COVID-19 epidemic: from data to mathematical models. Comment on "Data-driven mathematical modeling approaching for COVID-19: A survey" by Jacques Demongeot and Pierre Magal

François Hamel

Comment

COVID-19 epidemic: from data to mathematical models
Comment on "Data-driven mathematical modeling approaching for COVID-19: A survey"
by Jacques Demongeot and Pierre Magal

François Hamel [a]

[a] *Aix Marseille Univ, CNRS, I2M, Marseille, France*

In their extensive survey [5], based on the analysis of an impressive list of more than 250 references, the authors J. Demongeot and P. Magal review the most used mathematical models to describe the various phases of the COVID-19 waves in many places of the world. The COVID-19 outbreak is reminiscent of some old epidemics of infectious diseases, such as the plague epidemic in Europe in the 14th century, the London cholera epidemic in the 17th century, and the smallpox epidemic in the 18th century, which led Bernoulli and d'Alembert to create the first models for the prediction of further epidemic waves [1]. Epidemics are essentially made up of two distinct phases: an epidemic phase in which infections and cases first grow exponentially fast, and an endemic phase in which new cases are maintained at a low level.

Beyond the medical and social issues, modeling the COVID-19 epidemic faces many problems and questions: first of all, how to detect all cases, that is, from data of reported cases, how to estimate the total number of cases, including the unreported ones? The strong variations in the number of tests showed that the number of cases was strongly underestimated in the spring of 2020. The cumulative number of cases has been estimated following a flowchart of the model used in [8], based on the estimation of the proportions of asymptomatic infectious, undetected infectious, detectable infectious, reported, or removed, inside the exposed population.

Two main classes of mathematical models have been used and are reported in this survey. The first class is that of phenomenological models, which try to reproduce the data and their tendency with a limited number of parameters and without any differential equation. Such models were used for the first time by Lambert [9, 15], who took the exponential approximation of the initial growth of an epidemic in London in the middle of the 18th century. The other models are the mechanistic epidemic models, initiated by Bernoulli and d'Alembert [1]. They consist of differential equations modeling the evolution, as time t runs, of the number of susceptible $S(t)$ and infectious $I(t)$ inside the population. The most basic SI model is written as

$$\left\{ \begin{array}{rcl} S'(t) & = & -\tau(t)\, S(t)\, I(t), \\ I'(t) & = & \tau(t)\, S(t)\, I(t) - \nu\, I(t), \end{array} \right.$$

for $t \geq t_0$, with initial conditions $S(t_0) = S_0$ and $I(t_0) = I_0$ at time t_0. Here, $\tau(t)$ is the rate of transmission and ν is the inverse of the average duration of the asymptomatic infectious period. The cumulative number of reported cases, called $CR(t)$, then satisfies $CR'(t) = f\,\nu\, I(t)$, where $f \in [0, 1]$ is the fraction of reported individuals (mainly those with severe symptoms). For instance, $1/\nu$ is equal to a few days and $f = 0.9$, from clinical information.

At the beginning of the epidemic, the number of susceptibles and the transmission rate can be assumed to be nearly constant, equal to S_0 and τ_0, so the number of infectious follows an exponential growth $I(t) = I_0\, e^{\chi_2 (t-t_0)}$, where $\chi_2 = \tau_0 S_0 - \nu$, while $CR(t) = \chi_1\, (e^{\chi_2 (t-t_0)} - 1) + \chi_3$. The parameters χ_1 and χ_3 are evaluated from the data, and then give I_0 and t_0. This strategy has been applied to COVID-19 in mainland China [4]. However, after an early phase, the data and the exponential growth diverge and the transmission rate $\tau(t)$ can no longer assumed to be constant. This rate essentially depends on three factors: the coefficient of virulence of the infectious agent, the coefficient of susceptibility of the host, and the number of contacts per time unit between individuals [10, 12]. All these factors depend on time due to possible mutation of the infectious agent, the innate or acquired varying immunity of the host, and the mitigation measures and social behavior. They may also depend on the location, because of differences in environnemental and meteorological conditions and because of the heterogeneity of the population density. From the integration of the S and I equations, a theoretical formula for the transmission rate $\tau(t)$ can be derived, in terms of the cumulative number of reported cases $CR(t)$ and its first and second order derivatives, together with the parameters ν and f. A central question then arises: how to estimate the average duration of the asymptomatic infectious period $1/\nu$? Several estimates have been given [2]. It turns out that only

an upper bound (3.3 days in mainland China) can be shown by using SI models, since different values fit exactly to the data. Another strategy consists in assuming that the cumulative number of reported cases $CR(t)$ solves a Bernoulli-Verhulst [1, 14] model or a more general logistic equation [13], such as:

$$CR'(t) = \chi_2\, CR(t)\left(1 - \left(\frac{CR(t)}{CR_\infty}\right)^\theta\right).$$

Such equations give explicit formulas for $CR(t)$ and then for $\tau(t)$. The data can then be used to estimate the parameters χ_2, θ, $CR(t_0)$ and CR_∞ (for instance, $\chi_2 = 0.66$ and $\theta = 0.22$ in COVID-19 in China).

The modeling of multiple epidemic waves is also discussed and especially that of the regularization of the junction between epidemic and endemic phases. The best fit of the phenomenological model of [6] to the data of the cumulative reported cases of the COVID-19 epidemics in France are presented in [5]. The determination of the transmission rate $\tau(t)$ is used to calculate two other important quantities, namely the instantaneous reproduction number $R_e(t) = \tau(t)S(t)/\nu$ and the quasi-instantaneous reproduction number $R_e^0(t) = \tau(t)S_0/\nu$. Comparing these numbers with 1 indicates whether the epidemic tends to persist or die out in the long term. The time evolution of these parameters in various countries on three continents is presented in [5], after [3, 7, 11], showing also the necessity and the difficulty of regularizing the data. Other mechanistic models with more compartments are described in [5], including one with unreported infectious cases and a class of models with age groups with different transmission rates. The proof of the existence and uniqueness of optimal transmission rates minimizing the L^2 errors between the data and the model is given in [5].

The COVID-19 pandemic has led to an enormous amount of publications, of the order to 700.000 between early 2020 and August 2023! The authors list in [5] some topics and references for further reading, such as the dynamics of clusters, the analysis of other data (contact tracing, testing, wasted water), SIR models with additional asymptomatic, non-reported, vaccinated, or deceased patients, multigroups and multiscale models, re-infections, natural and hybrid immunity, vaccination and mitigation measures for the prediction of further evolution.

Everyone who wants an up-to-date and comparative analysis of the mathematical models of COVID-19 and the derivation of the parameters from the data, or who is interested in finding out more from a historical point of view, will certainly find matter for further thought in the excellent and thorough review [5].

Acknowledgement. This work has been supported by the French Agence Nationale de la Recherche (ANR), in the framework of the ReaCh project (ANR-23-CE40-0023-02).

References

[1] Bernoulli D. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. Paris: Mémoire Académie Royale des Sciences de Mathématique et de Physiques; 1760.

[2] Byrne A, McEvoy D, Collins A, Hunt K, Casey M, et al. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. BMJ Open 2020;10:e039856.

[3] Cori A, Ferguson N, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 2013;178:1505–12.

[4] Demongeot J, Griette Q, Magal P. SI epidemic model applied to COVID-19 data in mainland China. R Soc Open Sci 2020;7:201878.

[5] Demongeot J, Magal P. Data-driven mathematical modeling approaching for COVID-19: A survey. Phys Life Rev 2024;50:166.

[6] Griette Q, Demongeot J, Magal P. A robust phenomenological approach to investigate COVID-19 data for France. Math Appl Sci Eng 2021;2:149–60.

[7] Griette Q, Demongeot J, Magal P. What can we learn from COVID-19 data by using epidemic models with unidentified infectious cases? Math Biosci Eng 2022;19:537–94.

[8] Griette Q, Magal P. Clarifying predictions for COVID-19 from testing data: the example of New York State. Infect Dis Model 2021;6:273–83.

[9] Lambert J. Observationes variae in mathesin puram. Acta Helv Phys-Math-Anat-Bot-Med 1758;3:128–68.

[10] Magal P, Ruan S. Susceptible-infectious-recovered models revisited: from the individual level to the population level. Math Biosci 2014;250:26–40.

[11] Obadia T, Haneef R, Boëlle P. The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak 2012;12:1–9.

[12] Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (covid-19) in China. J Popul Econ 2020;33:1127–72.

[13] Tsoularis A, Wallace J. Analysis of logistic growth models. Math Biosci 2002;179:21–55.

[14] Verhulst P. Notice sur la loi que la population suit dans son accroissement. Corresp Math Phys 1838;10:113–29.

[15] Veron J. Les mathématiques de la population, de Lambert à Lotka. Math Sci Hum 2002;159:43–55.