A HAWKES PROCESS MODEL FOR THE PROPAGATION OF COVID-19: SIMPLE ANALYTICAL RESULTS

SUPPLEMENTAL MATERIAL

Juan Valentin Escobar[1]

[1] Instituto de Física, Departamento de Física Química, Universidad Nacional Autónoma de México, 04510 Coyoacán, Mexico City, Mexico

S1. METHODS

S1.1 Data collection. Daily new cases and new deaths were obtained from the website https://www.worldometers.info/coronavirus/. Daily counts were aggregated weekly starting from February 15th 2020, so that, for example, week #0 comprises the daily data from Feb 15th to Feb. 21st 2020.

S1.2 Estimation of γ and ρ from clinical trials. To be able to estimate the number of latent cases as well as to lay out different scenarios, we first need to know the constants ρ and γ. We do this by first considering the evolution of the patients-zero in the absence of contagion. The probability of being registered after a number t of weeks has elapsed since this population was introduced, is given by $\lambda\left(t\middle|\beta_k=0\right)=\alpha\gamma^t=(1-\gamma-\rho)\gamma^t$. From this equation, the percentage of asymptomatic cases is $A\equiv 1-\sum_{t=0}^{\infty}\left(1-\gamma-\rho\right)\gamma^t=\rho/(1-\gamma)$. Then, if the possibility of having asymptomatic cases is precluded (as in the case of clinical tests with isolated infected patients), the naturally normalized bare memory kernel is $\varphi(t)\equiv\lambda\left(t\middle|\beta_k=0,\rho=0\right)$, or

$$\varphi(t)=(1-\gamma)\gamma^t. \tag{S1}$$

The average time elapsed from contagion to being registered is then $T \equiv \sum_{t=0}^{\infty} \varphi_t = \gamma/(1-\gamma)$. Independent clinical studies on isolated infected patients have found that T = 13 days ≃ 2 weeks [1], which contemplates the fact that, on average, it takes people 2 days after developing dyspnea to go to a hospital. Since in our model t = 0 is the first week after exposure (see Fig. 2), this period corresponds to $T \simeq 1$, yielding γ = 0.5. Regarding the rate of asymptomatic cases, this is a topic of current debate [2,3]. But investigations on isolated patients over long periods of time have found $A \simeq 20\%$ to be a conservative estimate [4]. We use this value in the present study, but other scenarios could be easily explored by modifying it. Along with the estimation for γ made above, this implies ρ=0.1. Using clinical data to fix γ and ρ as we have just done implies that the only free parameter of the model is the number of patients-zero, λ_{-1}. The simplicity of our model carries with it another important condition that must be consistent with independent clinical results for COVID-19. Given that the γ and ρ branches are symmetric in the diagram of Fig. 2, the latency function for new registered cases is identical to the latency function for recovered ones. Consequently, the average time from infection to recovery (T_ρ) must equal T. A recent study [5] has found that the probability of being infectious at time t is given by a gamma distribution function with an average close to 2 weeks. While other studies have found that the period during which virus shedding lasts is longer than this value, its magnitude is not enough for contagion to continue [6]. Once again, this average of two weeks corresponds to T_ρ=1 in our model, which equals T, as required. Note that we are requiring these average values to coincide and not the full probability functions themselves, which would be a much stronger condition. The only distribution function we do require to equal the observed data is the activity λ_t during the relaxation period with constant β (Eq. 5, insets of Figs. 1a and 1b). Here we have made the simplifying assumption that symptomatic and asymptomatic clinical features are similar, even though some differences have been found for the characteristics of viral shedding in these populations [7]. We emphasize the fact that the condition $T = T_\rho$ is a consequence of the simplicity of our model that may not be satisfied by other infectious diseases.

SI.3 Estimation of ε from clinical trials. In the absence of contagion, the number of symptomatic (and thus, registered) recovered patients *vs.* time is given by $\lambda_{R,t} = \alpha\delta^t\varepsilon$ (see section S2). From the average time for a symptomatic patient to recover (T_δ), we obtain $\delta = T_\delta / (1+T_\delta)$. T_δ can be estimated again from the average time it takes for an individual to stop being infectious (close to 2 weeks), which translates into $T_\delta \simeq 1$, yielding δ = 0.5. Note that $T_\delta = T_\rho$. In other words, in our model both asymptomatic and symptomatic cases take same time to recover (on average), which is one of the simplifying assumptions we make, as stated in the previous subsection.

S2. Branching process for the number of new deaths.

Here we present in detail the branching process proposed that leads to the number of weekly new deaths of Eq. 7 of the main text from the weekly new cases:

$$\lambda_{D,t} = \lambda_{-1} P_t \left(1-\gamma-\rho\right) \Delta \sum_{i=0}^{t-1} \delta^i \prod_{k=0}^{t-i-2} \left(\gamma+\beta_k\right) \qquad (7, \text{S2})$$

Following the diagram presented in Fig. S1, we assume that a fraction ε of the registered population (normalized in this diagram) at time t will recover on week t + 1, while a fraction δ_t will continue to be ill. Then, the fraction that will die on week t + 1 is $1-\delta_t-\varepsilon \equiv \Delta_t$. The elements inside the blue circles at each time represent the fraction of the population that will die at time t from a single source of registered individuals at time t = 1. As shown in the main text, the number registered cases (and thus, the number of sources of new deaths) is given by:

$$\lambda_t = \lambda_{-1}\left(1-\gamma-\rho\right) \prod_{k=0}^{t-1} \left(\gamma+\beta_k\right) \qquad (2, \text{S3})$$

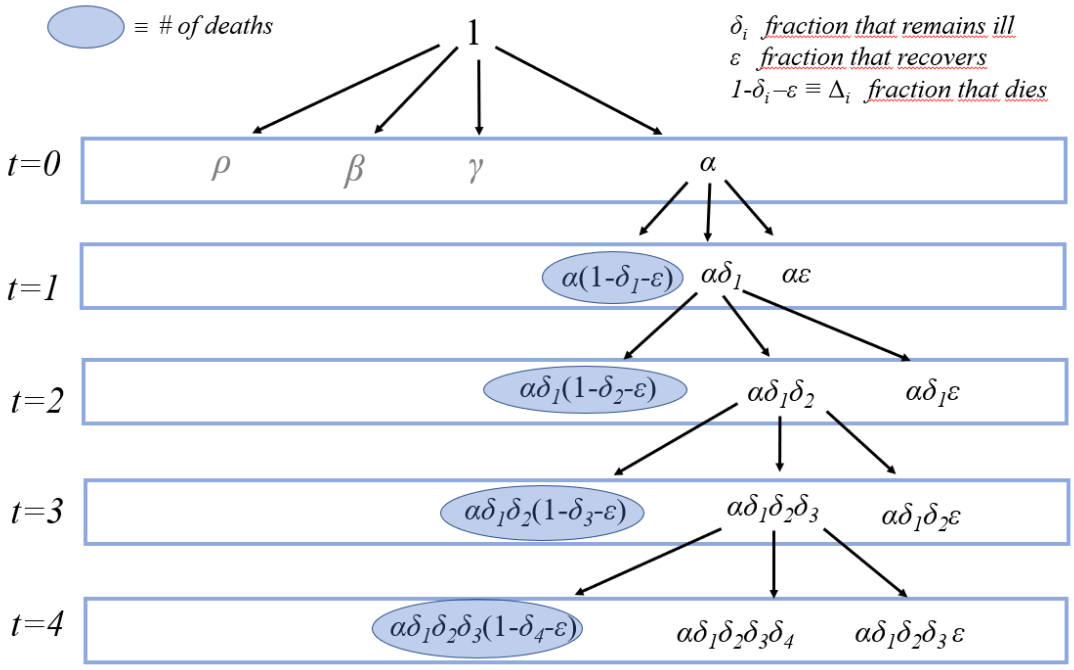

Fig. S1 Branching process that yields the number of new deaths from a single source of α registered cases as function of time from a normalized number of patients-zero.

Each element of the activity obtained with this equation will follow the same branching process as the one depicted in Fig. 1S. In Fig. 2S, we show explicitly this process up to t = 3, where the sum of every element at a given time (a given row) gives the number of deaths.

The pattern observed in Fig. S2 can be recovered with the set of equations:

$$F_{\tau,i} = \prod_{j=0}^{i-1} \delta_{\tau-j} \,, \tag{S4}$$

$$G_i = \prod_{j=0}^{i-1} \left(\gamma + \beta_j\right), \tag{S5}$$

$$\lambda_{D,t} = \Delta_t \left(\sum_{i=0}^{t-1} F_{t-1,i} G_{t-i-1} \right) . \qquad \text{(S6)}$$

If δ_k equals a constant δ for all times, Eq. S6 reduces to Eq. 7 of the main text (Eq. S2 above).

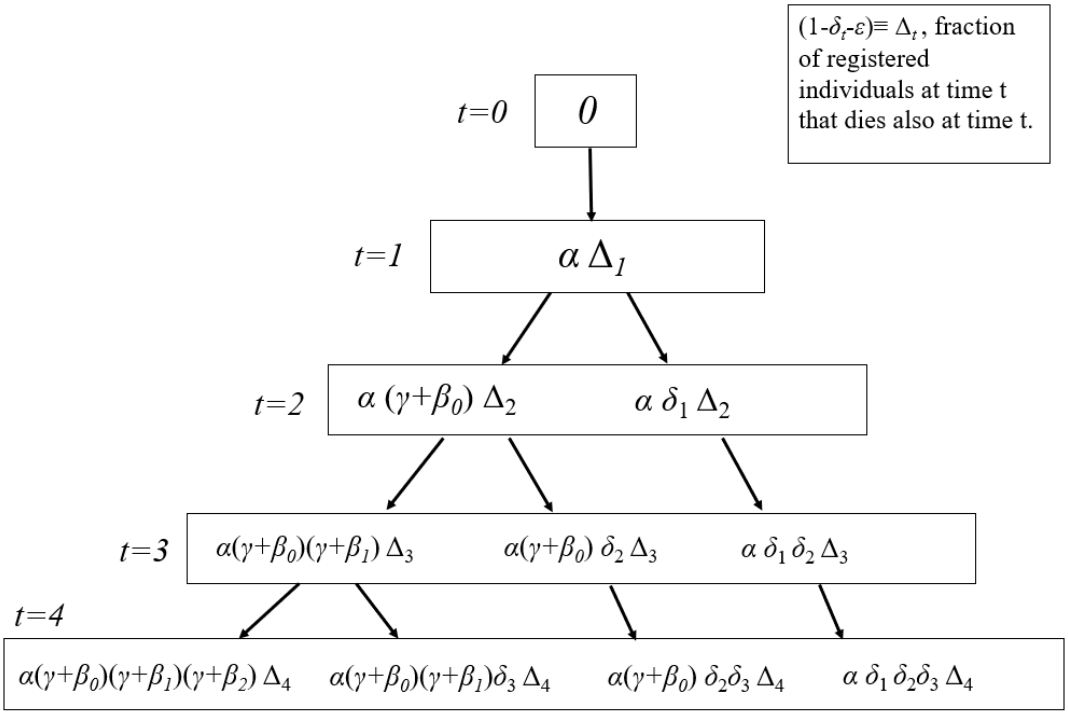

Fig. S2 Branching process yielding the number of new deaths $\lambda_{D,t}$ from all the sources generated from the total number of new registered cases given by λ_t. The sum of all elements for a given time (row) gives $\lambda_{D,t}$.

As mentioned in the Methods section, the number of recovered isolated patients (γ=0, β_k=0) at time t is obtained by adding all elements in the middle of each row of Fig. S1, yielding.

$$\lambda_{R,t} = \alpha \delta^t \varepsilon . \qquad \text{(S6)}$$

The average time to recover is then,

$$T_R = \sum_{t=0}^{\infty} \left(\frac{\delta^t}{1-\delta} \right) t = \frac{\delta}{1-\delta}, \tag{S7}$$

from which we obtain

$$\delta = \frac{T_R}{1+T_R}. \tag{S8}$$

S3. Parameters of the model of Fig. 4

As explained in the main text, the relation between the percentage of positive tests $P(t)$ and the percentage of registered cases that will be symptomatic is not known. However, as we argued, if P_t=1 then it is reasonable to assume that close to 100% of the tested people will present symptoms. Also, we expect that this percentage will be a monotonously increasing function of P_t. A simple relation that satisfies these conditions is:

$$\lambda_{D,t} \propto P_t^a . \tag{S9}$$

where a is some constant. The following figure shows three different realizations of Eq. 7 (which incorporates Eq. S9), for a=1/3 (blue trace), a=2/3 (black trace), a=3/4 (orange trace) and a=1 (green trace). The values of ε where chosen in each case as to best fit the data up to week #15. As can be seen in this figure, the parameter a=2/3 gives the best approximation to the data. While other powers improve slightly the fit, we stick to this rational number of simple form for the sake of clarity

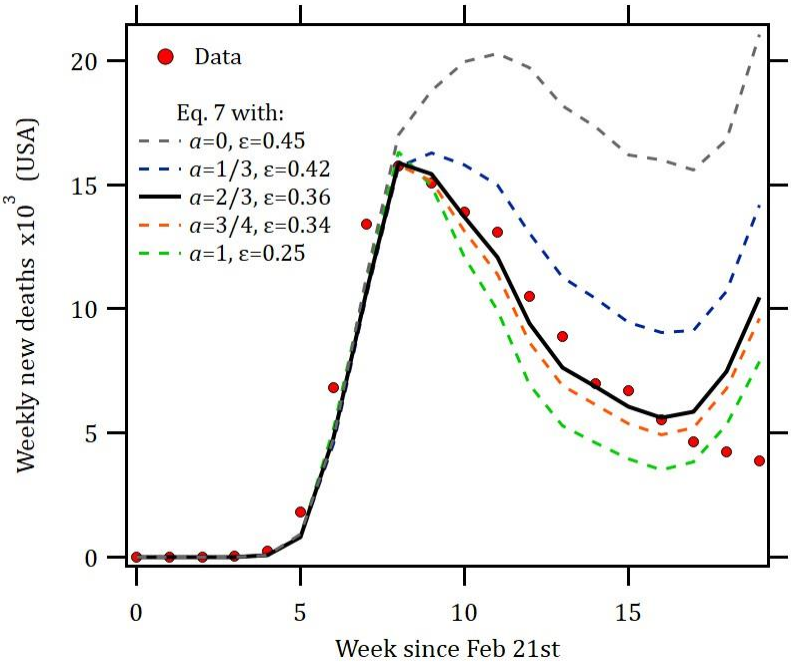

Fig. S3 Data for the weekly number of new deaths un the USA (circles), and realizations of Eq. 7 (S2) for the powers: a=0 (gray trace), a=1/3 (blue trace), a=2/3 (black trace), a=3/4 (orange trace) and a=1 (green trace). The case a=0 represents a constant death-rate from the number of new cases.

REFERENCES

[1] WANG D. W. *et al.*, Jama-Journal of the American Medical Association **323**, 1061 (2020). https://doi.org/doi:10.1001/jama.2020.1585

[2] DAY M., Bmj-British Medical Journal **368**, m1165 (2020). https://doi.org/10.1136/bmj.m1165

[3] PARK M. et al, Journal of Clinical Medicine **9**, 967 (2020). https://pubmed.ncbi.nlm.nih.gov/32244365/

[4] MIZUMOTO K., et al, Eurosurveillance **25**, 2, 2000180 (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180

[5] LONG Q. X. *et al.*, Nature Medicine. https://doi.org/10.1038/s41591-020-0869-5

[6] WOLFEL R. *et al.*, Nature **581** 7809 (2020). https://doi.org/10.1038/s41586-020-2196-x

[7] LONG Q. X. *et al.*, Nature Medicine Jun 18:1-5 (2020). https://doi.org/10.1038/s41591-020-0965-6