Supplementary Material: Epidemiological Model

The model described here builds on a previously published model of SARS-CoV-2 transmission introduced in Flaxman et al, 2020 [1], subsequently extended into a two-category framework in Faria et al, 2020 [2]. Replication code is available at `https://github.com/ImperialCollegeLondon/Delta_Variant_Delhi`.

Model Specification The model describes two categories, denoted $s \in \{1, 2\}$. The population-unadjusted reproduction number for the first category is defined as

$$R_{s=1,t} = \mu_0 \, 2 \, \sigma(X_t) \,, \tag{1}$$

where μ_0 is a scale parameter (3.3), σ is a logistic function, and X_t is a first autoregressive process. The population-unadjusted reproduction number of the second category is modelled as

$$R_{s=2,t} = \rho \mathbf{1}_{[t_2,\infty)} R_{1,t} \,, \tag{2}$$

with

$$\rho \sim \text{Gamma}(5, 5) \; \in [0, \infty) \,, \tag{3}$$

where ρ is a parameter defining the relative transmissibility of category 2 compared to category 1, and $\mathbf{1}_{[t_2,\infty)}$ is an indicator function taking the value of 0 prior to t_2, and 1 thereafter, highlighting that category 2 does not contribute to the observed epidemic evolution before its emergence. The prior for ρ is chosen because to be weakly informative, setting 90% of prior mass a between $\times 0.4$ and $\times 1.8$ increase in transmissibility, while maintaining a neutral to conservative default in the context of increased transmissibility since it has a mean of $\times 1$ and median of $\times 0.9$.

Infections $i_{s,t}$ arise for each category s at time t according to a discrete renewal process

$$i_{s,t} = \left(1 - \frac{n_{s,t}}{N}\right) R_{s,t} \sum_{\tau<t} i_{s,\tau} g_{t-\tau} \,, \tag{4}$$

where N is the total population size, $n_{s,t}$ is the total extent of population immunity to category s present at time t, and g is the generation interval distribution.

The susceptible depletion term for category s is modelled as

$$n_{s,t} = \sum_{\tau<t} i_{s,\tau} W_{t-\tau} + \beta_s \left(1 - \alpha_{s,t}\right) \sum_{\tau<t} i_{\setminus s,\tau} W_{t-\tau} \,. \tag{5}$$

The notation $\setminus s$ means not-s, and we make the assumption of symmetric cross-immunity, $\beta = \beta_s = \beta_{\setminus s}$. Cross-immunity is given the prior

$$\beta \sim \text{Beta}(2, 1) \,. \tag{6}$$

Immune escape or the evasion of cross-immunity of Delta as reported in the analysis, is defined as the complement of the cross-immunity parameter, that is $(1 - \beta)$. The prior for β has been chosen to reflect our default assumption that the mostly likely scenario is no evasion of cross-immunity. $W_{t-\tau}$ is the time-dependent waning of immunity elicited by previous infection, which is modelled as a Rayleigh survival-type function with Rayleigh parameter of sigma = 310, which produces 50% of individuals still immune after 1 year. The cross-immunity susceptible term $\alpha_{s,t}$ is modelled as

$$\alpha_{s,t} = \frac{(1 - \beta_s) \sum_{\tau<t} i_{s,\tau} W_{t-\tau}}{N - \beta_s \sum_{\tau<t} i_{s,\tau} W_{t-\tau}} \,. \tag{7}$$

Infections in Delhi are seeded for six days at the start of the epidemic from t_1 as

$$i_{1,t_1} \sim \text{Exponential}(1/\tau) \,, \tag{8}$$

with

$$\tau \sim \text{Exponential}(0.03) \,, \tag{9}$$

and the second category for six days from t_2, which is 14-02-2021 in the central scenario, as

$$i_{2,t_2} \sim \text{Normal}(1, 20^2) \ \in [1, \infty) \,. \tag{10}$$

Non-unit seeding of the B.1.617.2 variant and the diffuse prior represent our uncertainty in the precise date and magnitude of B.1.617.2's introduction/importation into Delhi.

The model generates deaths d_t from infections in terms of onset-to-death distribution π via the following mechanistic relationship given:

$$d_t = \sum_s \text{ifr}_s \sum_{\tau<t} i_{s,\tau} \pi_{t-\tau} \,. \tag{11}$$

The infection fatality ratios of each of the categories (ifr_s) are given moderately informative priors:

$$\text{ifr}_s \sim \text{Normal}(0.25, 0.02^2) \ \in [0, 100] \tag{12}$$

with our central estimate based on the results of Brazeau et al [3] and adjusted for the demography of the city. We allow for variation around the estimate of 0.25 however, with the prior providing some support for IFRs in the range 0.15% - 0.35%. This range is similar to estimates by Banaji for Mumbai, a city with comparable demographics, for which an IFR is reported with 95% confidence intervals of (0.15%, 0.33%).[4] A limitation of the model is the assumption of homogeneous exposure across subsets of the population.

Likelihood component 1. The observation model uses three types of data from four sources. In the first, the likelihood for the expected deaths D_t, is modelled as negative-binomially distributed,

$$D_t \sim \text{NegativeBinomial}(d_t(1-\omega), d_t + \frac{d_t^2}{\phi}) \,, \tag{13}$$

with mortality data d_t and diffuse dispersion prior

$$\phi \sim \text{Normal}(0, 5^2) \in [0, \infty) \,. \tag{14}$$

The underreporting factor ω describes the degree of death underascertainment, e.g. a value of 0.25 means 25% of COVID-19 deaths are not reported (due, e.g. to limited testing). The central scenario chosen for death underreporting is 50%, due to the wide range of values reported in available literature for Delhi and India.[5, 6, 7, 4, 8, 9] To mitigate the effect of the uncertainty in death underreporting, sensitivity tests are carried out a range of values, ω in {10%, 33%, 50%, 66%}, to ensure inferences are robust.

Likelihood component 2. The second likelihood is based on genomic data from individuals where infections were sequenced and where the sequence was uploaded to GISAID. Specifically, the proportion of sequenced genomes identified as B.1.617.2 at time t are modelled with a binomial likelihood

$$G_t^+ \sim \text{Binomial}(G_t^+ + G_t^-, \theta_t) \,, \tag{15}$$

with positive counts for B.1.617.2 denoted G_t^+ and counts for lineages not belonging to B.1.617.2 recorded as G_t^-. The success probability for B.1.617.2 positivity is modelled as the infection ratio

$$\theta_t = \frac{\tilde{i}_{2,t}}{\tilde{i}_{1,t} + \tilde{i}_{2,t}} \,, \tag{16}$$

where $\tilde{i}_{s,t}$ is given by

$$\tilde{i}_{s,t} = \sum_{\tau \leq t} i_{s,\tau} \, \kappa_{t-\tau} \,, \tag{17}$$

to account for the time varying PCR positivity displayed over the natural course of a COVID-19 infection. The distribution κ describes the probability of being PCR positive over time following infection, and is based on [10].

Likelihood component 3. Serological data are incorporated in our modelling framework, using results from the survey presented in this work, and from Velumani et al..[11] The observed seropositivity (S_t) on a given day, t, is modelled as follows

$$S_t^+ \sim \text{Binomial}(S_t^+ + S_t^-,\, \nu_t \sum_{\tau \leq t} i_{s,\tau} C_{t-\tau})\,, \tag{18}$$

where $C_{t-\tau}$ is the cumulative probability of an individual infected on day τ having seroconverted less seroreverted by time t. This distribution is empirical and based on [12]. The term ν_t is a multiplicative random effect specified to mitigate likely biases in the serological data

$$\nu_t = 2\sigma(Y_t)\,, \quad Y_t \sim \mu + \eta_t\,, \quad \mu \sim N(0,1)\,, \quad \eta_t \sim N(0,\delta)\,, \quad \delta \sim N^+(0,1)\,, \tag{19}$$

where σ is the logistic function.

Eq (4) can be modified to account for population effects (decreasing susceptible population over time) such that no over-shooting happens due to discretization[1]

$$i_{s,t} = (N - n_{s,t}) \left(1 - \exp\left(-\frac{i_{s,t}}{N}\right)\right). \tag{20}$$

The formula for $i_{s,t}$ is derived from a continuous time model on $[t-1, t]$. This is to avoid discrete time effects such as infections going above the total population N. Specifically, we assume that the infections $i(\Delta t)$ in $[t-1, t-1+\Delta t]$ are given by the differential equation $\partial i(\Delta t)/\partial \Delta t = i_t(1 - (n_{s,t} + i(\Delta t))/N)$, which has the solution $i(1) = i_t$ as above.

Modelling limitations Inferences from the model are subject multiple limitations, arising from biases in the data, and in our choices of priors. In particular, the level of underreporting in Delhi is poorly characterised. This uncertainty is mitigated by sensitivity testing, as described in the Supplementary Information. Similarly, uncertainty in the temporal waning of immunity, the date of when Delta was first introduced to Delhi, and the IFRs of SARS-CoV-2 variants in Delhi, all provide sources of bias that we currently only mitigate through sensitivity testing. Furthermore, we note serological data is likely to be systematically biased. This is partially addressed through the inclusion of an additional term in the serology likelihood that provides a multiplicative random effect.

Computational notes The analysis uses R version 3.6.3. Inferences are based on 2000 iterations of Hamiltonian Monte Carlo using 2 chains, with rhat statistics confirmed to be less than 1.02. The inference is performed using rstan version 2.21.2. Replication code and data are available at `https://github.com/ImperialCollegeLondon/Delta_Variant_Delhi`.

[1] See implementation details in the R package `epidemia`: https://imperialcollegelondon.github.io/epidemia/.

References

[1] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe". In: *Nature* 584.7820 (2020), pp. 257–261. DOI: https://doi.org/10.1038/s41586-020-2405-7.

[2] Nuno R Faria, Thomas A Mellan, Charles Whittaker, Ingra M Claro, Darlan da S Candido, Swapnil Mishra, Myuki A E Crispim, Flavia C S Sales, Iwona Hawryluk, John T McCrone, Ruben J G Hulswit, Lucas A M Franco, Mariana S Ramundo, Jaqueline G de Jesus, Pamela S Andrade, Thais M Coletti, Giulia M Ferreira, Camila A M Silva, Erika R Manuli, Rafael H M Pereira, Pedro S Peixoto, Moritz U G Kraemer, Nelson Gaburo Jr, Cecilia da C Camilo, Henrique Hoeltgebaum, William M Souza, Esmenia C Rocha, Leandro M de Souza, Mariana C de Pinho, Leonardo J T Araujo, Frederico S V Malta, Aline B de Lima, Joice do P Silva, Danielle A G Zauli, Alessandro C de S Ferreira, Ricardo P Schnekenberg, Daniel J Laydon, Patrick G T Walker, Hannah M Schlüter, Ana L P Dos Santos, Maria S Vidal, Valentina S Del Caro, Rosinaldo M F Filho, Helem M Dos Santos, Renato S Aguiar, José L Proença-Modena, Bruce Nelson, James A Hay, Mélodie Monod, Xenia Miscouridou, Helen Coupland, Raphael Sonabend, Michaela Vollmer, Axel Gandy, Carlos A Prete Jr, Vitor H Nascimento, Marc A Suchard, Thomas A Bowden, Sergei L K Pond, Chieh-Hsi Wu, Oliver Ratmann, Neil M Ferguson, Christopher Dye, Nick J Loman, Philippe Lemey, Andrew Rambaut, Nelson A Fraiji, Maria do P S S Carvalho, Oliver G Pybus, Seth Flaxman, Samir Bhatt, and Ester C Sabino. "Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil". en. In: *Science* (Apr. 2021).

[3] Nicholas Brazeau, Robert Verity, Sara Jenks, Han Fu, Charles Whittaker, Peter Winskill, Ilaria Dorigatti, Patrick Walker, Steven Riley, Ricardo P Schnekenberg, et al. "Report 34: COVID-19 infection fatality ratio: estimates from seroprevalence". In: (2020). URL: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-34-ifr/.

[4] Murad Banaji. "Estimating COVID-19 infection fatality rate in Mumbai during 2020". In: *medRxiv* (2021).

[5] Soumik Purkayastha, Rupam Bhattacharyya, Ritwik Bhaduri, Ritoban Kundu, Xuelin Gu, Maxwell Salvatore, Debashree Ray, Swapnil Mishra, and Bhramar Mukherjee. "A comparison of five epidemiological models for transmission of SARS-CoV-2 in India". In: *BMC infectious diseases* 21.1 (2021), pp. 1–23.

[6] Margarita Pons-Salort, Jacob John, Oliver J Watson, Nicholas F Brazeau, Robert Verity, Gagandeep Kang, and Nicholas C Grassly. "Reconstructing

the COVID-19 epidemic in Delhi, India: infection attack rate and reporting of deaths". In: *medRxiv* (2021).

[7] Manoj V Murhekar, Tarun Bhatnagar, Sriram Selvaraju, Kiran Rade, V Saravanakumar, Jeromie Wesley Vivian Thangaraj, Muthusamy Santhosh Kumar, Naman Shah, R Sabarinathan, Alka Turuk, et al. "Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020". In: *Indian Journal of Medical Research* 152.1 (2020), p. 48.

[8] Jayakrishnan Unnikrishnan, Sujith Mangalathu, and Raman V Kutty. "Estimating under-reporting of COVID-19 cases in Indian states: an approach using a delay-adjusted case fatality ratio". In: *BMJ open* 11.1 (2021), e042584.

[9] Raaj Kishore Biswas, Awan Afiaz, and Samin Huq. "Underreporting COVID-19: the curious case of the Indian subcontinent". In: *Epidemiology & Infection* 148 (2020).

[10] Joel Hellewell, Timothy William Russell, Rupert Beale, Gavin Kelly, Catherine Houlihan, Eleni Nastouli, Adam J Kucharski, SAFER Investigators, Field Study Team, Crick COVID-19 Consortium, et al. "Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections". In: *medRxiv* (2020). DOI: https://doi.org/10.1101/2020.11.24.20229948.

[11] Arokiaswamy Velumani, Chaitili Nikam, Wilson Suraweera, Sze Hang Fu, Hellen Gelband, Patrick E Brown, Isaac Bogoch, Nico Nagelkerke, and Prabhat Jha. "SARS-CoV-2 Seroprevalence in 12 Cities of India from July-December 2020". In: *medRxiv* (2021).

[12] Benny Borremans, Amandine Gamble, KC Prager, Sarah K Helman, Abby M McClain, Caitlin Cox, Van Savage, and James O Lloyd-Smith. "Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2 infection by time since symptom onset". In: *Elife* 9 (2020), e60122. DOI: 10.7554/eLife.60122.