 (connectivity) accelerates the acquisition of immunity in populations. More
59 susceptible and more connected individuals have a higher propensity to be infected
60 and thus are likely to become immune earlier. Due to this selective immunization by
61 natural infection, heterogeneous populations acquire herd immunity by natural
62 infection more efficiently than suggested by models that do not fully account for these
63 types of variation. We integrate continuous distributions of susceptibility or
64 connectivity in otherwise basic epidemiological models for COVID-19, which
65 necessarily account for non-pharmaceutical intervention effects, and generate three
66 types of results. First, at national levels the herd immunity threshold by natural
67 infection declines from around 70% to 20-30%. This is newly reported here for Spain
68 and in agreement with recent estimates for England and Scotland (Gomes et al 2021).
69 Second, these inferences can be made relatively early in the pandemic, such as
70 between first and second waves, provided the first wave is sufficiently large and
71 spatially synchronous. Third, we include a selection of results for Portugal to illustrate
72 how the inferential procedure degenerates when national data do not meet certain
73 conditions.

74 **Individual variation in SARS-CoV-2 transmission**

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

77 *Variation in susceptibility to infection*

78 Individual variation in susceptibility is integrated as a continuously distributed factor
79 that multiplies the force of infection upon individuals (Diekmann et al 1990) in the
80 form of an infinite system of ordinary differential equations (ODEs):

$$81 \quad \dot{S}(x) = -\lambda x S(x), \quad (1)$$

$$82 \quad \dot{E}(x) = \lambda x [S(x) + \sigma R(x)] - \delta E(x), \quad (2)$$

$$83 \quad \dot{I}(x) = \delta E(x) - \gamma I(x), \quad (3)$$

$$84 \quad \dot{R}(x) = (1 - \phi)\gamma I(x) - \sigma \lambda x R(x), \quad (4)$$

85 where $S(x)$ is the density of individuals with susceptibility x , $E(x)$ and $I(x)$ are the
86 densities of individuals who originally had susceptibility x and became exposed and
87 infectious, while $R(x)$ represents those who have recovered and have their
88 susceptibility reduced to a reinfection factor σ due to acquired immunity. Parameter δ
89 is the rate of progression from exposed to a period of maximal infectiousness (=
90 1/5.5 per day [McAloon et al. 2020]), γ is the rate of recovery from maximal
91 infectiousness (= 1/4 per day [Nishiura et al. 2020; Lauer et al. 2020; Li et al.
92 2020]), ϕ is the proportion of individuals who die as a result of infection (= 0.009
93 [Ward et al. 2021]), and:

$$94 \quad \lambda = \frac{\beta}{N} \int \rho E(y) + I(y) dy \quad (5)$$

95 is the average force of infection upon susceptible individuals in a population of
96 approximately constant size N and transmission coefficient β . Standardizing so that
97 susceptibility distributions have mean $\int x g(x) dx = 1$, given a probability density
98 function $g(x)$, the basic reproduction number, defined as the expected number of
99 secondary infections generated by an infected individual in a population that has no
100 specific immunity to the virus (Diekmann et al 1990), is:

$$101 \quad \mathcal{R}_0 = \beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right), \quad (6)$$

102 where ρ is a factor measuring the infectiousness of individuals in compartment E in
103 relation to those in I ($= 0.5$ [Wei et al. 2020; To et al. 2020; Arons et al. 2020; He et
104 al. 2020]). The coefficient of variation in individual susceptibility $CV =$
105 $\sqrt{\int (x - 1)^2 g(x) dx}$ is also treated as a parameter.

106 The basic reproduction number \mathcal{R}_0 is a theoretical framework. It cannot be exactly
107 measured directly but it can be approximated from the initial growth in case numbers.
108 However, as the virus spreads through the population, infected and immune
109 individuals accumulate, reducing the availability of susceptible hosts. As a result,
110 growth in case numbers deviates from being a direct indication of \mathcal{R}_0 but rather of a
111 so-called effective reproduction number \mathcal{R}_{eff} .

112 When susceptibility is given by a gamma distribution and acquired immunity is
113 totally protective, the effective reproduction number is:

$$114 \quad \mathcal{R}_{eff}(t) = \mathcal{R}_0 \left(\frac{S(t)}{N} \right)^{1+CV^2}, \quad (7)$$

115 where $S(t) = \int S(x, t) dx$ is the total number of susceptible individuals at time t , and
116 (Equations 1-4) reduce to a finite system of ODEs:

$$117 \quad \dot{S} = -\beta(\rho E + I) \left(\frac{S}{N} \right)^{1+CV^2}, \quad (8)$$

$$118 \quad \dot{E} = \beta(\rho E + I) \left(\frac{S}{N} \right)^{1+CV^2} - \delta E, \quad (9)$$

$$119 \quad \dot{I} = \delta E - \gamma I, \quad (10)$$

$$120 \quad \dot{R} = (1 - \phi)\gamma I, \quad (11)$$

121 where S , E , I and R are the total numbers of susceptible, exposed, infectious and
122 recovered individuals, respectively (Novozhilov 2008; Montalbán et al. 2020).

123 *Variation in connectivity*

124 In a directly transmitted infectious disease, such as COVID-19, variation in exposure
125 to infection is primarily governed by patterns of connectivity among individuals. Here
126 we incorporate this in the system (Equations 1-4) under the assumption that
127 individuals mix at random (Pastor-Satorras and Vespignani 2001; Miller et al. 2012),
128 while in Supplementary Information we conduct some sensitivity analyses to this
129 assumption. Under random mixing and heterogeneous connectivity, the force of
130 infection is written as:

$$131 \quad \lambda = \frac{\beta \int y[\rho E(y) + I(y)] dy}{N \int yg(y) dy}, \quad (12)$$

132 and the basic reproduction number is:

$$133 \quad \mathcal{R}_0 = (1 + CV^2)\beta \left(\frac{\rho}{\delta} + \frac{1}{\gamma} \right). \quad (13)$$

134 In this setup, when connectivity is given by a gamma distribution and acquired
135 immunity is totally protective, the effective reproduction number is:

$$136 \quad \mathcal{R}_{eff}(t) = \mathcal{R}_0 \left(\frac{S(t)}{N} \right)^{1+2CV^2}, \quad (14)$$

137 and (Equations 1-4) reduce to the finite system of ODEs:

$$138 \quad \dot{S} = -(1 + CV^2)\beta(\rho E + I) \left(\frac{S}{N} \right)^{1+2CV^2}, \quad (15)$$

139 $\dot{E} = (1 + CV^2)\beta(\rho E + I) \left(\frac{S}{N}\right)^{1+2CV^2} - \delta E,$ (16)

140 $\dot{I} = \delta E - \gamma I,$ (17)

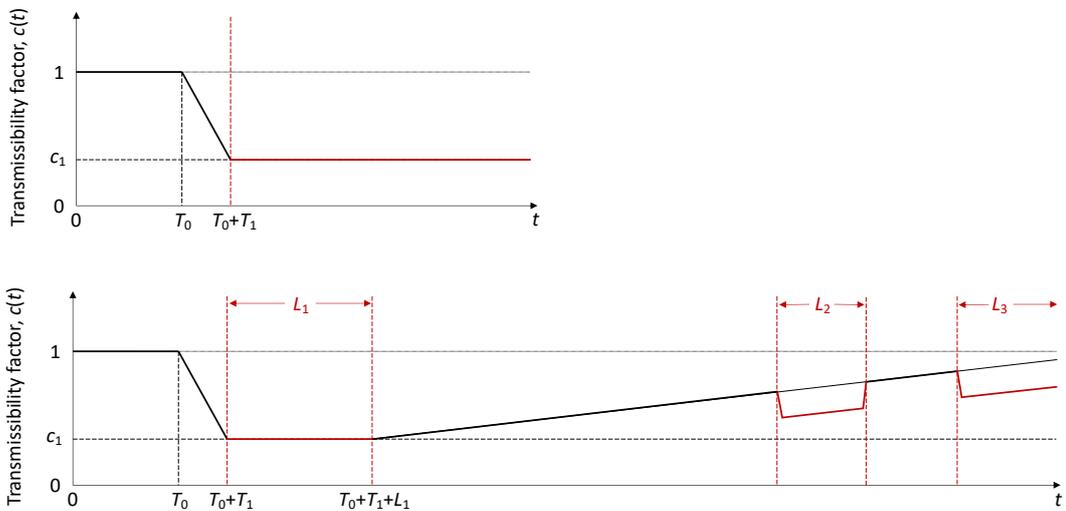
141 $\dot{R} = (1 - \phi)\gamma I,$ (18)

142 where S, E, I and R are the total numbers of susceptible, exposed, infectious and
 143 recovered individuals, respectively (Novozhilov 2008; Montalbán et al. 2020).

144 We have used the reduced systems (Equations 8-11) and (Equations 15-18) in
 145 combination with non-pharmaceutical interventions (NPIs) to analyse series of daily
 146 deaths in the study countries.

147 **Non-pharmaceutical interventions and other transmissibility modifiers**

148 NPIs designed to control transmission typically reduce β and hence \mathcal{R}_0 . Denoting the
 149 time-dependent reproduction number when control measures are in place by $\mathcal{R}_c(t)$,
 150 the modified effective reproduction number is obtained by replacing \mathcal{R}_0 with $\mathcal{R}_c(t)$ in
 151 (Equation 7) and (Equation 14) as appropriate. For the estimation of $\mathcal{R}_c(t)$ we
 152 introduce flexible transmissibility profiles $c(t)$ as illustrated in Figure 1.



153

154 **Figure 1: Transmissibility profile.** Schematic illustration of factor $c(t)$ representing
 155 the combined effects of NPIs, seasonality and viral evolution on the reproduction
 156 number $\mathcal{R}_c(t)$. T_0 is the time when \mathcal{R}_0 begins to decrease due to behavioural change
 157 or seasonality; T_1 is the time interval over which transmission reduces gradually until
 158 the first lockdown; $c_1 \leq 1$ is the average $c(t)$ achieved during the first lockdown; L_1 ,
 159 L_2 and L_3 , denote the length in days of the successive periods of strictest NPI
 160 measures (L_1 being the first lockdown). These profiles are adopted in fits until: 1 July
 161 2020 (top); 1 March 2021 (bottom).

162 *One-wave transmissibility profile*

163 When the model is applied to the first pandemic wave only (until 1 July 2020) we use
 164 the top profile in Figure 1. T_0 is the time when \mathcal{R}_0 begins to decrease due to
 165 behavioural change or seasonality; T_1 is the time interval over which transmission
 166 reduces gradually (assumed linearly, until the first lockdown; $c_1 \leq 1$ is the average
 167 $c(t)$ from the beginning of the first lockdown onwards (14 March in Spain, 19 March
 168 in Portugal). Mathematically this is constructed as:

$$169 \quad c(t) = \begin{cases} 1, & \text{if } 0 < t \leq T_0, \\ 1 - (1 - c_1) \cdot \frac{(1 - T_0)}{T_1}, & \text{if } T_0 < t \leq T_0 + T_1, \\ c_1, & \text{otherwise.} \end{cases} \quad (19)$$

170 *Two-wave transmissibility profile*

171 Applying the model over longer periods which capture multiple waves and multiple
 172 lockdowns requires additional features on the transmissibility profile. Denoting by L_1
 173 the duration of the first lockdown (29 days in Spain, 44 days in Portugal), we allow
 174 restrictions to be progressively relaxed at the end of this period by letting transmission
 175 begin a linear increase such that $c(t)$ reaches 1 in T_2 days, which may or may not be
 176 within the range of the study. Changes in other factors that affect transmission (such
 177 as seasonality or viral evolution) are inseparable from contact changes in this
 178 framework and are also accounted for by $c(t)$. Mathematically this is constructed as:

$$179 \quad c_0(t) = \begin{cases} 1, & \text{if } 0 < t \leq T_0, \\ 1 - (1 - c_1) \cdot \frac{(1 - T_0)}{T_1}, & \text{if } T_0 < t \leq T_0 + T_1, \\ c_1, & \text{if } T_0 + T_1 < t \leq T_0 + T_1 + L_1, \\ 1 - (1 - c_1) \cdot \frac{(T_0 + T_1 + L_1 + T_2 - t)}{T_1}, & \text{otherwise.} \end{cases} \quad (20)$$

180 Second and third lockdowns in the autumn and winter season are implemented as a
 181 further reduction in transmission (by factors c_2 and c_3 , respectively) over the
 182 stipulated time periods (L_2 and L_3 in the bottom panel of Figure 1):

$$183 \quad c_{\text{Spain}}(t) = \begin{cases} c_2 \cdot c_0(t), & \text{if } t \in [26 \text{ October } 2020, 22 \text{ December } 2020], \\ c_3 \cdot c_0(t), & \text{if } t \in [7 \text{ January } 2021, \text{ last data point}], \\ c_0(t), & \text{otherwise.} \end{cases} \quad (21)$$

$$184 \quad c_{\text{Portugal}}(t) = \begin{cases} c_2 \cdot c_0(t), & \text{if } t \in [9 \text{ November } 2020, 21 \text{ December } 2020], \\ c_3 \cdot c_0(t), & \text{if } t \in [14 \text{ January } 2021, \text{ last data point}], \\ c_0(t), & \text{otherwise.} \end{cases} \quad (22)$$

185 In Spain, second and third lockdowns were effectively a single intervention mildly interrupted
 186 by a short relaxation over Christmas, and hence we assume $c_2 = c_3$ in this country. This is
 187 contrasted by Portugal where the third lockdown was much stricter than the second.

188 **Herd immunity thresholds**

189 Individual variation in risk of acquiring infection is under selection by the force of
 190 infection, whether individual differences are due to biological susceptibility,
 191 exposure, or both. The most susceptible or exposed individuals are selectively
 192 removed from the susceptible pool as they become infected and eventually recover
 193 with immunity (some die), resulting in decelerated epidemic growth and accelerated
 194 acquisition of immunity in the population. The herd immunity threshold (HIT) defines
 195 the percentage of the population that needs to be immune to reverse epidemic growth
 196 and prevent future waves. In the absence of NPIs or other transmissibility modifiers,

197 if individual susceptibility or connectivity is gamma-distributed and mixing is
 198 random, basic HIT curves (\mathcal{H}_0) can be derived analytically (Montalbán et al 2020)
 199 from the model systems (Equations 1-4, with the respective forces of infections). In
 200 the case of variation in susceptibility to infection we obtain

$$201 \quad \mathcal{H}_0 = 1 - \left[\frac{1 - \sigma R_0}{(1 - \sigma)R_0} \right]^{\frac{1}{1+CV^2}}, \quad (23)$$

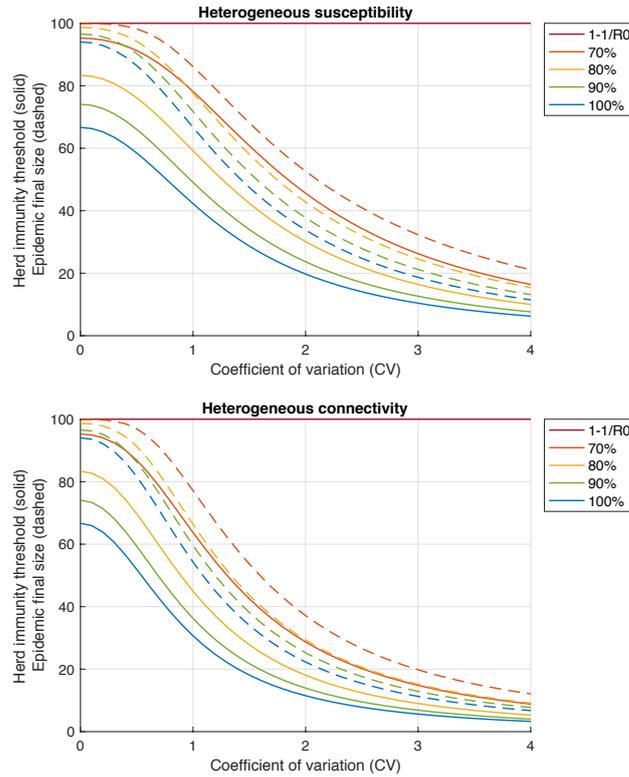
202 while variable connectivity results in

$$203 \quad \mathcal{H}_0 = 1 - \left[\frac{1 - \sigma R_0}{(1 - \sigma)R_0} \right]^{\frac{1}{1+2CV^2}}. \quad (24)$$

204 In less straightforward cases, such as when the characteristics follow a distribution
 205 other than gamma (Gomes et al. 2020), when mixing is not random (Supplementary
 206 Information), when both distributions in susceptibility and connectivity are
 207 considered, or when contact networks are rewired as result of NPIs or otherwise, \mathcal{H}_0
 208 can be obtained numerically.

209 Both (Equation 23) and (Equation 24) convey substantial declines in HIT as
 210 individual variation increases (Figure 2, Gomes et al. 2020), most strikingly over
 211 relatively low CV (from 0 up to 1 or 2). For concreteness, when $R_0 = 3$, $\mathcal{H}_0 = 67\%$
 212 for $CV = 0$, while $CV = 1$ brings \mathcal{H}_0 down to 42% for heterogeneous susceptibility
 213 and 30% for heterogeneous connectivity, and $CV = 2$ brings \mathcal{H}_0 further down to 20%
 214 and 11%, respectively. Accounting for reinfection (i.e. $\sigma > 0$ in the formulas) might
 215 moderate these reductions. In any case, such differences in HIT are an indication of
 216 strong sensitivity of the epidemic dynamics to the parameter CV. Since early in the
 217 COVID-19 pandemic we and others (Gomes et al. 2020, Britton et al 2020) have

218 advocated for attention to this issue. Given that the steeper decline in \mathcal{H}_0 appears to
 219 occur over a range of realistic CVs (Gomes et al. 2020, Gomes et al 2021) it is
 220 plausible that some aspects of the pandemic dynamics need to be reinterpreted in a
 221 broader framework.



222

223 **Figure 2: Herd immunity threshold and epidemic final size.** Herd immunity
 224 thresholds (solid curves) are calculated according to (Equation 23) for heterogeneous
 225 susceptibility and (Equation 24) for heterogeneous connectivity. Final sizes of the
 226 corresponding unmitigated epidemics are also shown (dashed). Curves are generated
 227 for different values of the efficacy of immunity conferred by natural infection ($1 - \sigma$)
 228 as displayed in the legend.

229 We emphasise, however, that \mathcal{H}_0 is a theoretical framework almost to the extent that
 230 \mathcal{R}_0 is a theoretical framework. It cannot be measured directly when epidemic
 231 trajectories are affected by interventions, but it can be inferred indirectly from
 232 epidemiological data. By construction, \mathcal{H}_0 changes if the parameters that determine
 233 its value change. Most notably, natural changes in \mathcal{R}_0 through time, which can
 234 happen due to seasonal forces or viral evolution, transfer to \mathcal{H}_0 according to

235 (Equation 23), (Equation 24), or even their homogeneity equivalent $1 - 1/\mathcal{R}_0$. As a
236 result, the percentage of the population immune required to prevent sustained
237 epidemic growth may deviate from the initial \mathcal{H}_0 . Notwithstanding, a model with
238 lower \mathcal{H}_0 outputs smaller epidemics than a model with higher \mathcal{H}_0 , all non-basic
239 processes being the same.

240 The phenomenon of variation and selection which accounts for lower HIT was widely
241 explained around mid-2020 (Hartnett 2020) and generated broad public interest in the
242 context of COVID-19. By early 2021, vaccines had become available, and the
243 narrative changed to imply that the HIT might be unachievable for COVID-19
244 (Aschwanden 2021). This apparent contradiction stems from the adoption of less
245 fundamental meaning of HIT by the latter. While Hartnett (2020) writes about using
246 the basic \mathcal{H}_0 to assess pandemic potential, stressing how that is weakened by
247 variation and selection by natural infection, Aschwanden (2021) endorses a more
248 common sensical view that reinfection may become frequent enough to make herd
249 immunity unachievable. In the light of the theory presented here – specifically
250 (Equation 23) and (Equation 24) – these views are transversal and not contradictory.

251 **Data**

252 We fit compartmental models to publicly available epidemiological data from the
253 coronavirus dashboards for Spain [<https://cneocovid.isciii.es/covid19>] and Portugal
254 [<https://covid19.min-saude.pt/ponto-de-situacao-atual-em-portugal>] to fit the models
255 and estimate parameters of interest. Namely, we fit model reconstructed mortality
256 timeseries assuming a fixed infection fatality ratio to datasets containing daily deaths,
257 $\{(k, \tilde{y}_k)\}_{k=1}^n$, where $k = 1$ is the day when the cumulative moving average of death
258 numbers exceed $5 \cdot 10^{-7}$ of the population (7 March in Spain, 19 March 2020 in

259 Portugal).

260 Model fits were carried out to the raw series of daily deaths until the 1 July 2020 in
261 the first instance (to cover the first wave of the epidemic in the study countries), and
262 until the 1 March 2021 in an extended analysis (as a compromise between having a
263 series sufficiently long to capture much of the second wave and not so long that it
264 would be affected by vaccination and require the vaccine to be modelled). We defined
265 the initial conditions as:

$$266 \quad R(0) = 0, \tag{25}$$

$$267 \quad I(0) = \frac{\tilde{y}_{1+\eta}}{1 - \exp(-\phi\gamma)}, \tag{26}$$

$$268 \quad E(0) = \frac{I(0)}{1 - \exp(-\gamma)}, \tag{27}$$

$$269 \quad S(0) = N - E(0) - I(0) - R(0), \tag{28}$$

270 and the population sizes N were obtained from the most recent respective census
271 (approx. 46.94 million in Spain, 10.28 million in Portugal, 3.57 million in the North
272 Region of Portugal, 3.66 million in Lisbon and Tagus Valley Region of Portugal).

273 **Model fitting and parameter estimating**

274 We assumed that reinfection was negligible throughout the study period. A study
275 conducted in England (Hall et al. 2021), between June 2020 and January 2021,
276 concluded that previous SARS-CoV-2 infection induced 84% effective immunity to
277 future infections. In (Gomes et al. 2121) we fit models to daily COVID-19 deaths in
278 England and Scotland, assuming no reinfection (i.e. 100% effective immunity) or
279 90% effective immunity, and found this alteration to have no significant effect on

280 projected model trajectories. Basically, the fit readjusts the parameters when
281 reinfection is added to the model in such a way that the output solutions remain
282 similar.

283 Parameter estimation was performed with the software MATLAB using the PESTO
284 (Parameter ESTimation Toolbox) package (Stapor et al. 2018). We assumed the
285 number of SARS-CoV-2 infections to be Poisson distributed.

286 We approximate the dynamics of COVID-19 deaths by estimating the set of
287 parameters θ that maximises the log-likelihood (LL) of observing the daily numbers
288 of reported deaths Y :

$$289 \quad LL(\theta|Y) = - \sum_{k=1}^n y(k, \theta) + \sum_{k=1}^n \tilde{y}(k) \ln(y(k, \theta)) - \sum_{k=1}^n \ln(\tilde{y}(k)!), \quad (29)$$

290 where $y(k, \theta)$ are the simulated model output numbers of COVID-19 deaths at day k
291 for the set of parameters θ , $Y = \{(k, \tilde{y}_k)\}_{k=1}^n$ are the numbers of daily reported
292 deaths, and n is the total number of days included in the analysis.

293 **Fitting models to one pandemic wave**

294 The models exploring heterogeneity in susceptibility (Equations 8-11) and
295 connectivity (Equations 15-18) both with transmissibility profile as in (Equation 19)
296 were fit to COVID-19 deaths recorded daily until 1 July 2020, in Spain and Portugal.
297 A homogeneous version obtained by setting $CV = 0$ in either model was also fitted.
298 Results for Spain are shown in Table 1 and Figure 3, and for Portugal in Table 2 and
299 Figure 4.

300 We estimate the basic reproduction number \mathcal{R}_0 with 95% credible intervals (CI)
301 around 3.5 – 3.8 in Spain and 2.3 – 3.4 in Portugal. For the minimal transmissibility
302 factor c_1 we estimate 0.21 – 0.27 when individual variation is allowed and 0.18 –
303 0.19 with the homogeneity constraint in Spain, while in Portugal we estimate the
304 wider intervals 0.26 – 0.39. For coefficients of variation in Spain, we estimate CV in
305 the range 1.3 – 2.4 under heterogeneous susceptibility and 1.0 – 1.7 under
306 heterogeneous connectivity. In Portugal, we obtain the much wider and uninformative
307 ranges 0.0 – 3.2.

308 Left plots in Figures 3 and 4 show the best fitting model solutions generated from the
309 median posterior estimates of each parameter in the respective countries as well as the
310 95% CI generated from 100,000 posterior samples. The basic herd immunity
311 thresholds \mathcal{H}_0 , calculated from \mathcal{R}_0 and CV estimates, are $\mathcal{H}_0 = 19\%$ (95% CI, 13-
312 32%) under heterogeneous susceptibility, $\mathcal{H}_0 = 19\%$ (95% CI, 13-36%) under
313 heterogeneous connectivity, and $\mathcal{H}_0 = 72\%$ (95% CI, 71-73%) when homogeneity is
314 imposed, in Spain. In Portugal, we obtain $\mathcal{H}_0 = 19\%$ (95% CI, 5-69%) under
315 heterogeneous susceptibility, $\mathcal{H}_0 = 13\%$ (95% CI, 3-69%) under heterogeneous
316 connectivity, and $\mathcal{H}_0 = 67\%$ (95% CI, 66-69%) when homogeneity is imposed.
317 Credible intervals for \mathcal{H}_0 in Portugal are wide and uninformative when individual
318 variation is allowed which is expected given the wide ranges obtained for CV. NPIs in
319 Portugal were initiated very early in the epidemic which resulted in transmissibility
320 reductions blending with \mathcal{R}_0 , making parameter identification a major challenge from
321 the data accessible to us.

322 In Spain, where the three models provide good fits to the data, model selection criteria
323 such as AIC (Akaike information criterion) support the heterogeneous

324 implementations. To better distinguish the various models, we run the respective
325 systems of equations forward, under a set of conventions, and compare the respective
326 projected outcomes. Right plots in Figures 3 were generated by taking the end
327 conditions of the left plots, moving all exposed and infectious individuals to
328 recovered (except a residual proportion to seed a new outbreak) and running each
329 model with the estimated \mathcal{R}_0 and CV until the susceptible pool has been effectively
330 depleted in all implementations. In this manner we can visualise how much more
331 infection appears to be ahead when models are constrained to be homogeneous (a
332 manifestation of their relatively higher \mathcal{H}_0 . Roughly, epidemics peak one order of
333 magnitude higher when models are homogeneous. This must have broad implications,
334 which we believe remain largely unappreciated, for how a population will experience
335 an epidemic, irrespective of how this basic scenario is adapted to specific factors such
336 behavioural patterns, seasonality, viral evolution, or vaccination.

337 The same analysis applied to Portugal, selects in favour of the homogeneity
338 assumption and larger projected waves. We recall, however, the great uncertainty
339 associated with these specific results. We investigate this further through fittings to
340 longer data series, in the first instance.

341 **Table 1: Model parameters for Spain (one wave).** Estimated by Bayesian inference
342 based on daily deaths until 1 July 2020. Model selection based on maximum log-
343 likelihood (LL) and Akaike information criterion (AIC). Best fitting models have
344 lower AIC scores (best in red, second best in blue). Herd immunity threshold (\mathcal{H}_0)
345 derived from estimated \mathcal{R}_0 and CV.

	Heterogeneous susceptibility		Heterogeneous connectivity		Homogeneous	
	Median	95% CI	Media n	95% CI	Media n	95% CI
<i>Estimated parameters</i>						
T_0	5.44	[3.52, 6.07]	5.40	[3.45, 6.07]	5.48	[3.73, 6.04]
c_1	0.24	[0.21, 0.27]	0.24	[0.21, 0.28]	0.18	[0.18, 0.19]
η	16*	[15.53, 16.47]	16*	[15.52, 16.48]	16*	[15.52, 16.47]
\mathcal{R}_0	3.59	[3.51, 3.79]	3.59	[3.51, 3.80]	3.53	[3.47, 3.71]

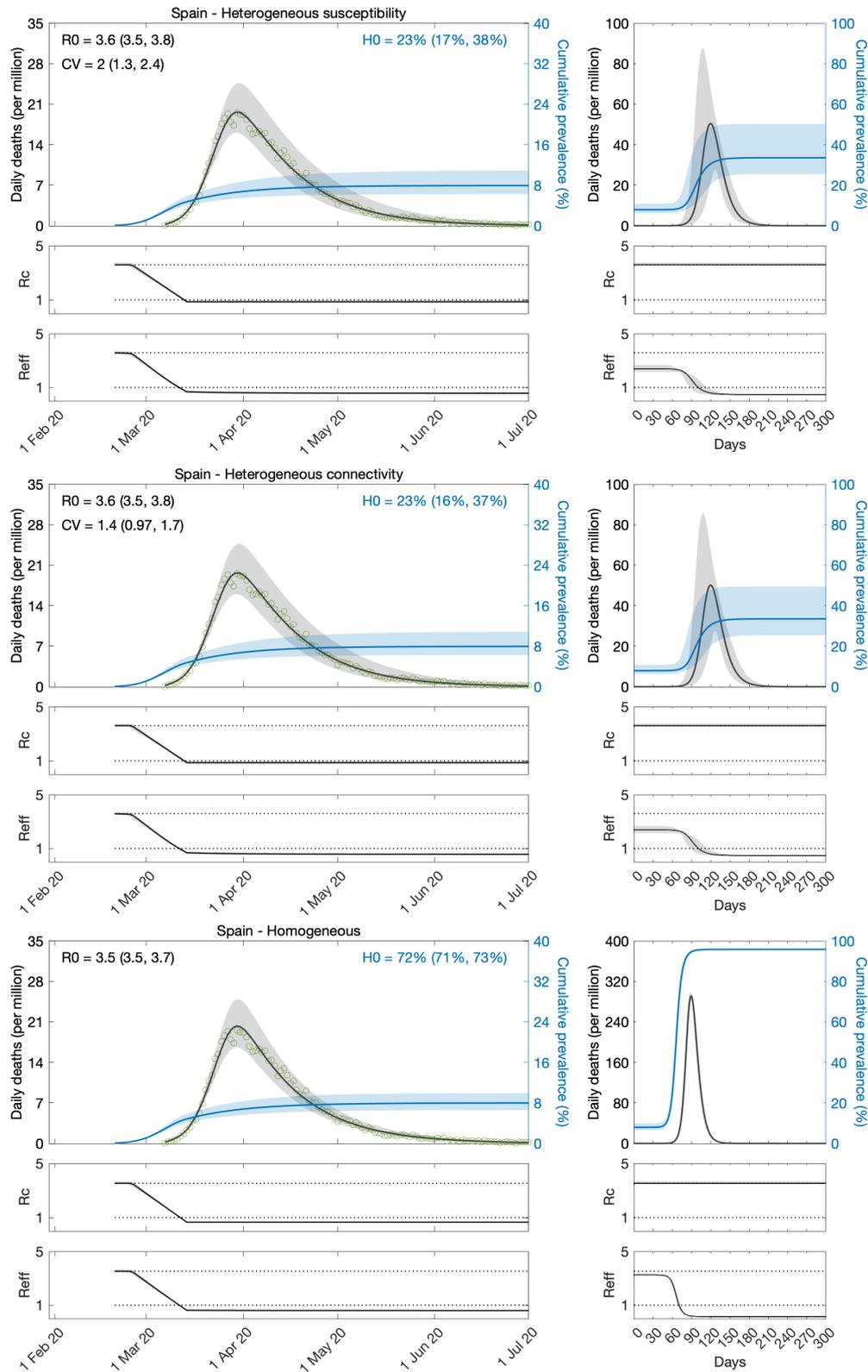
CV	1.98	[1.33, 2.44]	1.41	[0.97, 1.74]	0	-
<i>Derived from estimated parameters</i>						
\mathcal{H}_0	23%	[17%, 38%]	23%	[16%, 37%]	72%	[71%, 73%]
<i>Model selection</i>						
LL	-503.78		-503.50		-518.33	
AIC	1017.60		1017.00		1044.70	

346 * Rounded median resulting in integer number of days.

347 **Table 2: Model parameters for Portugal (one wave).** Estimated by Bayesian
348 inference based on daily deaths until 1 July 2020. Model selection based on maximum
349 log-likelihood (LL) and Akaike information criterion (AIC). Best fitting models have
350 lower AIC scores (best in red, second best in blue). Herd immunity threshold (\mathcal{H}_0)
351 derived from estimated \mathcal{R}_0 and CV.

	Heterogeneous susceptibility		Heterogeneous connectivity		Homogeneous	
	Median	95% CI	Median	95% CI	Median	95% CI
<i>Estimated parameters</i>						
T_0	0.05	[0.00, 0.10]	0.05	[0.00, 0.10]	0.05	[0.00, 0.10]
c_1	0.31	[0.26, 0.36]	0.32	[0.26, 0.39]	0.31	[0.26, 0.35]
η	20*	[14.75, 25.72]	20*	[13.27, 25.94]	21*	[15.56, 26.04]
\mathcal{R}_0	2.62	[2.29, 3.21]	2.66	[2.27, 3.39]	2.61	[2.28, 3.15]
CV	0.31	[0.01, 3.24]	0.38	[0.01, 3.04]	0	-
<i>Derived from estimated parameters</i>						
\mathcal{H}_0	58%	[7%, 69%]	53%	[4%, 70%]	62%	[56%, 68%]
<i>Model selection</i>						
LL	-287.10		-287.14		-287.34	
AIC	584.20		584.28		582.67	

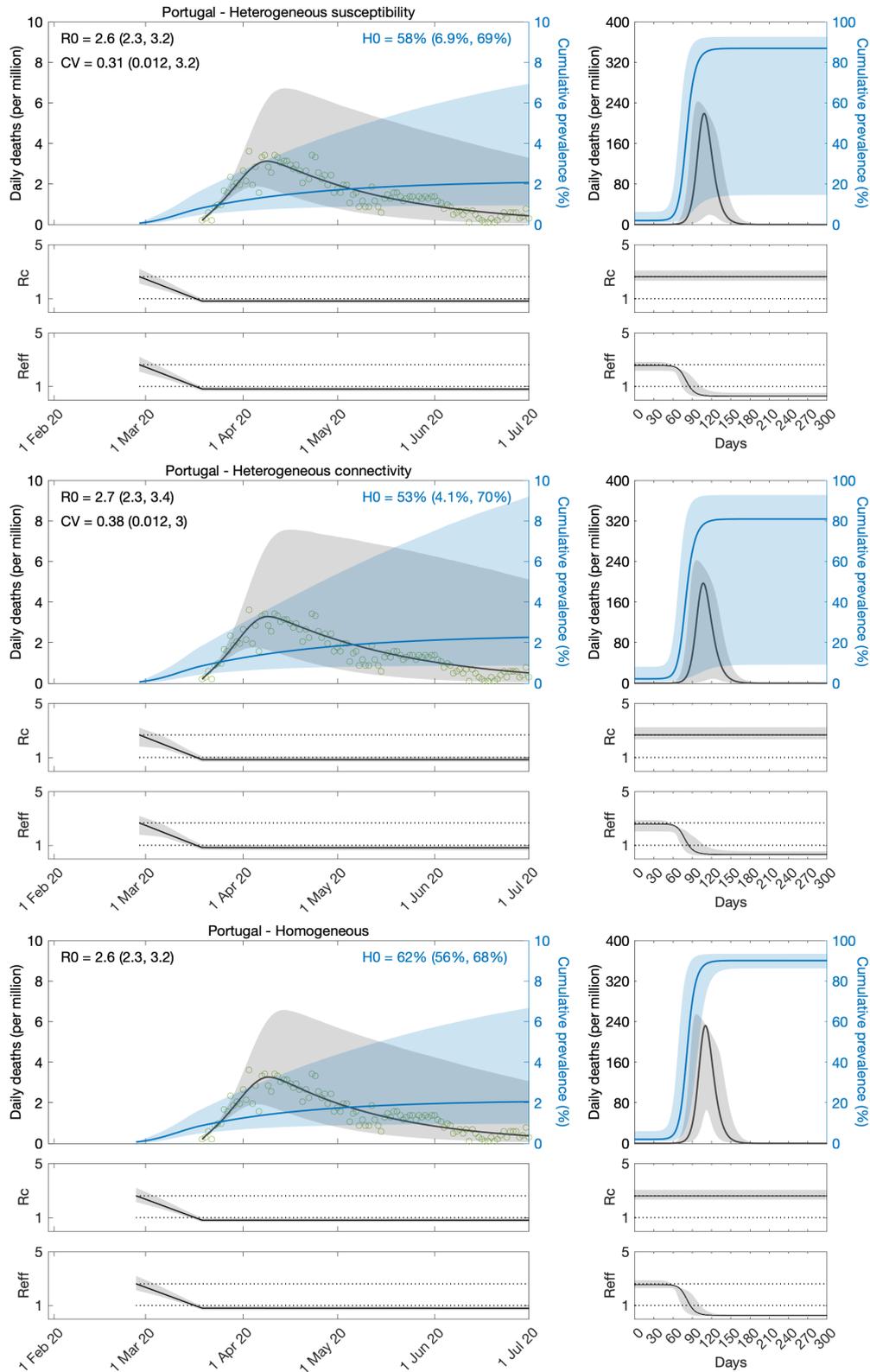
352 * Rounded median resulting in integer number of days.



353

354 **Figure 3: Estimating SARS-CoV-2 transmission in Spain by fitting one wave of**
 355 **COVID-19 deaths.** Variation in susceptibility (top panels); variation in connectivity
 356 (middle panels); and homogeneous model (bottom panels). Susceptibility or
 357 connectivity factors implemented as gamma distributions. Controlled (R_c) and
 358 effective (R_{eff}) reproduction numbers are displayed on shallow panels underneath the
 359 main plots. Basic reproduction number, coefficients of variation and transmissibility

360 profile parameters estimated by Bayesian inference as described in Methods
 361 (estimates in Table 1). Curves represent model reconstructions from the median
 362 posterior parameter estimates. Shades represent 95% credible intervals from 100,000
 363 posterior samples.



364

365 **Figure 4: Estimating SARS-CoV-2 transmission in Portugal by fitting one wave of**
366 **COVID-19 deaths.** Variation in susceptibility (top panels); variation in connectivity
367 (middle panels); and homogeneous model (bottom panels). Susceptibility or
368 connectivity factors implemented as gamma distributions. Controlled (R_c) and
369 effective (R_{eff}) reproduction numbers are displayed on shallow panels underneath the
370 main plots. Basic reproduction number, coefficients of variation and transmissibility
371 profile parameters estimated by Bayesian inference as described in Methods
372 (estimates in Table 2). Curves represent model reconstructions from the median
373 posterior parameter estimates. Shades represent 95% credible intervals from 100,000
374 posterior samples.

375 **Fitting models to two serial pandemic waves**

376 Here we take the models with heterogeneity in susceptibility (Equations 8-11) and
377 connectivity (Equations 15-18) and apply the transmissibility profile in (Equations
378 20-22) to both before fitting the model outputs to COVID-19 deaths recorded daily
379 until 1 March 2021, in Spain and Portugal. As before a homogeneous version
380 obtained by setting $CV = 0$ was also fitted. Results for Spain are shown in Table 3
381 and Figure 5, and for Portugal in Table 4 and Figure 6.

382 We estimate \mathcal{R}_0 with 95% CI around 3.9 – 4.4 in Spain and 2.7 – 4.6 in Portugal.
383 For the minimal transmissibility factor c_1 we estimate 0.20 – 0.21 when individual
384 variation is allowed and 0.17 with the homogeneity constraint in Spain, while in
385 Portugal we obtain 0.23 – 0.32 with individual variation and 0.16 – 0.18 without. For
386 coefficients of variation in Spain, we estimate CV around 2.0 under heterogeneous
387 susceptibility and 1.3 under heterogeneous connectivity. In Portugal, the estimates are
388 around 0.5.

389 Left plots in Figures 5 and 6 show best fitting model solutions as well as the
390 respective 95% CI. In Spain, basic herd immunity thresholds calculated from best-
391 fitting \mathcal{R}_0 and CV are $\mathcal{H}_0 = 25\%$ under heterogeneous susceptibility, $\mathcal{H}_0 = 27\%$
392 under heterogeneous connectivity, and $\mathcal{H}_0 = 76\%$ when homogeneity is imposed. In

393 Portugal, we obtain again wider and higher ranges: $\mathcal{H}_0 = 56\%$ (95% CI, 52-57%)
 394 under heterogeneous susceptibility, $\mathcal{H}_0 = 56\%$ (95% CI, 54-57%) under
 395 heterogeneous connectivity, and $\mathcal{H}_0 = 77\%$ (95% CI, 76-78%) when homogeneity is
 396 imposed.

397 Puzzling, as in the case of shorter data series, model selection supports the
 398 homogeneous model for Portugal while still favouring the incorporation of individual
 399 variation in Spain. Moreover, for each country, results are consistent whether we base
 400 our estimates on one or two waves of the national epidemic. This consistency was
 401 also verified in England and Scotland (Gomes et al 2021). This suggests that the
 402 earlier initiation of NPI in Portugal may not in itself explain why results for Portugal
 403 contrast with those for other countries studied.

404 **Table 3: Model parameters for Spain (two waves).** Estimated by Bayesian
 405 inference based on daily deaths until 1 March 2021. Model selection based on
 406 maximum log-likelihood (LL) and Akaike information criterion (AIC). Best fitting
 407 models have lower AIC scores (best in red, second best in blue). Herd immunity
 408 threshold (\mathcal{H}_0) derived from estimated \mathcal{R}_0 and CV.

	Heterogeneous susceptibility		Heterogeneous connectivity		Homogeneous	
	Median	95% CI	Median	95% CI	Median	95% CI
<i>Estimated parameters</i>						
T_0	3.22	[3.18, 3.31]	6.06	[6.06, 6.07]	5.99	[5.98, 5.99]
T_2	448.42	[446.58, 452.53]	429.91	[427.04, 433.98]	783.88	[777.04, 784.12]
c_1	0.20	[0.20, 0.20]	0.21	[0.21, 0.21]	0.17	[0.17, 0.17]
c_2	0.71	[0.71, 0.71]	0.68	[0.68, 0.68]	0.50	[0.50, 0.50]
η	12*	[12.42, 12.50]	12*	[11.54, 12.45]	10*	[10.04, 10.07]
\mathcal{R}_0	4.36	[4.32, 4.37]	3.93	[3.92, 3.95]	4.11	[4.10, 4.11]
CV	2.02	[2.01, 2.02]	1.30	[1.28, 1.31]	0	-
<i>Derived from estimated parameters</i>						
\mathcal{H}_0	25%	[25%, 25%]	27%	[27%, 27 %]	76%	[76%, 76%]
<i>Model selection</i>						
LL	-3511.40		-3269.90		-5826.40	
AIC	7036.80		6553.90		11665.00	

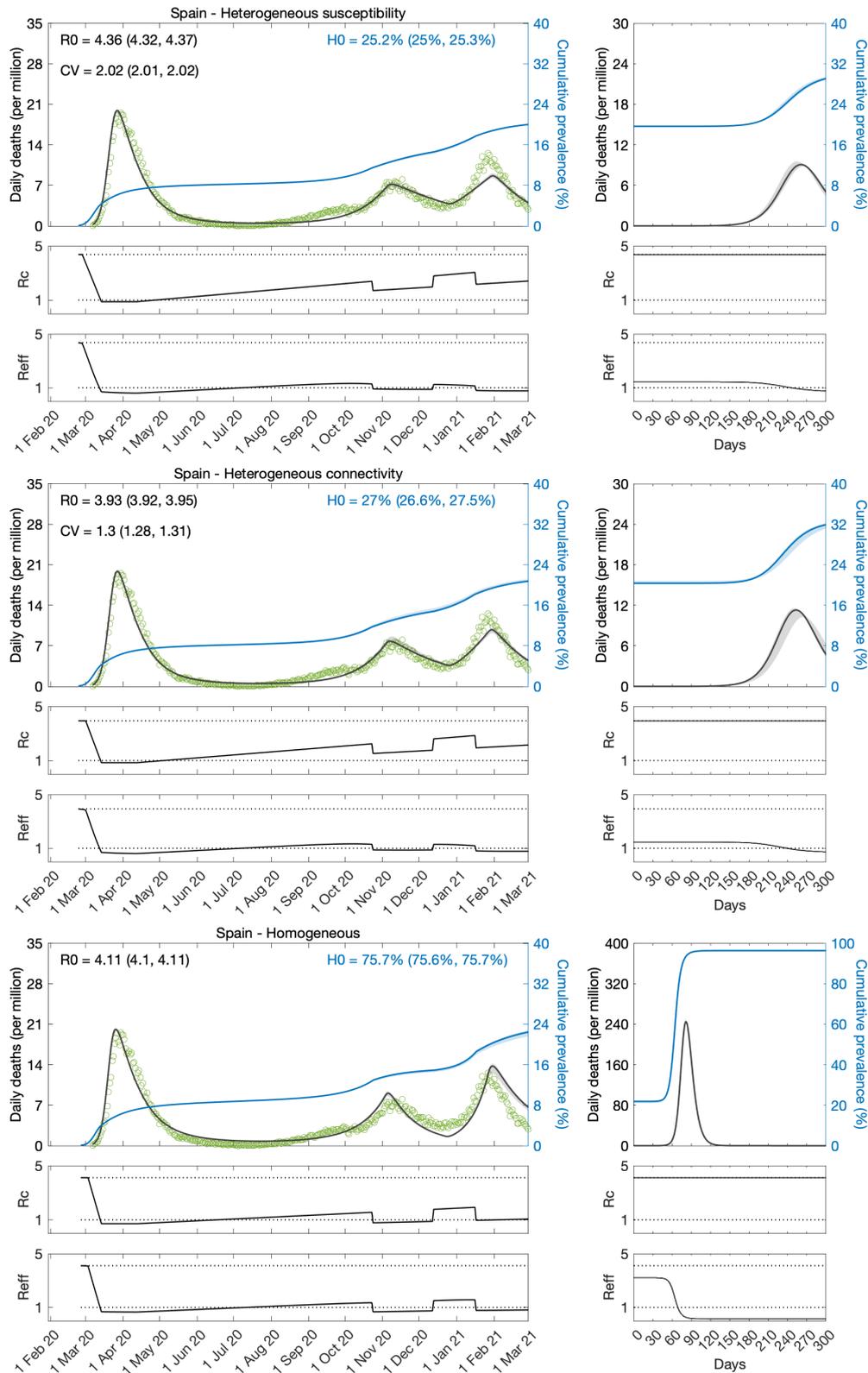
409 * Rounded median resulting in integer number of days.

410 **Table 4: Model parameters for Portugal (two waves).** Estimated by Bayesian
 411 inference based on daily deaths until 1 March 2021. Model selection based on
 412 maximum log-likelihood (LL) and Akaike information criterion (AIC). Best fitting

413 models have lower AIC scores (best in red, second best in blue). Herd immunity
 414 threshold (\mathcal{H}_0) derived from estimated \mathcal{R}_0 and CV.

	Heterogeneous susceptibility		Heterogeneous connectivity		Homogeneous	
	Median	95% CI	Median	95% CI	Median	95% CI
<i>Estimated parameters</i>						
T_0	0.03	[0.01, 0.05]	0.04	[0.00, 0.10]	0.05	[0.00, 0.10]
T_2	585.14	[573.66, 672.90]	644.26	[615.71, 695.38]	899.67	[872.69, 946.07]
c_1	0.27	[0.27, 0.32]	0.24	[0.23, 0.25]	0.17	[0.16, 0.18]
c_2	0.71	[0.71, 0.87]	0.71	[0.70, 0.77]	0.66	[0.64, 0.68]
c_3	0.34	[0.27, 0.48]	0.42	[0.39, 0.48]	0.35	[0.34, 0.36]
η	16*	[15.52, 17.39]	14*	[12.76, 15.34]	12*	[11.54, 12.33]
\mathcal{R}_0	2.94	[2.66, 2.96]	3.25	[3.14, 3.32]	4.36	[4.20, 4.58]
CV	0.55	[0.54, 0.59]	0.47	[0.47, 0.49]	0	-
<i>Derived from estimated parameters</i>						
\mathcal{H}_0	56%	[52%, 57%]	56%	[54%, 57%]	77%	[76%, 78%]
<i>Model selection</i>						
LL	-1550.80		-1376.80		-1234.30	
AIC	3117.50		2769.50		2482.50	

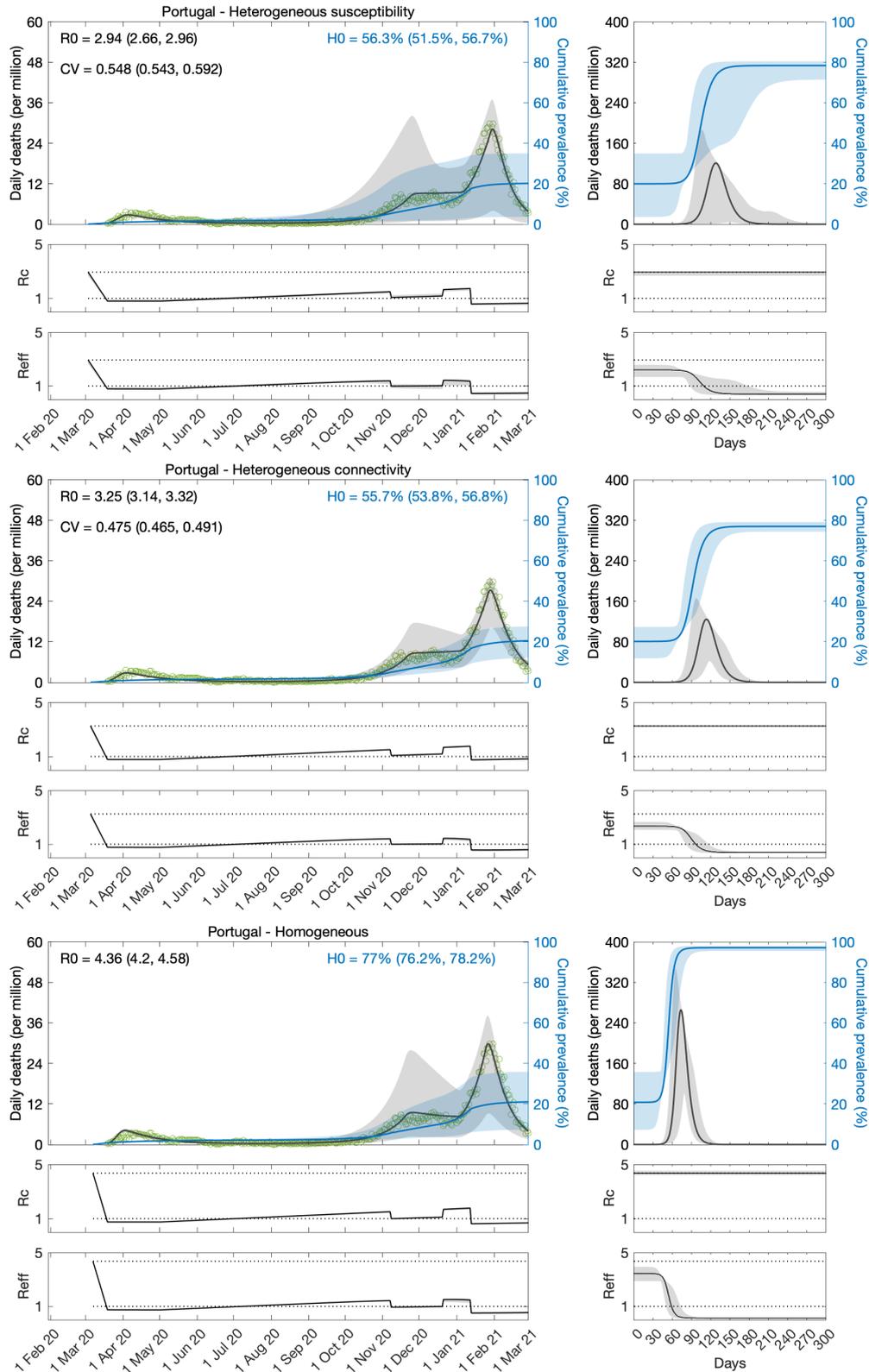
415 * Rounded median resulting in integer number of days.



416

417 **Figure 5: Estimating SARS-CoV-2 transmission in Spain by fitting two serial**
 418 **waves of COVID-19 deaths.** Variation in susceptibility (top panels); variation in
 419 connectivity (middle panels); and homogeneous model (bottom panels). Susceptibility
 420 or connectivity factors implemented as gamma distributions. Controlled (R_c) and
 421 effective (R_{eff}) reproduction numbers are displayed on shallow panels underneath the
 422 main plots. Basic reproduction number, coefficients of variation and transmissibility

423 profile parameters estimated by Bayesian inference as described in Methods
 424 (estimates in Table 3). Curves represent model reconstructions from the median
 425 posterior parameter estimates. Shades represent 95% credible intervals from 100,000
 426 posterior samples.



427

428 **Figure 6: Estimating SARS-CoV-2 transmission in Portugal by fitting two serial**
429 **waves of COVID-19 deaths.** Variation in susceptibility (top panels); variation in
430 connectivity (middle panels); and homogeneous model (bottom panels). Susceptibility
431 or connectivity factors implemented as gamma distributions. Controlled (R_c) and
432 effective (R_{eff}) reproduction numbers are displayed on shallow panels underneath the
433 main plots. Basic reproduction number, coefficients of variation and transmissibility
434 profile parameters estimated by Bayesian inference as described in Methods
435 (estimates in Table 4). Curves represent model reconstructions from the median
436 posterior parameter estimates. Shades represent 95% credible intervals from 100,000
437 posterior samples.

438 **Fitting models to regional data in Portugal**

439 Intrigued by the puzzling results for Portugal at country level, we gathered regional
440 data. We found that the epidemic was considerably different between the two largest
441 regions: North, home to roughly one third of the Portuguese population; and Lisbon
442 and Tagus Valley, home to another third. Asynchrony of epidemic dynamics between
443 regions of similar sizes may require disaggregated analyses. We then decided to fit
444 the mortality data for the two regions simultaneously, estimating common parameters
445 to describe country-wide lockdowns, and regions-specific parameters to describe the
446 basic transmission dynamics (\mathcal{R}_0 and CV). The results are provided in Tables 5, 6 and
447 Figures 7-9. According to these analyses, the best-fitting models include
448 heterogeneity.

449 Basic herd immunity thresholds inferred by fits to the longer series (until 1 March
450 2021) are around 25 – 34% in the North (similar as Spain, England, and Scotland)
451 and 47 – 52% in Lisbon and Tagus Valley. The higher \mathcal{H}_0 in the capital region
452 results from the estimation of a lower CV. This may be real and due to the more urban
453 character of Lisbon and Tagus Valley or, in contrast, it may be a spurious result of the
454 lack of a first wave in the region to inform the model. It would be interesting to

455 replicate the regional analysis in other countries to investigate to what extent more
 456 urban regions have higher HIT.

457 Fits to the shorter series (until 1 July 2020) are again inconclusive. Not only the
 458 uncertainty around parameter estimates is large but some estimates are implausible.
 459 First, \mathcal{R}_0 around 2 or less is on the low end of consensus estimates (Flaxman et al.
 460 2020; Keeling et al. 2020; Viana et al. 2021; Wood 2021). More strikingly, the
 461 algorithm is incapable of estimating CV from these regional series, resulting in
 462 convergence to the upper and lower limits of the prior distribution (uniform between
 463 $0.01 - \sqrt{12}$) in the North and Lisbon regions, respectively.

464 In summary, the results reported in this section support two notions. First,
 465 asynchronous dynamics may compromise parameter estimation based on model
 466 fittings to aggregated data. This should depend on whether the asynchrony in question
 467 is between similar-sized regions. Second, the early estimation of parameters for
 468 heterogeneous models (especially CV) may require a larger first wave than that
 469 required by models that are either homogeneous (Flaxman et al. 2020; Wood 2021) or
 470 have their heterogeneity informed directly by specific data (Keeling et al. 2020). The
 471 downside of these other approaches, however, is that heterogeneity is either absent or
 472 possibly incomplete, resulting in reduced selection and biased estimates. Among the
 473 studies we have completed, England, Scotland and Spain had sufficiently sized first
 474 waves to inform the inference of CV while Portugal did not.

475 **Table 5: Model parameters for the North and Lisbon regions of Portugal (one**
 476 **wave).** Estimated by Bayesian inference based on daily deaths until 1 July 2020.
 477 Model selection based on maximum log-likelihood (LL) and Akaike information
 478 criterion (AIC). Best fitting models have lower AIC scores (best in red, second best in
 479 blue). Herd immunity threshold (\mathcal{H}_0) derived from estimated \mathcal{R}_0 and CV.

Heterogeneous susceptibility		Heterogeneous connectivity		Homogeneous	
Median	95% CI	Median	95% CI	Median	95% CI

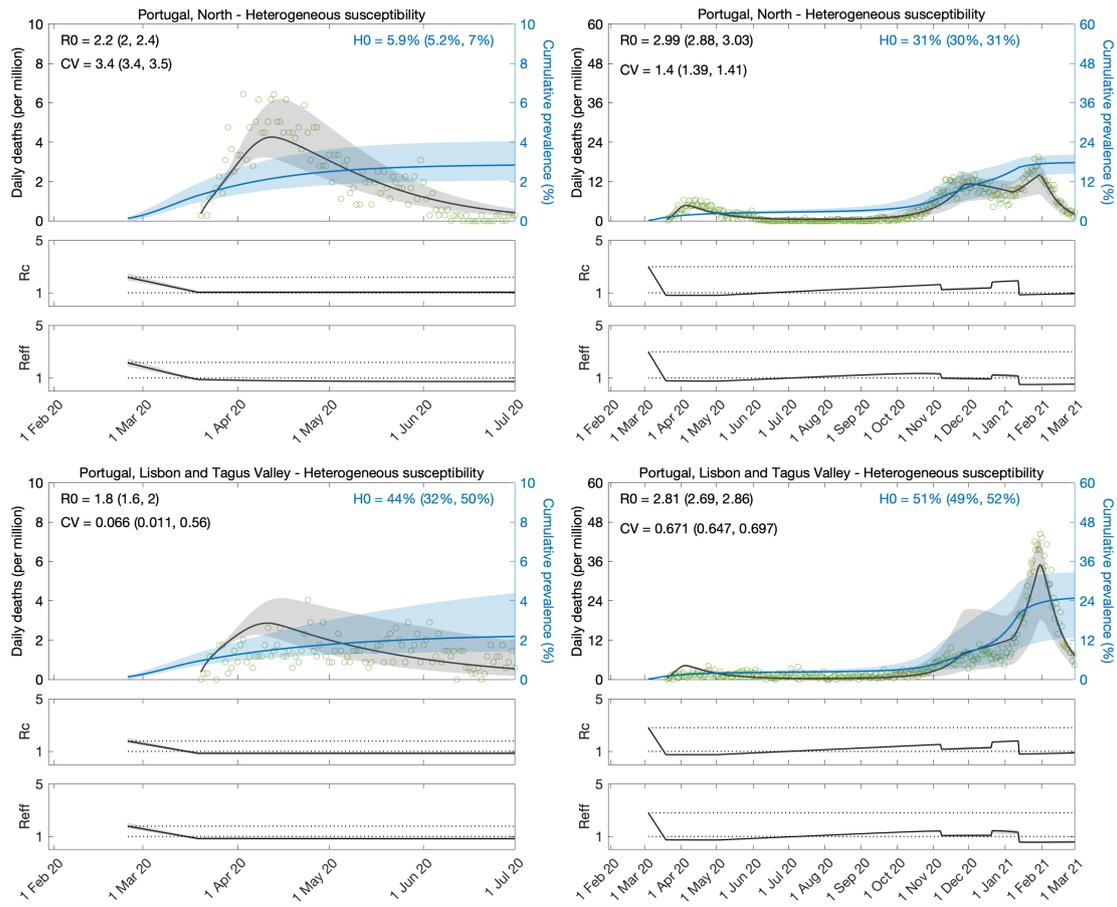
<i>Common parameters</i>						
T_0	0.05	[0.00, 0.10]	0.05	[0.00, 0.10]	0.05	[0.00, 0.10]
c_1	0.48	[0.43, 0.52]	0.55	[0.47, 0.61]	0.43	[0.40, 0.47]
η	24*	[19.45, 28.99]	21*	[14.18, 27.37]	29*	[28.50, 28.50]
<i>North</i>						
\mathcal{R}_0	2.17	[2.00, 2.45]	2.43	[2.17, 2.92]	1.91	[1.80, 2.09]
CV	3.44	[3.38, 3.46]	3.45	[3.39, 3.46]	0	-
\mathcal{H}_0	5.9%	[5.2%, 7.0%]	3.5%	[3.0%, 4.4%]	48%	[44%, 52%]
<i>Lisbon and Tagus Valley</i>						
\mathcal{R}_0	1.78	[1.65, 2.00]	1.65	[1.49, 1.96]	1.85	[1.74, 2.02]
CV	0.07	[0.01, 0.56]	0.07	[0.01, 0.57]	0	-
\mathcal{H}_0	44%	[32%, 50%]	39%	[21%, 49%]	46%	[43%, 51%]
<i>Model selection</i>						
LL	-520.56		-479.75		-578.50	
AIC	1055.10		973.49		1167.00	

480 * Rounded median resulting in integer number of days.

481 **Table 6: Model parameters for the North and Lisbon regions of Portugal (two**
482 **waves).** Estimated by Bayesian inference based on daily deaths until 1 March 2021.
483 Model selection based on maximum log-likelihood (LL) and Akaike information
484 criterion (AIC). Best fitting models have lower AIC scores (best in red, second best in
485 blue). Herd immunity threshold (\mathcal{H}_0) derived from estimated \mathcal{R}_0 and CV.

	Heterogeneous susceptibility		Heterogeneous connectivity		Homogeneous	
	Median	95% CI	Median	95% CI	Median	95% CI
<i>Common parameters</i>						
T_0	0.06	[0.00, 0.10]	0.09	[0.06, 0.10]	0.05	[0.03, 0.05]
T_2	498.77	[489.85, 515.47]	499.41	[462.35, 547.99]	960.64	[960.28, 961.42]
c_1	0.27	[0.26, 0.29]	0.25	[0.25, 0.25]	0.17	[0.17, 0.17]
c_2	0.75	[0.73, 0.82]	0.71	[0.69, 0.75]	0.73	[0.73, 0.73]
c_3	0.44	[0.42, 0.45]	0.49	[0.45, 0.50]	0.36	[0.36, 0.36]
η	16*	[15.36, 16.46]	15*	[13.75, 15.16]	12*	[11.62, 11.64]
<i>North</i>						
\mathcal{R}_0	2.99	[2.88, 3.03]	3.12	[3.05, 3.18]	4.34	[4.34, 4.35]
CV	1.40	[1.39, 1.41]	0.99	[0.93, 1.20]	0	-
\mathcal{H}_0	31%	[30%, 31%]	32%	[25%, 34%]	77%	[77%, 77%]
<i>Lisbon and Tagus Valley</i>						
\mathcal{R}_0	2.81	[2.69, 2.86]	2.95	[2.87, 3.01]	4.39	[4.38, 4.40]
CV	0.67	[0.65, 0.70]	0.54	[0.53, 0.57]	0	-
\mathcal{H}_0	51%	[49%, 52%]	49%	[47%, 51%]	77%	[77%, 77%]
<i>Model selection</i>						
LL	-2313.50		-2471.00		-2764.30	
AIC	4647.10		4961.90		5544.60	

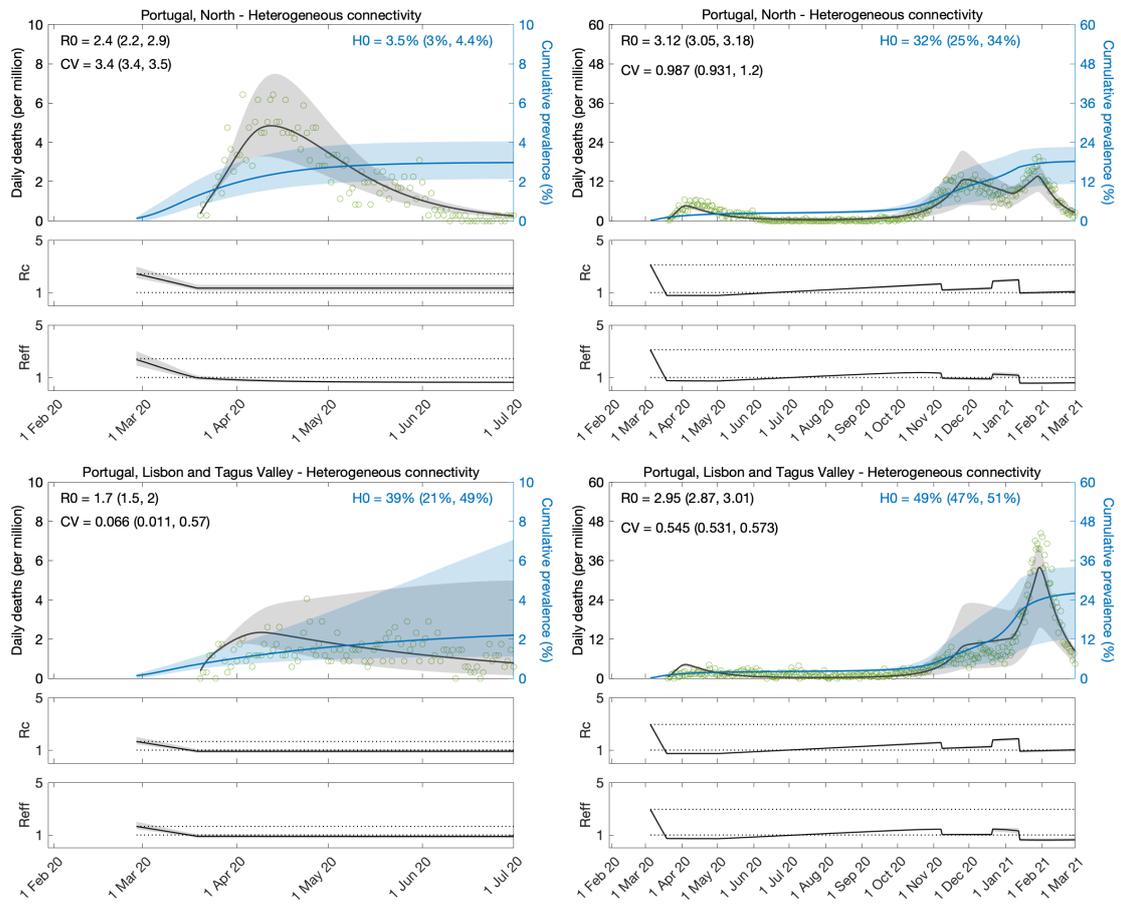
486 * Rounded median resulting in integer number of days.



487

488 **Figure 7: Estimating SARS-CoV-2 transmission in the two larger regions of**
 489 **Portugal.** Heterogeneous susceptibility implemented as a gamma distribution.
 490 Controlled (R_c) and effective (R_{eff}) reproduction numbers are displayed on shallow
 491 panels underneath the main plots. Basic reproduction number, coefficients of
 492 variation and transmissibility profile parameters estimated by Bayesian inference as
 493 described in Methods (estimates in Table 5). Curves represent model reconstructions

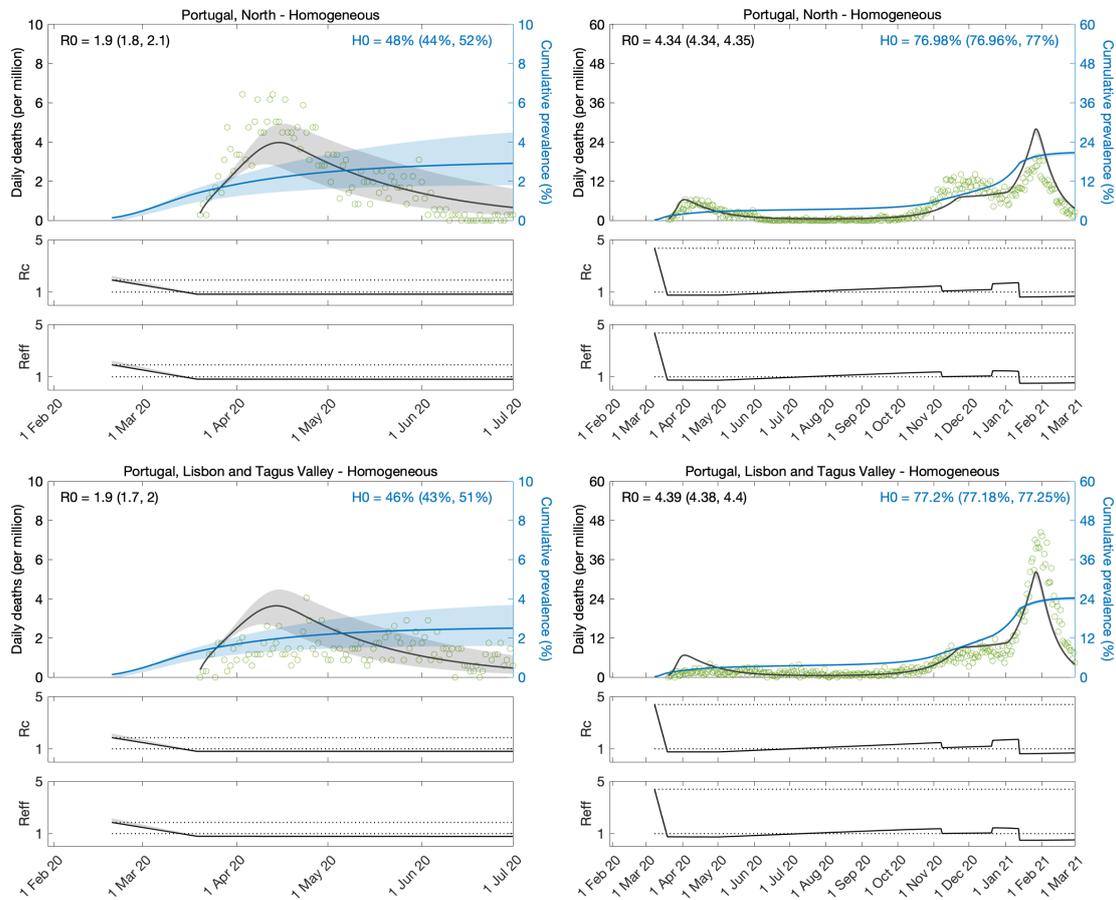
494 from the median posterior parameter estimates. Shades represent 95% credible
 495 intervals from 100,000 posterior samples.



496

497 **Figure 8: Estimating SARS-CoV-2 transmission in the two larger regions of**
 498 **Portugal.** Heterogeneous connectivity implemented as a gamma distribution.
 499 Controlled (R_c) and effective (R_{eff}) reproduction numbers are displayed on shallow
 500 panels underneath the main plots. Basic reproduction number, coefficients of
 501 variation and transmissibility profile parameters estimated by Bayesian inference as
 502 described in Methods (estimates in Table 5). Curves represent model reconstructions

503 from the median posterior parameter estimates. Shades represent 95% credible
 504 intervals from 100,000 posterior samples.



505

506 **Figure 9: Estimating SARS-CoV-2 transmission in the two larger regions of**
 507 **Portugal.** Homogeneous model. Controlled (R_c) and effective (R_{eff}) reproduction
 508 numbers are displayed on shallow panels underneath the main plots. Basic
 509 reproduction number, coefficients of variation and transmissibility profile parameters
 510 estimated by Bayesian inference as described in Methods (estimates in Table 5).
 511 Curves represent model reconstructions from the median posterior parameter
 512 estimates. Shades represent 95% credible intervals from 100,000 posterior samples.

513 **Discussion**

514 We fit SEIR models, with inbuilt distributions of individual susceptibility or exposure
 515 to infection, to daily series of COVID-19 deaths in Spain and Portugal. In similarity
 516 with other studies (Flaxman et al. 2020, Keeling et al. 2020, Bertuzzo et al. 2021,
 517 López & Rodó 2021, Viana et al. 2021), we estimated relevant transmission
 518 parameters, such as the basic reproduction number \mathcal{R}_0 , and a time dependent

519 transmissibility profile $c(t)$ which multiplies \mathcal{R}_0 to account for effects of NPIs,
520 seasonality, viral evolution, or any unexplicit factor that modifies the ability of the
521 virus to infect new hosts. In contrast, our sets of estimated parameters include
522 coefficients of variation that characterise distributions of unmeasured individual
523 susceptibility or connectivity while more common practices are to impose measured
524 variation into the model and ignore the unmeasured.

525 The two approaches have different advantages and disadvantages, and which to adopt
526 depends on study objectives. Using directly measured variation to inform the model
527 allows phenomena to be represented on a micro scale as desired for designing targeted
528 interventions, for instance, but ignoring unmeasured variation may create biases on
529 the macro scale if relevant selection is occurring on omitted characteristics. This may
530 result in the model overpredicting the impact of uniform interventions, for example.

531 The approach of inferring the amount of variation that is under selection by fitting the
532 model on a macro scale is conceived to eliminate those biases, but not suitable for
533 informing targeted actions. In addition, the latter approach appears to require larger
534 infected numbers than the former. In a global crisis such as the current COVID-19
535 pandemic the two approaches can have a role, helping to overcome each other's
536 limitations.

537 The inference of selectable variation presented here is relatively new to infectious
538 disease modelling. Prior to attempting this in the context of the COVID-19 pandemic,
539 we and others have been accelerating progress in systems that were either
540 experimentally controlled (Dwyer et al. 1997; Ben-Ami et al. 2008; Zwart et al. 2011;
541 Pessoa et al. 2014; Langwig et al. 2017; King et al. 2018) or already endemic (Smith
542 et al. 2005; Bellan et al. 2015; Corder et al. 2020). Several aspects of the pandemic

543 made it more challenging. There was an urgency for early results from inherently
544 scarce data, a challenge amplified by non-pharmaceutical interventions designed to
545 suppress the epidemic. In countries where interventions began earlier, such as
546 Portugal, it has been impossible to reach conclusive results without refining data and
547 methods further. In Spain (this study) and England and Scotland (Gomes et al 2021),
548 on the other hand, interventions started later, and results were consistent internally
549 and with each other. Moreover, comparing our coefficients of variation for individual
550 connectivity to those measured directly by contact surveys, we found them to be less
551 than 40% higher. This suggests that there was more selectable variation than the
552 surveys had captured, as expected, but not enough to seem implausible.

553 In conclusion, we believe that the results presented here constitute a major step in the
554 establishment of a novel approach to improve the accuracy and predictability of
555 infectious disease modelling. It holds good promise for use in conjunction with other
556 tools to help devise future pandemic policies.

557 **Acknowledgements**

558 We thank Jan Hasenauer, Rodrigo Corder and Antonio Montalbán for technical
559 advice and contributions to related research. M.U.F. received funding from Conselho
560 Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

561 **Author contributions**

562 M.G.M.G. conceived the study. R.A. and M.G.M.G. performed the analyses. All
563 authors interpreted the data and wrote the paper.

564 **Competing interests**

565 The authors declare no competing interests.

566

567 **References**

- 568 1. Keyfitz, N. & Littman, G. (1979) Mortality in a heterogeneous population. *Popul. Stud.*
569 **33**, 333-342.
- 570 2. Vaupel, J., Manton, K. & Stallard, E. (1979) Impact of heterogeneity in individual frailty
571 on the dynamics of mortality. *Demography* **16**, 439-454.
- 572 3. Vaupel, J., & Yashin, A. (1985) Heterogeneity ruses – some surprising effects of
573 selection on population dynamics. *Am. Stat.* **39**, 176-185.
- 574 4. Kendall, B. E. & Fox, G. A. (2002) Variation among individuals and reduced
575 demographic stochasticity. *Conserv. Biol.* **16**, 109-116.
- 576 5. Jenouvrier, S, Aubry, L. M., Barbraud, C, Weimerskirch, H & Caswell, H. (2018)
577 Interacting effects of unobserved heterogeneity and individual stochasticity in the life
578 history of the southern fulmar. *J. Anim. Ecol.* **87**, 212-222.
- 579 6. Steiner, U. K. & Tuljapurkar, S. (2012) Neutral theory for life histories and individual
580 variability in fitness components. *Proc. Natl. Acad. Sci U. S. A.* **109**, 4684-4689.
- 581 7. Gomes, M. G. M., King, J. G., Nunes, A., Colegrave, N. & Hoffmann, A. (2019) The
582 effects of individual nonheritable variation on fitness estimation and coexistence. *Ecol.*
583 *Evol.* **16**, 8995-9004.
- 584 8. Aalen, O. O., Valberg, M., Grotmol, T. & Tretli, S. (2015) Understanding variation in
585 disease risk: the elusive concept of frailty. *Int. J. Epidemiol.* **4**, 1408-1421.
- 586 9. Stensrud, M. J. & Valberg, M. (2017) Inequality in genetic cancer risk suggests bad
587 genes rather than bad luck. *Nat. Commun.* **8**, 1165.
- 588 10. Anderson, R. M., Medley, G. F., May, R. M. & Johnson, A. M. (1986) A preliminary
589 study of the transmission dynamics of the human immunodeficiency virus (HIV), the
590 causative agent of AIDS. *IMA J. Math. Appl. Med. Biol.* **3**, 229-263.
- 591 11. Dwyer, G., Elkinton, J. S. & Buonaccorsi, J. P. (1997) Host heterogeneity in
592 susceptibility and disease dynamics: Tests of a mathematical model. *Am. Nat.* **150**, 685-
593 707.

- 594 12. Smith, D. L., Dushoff, J., Snow, R. W. & Hay, S. I. (2005) The entomological
595 inoculation rate and *Plasmodium falciparum* infection in African children. *Nature* **438**,
596 492-495.
- 597 13. Bellan, S. E., Dushoff, J., Galvani, A. P. & Meyers, L. A. (2015) Reassessment of HIV-1
598 acute phase infectivity: accounting for heterogeneity and study design with simulated
599 cohorts. *PLOS Med.* **12**, e1001801.
- 600 14. Gomes, M. G. M., et al. (2019) Introducing risk inequality metrics in tuberculosis policy
601 development. *Nat. Commun.* **10**, 2480.
- 602 15. Corder, R. M., Ferreira, M. U. & Gomes, M. G. M. (2020) Modelling the epidemiology
603 of residual *Plasmodium vivax* in a heterogeneous host population: a case study in the
604 Amazon Basin. *PLOS Comput. Biol.* **16**, e1007377.
- 605 16. Halloran, M. E., Longini, I. M. Jr. & Struchiner, C. J. (1996) Estimability and
606 interpretability of vaccine efficacy using frailty mixing models. *Am. J. Epidemiol.* **144**,
607 83-97.
- 608 17. O'Hagan, J. J., Hernán, M. A., Walensky, R. P. & Lipsitch, M. (2012) Apparent
609 declining efficacy in randomized trials: Examples of the Thai RV144 HIV vaccine and
610 CAPRISA 004 microbicide trials. *AIDS* **26**, 123.
- 611 18. Gomes, M. G. M., et al. (2014) A missing dimension in measures of vaccination
612 impacts. *PLOS Pathog.* **10**, e1003849.
- 613 19. Gomes, M. G. M., Gordon, S. B. & Lalloo, D. G. (2016) Clinical trials: the mathematics
614 of falling vaccine efficacy with rising disease incidence. *Vaccine* **34**, 3007.
- 615 20. Langwig, K. E., et al. (2017) Vaccine effects on heterogeneity in susceptibility and
616 implications for population health management, *mBio* **8**, e00796-17.
- 617 21. Pessoa, D., et al. (2016) Unveiling time in dose-response models to infer host
618 susceptibility to pathogens. *PLOS Comput. Biol.* **10**, e1003773.
- 619 22. King, J. G., Souto-Maior, C., Sartori, L. M., Maciel-de-Freitas, R. & Gomes, M. G. M.
620 (2018) Variation in *Wolbachia* effects on *Aedes* mosquitoes as a determinant of
621 invasiveness and vectorial capacity. *Nat. Commun.* **9**, 1-8.

- 622 23. Gomes, M. G. M., et al. (2020) Individual variation in susceptibility or exposure to
623 SARS-CoV-2 lowers the herd immunity threshold. *medRxiv*
624 10.1101/2020.04.27.20081893.
- 625 24. World Health Organization (2021) COVID-19 weekly epidemiological update:
626 [https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---](https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---13-july-2021)
627 [13-july-2021](https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---13-july-2021).
- 628 25. Gomes, M. G. M., Ferreira, M. U., Chikina, M., Pegden, W. & Aguas, R. (2021) Frailty
629 variation models for susceptibility and exposure to SARS-CoV-2. *medRxiv*
630 10.1101/2021.05.25.21257766.
- 631 26. Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. (1990) On the definition and
632 computation of the basic reproduction ratio R_0 in models for infectious diseases in
633 heterogeneous populations. *J. Math. Biol.* **28**, 365-382.
- 634 27. McAloon, C., et al. (2020) Incubation period of COVID-19: a rapid systematic review
635 and meta-analysis of observational research. *BMJ Open* **10**, e039652.
- 636 28. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. (2020) Serial interval of novel
637 coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284-6.
- 638 29. Lauer, S. A., et al. (2020) The Incubation Period of Coronavirus Disease 2019 (COVID-
639 19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern.*
640 *Med.* **172**, 577-582.
- 641 30. Li, Q., et al. (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-
642 infected pneumonia. *N. Engl. J. Med.* **382**, 1199-1207.
- 643 31. Wei, W. E., et al. (2020) Presymptomatic Transmission of SARS-CoV-2 — Singapore,
644 January 23–March 16, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 411-415.
- 645 32. To, K. K. W., et al. (2020) Temporal profiles of viral load in posterior oropharyngeal
646 saliva samples and serum antibody responses during infection by SARS-CoV-2: an
647 observational cohort study. *Lancet Infect. Dis.* **20**, 565-74.
- 648 33. Arons, M. M., et al. (2020) Presymptomatic SARS-CoV-2 Infections and Transmission
649 in a Skilled Nursing Facility. *N. Engl. J. Med.* **382**, 2081-2090.

- 650 34. He, X., et al. (2020) Temporal dynamics in viral shedding and transmissibility of
651 COVID-19. *Nat. Med.* **26**, 672-675.
- 652 35. Ward, H., et al. (2021) SARS-CoV-2 antibody prevalence in England following the first
653 peak of the pandemic. *Nat. Commun.* **12**, 905.
- 654 36. Novozhilov, A. S. (2008) On the spread of epidemics in a closed heterogeneous
655 population. *Math. Biosci.* **215**, 177-185.
- 656 37. Montalbán, A., Corder, R. M. & Gomes, M. G. M. (2020) Herd immunity under
657 individual variation and reinfection. *arXiv* 200800098v2.
- 658 38. Pastor-Satorras, R. & Vespignani, A. (2001) Epidemic dynamics and endemic states in
659 complex networks. *Phys. Rev. E* **63**, 066117.
- 660 39. Miller, J. C., Slim, A. C. & Volz, E. M. (2012) Edge-based compartmental modelling for
661 infectious disease spread. *J. R. Soc. Interface* **9**, 890-906.
- 662 40. Britton, T., Ball, F. & Trapman, P. (2020) A mathematical model reveals the influence
663 of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846-849.
- 664 41. Hartnett, K. (2020) The tricky math of herd immunity for COVID-19. *Quanta Magazine*.
665 [https://www.quantamagazine.org/the-tricky-math-of-covid-19-herd-immunity-](https://www.quantamagazine.org/the-tricky-math-of-covid-19-herd-immunity-20200630/)
666 [20200630/](https://www.quantamagazine.org/the-tricky-math-of-covid-19-herd-immunity-20200630/)
- 667 42. Aschwanden, C. (2021) Five reasons why COVID herd immunity is probably
668 impossible. *Nature* **591**, 520-522. <https://www.nature.com/articles/d41586-021-00728-2>
- 669 43. Hall, et al. (2021) SARS-CoV-2 infection rates of antibody-positive compared with
670 antibody-negative health-care workers in England: a large, multicentre, prospective
671 cohort study (SIREN). *Lancet* **397**, 1459-1469.
- 672 44. Flaxman, S., et al. (2020) Estimating the effects of non-pharmaceutical interventions on
673 COVID-19 in Europe. *Nature* **584**, 257-261.
- 674 45. Keeling, M. J., et al. (2020) Fitting to the UK COVID-19 outbreak, short-term forecasts
675 and estimating the reproductive number. *medRxiv* 10.1101/2020.08.04.20163782.
- 676 46. Viana, J. et al. (2021) Controlling the pandemic during the SARS-CoV-2 vaccination
677 rollout. *Nat. Commun.* **12**, 3674.

- 678 47. Wood, S. N. (2021) Inferring UK COVID-19 fatal infection trajectories from daily
679 mortality data: Were infections already in decline before the UK lockdowns? *Biometrics*
680 10.1111/biom.13462
- 681 48. Bertuzzo, E. et al. (2020) The geography of COVID-19 spread in Italy and implications
682 for the relaxation of confinement measures. *Nat. Commun.* **11**, 4264.
- 683 49. López, L. & Rodó, X. (2021) A modified SEIR model to predict the COVID-19
684 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics.
685 *Results Phys.* **21**, 103746.