# Supplementary Notes

Wiep Van der Toorn [1,2], Djin-Ye Oh [3], Daniel Bourquain [4], Janine Michel [4], Eva Krause [4], Andreas Nitsche [4], Max von Kleist [1,2,5,*]

**1,2** Systems Medicine of Infectious Disease (P5) and Bioinformatics (MF1), Methodology and Research Infrastructure, Robert Koch-Institute, Berlin, Germany
**3** FG17 Influenza and other respiratory viruses, Department of infectious disease, Robert Koch-Institute, Berlin, Germany
**4** ZBS1 highly pathogenic viruses, Center for biological Threats and special pathogens, Robert Koch-Institute, Berlin, Germany
**5** German COVID Omics Initiative (deCOI)

*kleistm@rki.de

## SN.1 Mathematical formulations

Below we describe the model structure, as well as its equations. In essence, we distinguish different *phases* of the infection dynamics by whether the virus is detectable, the individual is infectious and may have symptoms. Notably, asymptomatic infections are also included. For asymptomatic individuals we assume the same infection dynamics without displaying symptoms. The considered *phases* and their attributes are depicted in Table SN.1.

**Table SN.1.** Phases of the underlying Markov model. $^{+}$Asymptomatically infected individuals are assumed to have the same infection dynamics without showing symptoms

| j | phase | detectable | infectious | symptoms |
|---|---|---|---|---|
| 1 | pre-detection | no | no | no |
| 2 | pre-symptomatic | yes | yes | no |
| 3 | infectious (& symptomatic) | yes | yes | $(\text{yes})^{+}$ |
| 4 | post-infectious (& symptomatic) | yes | no | $(\text{yes})^{+}$ |
| 5 | post-detectable | no | no | no |

We then model the disease progression semi-mechanistically using a Markov jump process formalisms (discrete state, continuous time). We use the Markov jump formalism, because it allows to model inter-individual differences in disease progression. In this framework, an infected individual will progress to the next phase of the disease after a random waiting time. We solve for all individuals with their respective random waiting times simultaneously, deriving a probability distribution that evolves through time. This allows to model inter-individual differences in e.g. time-to-detectability or time-to-infectiousness. Overall, we will adjust the model, such that it accurately computes the *probability* that patients are PCR-positive, infectious and symptomatic at any time $t$ after infection.

### SN.1.1 Model equations

We model the disease progression as a transit model with $j = \{1 \dots 5\}$ phases as depicted above in Table SN.1, each of which consists of $i = \{1 \dots n_j\}$ compartments. The rates of transition from one compartment to the next within a phase are chosen to be identical, while the number of compartments for each phase is fitted, so that the residence dynamics match clinically observed dynamics, as discussed in the next subsection.

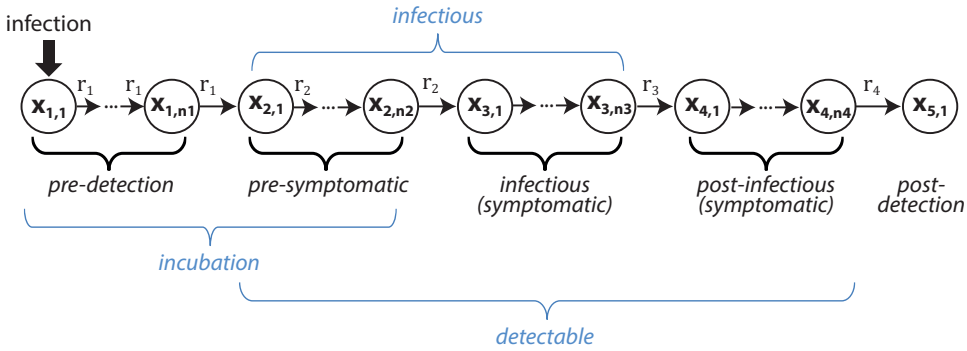


**Fig SN.1. Schematic of the model.** The model of disease progression depicting the distinct *phases*. Each phase consists of different sub-compartments.

The equations that model the probability that individuals are in phase $j = 1...5$ at time $t$ is given by:

$$\frac{d}{dt}p_t(x_{1,1}) = -r_1 \cdot p_t(x_{1,1}) \tag{SN.1}$$

$$\dots = \dots \tag{SN.2}$$

$$\frac{d}{dt}p_t(x_{1,n_1}) = r_1 \cdot p_t(x_{1,n_1-1}) - r_1 \cdot p_t(x_{1,n_1}) \tag{SN.3}$$

$$\frac{d}{dt}p_t(x_{2,1}) = r_1 \cdot p_t(x_{1,n_1}) - r_2 \cdot p_t(x_{2,1}) \tag{SN.4}$$

$$\dots = \dots \tag{SN.5}$$

$$\frac{d}{dt}p_t(x_{m,1}) = r_{m-1} \cdot p_t(x_{m-1,n_{m-1}}), \tag{SN.6}$$

where $p_t(x_j) = \sum_{i=1}^{n_j} p_t(x_{j,i})$ and the last state ("post-detection phase") is an absorbing state. In matrix notation, the model is given by

$$\frac{d}{dt}p_t(\mathbf{x}) = A \cdot p_t(\mathbf{x}) \tag{SN.7}$$

with e.g.:

$$A(\mathrm{SCR}) = \begin{pmatrix} -r_1 & 0 & \cdots & & & & \cdots & 0 \\ r_1 & -r_1 & & & & & & \vdots \\ 0 & r_1 & -r_2 & & & & & \\ \vdots & & \ddots & \ddots & & & & \\ & & & (1 - f_s \cdot \mathrm{SCR}) \cdot r_2 & -r_3 & & & \\ & & & & \ddots & \ddots & & \\ \vdots & & & & & r_{m-1} & -r_{m-1} & \vdots \\ 0 & \cdots & & & \cdots & 0 & r_{m-1} & 0 \end{pmatrix} \tag{SN.8}$$

where $f_s$ is the 'fraction symptomatic' (user input) and SCR is a boolean variable that defines whether 'symptom screening' is performed. Therefore, the system can be solved analytically with

$$\boxed{p_t(\mathbf{x}) = e^{t \cdot A} \cdot p_{t_0}(\mathbf{x})} \tag{SN.9}$$

where $e^{t \cdot A}$ denotes the matrix exponential defined as

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \tag{SN.10}$$

and $p_{t_0}(\mathbf{x})$ denotes the initial condition of the system.

### SN.1.2 Parameters & model structure

As mentioned above, we assume that the transition rates between compartments $i = \{1 \dots n_j\}$ of the same phase are identical, $r_{j,i} = r_j$. The *mean residence time* $\tau_j$ (= average duration that an infected individual stays in a phase) is then trivially related to the transition rate in this phase:

$$r_j = \frac{n_j}{\tau_j}. \tag{SN.11}$$

Conversely, $\tau_j = \frac{n_j}{r_j}$.

Given a *mean residence time*, i.e. the average duration of a phase $\tau_j$, it is therefore possible to change the shape (skewness) of the transitioning times by adjusting for the number of compartments in that phase until the model reflects the clinical time course accurately enough. An example is shown in Fig. SN.2 below, where addition of compartments introduces a "shoulder" without affecting the *mean duration* of that phase. In the following, we will estimate model parameters from clinical time course

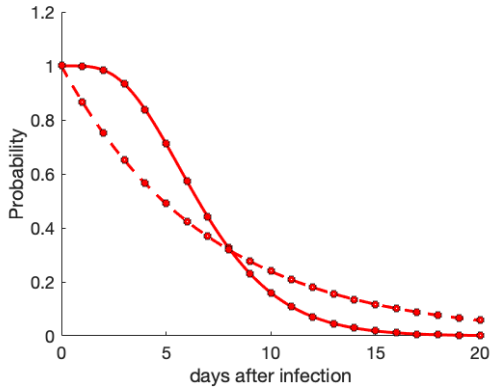


**Fig SN.2. Illustration of how introducing compartments can alter the shape of the residence time without affecting its mean.** Example with $\tau = 5$ days. Dashed line: one sub-compartment with rate $r = 1/5$. Solid line: five sub-compartments with rates $r = 5/5 = 1$.

data, by estimating $n_j$ and $\tau_j$ simultaneously. Moreover, we will discuss our model parameters $\tau_j$ in the context of published data.

### SN.1.3 Calculation of relative risk and risk reduction

Quarantine, testing and isolation are intended to decrease the risk that an infected individual is able to spread the disease. Mathematically, the residual risk is therefore related to the duration- and extent of infectiousness of an individual who is released from quarantine/isolation. The *relative risk* denotes the reduction in that risk relative to 'no intervention', akin to [1, 2]:

$$relative\ risk(t) = \frac{\int_t^\infty P_s(\text{inf}|\text{NPI})\ ds}{\int_0^\infty P_s(\text{inf}|\emptyset)\ ds} \tag{SN.12}$$

where $t$ refers to the duration of a quarantine or isolation period and $P_s(\text{inf}|\text{NPI})$ and $P_s(\text{inf}|\emptyset)$ refer to the probability of being infectious at time $s$ under a non-pharmaceutical intervention (NPI) vs. 'no intervention' $\emptyset$ (or a reference intervention'). Incomplete adherence to the NPI with fraction $w$ is computed as $relative\ risk(t) = \frac{w \cdot \int_t^\infty P_s(\text{inf}|\text{NPI})\ ds + (1-w) \cdot \int_0^\infty P_s(\text{inf}|\emptyset)\ ds}{\int_0^\infty P_s(\text{inf}|\emptyset)\ ds}$.

The denominator of eq. (SN.12) can easily be computed by e.g. augmenting the matrix A, such that

$$\widetilde{A} = \left[\begin{array}{ccc|c} & \mathbf{A} & & 0 \\ & & & \vdots \\ \hline 0 & \mathbf{y} & 0 & 0 \end{array}\right]$$

with $y_{j,\cdot} = 1$ in the last row of the matrix for all 'infectious states' (compare Table SN.1). Then one can solve for

$$p_\infty(\widetilde{\mathbf{x}}|\emptyset) = e^{s\cdot\widetilde{A}(\emptyset)} \cdot \begin{pmatrix} p_{t_0}(\mathbf{x}) \\ 0 \end{pmatrix}; \quad s \to \infty, \tag{SN.13}$$

where the pre-intervention risk can be user-defined or be calculated using the "prevalence estimator" functionality of the software and $\widetilde{A}(\emptyset)$ defines the augmented $A$ matrix without symptom screening. The risk in the 'no-intervention case' is the last entry of the derived vector, i.e. $p_\infty(\widetilde{x}_{N+1}|\emptyset)$.

Likewise, the risk after the non-pharmaceutical intervention strategy can be solved by

(i) first determining $p_t(\mathbf{x}|\text{NPI})$ for the intervention:

$$p_t(\mathbf{x}|\text{NPI}) \quad = \quad \left(\prod_i diag\big(\text{FOR}(\mathbf{x})\big)\cdot e^{\Delta t_i \cdot A(\text{SCR})}\right) \cdot e^{(t_{\text{end}}-t_n)\cdot A(\text{SCR})} \cdot p_{t_0}(\mathbf{x}) \tag{SN.14}$$

where $\Delta t_i \in [t_1, t_2 - t_1, \ldots t_n - t_{n-1}]$ denotes the time spans between the start of the quarantine/isolation and the first test at time $t_1$ and between any consecutive tests until $t_n$ (last test) and $A(\text{SCR})$ denotes the $A$ matrix for- or without symptom screening. Parameter $t_{\text{end}}$ denotes the end of the quarantine/isolation. The matrix $diag\big(\text{FOR}(\mathbf{x})\big)$ is defined

$$diag\big(\text{FOR}(\mathbf{x})\big) = \begin{pmatrix} \ddots & & \\ & \text{FOR}(x_{j,i}) & \\ & & \ddots \end{pmatrix} \tag{SN.15}$$

is a matrix with the the state-dependent false omission rates as its diagonal entries.

(ii) the risk is determined by computing:

$$p_\infty(\widetilde{\mathbf{x}}|\text{NPI}) = e^{s\cdot\widetilde{A}(\emptyset)} \cdot \begin{pmatrix} p_t(\mathbf{x}|\text{NPI}) \\ 0 \end{pmatrix}; \quad s \to \infty \tag{SN.16}$$

and given in the last entry of the derived vector, i.e. $p_\infty(\widetilde{x}_{N+1}|\text{NPI})$.

The $fold\ risk\ reduction(t)$ is calculated

$$fold\ risk\ reduction(t) = \frac{1}{relative\ risk(t)}, \tag{SN.17}$$

### SN.1.4 Prevalence estimation

In the software we provide a tool to estimate the prevalence from the recent incidence history. This setting is used to evaluate testing and quarantine strategies for incoming travellers. In doing so, we assume that an incoming traveller is exposed to the same infection risks that are present in the country where the person is travelling from. This allows to a) calculate the pre-test risk and b) to assess whether an individual is more likely to have acquired the infection recently or in the past (compare Fig. 4 in the main manuscript). Regarding the latter, we distinguish settings where there is an increasing trend in infections, vs. those ones where there is a current decline in infections. In the model, incoming travellers from the former setting are more likely to have acquired an infection recently, compared to travellers coming from the latter setting.

To calculate the pre-test risk, we use the infection dynamics model from eq. (SN.9) together with the incidence reports that the user supplies and the probability of case detection/reporting $P(\text{detect})$ for the country of interest. Using these ingredients we can calculate the pre-test risk as:

$$p_{t_0}(\mathbf{x}) = \sum_{s=-T}^{t_0} e^{(t_0-s)\cdot A(\emptyset)} \cdot p_s(\mathbf{x}), \tag{SN.18}$$

where $(-T)$ is the time horizon before the current date. In the software we evaluate the preceding 5 weeks. The initial condition $p_s(\mathbf{x})$ for day $s$ prior to today is computed from the incidence reports for the country of interest. For example, if $s \in$ week 3 prior to this week, then

$$\sum_j p_s(x_j) = \frac{\pi_{-\Delta t_3}(\text{infect})}{7 \cdot P(\text{detect})} \tag{SN.19}$$

where $\pi_{-\Delta t_3}(\text{infect})$ denotes the number of reported cases per week and 100,000 inhabitants (the incidence) in week 3 prior to this week. The individual probabilities assigned to the different phases are computed according to

$$\frac{p_s(x_j)}{\sum_j p_s(x_j)} = \frac{\tau_j}{\sum_j \tau_j}, \tag{SN.20}$$

where $\tau_j$ denotes the *mean residence time* in each phase. The initial probabilities within the sub-compartments of each phase are uniformly distributed.

## SN.2 Estimation of default parameters from published data

### SN.2.1 Set up

We estimate the default model parameters $n_j$ (number of sub-compartments for phase $j$) and $\tau_j$ (*mean residence time* in phase $j$) in three steps, based on different types of clinical studies:

- Step 1: First, parameters for the *incubation* phase (= *pre-detection + pre-symptomatic*, compare Fig. SN.1) of the model ($j = 1, 2$) are fitted. For this, we use a meta-analysis by Wei et al. [3] that encompasses fifty-six studies on the incubation period of SARS-CoV2.
- Step 2: The optimal parameters for the *symptomatic* phase of the model ($j = 3$) are estimated independently of Step 1. We estimate the parameters based on the consensus of five studies that report on viral load, Ct values and relative infectivity, three of which are published [4–6], in addition to *in-house* data and data that has been kindly provided to us by the consiliary lab for Corona viruses Germany (Drosten Lab Charite Berlin)) [7].
- Step 3: Lastly, the parameters of the *post-infectious* phase of the model ($j = 4$) are estimated. We fix the estimated parameters for the proceeding three phases $j = 1, 2, 3$ and fit the total model to the time-dependent PCR sensitivity profile reported by Kucirca et al. [8], effectively estimating the *mean residence time* in the *post-symptomatic* phase.

**Table SN.2.** Table depicting the three steps in the estimation procedure to derive default parameters from available clinical data. Estimated parameters for the respective step are indicated as "estim.", whereas fixed parameters are denoted as "fixed". Parameters that are carried forward from the proceeding estimation step are highlighted by checkmarks (✓). *fixed to 1.

| | pre-detection | | pre-symptomatic | | infectious | | post-infectious | |
|---|---|---|---|---|---|---|---|---|
| | $n_1$ | $\tau_1$ | $n_2$ | $\tau_2$ | $n_3$ | $\tau_3$ | $n_4$ | $\tau_4$ |
| Step 1 | estim. | estim. | fixed* | estim. | | | | |
| Step 2 | | | | | estim. | estim. | | |
| Step 3 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | fixed* | estim. |

### SN.2.2 Incubation (time to symptom onset)

We first estimate the parameters of the *incubation* phase based on the cumulative distribution of the time-to-symptom onset for general transmissions reported by Wei et al. [3].

We extract the mean cumulative distribution $y_t$ to which we fit our model in the temporal range $t \in [0, 30]$ days post infection. We then optimise the arguments $n_1$, $\tau_1$ and $\tau_2$ by minimising the squared deviation of our model predictions $p_t(x_j)$ in the following sense:

$$n_1^*, \tau_1^*, \tau_2^* = \underset{n_1,\tau_1,\tau_2}{\operatorname{argmin}} \left( y_t - \left( 1 - \sum_{j=1}^{2} p_t(x_j) \right) \right)^2, \tag{SN.21}$$

with initial conditions $p_{t_0}(x_{1,1}) = 1$ and $p_{t_0}(x_{j,i}) = 0$ for all $(j, i) \neq (1, 1)$. Fixing $n_2 = 1$ (compare Table SN.2), we obtain the optimal parameters: $n_1^* = 5$, $\tau_1^* = 2.86$ and $\tau_2^* = 3.91$ days.

To estimate lower and upper extreme values for $\tau_1$ and $\tau_2$, we fix $n_1 = 5$, as well as the ratio between the residence time of the pre-detection phase and the total incubation period, $\tau_1 = 0.422 \cdot (\tau_1 + \tau_2)$. We then optimize $\tau_1^{\text{upper/lower}}$ and $\tau_2^{\text{upper/lower}}$ for the minimum least squares deviation between our prediction and respectively the lower- and upper bounds reported in Wei et al. [3]. From this we obtain $\tau_1^{\text{upper/lower}} = (2.38, 3.37)$ and $\tau_2^{\text{upper/lower}} = (3.27, 4.62)$ days. The corresponding fits are depicted in Fig. SN.3

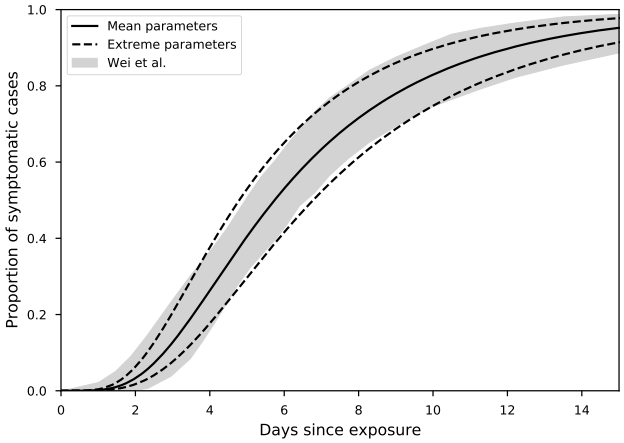


**Fig SN.3. Fitting of the *incubation* period.** Data reported by Wei et al. [3] (meta-analysis of 56 studies) is depicted by the grey shaded areas. Model predictions with optimal parameters $n_1 = 5$, $n_2 = 1$, $\tau_1 = 2.86$ (2.38, 3.37) days and $\tau_2 = 3.91$ (3.27, 4.62) days are shown as black lines: Dashed line: predicted *mean* duration of the incubation period. Dashed line: upper/lower bounds.

### SN.2.3 Infectiousness after symptom onset

Next, we estimate parameters of the *infectious* phase after symptom onset (or peak virus load) from viral load kinetics by Ejima et al., van Kampen et al., Jones et al. [4, 5, 7], *in-house* data and based on the relationship between culture positivity and time since symptom onset reported by Singanaygam et al. [6].

The *in-house* Ct values are transformed to viral kinetics using the methods exemplified in section SN.3. For each data set, we then fit a linear equation to the reported viral kinetics. We first use a sliding window technique (window size of 30 consecutive data points) to extract the average slope of the $\log_{10}$ viral load values to which we then fit a linear equation by minimizing the least squares deviation, similar to the method exemplified in section SN.3. For each fit, the slope $a$ and intercept $b$ are used to simulate viral load data in the temporal range $t \in [0, 21]$ using

$$\log_{10}\big(VL(t)\big) = -t \cdot a + b + \varepsilon, \tag{SN.22}$$

where $\varepsilon \sim \mathcal{N}\big(0, \sigma^2\big)$ is an additive error.

The simulated viral loads in turn are used to construct relative infectivity profiles based on the attack rate curve as exemplified in section SN.3, using the optimal parameters for the attack rate curve $z_t$ reported therein.

We fit the infectious phase of the model to each relative infectivity data set separately. We enforce one global $n_3$ parameter (number of sub-compartments) for all data sets, while allowing distinct values for $\tau_3$ (*mean residence time*) for each individual data set.

Using initial conditions $p_{t_0}(x_{3,1}) = 1$ and $p_{t_0}(x_{j,i}) = 0$ for all $(j, i) \neq (3, 1)$, we minimise the sum of weighted squared deviations:

$$n_3^*, \tau_3^* = \underset{n_3, \tau_3}{\operatorname{argmin}} \quad \sum_S \frac{\big(z_t - p_t(x_3)\big)^2}{N_S}, \tag{SN.23}$$

where $N_S$ is the number of observations in the respective study $S$. We obtain optimal parameters $n_3 = 13$ and $\tau_3 = 7.5$ (taken to be the median of the five fitted values for the individual studies).

To estimate lower and upper extreme values for $\tau_3$, we fix $n_3 = 13$ and use the same method minimizing the least squares deviation for the lower and upper bounds of the error range of all five data

sets combined. From this, we estimate $\tau_3^{\text{upper/lower}} = (2.79, 11.47)$. In Fig. SN.4A, we show the optimisation of parameter $n_3$, whereas Fig. SN.4B depicts the fits to the individual studies and Fig. SN.4C shows the the summary of all data together with model predictions using the default parameters $n_3 = 13$ and $\tau_3 = 7.5$ $(2.79, 11.47)$.

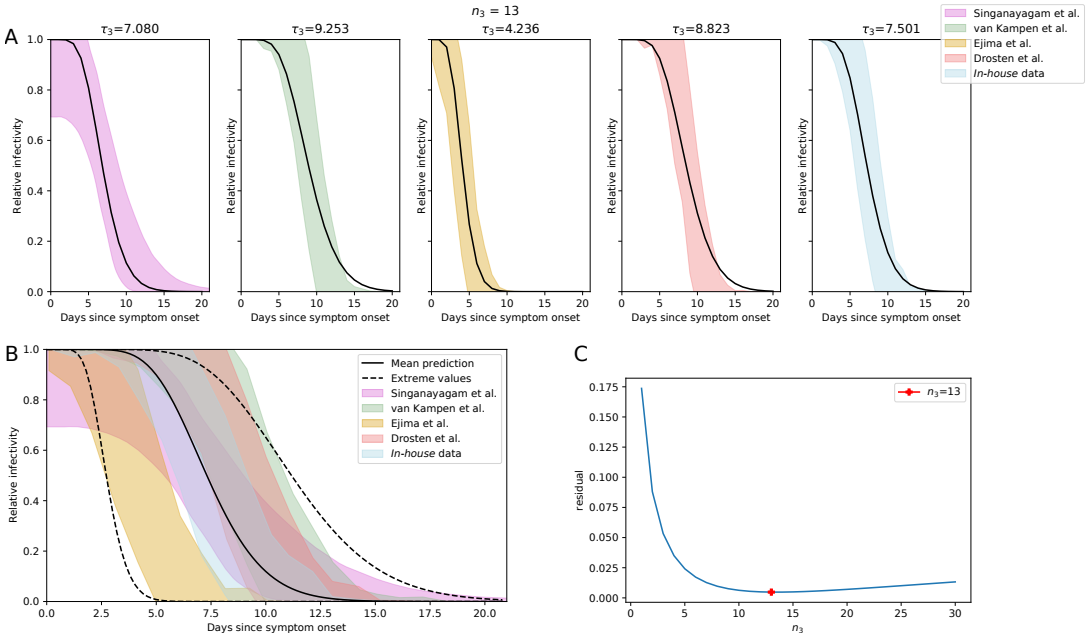


**Fig SN.4. Fitting of the *infectious* phase post symptom-onset. A.** Fit to the individual data sets. Areas denote the reported ranges of the *relative* infectiousness. Lines represent the respective model predictions. **B.** The plot depicts the infectiousness time courses from all data sets used in the estimation procedure as shaded areas. The solid and dashed lines depict model predictions using the final default parameters $n_3 = 13$ and $\tau_3 = 7.50$ $(2.79, 11.47)$ days.**C.** Estimation of the optimal number of sub compartments for the infectious phase. The optimum $n_3 = 13$ is marked by a red cross.

### SN.2.4 Time-dependent assay sensitivity and detection probability: PCR

Currently, PCR-based diagnostic tests are the gold standard to detect a SARS-CoV2 infection. Since the utilised primers are highly specific for SARS-CoV2, we set the default specificity to Spec = 0.999 to account for errors such as mislabeling of samples. The PCR also has an *analytical* sensitivity of nearly 100% if sufficient viral material is contained in the sample. The *clinical* sensitivity, however, depends on time since infection [8] and is capped at a maximum sensitivity

$$\text{Sens}_{\max} = \max_t \big[P_t(\text{PCR}^+|\text{SARSCoV2}^+)\big] \approx 80\%.$$

While the former is a result of the viral dynamics, the latter has to do with the pre-analytics, i.e. whether the health personal is able to get hold of sufficient virus material during swab sampling. Because the software is developed primary for comparing NPI strategies for the public, samples for PCR tests are assumed to be from the upper respiratory tract.

In the software, the maximum sensitivity is an input parameter, while the *mean residence time* in the *post-infectious* phase $\tau_4$ is chosen, such that the temporal change in the *false omission rate*, as reported in Kucirca et al. [8], as well as the decrease of detection probability from symptom onset, as reported in Borremans et al. [9], are captured accurately (compare Fig. SN.5 below).

To estimate the optimal $\tau_4$ value, we fix the optimal parameters for all previous phases and set $n_4 = 1$. The time-dependent assay *false omission rate* is computed as:

$$\text{FOR}_t = \text{Spec} \cdot p_t(x_1) + (1 - \text{Sens}_{\max}) \cdot \sum_{j=2}^{4} p_t(x_j) \tag{SN.24}$$

with $p_{t_0}(x_{1,1}) = 1$ and $p_{t_0}(x_{j,i}) = 0$ for all $(j,i) \neq (1,1)$. The relative detection probability from symptom onset is computed as:

$$P_t(\text{detect}) = \frac{\big(1 - diag\big(\text{FOR}(\mathbf{x})\big)\big) \cdot e^{t \cdot A} \cdot p_{t_0}(\mathbf{x})}{\big(1 - diag\big(\text{FOR}(\mathbf{x})\big)\big) \cdot e^{t_0 \cdot A} \cdot p_{t_0}(\mathbf{x})} \tag{SN.25}$$

with $p_{t_0}(x_{3,1}) = 1$ and $p_{t_0}(x_{j,i}) = 0$ for all $(j,i) \neq (3,1)$. We estimate $\tau_4$ by simultaneously fitting $P_t(\text{detect})$ to the detection probability profile for the upper respiratory tract from Borremans et al. [9], and $\text{FOR}_t$ to the temporal PCR *false omission rate* profile published by Kucruca et al. [8]. This gives us $\tau_4 = 8$ days. We fit no upper and lower bounds for $\tau_4$.

The model predicted $\text{FOR}_t$ and $P_t(\text{detect})$ together with the published data from Kucirca et al. [8], respectively Borremans et al. [9] is shown in Fig. SN.5.

### SN.2.5 Antigen-based rapid diagnostic tests (RDT)

Rapid diagnostic testing (RDT) through antigen detection is currently in development. Early validation results from two commercially available products compared their analytic performance with PCR [10]. In summary, the antigen tests show a comparable sensitivity with respect to PCR for samples with low CT values (large number of virus material) and appear to be less sensitive at high CT values (low virus content in sample). Overall, sensitivities of $P(\text{RDT}^+|\text{PCR}^+) = 85$–$89\%$ with respect to PCR were reported for the two evaluated testing systems. Specificity was 99.7–100% with respect to PCR. Since no data about the temporal changes of this relative sensitivity is available to date, we assumed it to be comparable to the PCR and set the default parameter for the maximum sensitivity of RDT assays to

$$\begin{aligned} \text{Sens}_{\max} = \max_t \big[P_t(\text{RDT}^+|\text{SARSCoV2}^+)\big] &= \max_t \big[P(\text{RDT}^+|\text{PCR}^+) \cdot P_t(\text{PCR}^+|\text{SARSCoV2}^+) \\ &\quad + P(\text{RDT}^+|\text{PCR}^-) \cdot P_t(\text{PCR}^-|\text{SARSCoV2}^+)\big] \\ &\approx 70\% \end{aligned} \tag{SN.26}$$

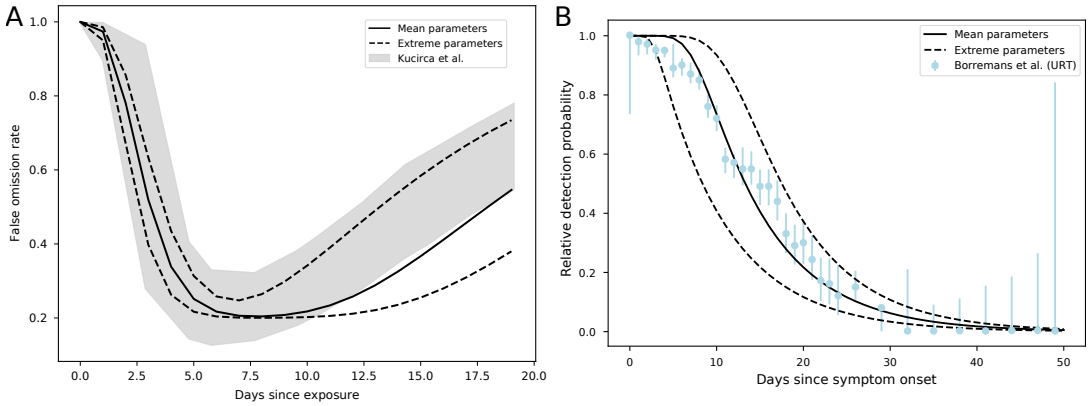


**Fig SN.5. Fitting of the temporal PCR sensitivity and detection probability**. **A.** The grey ranges show the temporal *false omission rate* derived by Kucirca et al. [8], whereas the solid and dashed lines indicate the predicted profile using the mean- and upper/lower parameter estimate. **B** Time-dependent PCR sensitivity after symptom onset from Borremans et al. [9] (error bars) together with model simulated PCR sensitivity with optimal parameters (solid line) and extreme parameters (dashed lines).

### SN.2.6 Summary of default parameters

While residence times in each phase $\tau_j$ can be changed arbitrarily in the software, the number of sub-compartments $n_j$ for each phase remains fixed. Table SN.3 summarises all model default parameters of the model.

**Table SN.3.** Summary of the model's default parameters. $\tau_j$ and $n_j$ denote the *mean residence time* (and interval) in phase $j$ in days and the number of sub-compartments respectively.

| pre-detection | | pre-symptomatic | | infectious | | post-infectious | |
|---|---|---|---|---|---|---|---|
| $\tau_1$ | $n_1$ | $\tau_2$ | $n_2$ | $\tau_3$ | $n_3$ | $\tau_4$ | $n_4$ |
| 2.86 (2.38;3.37) | 5 | 3.91 (3.27; 4.62) | 1 | 7.5 (2.79; 11.47) | 13 | 8 | 1 |

## SN.3 Analysis of infectivity profiles

We will first pursue a mechanistic modelling which allows to put viral kinetics, as well as immune-related virus neutralisation into context. We fit this model to *in house* data that reports viral kinetics, as well as culture-positivity. Using the mechanistic modelling, we are then enabled to deduce an attack rate curve $z_t$ from viral load profiles, for studies that do not report infectivity explicitly (inputing some parameters learned from the *in house* data).

The mechanistic model is then used to derive the attack rate curve $z_t$ for each considered data set where this curve is not explicitly stated and used for deriving default parameters for our Markov model (Fig. SN.1) as outlined in section SN.2.

### SN.3.1 Mechanistic modelling

We will model the attack rate as a function of the viral kinetics in the potential transmitter, as well as virus neutralisation within the host. Under the reasonable assumption of statistical independence the probability of infection (attack rate) can be written as

$$z_t \quad = \quad 1 - c_1^{\left(V_{eI}(t)\right)} \tag{SN.27}$$

where $0 << c_1 < 1$ is the probability of non-infection after exposure to a single infectious virus. The exposure with *infectious* virus $V_{eI}(t)$ can be thought of as a Bernulli process (compare below) and hence is a binomially distributed random number. The expected exposure to *infectious* virus is then $\mathbb{E}\big(V_{eI}(t)\big) = F_I(t) \cdot V_e(t)$, which we will use henceforth.

#### SN.3.1.1 Virus exposure $V_e(t)$

The virus exposure $V_e(t)$ can be assumed to be a function of the viral load $VL(t)$ in the upper respiratory tract of an infected individual. Viral kinetics have been elaborated in several studies, e.g. [4, 7]. Typically, the viral titers increase exponentially, reaching a set point at about symptom onset and decrease exponentially thereafter. Viral exposure $V_e(t)$ emanating from an infected individual can be thought of as being a fraction of the viral load in the upper respiratory tract. This can be modelled, akin to [11], as a Bernulli Process, hence,

$$V_e(t) \sim \mathcal{B}(p, VL(t)) \tag{SN.28}$$

where $\mathcal{B}$ is the binomial distribution and $p$ the success probability (fraction of virus load that is exposed to the recipient). The binomial distribution has expectation value $\mathbb{E}\big(V_e(t)\big) = p \cdot VL(t)$, which we will use henceforth.

#### SN.3.1.2 Fraction of infectious virus $F_I(t)$

It has been shown in several studies [5, 6, 12] that the infectiousness of virus from patient samples decreases as a function of the time since symptom onset. This is generally believed to be a result of neutralization by antibodies [13, 14]. Notably, neutralisation in the upper respiratory tract may occur before it is detectable in the blood plasma. Let us assume that the immune system response $IR(t)$ increases exponentially after infection. Then, the fraction of infectious virus $F_I(t)$ can be modelled by Emax kinetics [15]

$$F_I(t) = \frac{1}{1 + \widehat{IR}(t)} \tag{SN.29}$$

where $\widehat{IR}(t)$ is proportional to the immune response. An interpretation would be $\widehat{IR}(t) = Ab(t)/\mathrm{IC}_{50}$ (concentration of neutralising antibodies divided by their fifty percent inhibitory concentration).

### SN.3.2 The shape of the attack rate curve

In case of an exponential increase of the immune response, $\widehat{IR}(t_e) = \widehat{IR}(t_{0,e}) \cdot e^{\left(t_e \cdot c_3\right)}$, where $t_{0,e}$ is the time *of* exposure and $t_e$ is the time *after* exposure and hence $t_e = t + \tau_{\text{inc}}$.

$$z_t = 1 - c_1^{\left(V_{eI}(t)\right)} = 1 - c_1^{p \cdot VL(t) \cdot F_I(t)} \tag{SN.30}$$

$$z_t = 1 - c_1^{p\left[\frac{VL(t)}{1+\left(\widehat{IR}(t_{0,e}) \cdot e^{((t+\tau_{\text{inc}}) \cdot c_3)}\right)}\right]} \tag{SN.31}$$

$$= 1 - c_1^{p\left[\frac{VL(t)}{1+g(c_3) \cdot e^{(t \cdot c_3)}}\right]} \tag{SN.32}$$

where $g(c_3) = \left(\widehat{IR}(t_0^e)\right) \cdot e^{(\tau_{\text{inc}} \cdot c_3)}$.

### SN.3.3 Estimation of infectivity profiles from *in house* data

#### SN.3.3.1 Viral dynamics

We first estimate the viral decay kinetics based on the Ct values. The data is depicted below in Fig. SN.6, left. We first use a sliding window technique to extract the average slope of the Ct values (blue line) to which we then fit a linear equation in the temporal range $t \in [0\ \ 20]$ by minimizing the least squares deviation.

$$y_t = m \cdot t + y_0 \tag{SN.33}$$

$$\text{with} \tag{SN.34}$$

$$Ct(t) = y_t + \varepsilon \tag{SN.35}$$

with optimal parameter $m^* = 0.43$. We assumed an additive error (which justified the least squares regression) and estimated $\varepsilon \sim \mathcal{N}\left(0, \sigma^2\right)$, where $\sigma^2 = 4.44$. The resulting simulated Ct values are depicted in Fig. SN.6, middle.

As illustrated by [16], Ct values are linearly correlated with $\log_{10}$ viral loads (genome equivalents/$\mu$L), i.e.:

$$-c \cdot \log_{10}(VL(t)) + k = Ct(t) \tag{SN.36}$$

$$\log_{10}(VL(t)) = \frac{1}{c} Ct(t) + \frac{k}{c} \tag{SN.37}$$

For the *in house* PCR assay we have $c = 3.35$ for the Charité E protein primers used in our study, Fig. SN.1 and therefore $\frac{1}{c}$ is very close to the idealised value of $\frac{1}{\log_2(10)} \approx 0.3$.

Finally, the slope parameter for the viral loads (combining the Ct(t) slope and the linear regression parameter) is $\tilde{m} = m/c = 0.133$ (day$^{-1}$). Similarly, the variance gets scaled to $\tilde{\sigma}^2 = 4.44/c = 1.32$, now being an exponential error, i.e. $\epsilon \sim \mathcal{N}\left(0, 1.32\right)$. The actual viral loads depend on the intercept, which in turn depends on the extraction of viral samples, as well as on the assay equipment. We therefore chose to set the average viral loads to published values [12], i.e. $\mathbb{E}\left(VL(t_0)\right) \approx 6 \cdot 10^7$ (copies/swab)$\Rightarrow \tilde{y}_0 = 7.77$.

#### SN.3.3.2 Attack rate

We use the equation for the attack rate (compare eq.(SN.32)).

$$z_t = 1 - c_1^{p\left[\frac{VL(t)}{1+g(c_3) \cdot e^{(t \cdot c_3)}}\right]} \tag{SN.38}$$

where $g(c_3) = \widehat{IR}(t_0^e) \cdot e^{(\tau_{\text{inc}} \cdot c_3)}$. We will set $\widehat{IR}(t_0^e) = 0.01$.

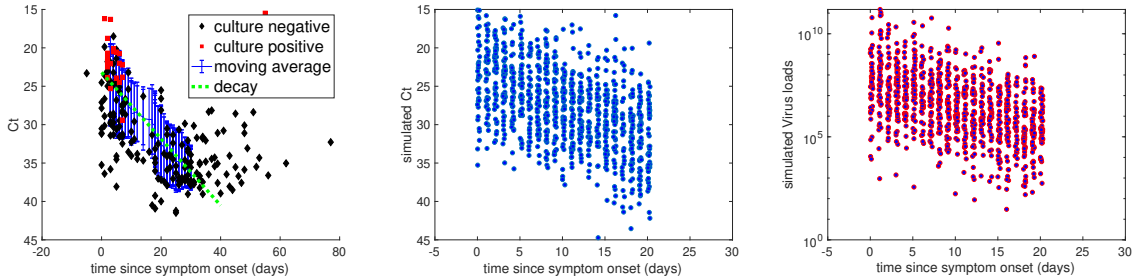


**Fig SN.6. Own data and data fitting. Left:** Ct values from Sars-CoV2 infected individuals after symptom onset. black dots denote culture negative- and red dots culture positive samples. The solid blue line denotes the sliding average and and the dashed green line is a fitted increase in Ct with slope 0.4 (day$^{-1}$) and intercept 24. **Middle:** Simulated Ct values. **Right:** Simulated genome equivalents/mL.

This leaves us with three free parameters: $10^{-6} < p < 0.01$ and $0.01 < c_3 < 3$ and $0.99 \leq c_1 < 0.9999$. We now minimize

$$\{p^*, c_3^*, c_1^*\} = \text{argmin}\Big(\big(\psi \cdot w - z_t\big)^2\Big), \tag{SN.39}$$

where $\psi \in [0, 1]$ denotes the data (culture negative or culture positive). We assessed different values for the hyperparameter (weight) $w$, see below in Fig. SN.7. The weighing parameter $w$ is intended to put more importance on the culture positive samples, as there may be false negative cultures due to transportation and storage of samples. This also puts more emphasis on increasing the sensitivity of the method (less false negative predictions), as it increases the methods' safety margin. After hyper parameter scan we selected $w = 3$.

#### SN.3.3.3 Summary and optimal parameters.

Using the methods described above, we get the following optimal parameters: $p = 0.01$, $c_3 = 0.741$, $c_1 = 0.99$. Using these values the model has a sensitivity of $sens = 96\%$ and a specificity of $spec = 0.57\%$.

Sensitivity was computed as:

$$sens. = P(pred.\ positive|culture\ positive) = \frac{A(t)|\text{culture pos.}}{\#\text{culture pos.}} \tag{SN.40}$$

Specificity was computed using

$$spec. = P(pred.\ negative|culture\ negative) = 1 - \frac{A(t)|\text{culture neg.}}{\#\text{culture neg.}} \tag{SN.41}$$

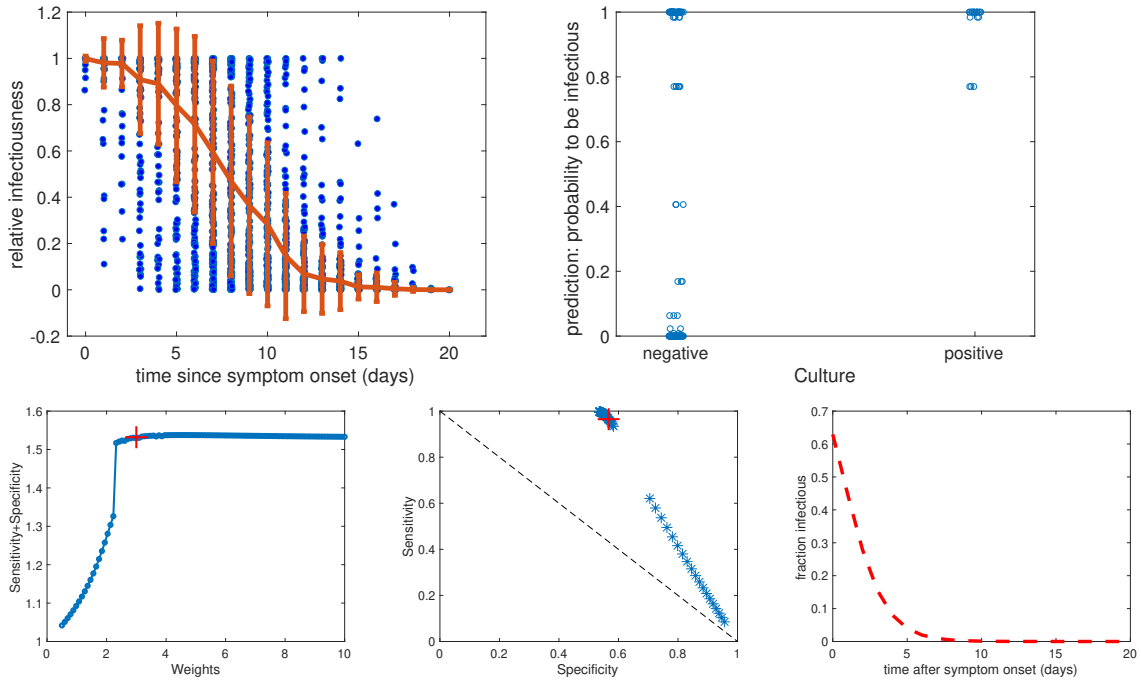


**Fig SN.7. Fitting to *in house* data. Upper Left:** Predicted relative infectivity based on sampled viral loads (compare Fig. SN.6 and estimated parameters (eq. (SN.39)). Blue dots: Samples infectivity values, red line and error bars: mean infectivity ± standard deviation. **Upper Right:** Scanning of hyper parameters (weights). Red cross: chosen weight. **Lower Left:** Sensitivity and specificity for different hyper parameters. **Lower Middle:** Predicted attack rate vs. positive culture. **Lower Right:** Predicted fraction of infectious virus, using $\frac{1}{1+g(c_3)\cdot e^{(t\cdot c_3)}}$.

# References


1. Duwal S, Dickinson L, Khoo S, von Kleist M. Hybrid stochastic framework predicts efficacy of prophylaxis against HIV: An example with different dolutegravir prophylaxis schemes. PLoS Comput Biol. 2018;14(6):e1006155. doi:10.1371/journal.pcbi.1006155.

2. Duwal S, Dickinson L, Khoo S, von Kleist M. Mechanistic framework predicts drug-class specific utility of antiretrovirals for HIV prophylaxis. PLoS Comput Biol. 2019;15(1):e1006740. doi:10.1371/journal.pcbi.1006740.

3. Wei Y, Wei L, Liu Y, Huang L, Shen S, Zhang R, et al. A systematic review and meta-analysis reveals long and dispersive incubation period of COVID-19. MedRXIV (`https://doiorg/101101/2020062020134387`). 2020;.

4. Ejima K, Kim KS, Ito Y, Iwanami S, Ohashi H, Watashi YKA, et al. Inferring Timing of Infection Using Within-host SARS-CoV-2 Infection Dynamics Model: Are "Imported Case" Truly Imported. medRxiv (`https://doiorg/101101/2020033020040519`). 2020;.

5. van Kampen JJA, van de Vijver DAMC, Fraaij PLA, Haagmans BL, Lamers MM, Okba N, et al. Shedding of infectious virus in hospitalized patients with coronavirus disease-2019 (COVID-19): duration and key determinants. MedRxiv (`https://doiorg/101101/2020060820125310`). 2020;.

6. Singanayagam A, Patel M, Charlett A, Lopez Bernal J, Saliba V, Ellis J, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. Euro Surveill. 2020;25(32). doi:10.2807/1560-7917.ES.2020.25.32.2001483.

7. Jones TC, Biele G, Mühlemann B, Veith T, Schwarzer R, Zuchowski M, et al. Analysis of SARS-CoV-2 viral load and infectivity from 9009 RT-PCR-positive cases in Germany. submitted. 2020;.

8. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. Ann Intern Med. 2020;173(4):262–267. doi:10.7326/M20-1495.

9. Borremans B, Gamble A, Prager KC, Helman SK, McClain AM, Cox C, et al. Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2 infection by time since symptom onset. Elife. 2020;9. doi:10.7554/eLife.60122.

10. Kaiser L, Eckerle I, Schibler M, Berger A, RDT Study Team. Validation Report: SARS-CoV2 Antigen Rapid Diagnostic Test. Universite de Geneve, Hospitaux Universitaires Geneve; 2020 (accessible at: `https://www.hug.ch/sites/interhug/files/structures/laboratoire_de_virologie/documents/Centre_maladies_virales_infectieuses/ofsp_rdt_report_gcevd_27.10.2020.pdf`).

11. Duwal S, Sunkara V, von Kleist M. Multiscale Systems-Pharmacology Pipeline to Assess the Prophylactic Efficacy of NRTIs Against HIV-1. CPT Pharmacometrics Syst Pharmacol. 2016;5(7):377–87. doi:10.1002/psp4.12095.

12. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020;581(7809):465–469. doi:10.1038/s41586-020-2196-x.

13. Meyer B, Reimerink J, Torriani G, Brouwer F, Godeke GJ, Yerly S, et al. Validation and clinical evaluation of a SARS-CoV-2 surrogate virus neutralisation test (sVNT). Emerg Microbes Infect. 2020;9(1):2394–2403. doi:10.1080/22221751.2020.1835448.

14. Liu L, To KKW, Chan KH, Wong YC, Zhou R, Kwan KY, et al. High neutralizing antibody titer in intensive care unit patients with COVID-19. Emerg Microbes Infect. 2020;9(1):1664–1670. doi:10.1080/22221751.2020.1791738.

15. Shen L, Peterson S, Sedaghat AR, McMahon MA, Callender M, Zhang H, et al. Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs. Nat Med. 2008;14(7):762–6. doi:10.1038/nm1777.

16. Vogels CBF, Brito AF, Wyllie AL, Fauver JR, Ott IM, Kalinich CC, et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. Nat Microbiol. 2020;5(10):1299–1305. doi:10.1038/s41564-020-0761-6.