

## Supplementary Materials for

### **Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold**

**Authors:** M. Gabriela M. Gomes<sup>\*</sup>, Rodrigo M. Corder, Jessica G. King, Kate E. Langwig, Caetano Souto-Maior, Jorge Carneiro, Guilherme A. Gonçalves, Carlos Penha-Gonçalves, Marcelo U. Ferreira, Ricardo Aguas.

Correspondence to: [gabriela.gomes@strath.ac.uk](mailto:gabriela.gomes@strath.ac.uk).

#### **This PDF file includes:**

Materials and Methods  
Figs. S1 to S22  
Tables S1 to S4

## Materials and Methods

### Model structure and underlying assumptions

The model presented here is a differential equation SEIR model, where susceptible individuals become exposed at a rate dependent on their susceptibility, the number of potentially infectious contacts they engage in, and the total number of infectious people in the population, per time unit. Upon exposure, individuals enter an asymptomatic incubation phase, during which they slowly become infectious (24-27). Thus, infectivity of exposed individuals is made to be  $1/2$  of that of infectious ones ( $\rho = 0.5$ ). After a few days, individuals develop symptoms – on average, 4 days after the exposure to the virus ( $\delta = 1/4$ ) – and become fully infectious (28-30). They recover, i.e., they lose their infectiousness 4 days after that ( $\gamma = 1/4$ ), on average (31). We have assumed that no reinfection can occur after recovery, which means absolute long-lasting immunity.

We implemented non-pharmaceutical (NP) interventions as a gradual decrease in viral transmissibility in the population and thus a lowering of the effective reproduction number ( $R_{eff}$ ). Once containment measures are put in place in each country, we postulate it takes 14 days until the maximum effectiveness of social distancing measures is reached. In the simulations presented throughout we have held this condition (maximum “lockdown” efficacy) for 30 days, after which period, social distancing measures are progressively relaxed, slowly returning to its original value (normality) after 1-year. Both the implementation and relaxing of the social distancing measures are imposed to be linear in this model.

### Bayesian Inference

The model laid out above is amenable to theoretical exploration as presented in the main manuscript and provides a perfect framework for inference. Fundamentally, to be able to reproduce the inception of any epidemic, we would need to estimate when local transmission started to occur ( $t_0$ ), and the pace at which individuals infected each other in the very early stages of the epidemic ( $R_0$ ). All countries, to different extents and at different timepoints of the epidemic, enforced some combination of social distancing measures. To fully understand the interplay between herd immunity and the impact of NP interventions, we then set out to estimate the time at which social distancing measures started to have an impact on daily incidence ( $t_0^d$ ), and what is their maximum effectiveness ( $d$ ), for each country.

In order to preserve identifiability, we made two simplifying assumptions: (i) the fraction of infectious individuals reported as COVID-19 cases (reporting fraction) is 10% (32); (ii) local transmission starts ( $t_0$ ) when countries report 1 case per 30 million population in one day (1 infection per 3 million population in one day) which in Italy, for example, happened on the 31 January 2020.

Parameter estimation was performed with the software Matlab, using PESTO (Parameter Estimation Toolbox) (33). We assume that the residuals between the model output and the daily reported cases are normally distributed with unknown standard deviations  $\sigma$ . We first fixed the beginning of the local transmission (parameter  $t_0$ ) for values from 0 to 30 days with respect to the first notified case in Italy. Next, we optimized the model for the other parameters  $\{R_0, t_0^d, d\}$  by maximizing the logarithm of the likelihood ( $L$ ) (Equation 1) of observing the daily reported number of cases in each country  $D = \{(k, \tilde{y}_k)\}_{k=0}^n$ , that is,

$$L(D, \theta) = \prod_{k=0}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\tilde{y}_k - y(k))^2}{2\sigma^2}\right) \quad (1)$$

in which  $y(k, \theta)$  is the simulated model output number of COVID-19 cases at day  $k$  (with respect to the first notified case in Italy),  $\theta = \{R_0, t_0^d, d\}$  is the set of parameters and  $n$  is the total number of days included in the analysis. We considered cases reported until April 30<sup>th</sup> ( $n = 91$ ). To ensure that the observed maximum is global, we performed 10 multi-starts optimization for each  $t_0$  and initialised  $\theta$  by randomly sampling values following a Latin hypercube. For each country, the best model was selected based on the set of parameters  $\{t_0, R_0, t_0^d, d\}$  that maximized the log-likelihood  $L$  (Equation 1). We, therefore, fixed  $t_0$  to that obtained in the optimization process and performed Monte-Carlo Markov Chain simulations of  $10^4$  iterations to obtain the median estimates and respective 95% credible intervals for parameters  $\{R_0, t_0^d, d\}$ . We used uniformly distributed priors with ranges  $\{1-5, 1-100, 0-1\}$ .

The procedure was applied to 11 countries (Australia, Austria, Brazil, Canada, Germany, Italy, Netherlands, Portugal, Spain, Turkey and United Kingdom). Parameter estimates displayed in Tables S1 (basic reproduction number,  $R_0$ ), S2 (maximum interventions impact,  $d$ ), S3 (time at which social distancing measures started to have an impact on daily incidence,  $t_0^d$ ) and S4 (time of the beginning of local transmission,  $t_0$ ), and Figures S1-S11, are comparable to those obtained in other studies (34-40). Fits of the heterogeneous models to case reporting data for the respective countries are shown in the main text for Italy and Austria (Figure 1 for heterogeneous susceptibility and Figure 2 for heterogeneous connectivity) and Figures S12-S20 for the remaining 9 countries. The dynamics of the homogeneous model are illustrated by a single figure for Italy and Austria (Figure S21).

Figure 22 is provided to assess robustness of our main result considering declining herd immunity thresholds with increasing individual variation in susceptibility or exposure (Figure 3) to change the distribution from gamma to lognormal.

Table S1. Estimated 95% credible intervals around the mean estimate for  $R_0$  using different models.

	Homogeneous			Heterogeneous susceptibility			Heterogeneous contacts		
	CV=1								
<b>Italy</b>	3.010	3.043	3.074	3.014	3.046	3.076	3.154	3.190	3.225
<b>Portugal</b>	2.664	2.706	2.748	2.666	2.706	2.747	2.687	2.731	2.775
<b>Spain</b>	2.823	2.853	2.882	2.785	2.812	2.839	2.948	2.978	3.013
<b>Austria</b>	3.028	3.069	3.112	3.029	3.069	3.110	3.016	3.058	3.098
<b>Brazil</b>	2.587	2.841	3.545	2.584	2.741	3.078	2.565	2.744	3.528
<b>UK</b>	2.424	2.457	2.491	2.439	2.471	2.505	2.524	2.560	2.596
<b>Netherlands</b>	2.617	2.659	2.701	2.623	2.662	2.703	2.739	2.782	2.823
<b>Turkey</b>	2.284	2.298	2.312	2.285	2.298	2.312	2.371	2.388	2.404
<b>Canada</b>	2.545	2.570	2.597	2.547	2.570	2.596	2.533	2.556	2.580
<b>Germany</b>	2.662	2.692	2.722	2.664	2.694	2.725	2.781	2.816	2.847
<b>Australia</b>	2.979	3.017	3.058	2.979	3.019	3.058	2.967	3.005	3.044
	CV=3								
<b>Italy</b>	3.010	3.042	3.074	3.036	3.068	3.102	3.502	3.549	3.598
<b>Portugal</b>	2.664	2.705	2.748	2.801	2.851	2.908	2.935	2.987	3.048
<b>Spain</b>	2.824	2.853	2.883	3.013	3.053	3.097	3.230	3.278	3.335
<b>Austria</b>	3.028	3.069	3.110	2.984	3.022	3.065	3.154	3.194	3.233
<b>Brazil</b>	2.624	2.801	3.296	2.510	3.274	4.626	2.477	2.630	3.134
<b>UK</b>	2.426	2.456	2.489	2.440	2.472	2.506	2.770	2.811	2.857
<b>Netherlands</b>	2.620	2.660	2.702	2.635	2.675	2.716	3.017	3.062	3.106
<b>Turkey</b>	2.283	2.298	2.311	2.289	2.303	2.318	2.570	2.588	2.607
<b>Canada</b>	2.545	2.571	2.598	2.585	2.611	2.641	2.735	2.763	2.794
<b>Germany</b>	2.663	2.692	2.723	2.671	2.700	2.735	3.045	3.089	3.129
<b>Australia</b>	2.979	3.019	3.059	2.978	3.020	3.061	3.023	3.055	3.088

Table S2. Estimated 95% credible intervals around the mean estimate for  $d$  (maximum intervention impact) using different models.

	Homogeneous			Heterogeneous susceptibility			Heterogeneous contacts		
	CV=1								
<b>Italy</b>	0.282	0.288	0.294	0.289	0.296	0.302	0.282	0.288	0.293
<b>Portugal</b>	0.291	0.310	0.331	0.297	0.317	0.335	0.299	0.319	0.339
<b>Spain</b>	0.288	0.299	0.310	0.304	0.316	0.326	0.298	0.308	0.319
<b>Austria</b>	0.128	0.160	0.188	0.129	0.162	0.189	0.131	0.165	0.192
<b>Brazil</b>	0.424	0.528	0.589	0.489	0.547	0.590	0.426	0.549	0.599
<b>UK</b>	0.392	0.408	0.423	0.397	0.414	0.430	0.392	0.408	0.422
<b>Netherlands</b>	0.331	0.345	0.356	0.338	0.350	0.362	0.329	0.341	0.352
<b>Turkey</b>	0.341	0.360	0.377	0.345	0.365	0.381	0.338	0.357	0.376
<b>Canada</b>	0.414	0.423	0.431	0.419	0.427	0.435	0.424	0.433	0.441
<b>Germany</b>	0.251	0.268	0.283	0.255	0.272	0.288	0.248	0.266	0.281
<b>Australia</b>	0.057	0.099	0.133	0.055	0.101	0.135	0.055	0.099	0.136
	CV=3								
<b>Italy</b>	0.282	0.288	0.294	0.345	0.351	0.358	0.327	0.333	0.340
<b>Portugal</b>	0.290	0.311	0.332	0.326	0.347	0.368	0.335	0.356	0.376
<b>Spain</b>	0.289	0.300	0.310	0.372	0.388	0.401	0.398	0.415	0.431
<b>Austria</b>	0.130	0.160	0.187	0.140	0.179	0.213	0.147	0.182	0.214
<b>Brazil</b>	0.462	0.532	0.585	0.346	0.472	0.619	0.496	0.584	0.634
<b>UK</b>	0.392	0.408	0.424	0.462	0.479	0.493	0.443	0.456	0.468
<b>Netherlands</b>	0.332	0.344	0.356	0.384	0.396	0.409	0.360	0.371	0.381
<b>Turkey</b>	0.340	0.360	0.377	0.375	0.396	0.418	0.359	0.377	0.393
<b>Canada</b>	0.414	0.422	0.431	0.443	0.452	0.461	0.435	0.443	0.451
<b>Germany</b>	0.250	0.267	0.283	0.284	0.304	0.321	0.272	0.289	0.303
<b>Australia</b>	0.054	0.097	0.134	0.057	0.101	0.137	0.058	0.101	0.138

Table S3. Estimated 95% credible intervals around the mean estimate for time at which social distancing measures started to have an impact on daily incidence ( $t_0^d$ ), using different models.

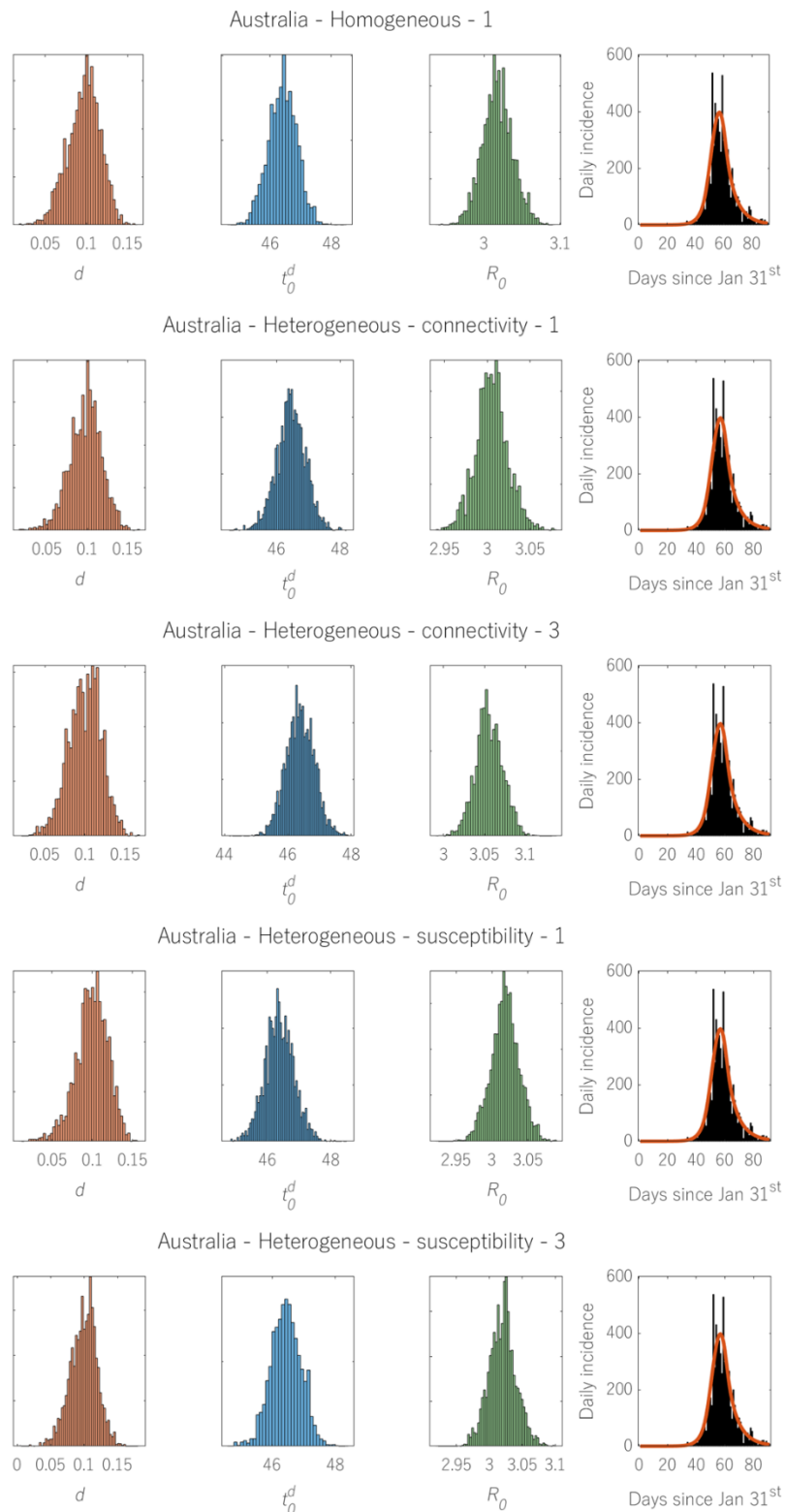
	Homogeneous			Heterogeneous susceptibility			Heterogeneous contacts		
	CV=1								
<b>Italy</b>	38.054	38.854	39.649	37.970	38.774	39.537	37.655	38.423	39.222
<b>Portugal</b>	47.785	49.404	50.968	47.900	49.421	50.915	47.683	49.275	50.827
<b>Spain</b>	44.860	45.840	46.748	44.994	45.904	46.798	44.461	45.459	46.359
<b>Austria</b>	43.784	44.838	45.838	43.874	44.901	45.892	43.873	44.875	45.928
<b>Brazil</b>	31.974	43.945	51.888	38.609	46.791	52.339	29.257	45.175	52.513
<b>UK</b>	52.964	54.809	56.603	53.025	54.730	56.496	52.447	54.234	56.008
<b>Netherlands</b>	45.520	46.974	48.543	45.466	46.908	48.420	45.095	46.424	47.952
<b>Turkey</b>	60.215	61.072	62.058	60.280	61.082	62.044	59.665	60.673	61.754
<b>Canada</b>	49.969	51.098	52.065	50.050	51.080	51.993	49.897	51.005	52.046
<b>Germany</b>	46.074	47.198	48.226	46.017	47.173	48.216	45.667	46.685	47.833
<b>Australia</b>	45.483	46.415	47.324	45.496	46.375	47.305	45.508	46.453	47.335
	CV=3								
<b>Italy</b>	38.098	38.857	39.646	37.277	38.102	38.952	35.812	36.775	37.645
<b>Portugal</b>	47.748	49.406	50.928	47.119	48.855	50.483	46.347	48.328	50.110
<b>Spain</b>	44.901	45.844	46.785	43.647	44.888	46.155	42.209	43.828	45.166
<b>Austria</b>	43.857	44.841	45.818	44.182	45.320	46.324	43.906	44.999	46.027
<b>Brazil</b>	36.067	46.545	51.863	20.207	32.229	51.566	28.565	44.696	52.614
<b>UK</b>	53.134	54.836	56.530	52.118	54.104	56.005	49.876	51.899	53.778
<b>Netherlands</b>	45.459	46.988	48.464	44.947	46.422	47.988	43.744	45.000	46.336
<b>Turkey</b>	60.250	61.112	62.117	60.184	61.311	62.332	58.878	59.894	60.895
<b>Canada</b>	49.917	51.030	52.097	49.331	50.489	51.549	48.578	49.775	50.925
<b>Germany</b>	46.049	47.179	48.238	46.014	47.258	48.406	44.699	45.877	47.163
<b>Australia</b>	45.457	46.386	47.304	45.556	46.437	47.330	45.525	46.383	47.273



Table S4. Estimated time for the beginning of local transmission ( $t_0$ ) using different models.

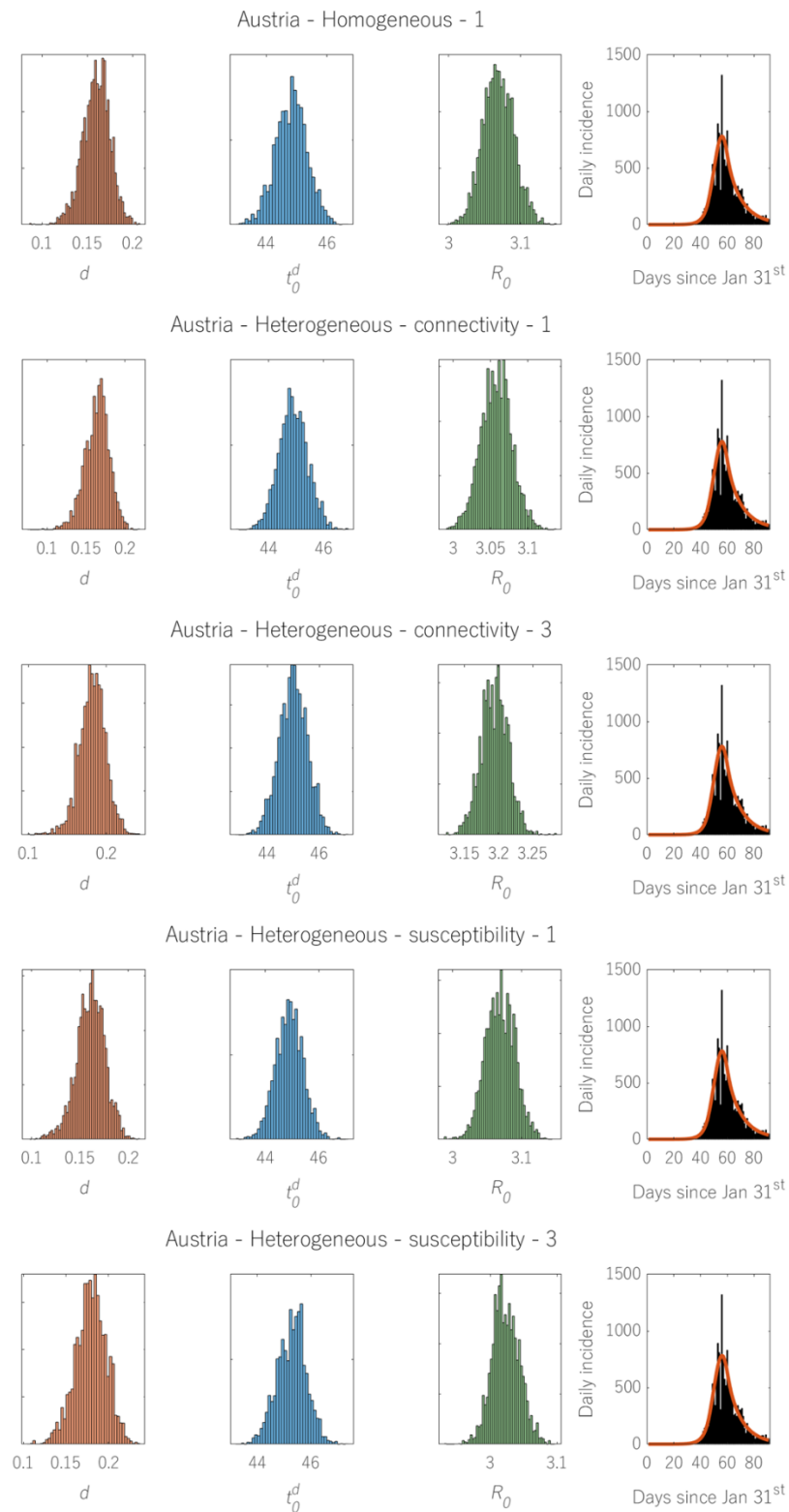
	Homogeneous	Heterogeneous susceptibility	Heterogeneous contacts
CV=1			
Italy	1	1	1
Portugal	0	0	1
Spain	1	1	1
Austria	7	7	4
Brazil	19	19	15
UK	1	1	1
Netherlands	1	1	1
Turkey	1	1	1
Canada	6	6	2
Germany	1	1	1
Australia	14	14	11
CV=3			
Italy	1	1	1
Portugal	0	7	1
Spain	1	5	1
Austria	7	6	1
Brazil	20	17	5
UK	1	1	1
Netherlands	1	1	1
Turkey	1	1	1
Canada	6	7	1
Germany	1	1	1
Australia	14	14	6

Figure S1



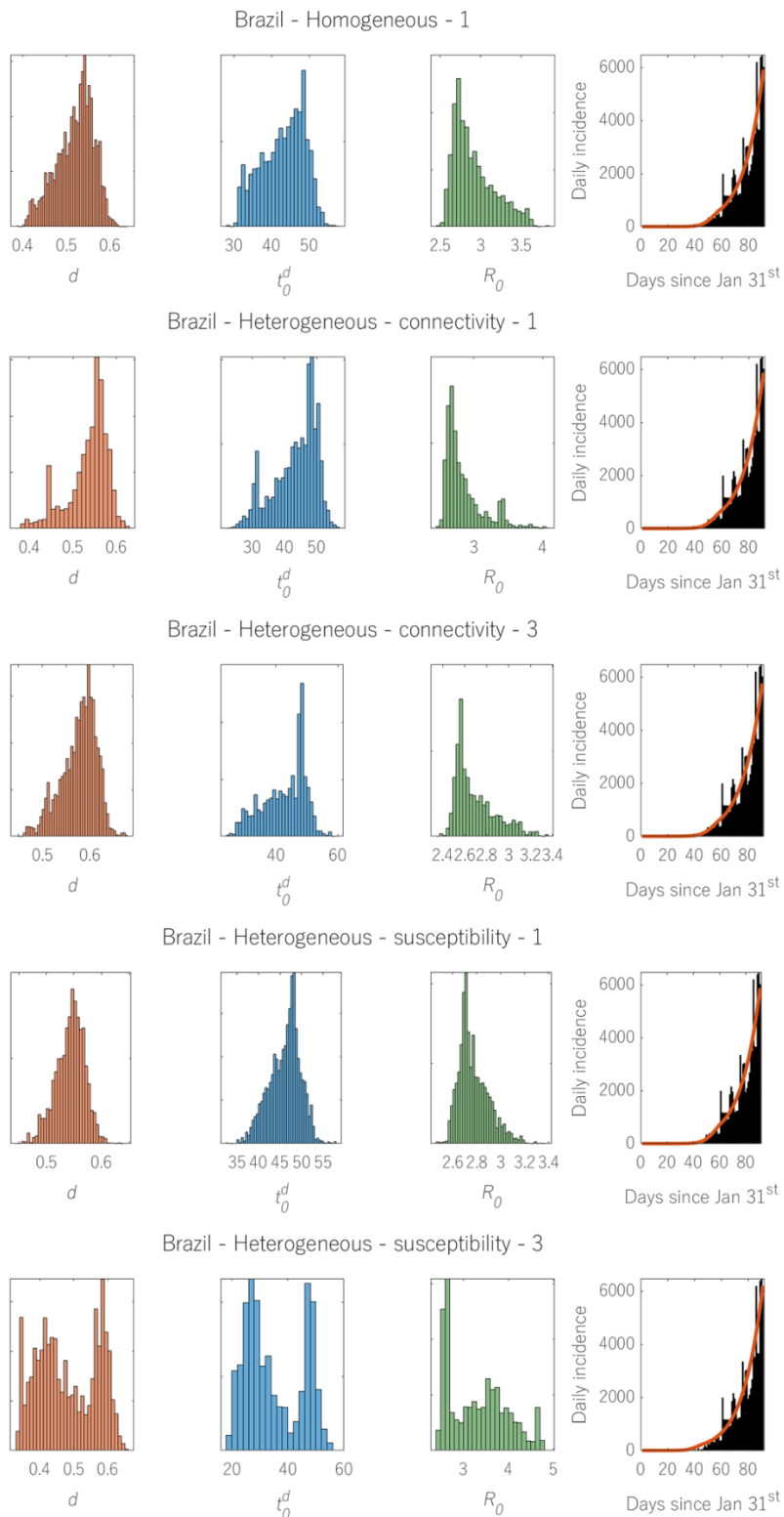
Parameter posterior distributions and corresponding model fit (right most plot) for Australia, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S2



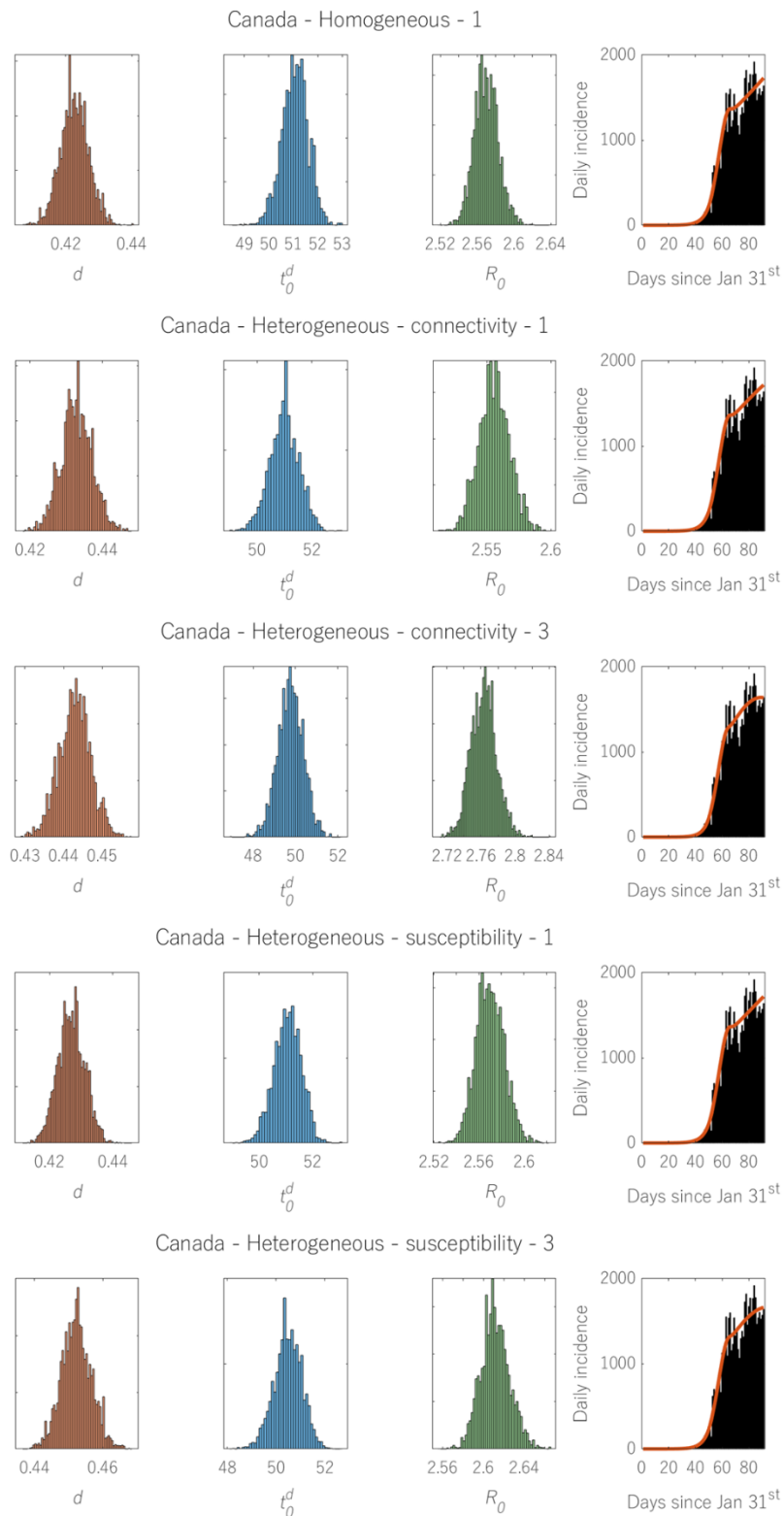
Parameter posterior distributions and corresponding model fit (right most plot) for Austria, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S3



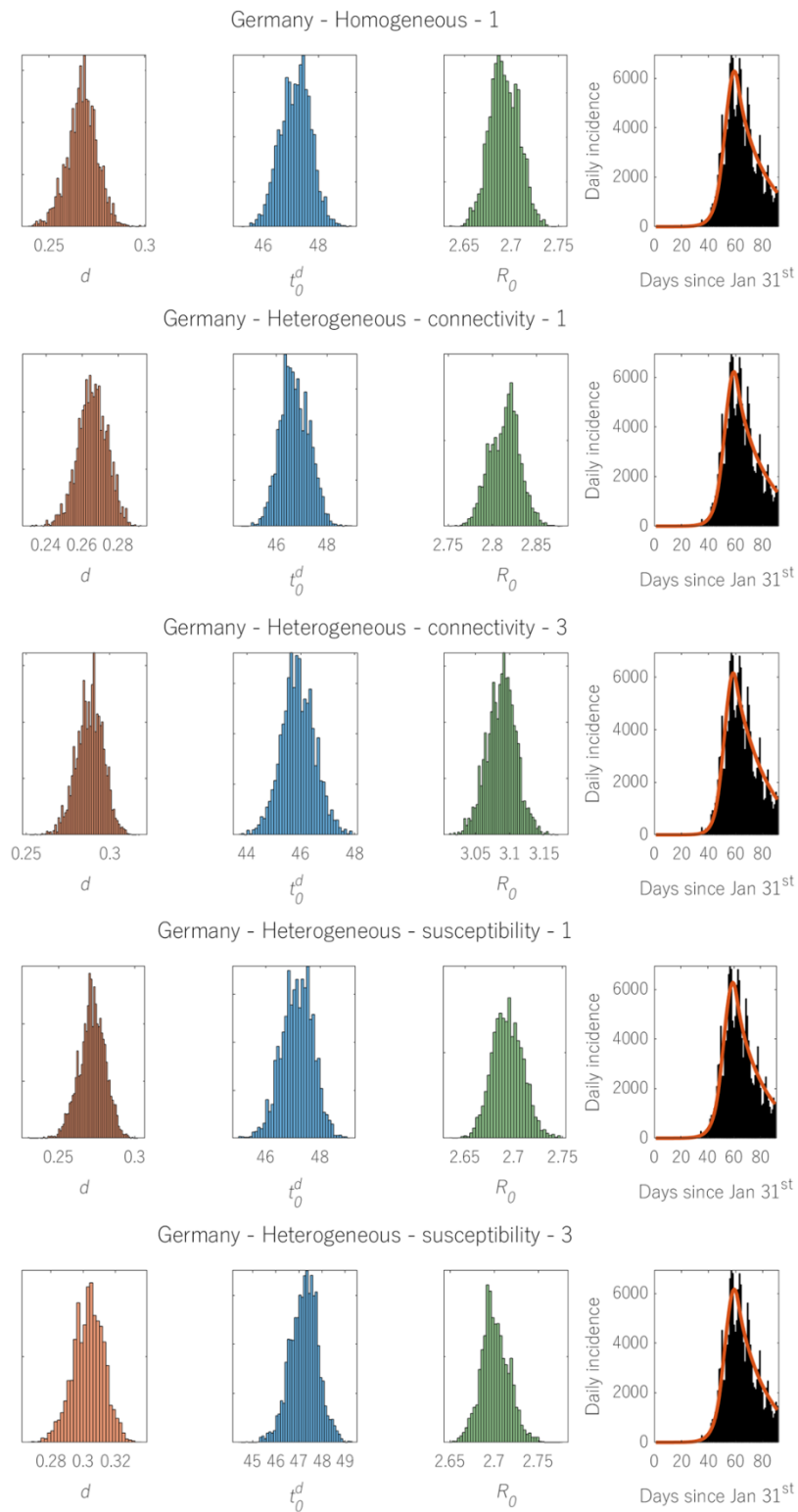
Parameter posterior distributions and corresponding model fit (right most plot) for Brazil, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S4



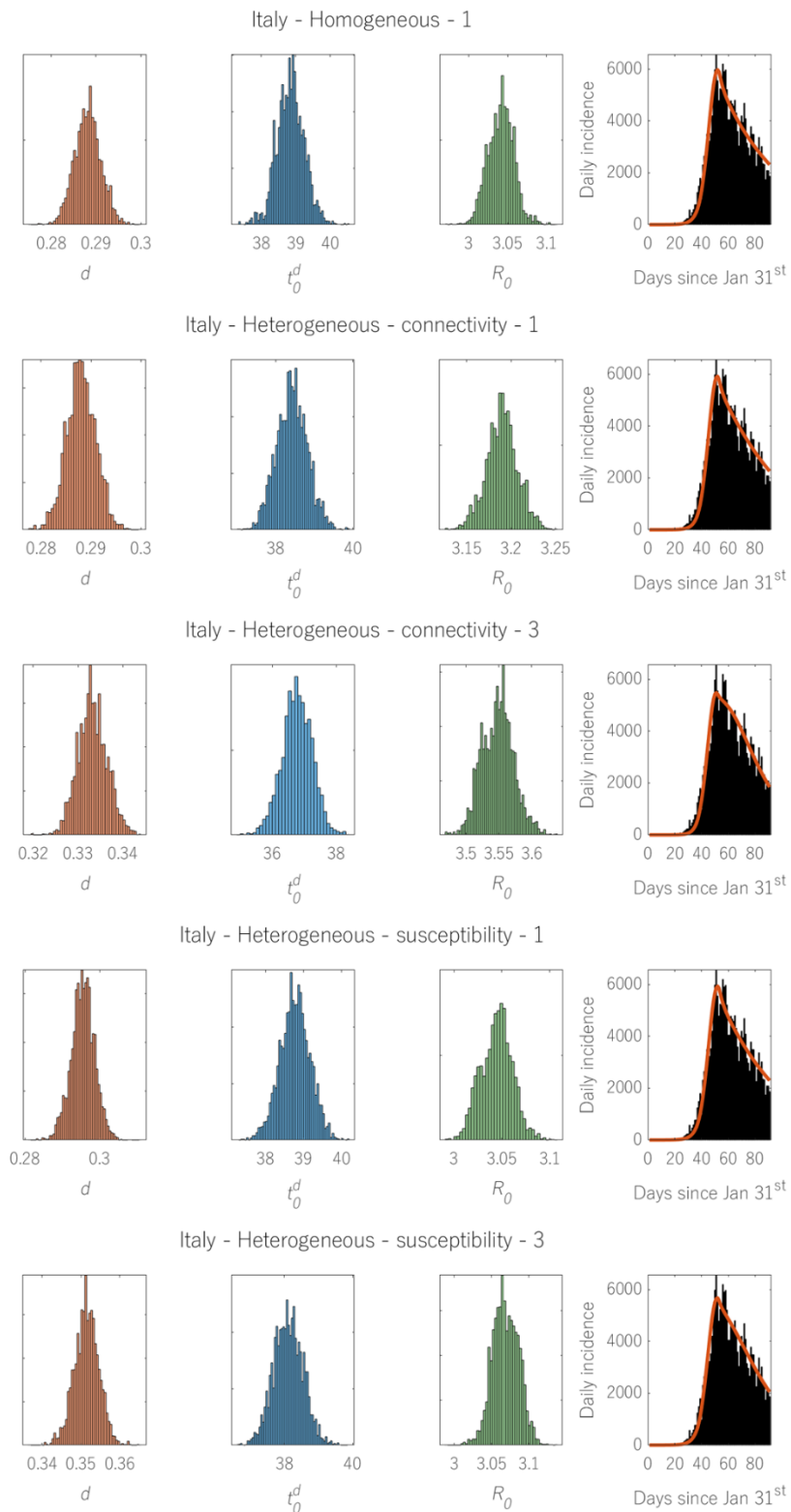
Parameter posterior distributions and corresponding model fit (right most plot) for Canada, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S5



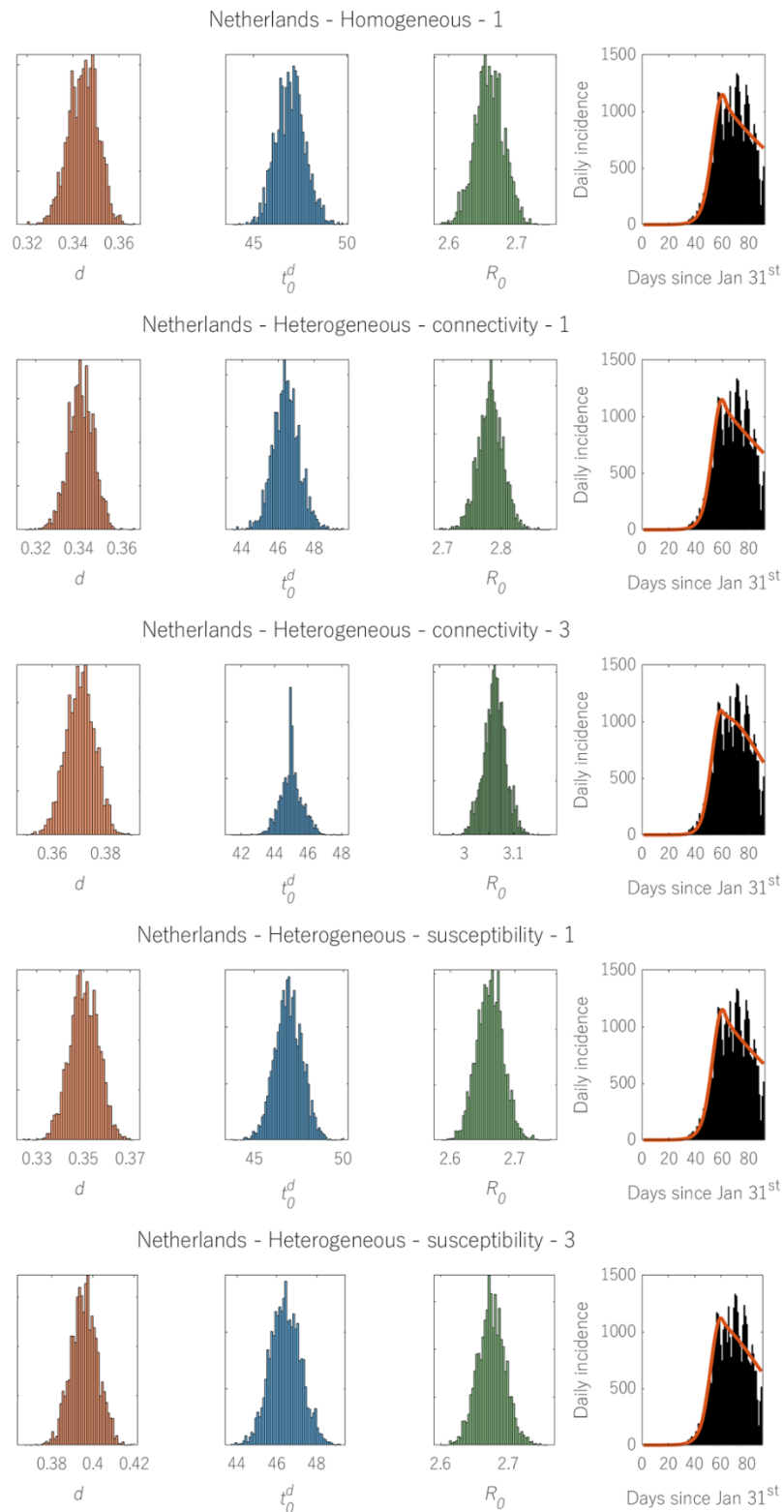
Parameter posterior distributions and corresponding model fit (right most plot) for Germany, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S6



Parameter posterior distributions and corresponding model fit (right most plot) for Italy, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

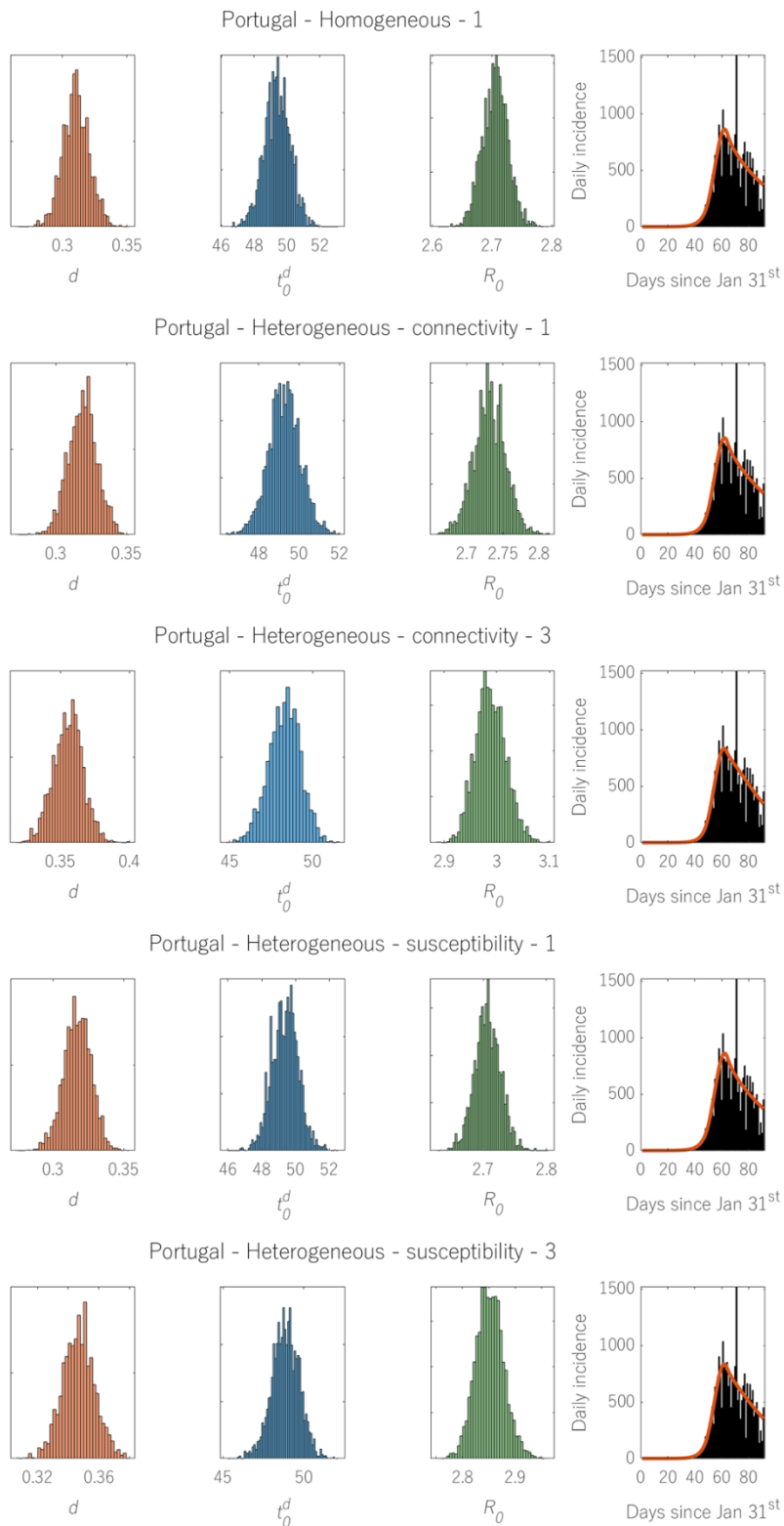
Figure S7



Parameter posterior distributions and corresponding model fit (right most plot) for the Netherlands, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

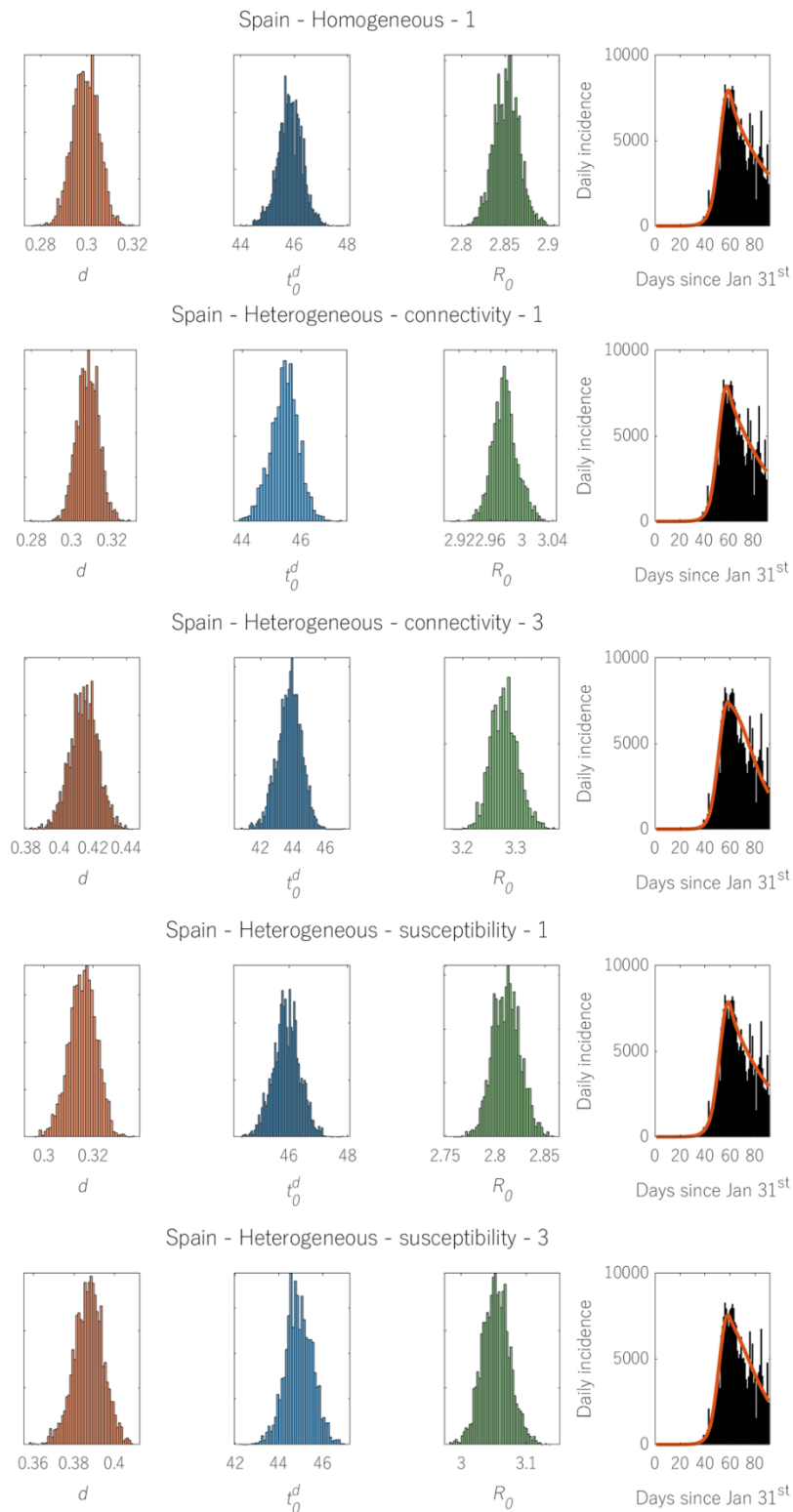


Figure S8



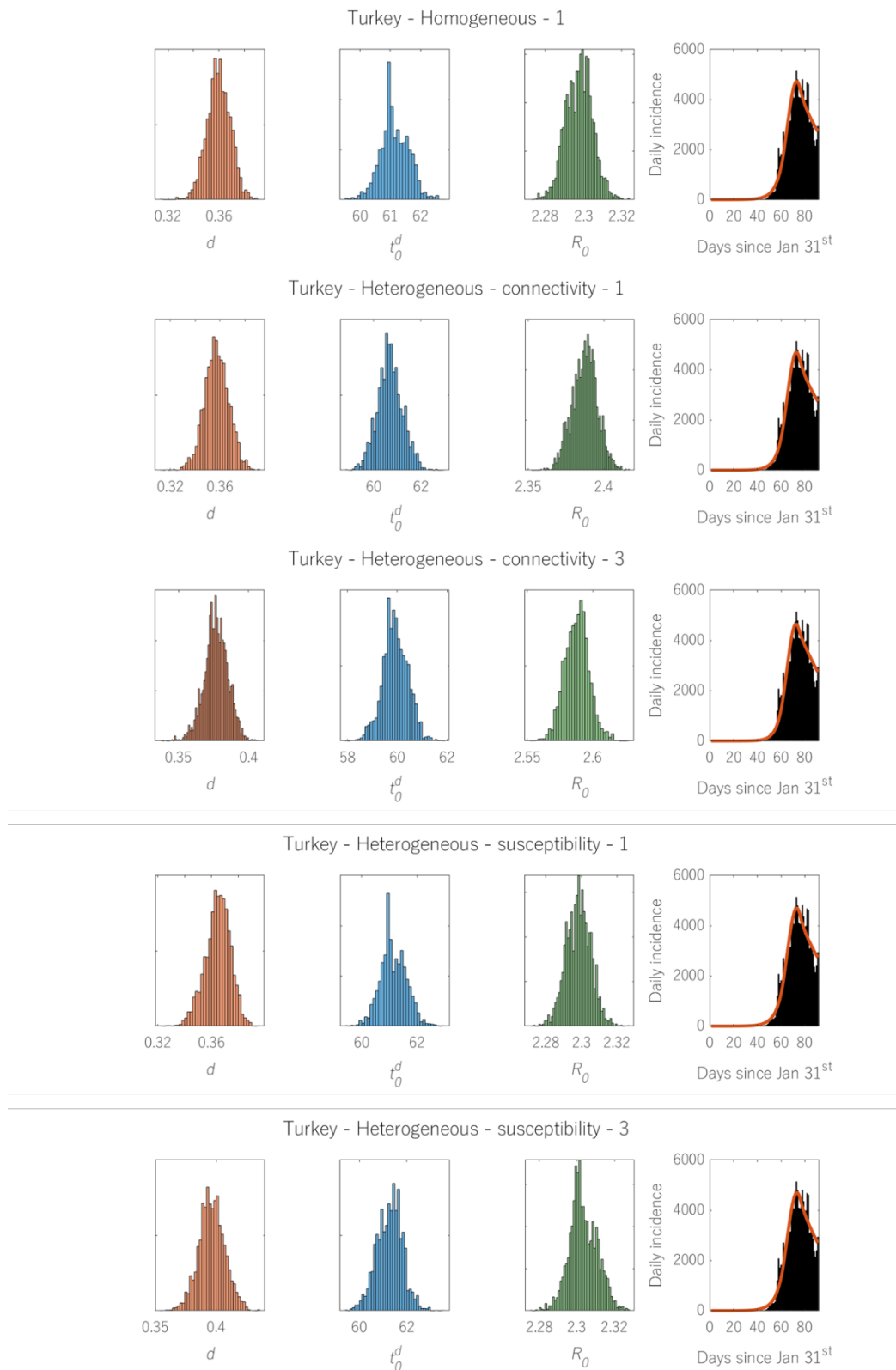
Parameter posterior distributions and corresponding model fit (right most plot) for Portugal, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S9



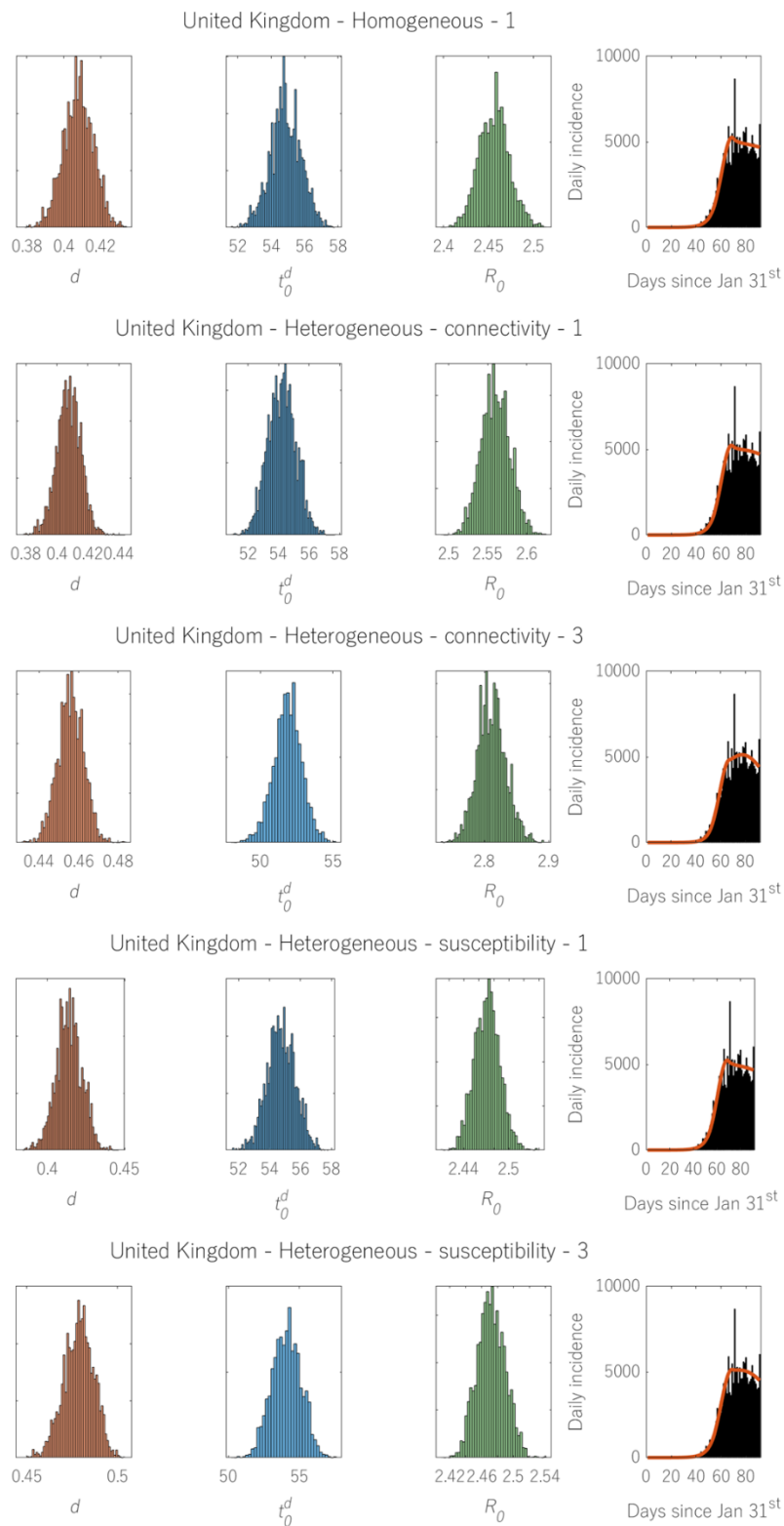
Parameter posterior distributions and corresponding model fit (right most plot) for Spain, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S10



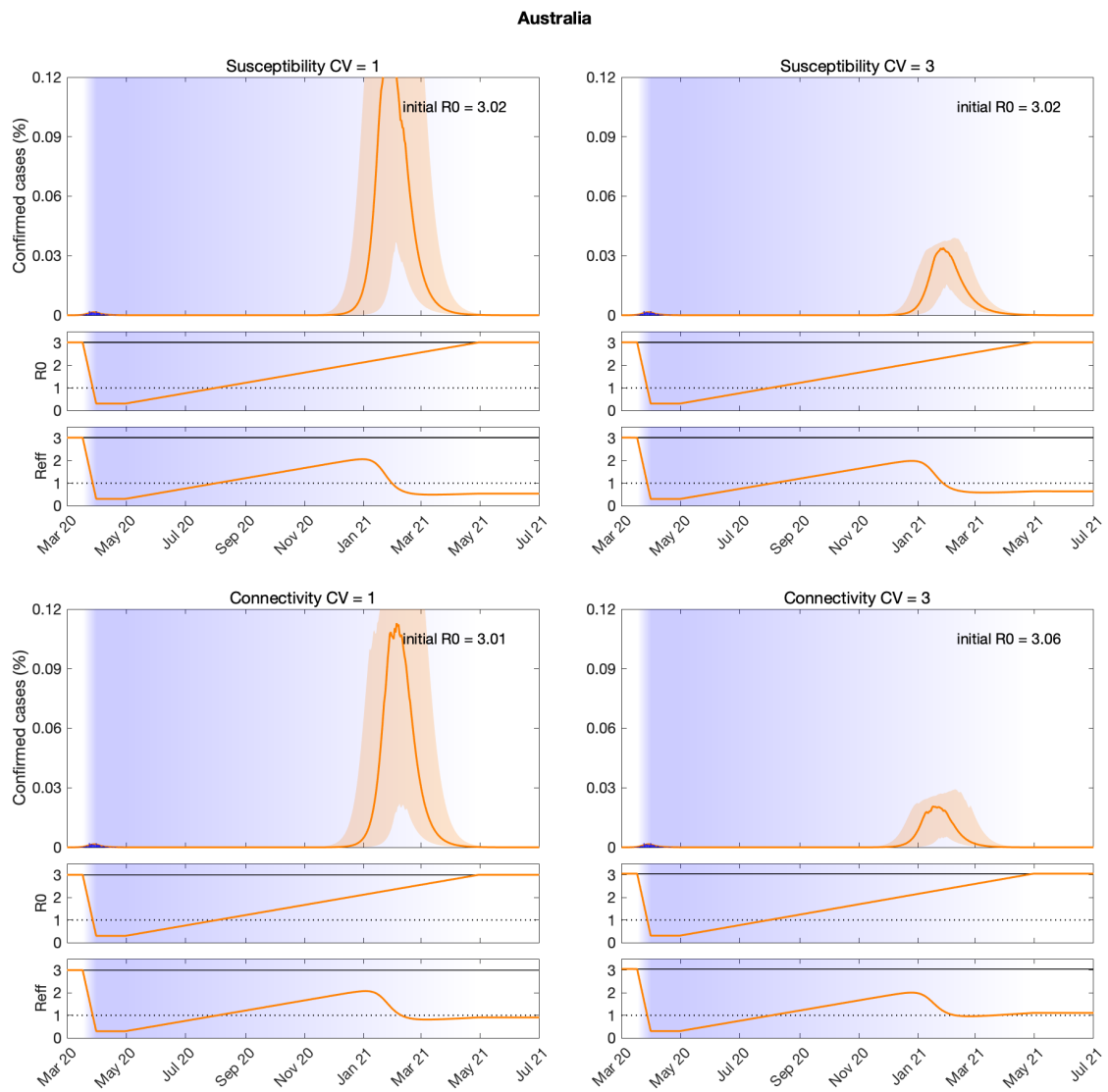
Parameter posterior distributions and corresponding model fit (right most plot) for Turkey, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S11



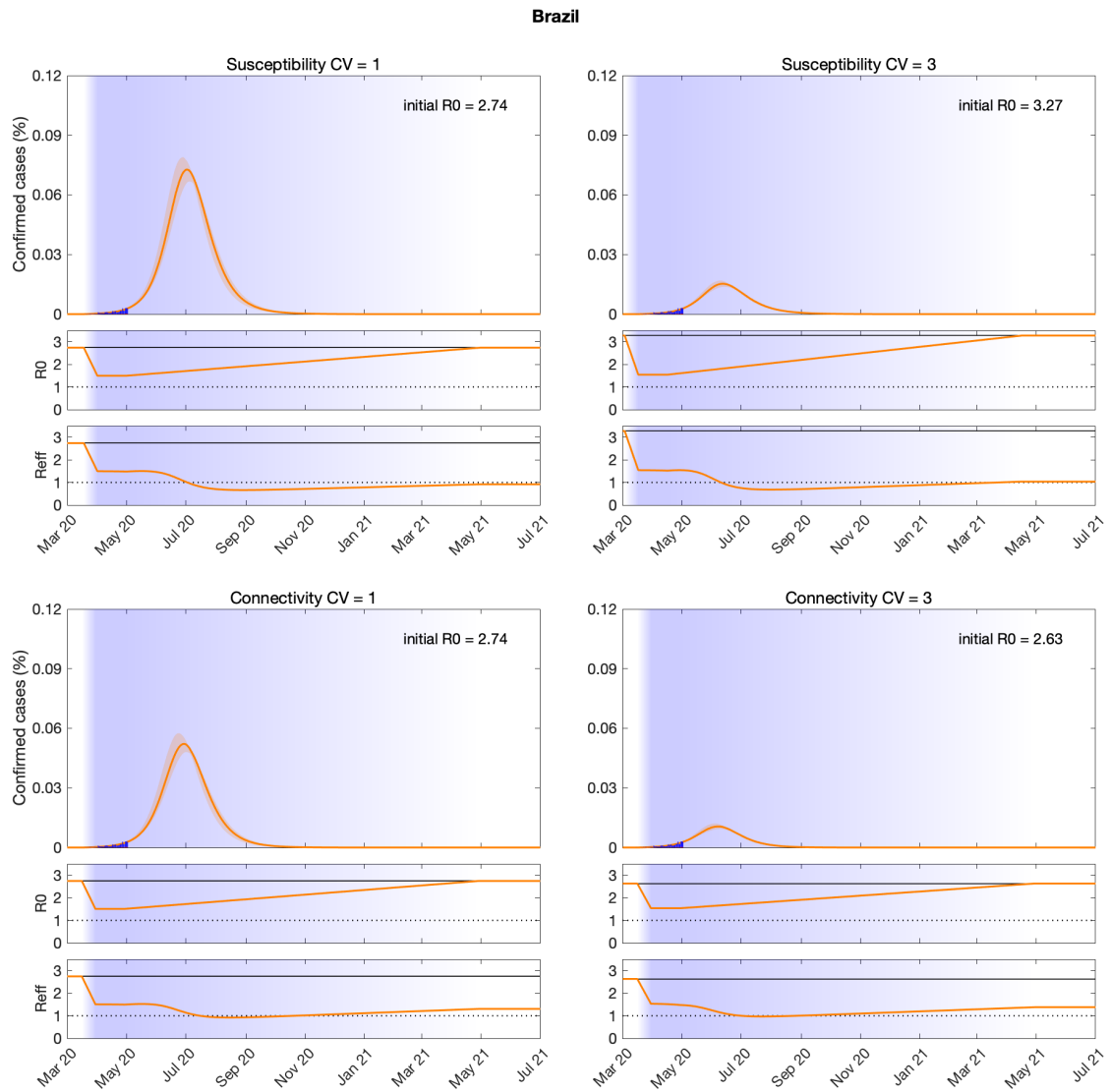
Parameter posterior distributions and corresponding model fit (right most plot) for the UK, using different models. Black bars represent the data (reported daily covid-19 cases) and the red line shows the model prediction using the median posterior parameter estimates. Note that the number in each panel refers to CV.

Figure S12



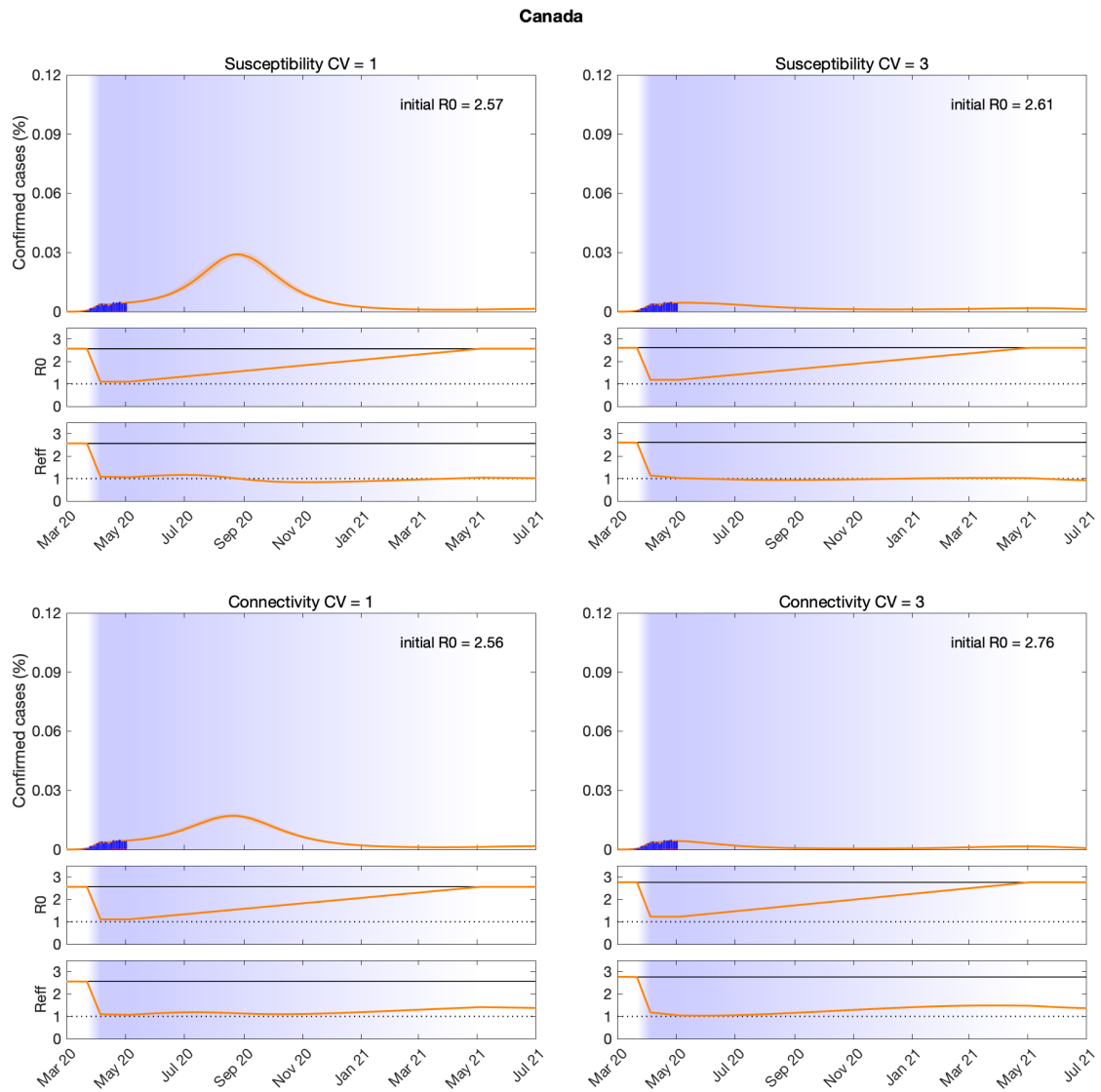
Suppressed wave and subsequent dynamics in Australia. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S13



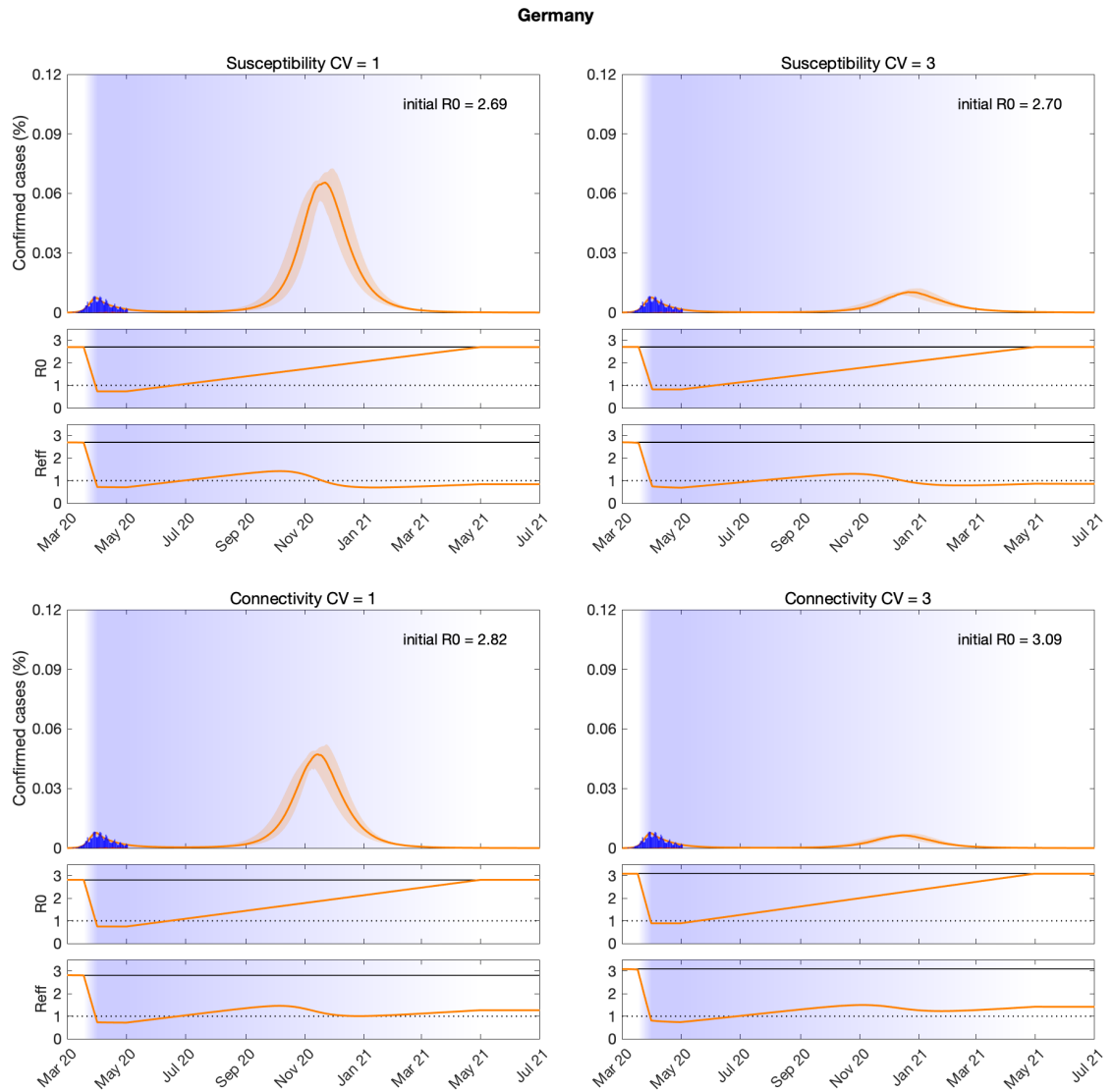
Suppressed wave and subsequent dynamics in Brazil. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S14



Suppressed wave and subsequent dynamics in Canada. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

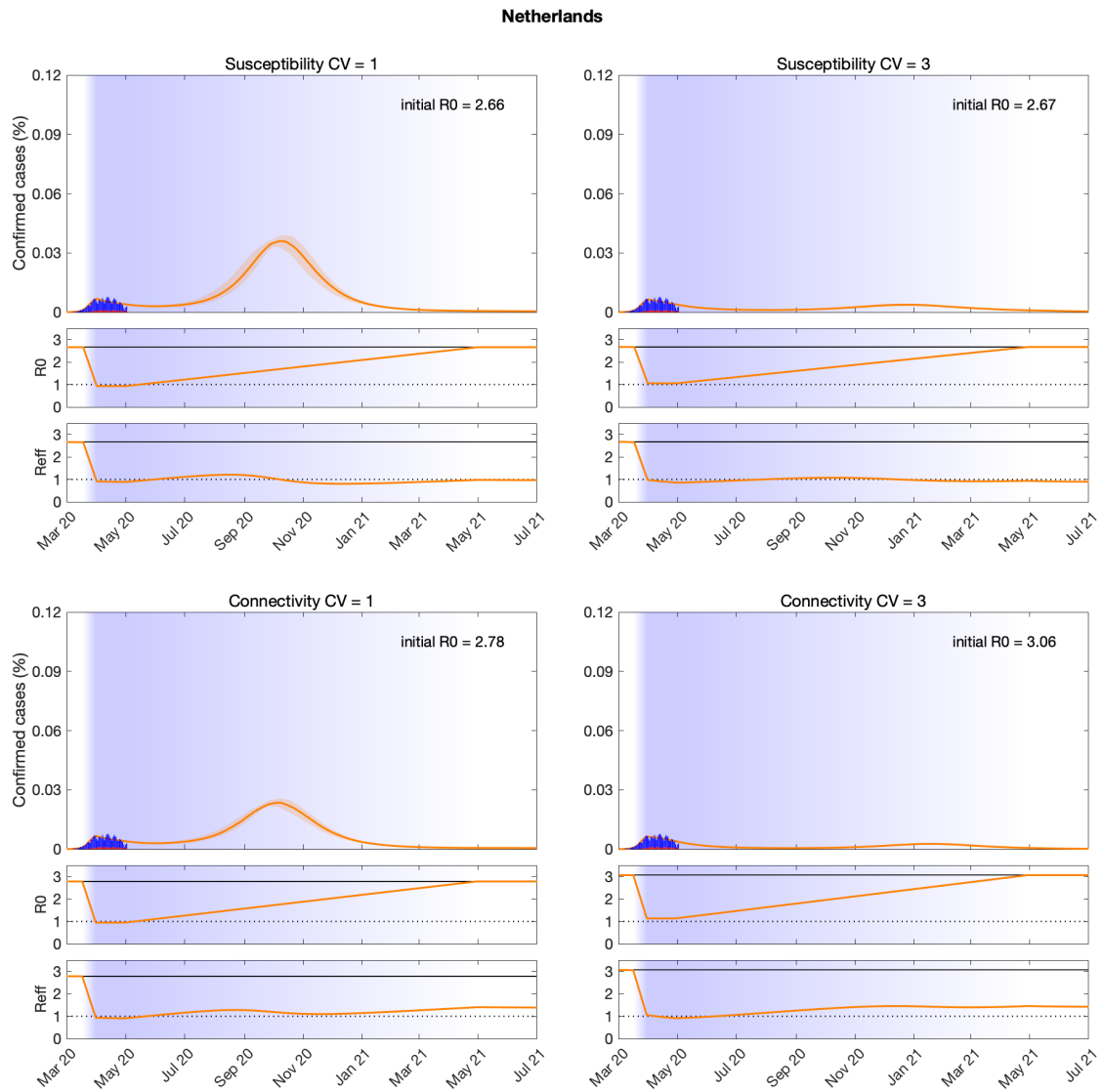
Figure S15



Suppressed wave and subsequent dynamics in Germany. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

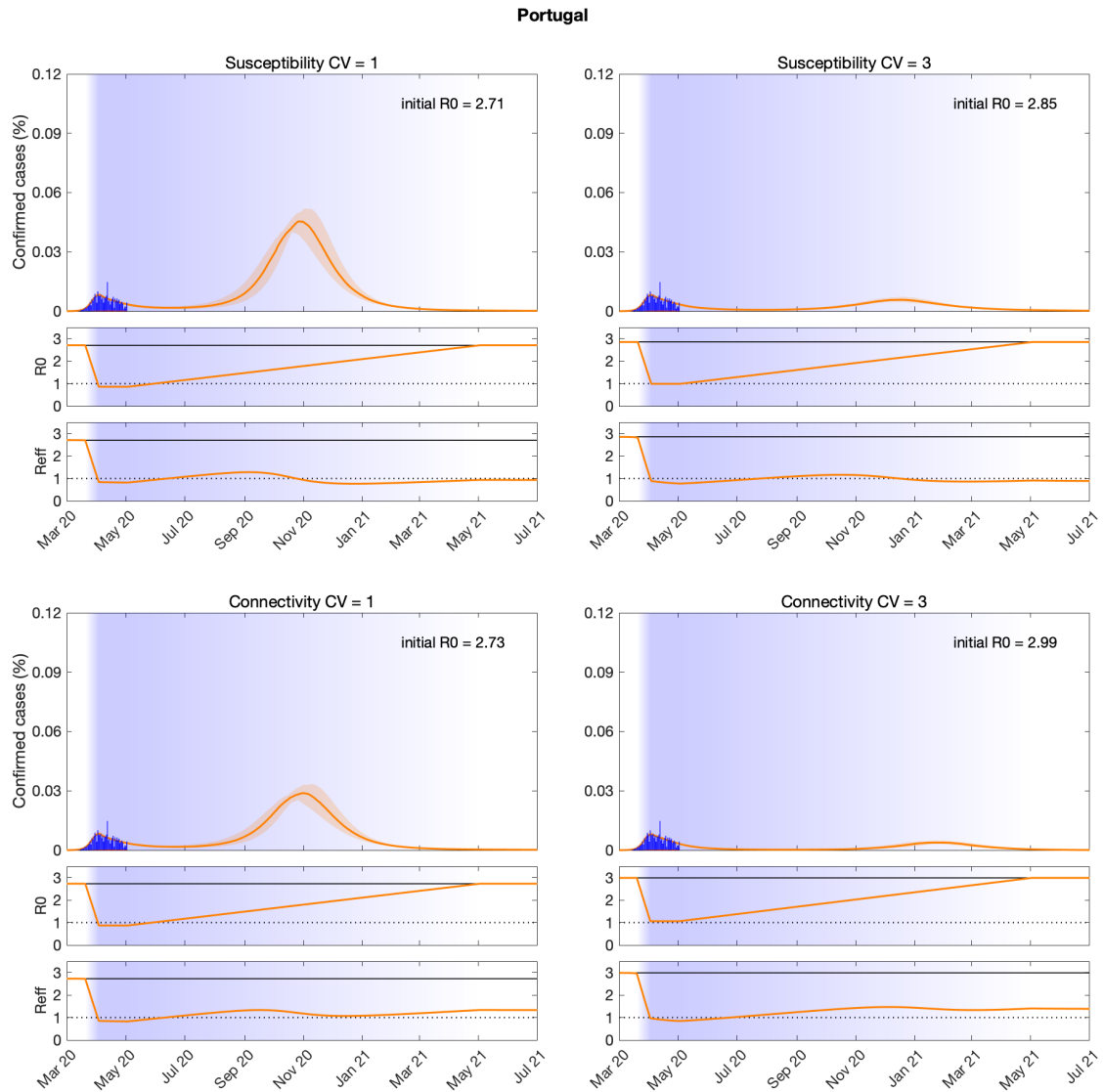


Figure S16



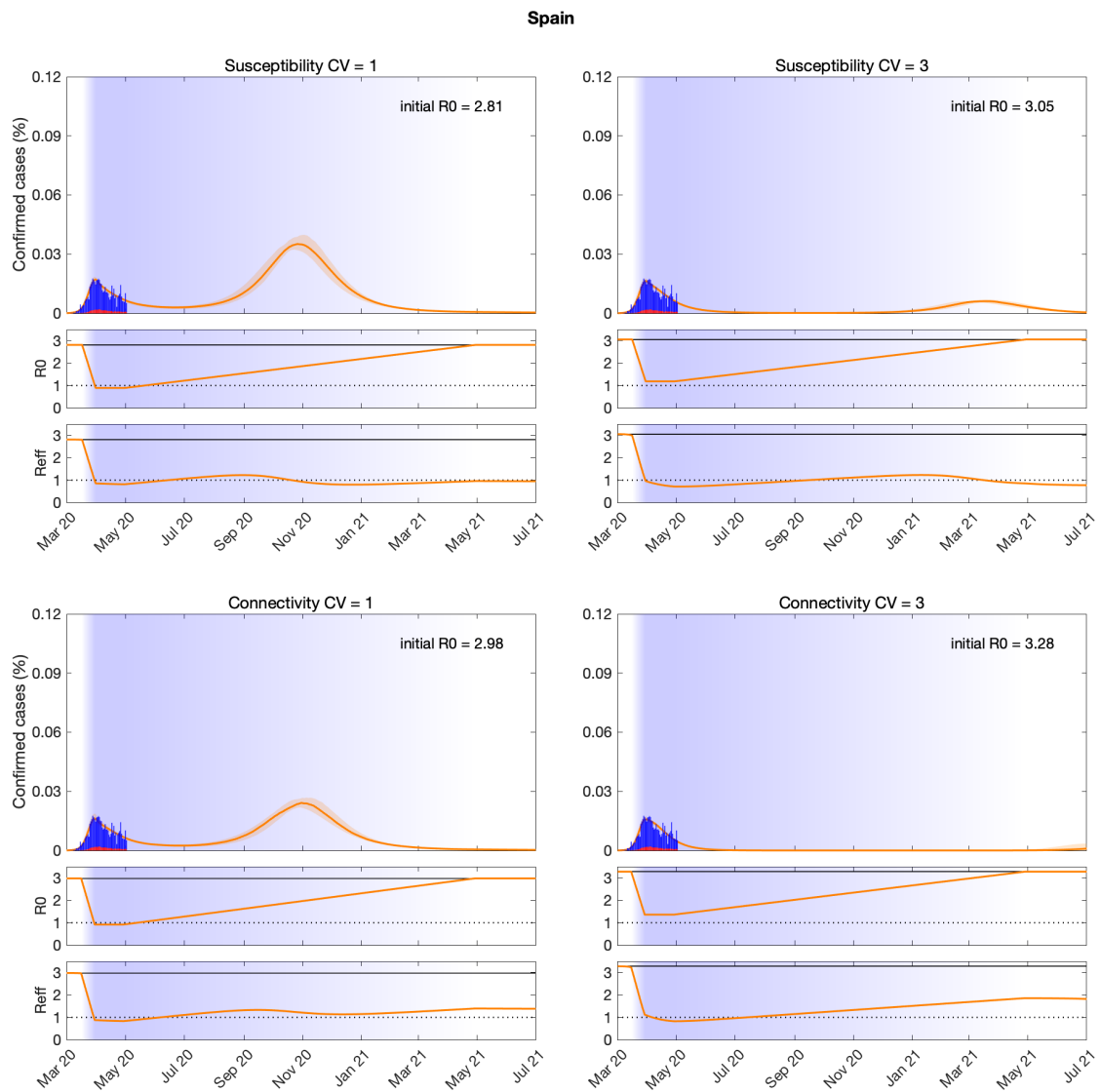
Suppressed wave and subsequent dynamics in Netherlands. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S17



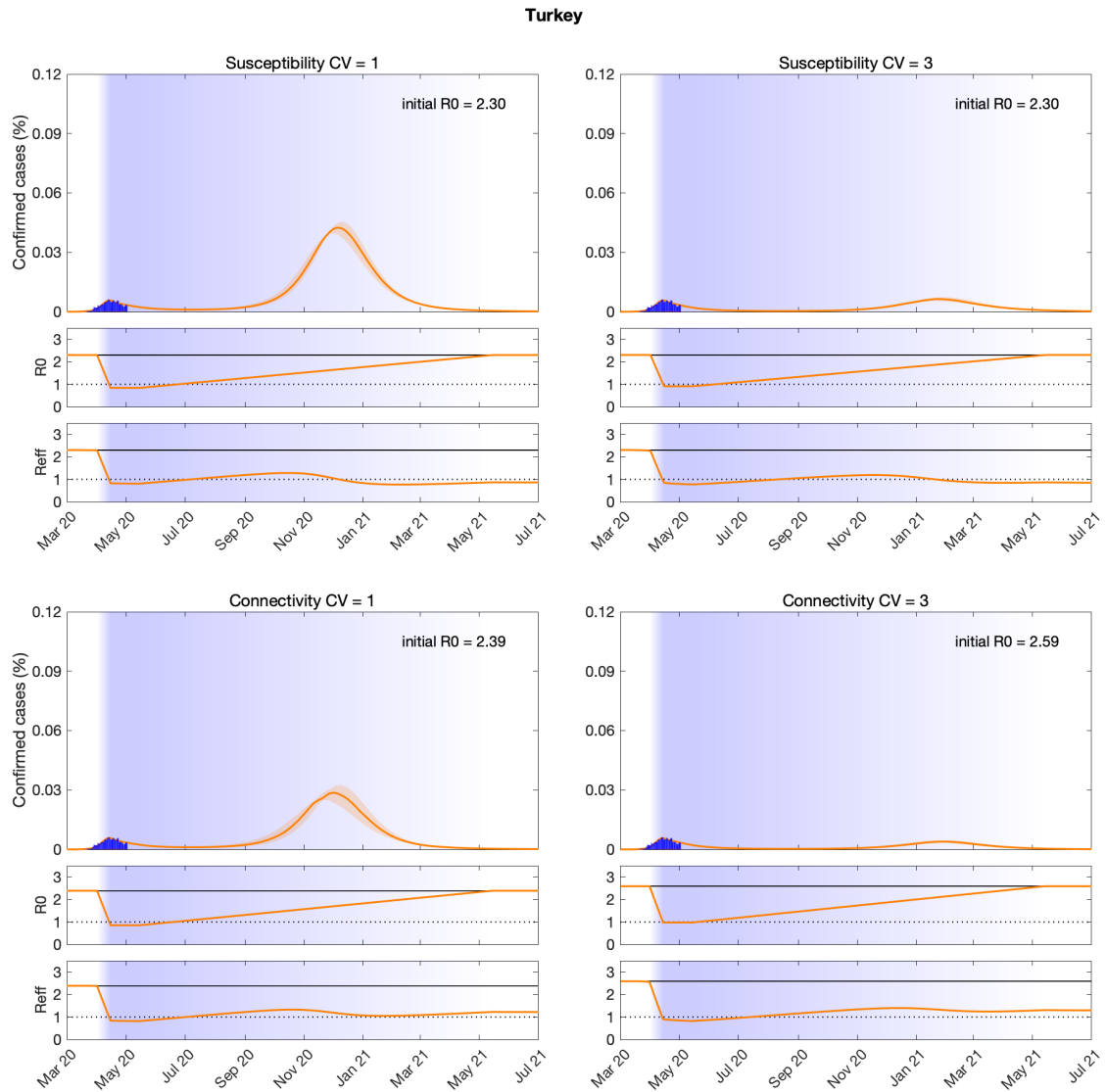
Suppressed wave and subsequent dynamics in Portugal. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S18



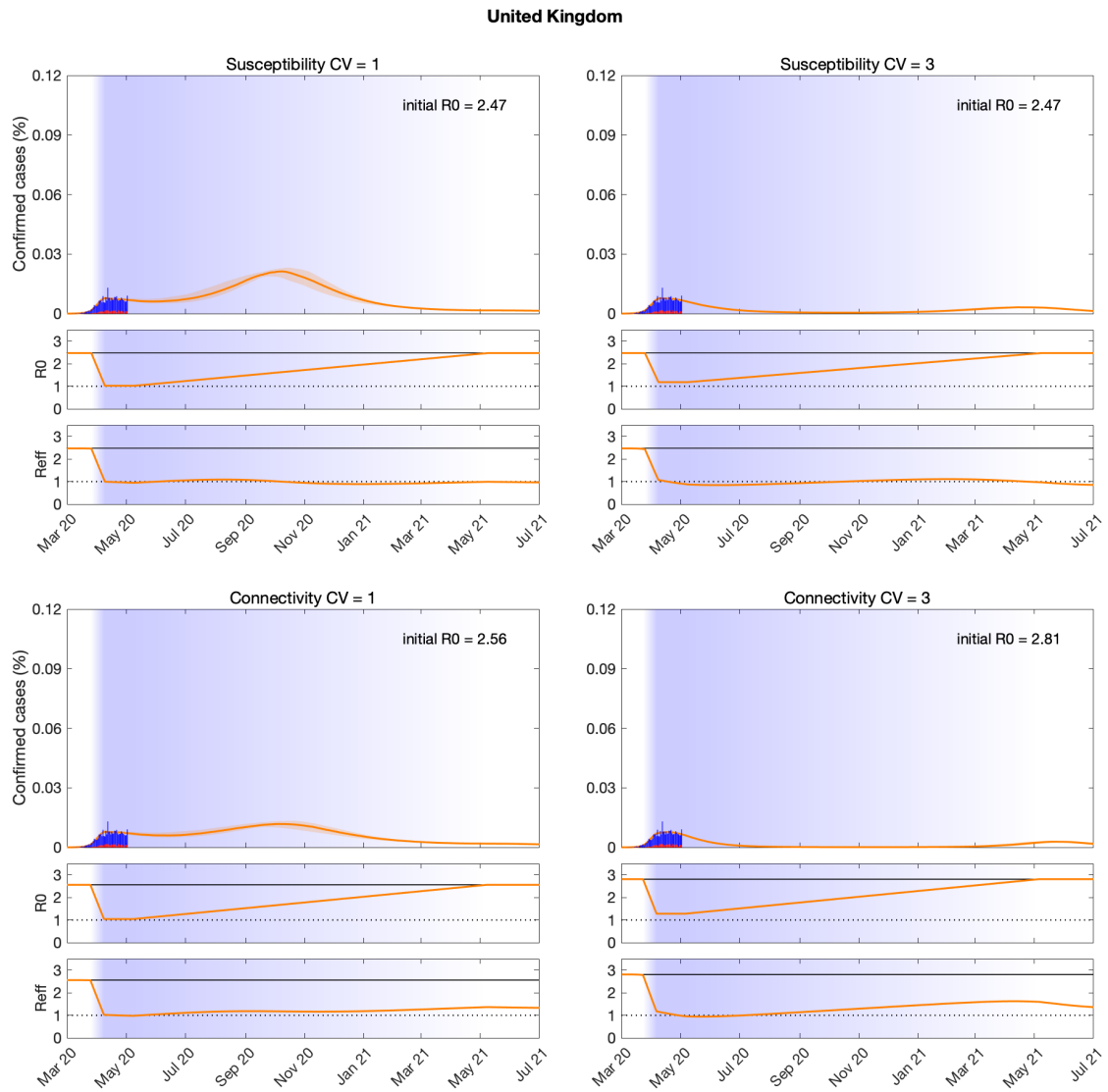
Suppressed wave and subsequent dynamics in Spain. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S19



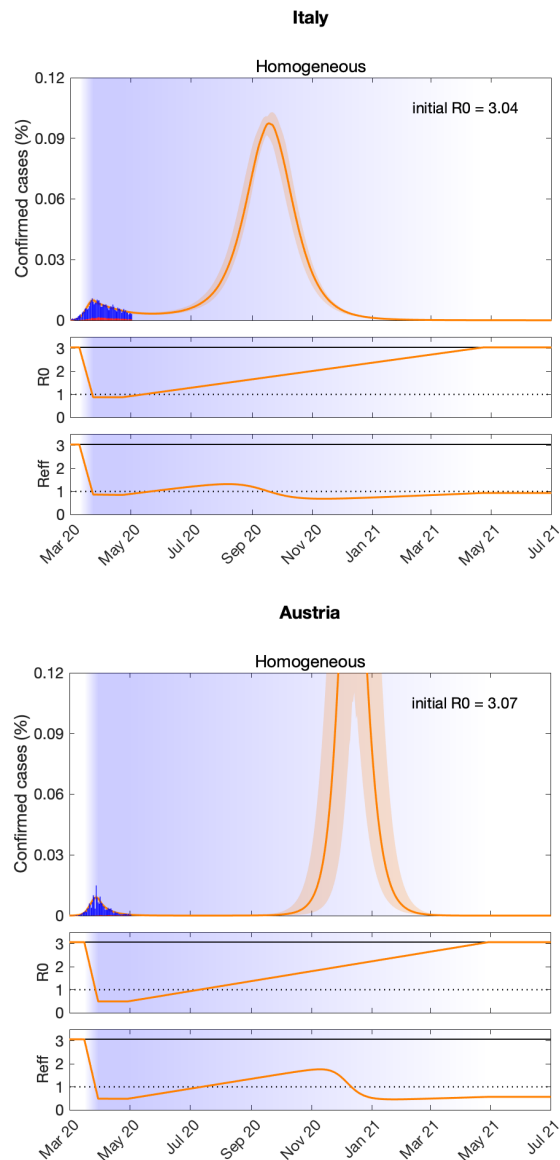
Suppressed wave and subsequent dynamics in Turkey. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S20



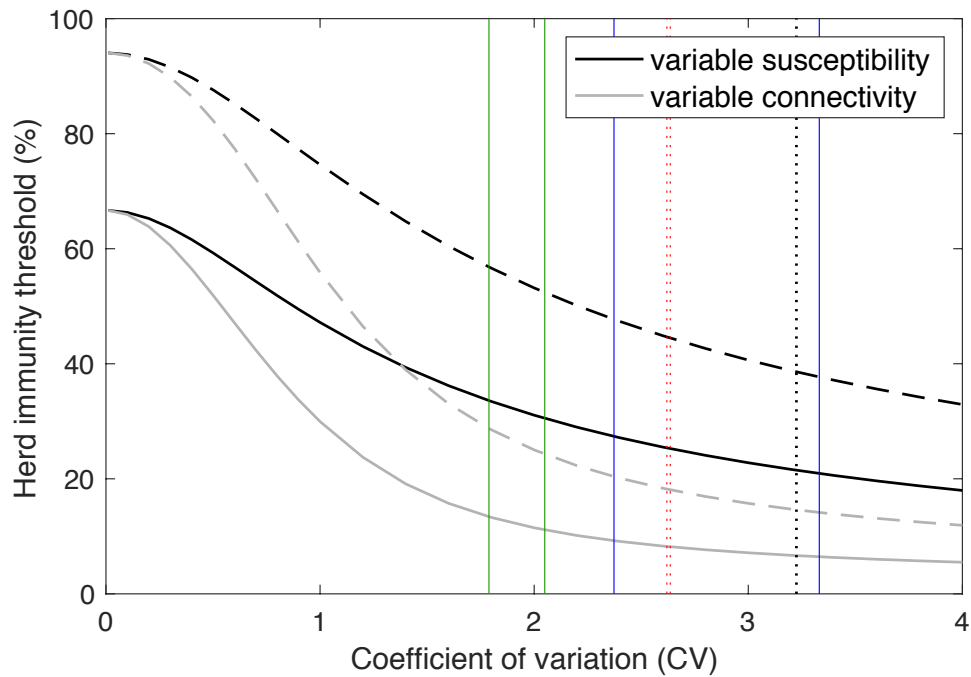
Suppressed wave and subsequent dynamics in United Kingdom. Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S21



Suppressed wave and subsequent dynamics according to a homogeneous SEIR model applied to Italy and Austria (heterogeneity in infectiousness would not change these dynamics). Blue bars are confirmed cases and overlaid red bars represent deaths. Basic ( $R_0$ ) and effective ( $R_{eff}$ ) reproduction numbers are displayed on bottom panels. Blue shades represent social distancing periods. Curves represent median model predictions from  $10^4$  posterior samples. Orange shades represent 95% credible intervals.

Figure S22



Herd immunity threshold (solid curves) and final size of uncontrolled epidemics (dashed curves) with lognormal distributed susceptibility or connectivity assuming  $R_0 = 3$ . Vertical lines indicate coefficients of individual variation for several infectious diseases according to literature: (solid green) susceptibility or exposure to malaria [Amazon 1.8 (6), Africa 2.4 (7)]; (solid blue) susceptibility or exposure to tuberculosis [Portugal 2.4, Brazil 3.3 (8)]; (dotted red) infectiousness for SARS-CoV-1 [Singapore 2.62, Beijing 2.64 (9)]; (dotted black) infectiousness for SARS-CoV-2 [3.2 (10)].