Title: Prospective Prediction of Future SARS-CoV-2 Infections Using Empirical Data on a National Level to Gauge Response Effectiveness

Natalia Blanco¹, Kristen Stafford¹2, Marie-Claude Lavoie¹, Axel Brandenburg³, Maria W. Górsna⁴, Matthew Merski⁴†

¹Center for International Health, Education, and Biosecurity, Institute of Human Virology - University of Maryland School of Medicine, Baltimore, Maryland USA
²Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, Maryland USA
³Nordita, KTH Royal Institute of Technology and Stockholm University, SE-10691 Stockholm, Sweden
⁴Structural Biology Group, Biological and Chemical Research Centre, Department of Chemistry, University of Warsaw, Warsaw, Poland

Corresponding author(s): Matthew Merski, Maria Gorna

Address: Biological and Chemical Research Centre, Żwirki i Wigury 101 02-089 Warsaw, Poland

Telephone: (+48) 22 55 26 642

Email: merski@gmail.com; mgorna@chem.uw.edu.pl

Running head: Prediction of Future SARS-CoV-2 Infections

Conflict of interest: None

Sources of funding: The work was supported by the National Science Centre, Poland [grant agreement 2014/15/D/NZ1/00968 to M.W.G.] & the EMBO Installation Grant, European Molecular Biology Organization to M.W.G.

Data availability: All data used in this manuscript are publically available at

https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide

Acknowledgements: We thank Dr. Marcin Ziemniak for useful discussions.
**ABSTRACT:**

**Background:** Predicting an accurate expected number of future COVID-19 cases is essential to properly evaluate the effectiveness of any treatment or preventive measure. This study aimed to identify the most appropriate mathematical model to prospectively predict the expected number of cases without any intervention.

**Methods:** The total number of cases for the COVID-19 epidemic in 28 countries was analyzed and fitted to several simple rate models including the logistic, Gompertz, quadratic, simple square, and simple exponential growth models. The resulting model parameters were used to extrapolate predictions for more recent data.

**Results:** While the Gompertz growth models (mean $R^2 = 0.998$) best fitted the current data, uncertainties in the eventual case limit made future predictions with logistic models prone to errors. Of the other models, the quadratic rate model (mean $R^2 = 0.992$) fitted the current data best for 25 (89%) countries as determined by $R^2$ values. The simple square and quadratic models accurately predicted the number of future total cases 37 and 36 days in advance respectively, compared to only 15 days for the simple exponential model. The simple exponential model significantly overpredicted the total number of future cases while the quadratic and simple square models did not.

**Conclusions:** These results demonstrated that accurate future predictions of the case load in a given country can be made significantly in advance without the need for complicated models of population behavior and generate a reliable assessment of the efficacy of current prescriptive measures against disease spread.

**Keywords:** SARS-CoV-2, COVID-19, expected number of cases, exponential growth, logistic model, mathematical model
INTRODUCTION:

On March 11, 2020 the World Health Organization (WHO) declared the novel coronavirus outbreak (SARS-CoV-2 causing COVID-19) as a pandemic more than three months after the first cases of pneumonia were reported in Wuhan, China in December, 2019. From Wuhan the virus rapidly spread globally, currently leading to ten million confirmed cases and half a million deaths around the world. Although coronaviruses have a wide range of hosts and cause disease in many animals, SARS-CoV-2 is the seventh named member of the Coronaviridae known to infect humans. An infected individual will start presenting symptoms an average of 5 days after exposure but approximately 42% of infected individuals remain asymptomatic. Furthermore, almost six out of 100 infected patients die globally due to COVID-19.

Currently, treatment and vaccine options for COVID-19 are limited. There is currently no effective or approved vaccine for SARS-CoV-2 although a report from April 2020 noted 78 active vaccine projects, most of them at exploratory or pre-clinical stages. As the virus is transmitted mainly from person to person, prevention measures include social distancing, self-isolation, hand washing, and use of masks. Strict measures of quarantine have been shown as the most effective mitigation measures, reducing up to 78% of expected cases compared to no intervention. Nevertheless, to evaluate the actual effectiveness of any mitigation measure it is necessary to accurately predict the expected number of cases in the absence of intervention.

While there has been some early concern about the ability of SARS-CoV-2 to spread at an apparent near exponential rate, real limitations in available resources (i.e. susceptible population) will reduce the spread to a logistic growth rate. Logistic growth produces a sigmoidal curve (Figure 1) where the total number of cases \(N\) eventually asymptotically approaches the population carrying capacity \(N_M\), which for viral epidemics is analogous to the fraction of the population that will be infected before “herd immunity” is achieved. This is represented in derivative form by the generalized logistic function (Equation 1):

\[
\frac{dN}{dt} = rN^\alpha \left[ 1 - \left( \frac{N}{N_M} \right)^\beta \right]^\gamma
\]

Eqn. 1)
where $\alpha$, $\beta$, & $\gamma$ are mathematical shape parameters that define the shape of the curve, and $r$ is the general rate term, analogous to the standard epidemiological parameter, $R_0$, the reproductive number, which is a measure of the infectivity of the virus itself $^{13,14}$. For a logistic curve where $\alpha = \frac{1}{2}$ and $\beta = \gamma = 0$, one gets quadratic growth$^{15}$ with $N = (rt/2)^2$, while for $\alpha = \beta = \gamma = 1$, this equation can be re-arranged to quadratic form (Equation 2)$^{11}$ and integrated (Equation 3):

$$\text{Eqn. 2)} \quad \frac{dN}{dt} = \frac{-rN^2}{N_M} + rN$$

$$\text{Eqn. 3)} \quad N(t) = \frac{N_0N_M}{(N_M-N_0)e^{-rt} + N_0}$$

where $N_0$ represents the initial number of cases within the population. The shape parameters of logistic functions magnify the uncertainty in fitting these curves especially during the early part of the epidemic. These difficulties are further exacerbated by the additional uncertainty in estimation of $R_0$ for SARS-CoV-2, which is directly linked to the current uncertainty in the population carrying capacity $N_M$. And while the basic logistic function gives rise to a symmetrical sigmoidal curve, asymmetrical curves such as the Gompertz growth function (Equation 4)$^{16,17}$:

$$\text{Eqn. 4)} \quad N(t) = N_0e^{\left(\frac{\ln(N_M/N_0)}{(1-e^{-rt})}\right)}$$

which emerges$^{11}$ from Equation 1 for $\alpha = \gamma = 1$, where $r$ is replaced by $\beta r$ and $\beta$ approaches 0. With this function the rate of spread slows significantly after passing the mid-point resulting in long-tailed epidemics (Figure 1).

Traditionally the number of cases that will occur in an epidemic like COVID-19 is modeled with an SEIR model (Susceptible, Exposed, Infected, Recovered/Removed), in which the total population is divided into four categories: susceptible - those who can be infected, exposed – those who in the incubation period but not yet able to transmit the virus to others, infectious - those who are capable of spreading disease to the susceptible population, and recovered/removed – those who have finished the disease course and are not susceptible to re-infection or have died. For a typical epidemic, the ability for infectious individuals to spread the disease is proportional to the fraction of the population in the susceptible category with “herd immunity”$^{12,13}$ and extinction of the epidemic occurs...
once a limiting fraction of the population has entered into the Recovered/Removed category\(^\text{13}\). However, barriers to transmission, either natural\(^\text{18}\) or artificial (\textit{i.e.} quarantines, vaccines)\(^\text{13}\) can extinguish the epidemic before the community is fully infected. Artificial barriers such as mandatory quarantining in China\(^\text{19,20}\) or aggressive contact tracing in South Korea\(^\text{21}\) currently seem to have largely stemmed the spread of SARS-CoV-2. Numerous political, social, and material factors prevent the implementation of either of these responses in many other countries\(^\text{22}\), but it should be possible to find alternative approaches which can be equally effective. For an epidemic as serious as COVID-19, it behooves medical, scientific, and policy experts to determine as rapidly as possible which community responses are effective and achievable within different populations. However, gauging the effectiveness of these responses requires an accurate prediction\(^\text{23}\) of the number of future cases, and overestimation of the expected number of cases will make neutral or even harmful responses appear to be effective when those overestimated cases fail to occur. Thus, accurate prospective predictions (\textit{i.e.} before knowing the actual outcome) are preferable to retrospective analysis in which effectiveness is gauged after the results of the prescriptive actions are known\(^\text{24,25}\).

This study aimed to evaluate if a simple model was able to correctly prospectively predict the total number of cases at a future date. We found that fitting the case data to a quadratic (parabolic) rate curve\(^\text{15}\) for the early points in the epidemic curves (before the mitigation efforts began to have effects) was easy, efficient, and made good predictions for the number of cases at future dates despite significant national variation in the start of the infection, mitigation response, or economic condition.

**METHODS:** Data on the number of COVID-19 cases was downloaded from the European Centre for Disease Prevention and Control (ECDC) on June 1, 2020\(^\text{26}\). Countries that had reported the highest numbers of cases in mid-March 2020 (and Russia) were chosen as the focus of our analysis to minimize statistical error due to small numbers. The total number of cases for each country was calculated as a simple sum of that day plus all previous days. Days that were missing from the record were assigned values of zero. The early part of the curve was fit and statistical parameters were generated using Prism 8 (GraphPad) using the non-linear regression module using the program standard centered second order polynomial (quadratic), exponential growth, and the Gompertz growth
model as defined by Prism 8, and a simple user-defined simple square model \( N = At^2 + C \) where \( N \) is the total number of cases, \( A \) and \( C \) are the fitting constants, and \( t \) is the number of days from the beginning of the epidemic curve. The beginning of the curve (SI Table 1) was defined empirically among the first days in which the number of cases began to increase regularly. Typically, this occurred when the country had reported less than 100 total cases. The early part of the curve was defined by manual examination looking for changes in the curve shape and later confirmed by \( R^2 \) values for the quadratic model. Prospective predictions for the number of cases were done by fitting the total number of COVID-19 cases for each day starting with day 5 and then extrapolating the number of cases using the estimated model parameters to predict the number of cases for the final day for which data was available (June 1, 2020) or to the last day before significant decrease in the \( R^2 \) value for the quadratic fit. Fit parameters for the Gompertz growth model were not used to make predictions if the fit itself was ambiguous. Acceptable predictions were defined as being within a factor of two from the actual number (i.e. predictions within 50-200% of the actual total).

RESULTS:

A simple exponential growth model is a poor fit for the SARS-COV-2 pandemic:

The total number of cases for each of 28 countries was plotted with time and several model equations were fit to the early part of the data before mitigating effects from public health policies began to change the rate of disease spread. In total, 20 (71 %) countries showed mitigation of disease spread by June 1 (Figure 2). When the early, pre-mitigation portion of the data was examined for all 28 countries, the Gompertz growth model had the best statistical parameters (mean \( R^2 = 0.998 \pm 0.0028 \), Table 1) although a fit could not be obtained for the data from 2 countries and many of the fit values for \( N_M \) were unrealistic compared to national populations (e.g. China and India had predicted \( N_M \) values corresponding to 0.014 % and 0.33 % of their populations respectively26 (SI Table 2)). Fitting was also incomplete for the generalized logistic model for all 28 countries underlining the difficulty in applying this model. On the other hand, the simple models were able to robustly fit all the current data, with the quadratic (parabolic) model performing the best (mean \( R^2 = 0.992 \pm 0.004 \)) and the exponential model the worst (mean \( R^2 = 0.957 \pm 0.022 \))(Table 1). In only three (11 %) countries did
the exponential model have the best overall \( R^2 \) value among the simple models. Furthermore, the trend of the overall superiority of the Gompertz model followed by the quadratic was also observed in the standard error of the estimate statistic as well. The mean standard error of the estimate (Sy.x, analogous to the root mean squared error for fits of multiple parameters) value for the 28 countries was 1699 for the Gompertz model, 5613 for the quadratic model, 8572 for the simple square model and 11257 for the exponential model (Table 1). Likewise, plots of the natural log of the total number of cases in the early parts of the epidemic (\( \ln(N) \)) with time are significantly less linear (as determined by \( R^2 \)) than equivalent plots of the square root of the total number of cases (\( N^{1/2} \)) (SI Table 3, SI Figs 1, 2).

**Quadratic growth models provide improved fits to the early portion of the epidemic courses:**

While logistic growth models have been widely used to model epidemics\(^{16,27}\), uncertainties in estimates of \( R_0 \) (and therefore the population carrying capacity \( N_M \)) make prospective predictions of the course of the epidemic difficult\(^{14,27}\). (Figure 3, Table 2, SI Table 4). Here we define predictions as accurate when they are within a factor of two (50-200\%) of the actual outcome. For most countries, the simple exponential model massively overpredicts the number of future cases. Predictions generated more than 14 days prior were more than double the actual number of cases for 17 (61\%) countries examined. In fact, for 15 (54\%) countries, the exponential model made at least one overprediction by a factor of greater than 10,000 fold, while the quadratic and simple square models make no overprediction by more than a factor of 3.3 and 2.1, respectively (i.e. using the first 10 days of data from Portugal the exponential model predicts 34 million cases while the quadratic, simple square, and Gompertz growth models predict 24957, 20358 and 18953 cases respectively while 23683 total cases were observed while the total population of Portugal in 2018 was 10.3 million\(^{26}\)).

Predictions using the quadratic and simple square models were much more accurate. Only in four (14\%) countries does the quadratic model ever overpredict the final number of cases by more than a factor of two while the simple square model overpredict by a factor of two for only one (4\%) country (SI Table 4). For the quadratic model, the mean maximum daily overprediction was a factor of 1.6-fold (median 1.3 fold) while for the simple square model the mean maximum daily overprediction
was 1.3-fold (median 1.1 fold). Both of these models produced much more accurate predictions than the simple exponential model (Table 2).

**DISCUSSION:**

The start of the global SARS-CoV-2 pandemic has resulted in an unprecedented set of national responses. These responses have varied considerably from a strict lockdown in China\textsuperscript{10}, to aggressive contact tracing in South Korea\textsuperscript{28}, to mandatory shelter in place restrictions in France\textsuperscript{29}, to giving citizens information and allowing them more freedom to make choices as in Sweden\textsuperscript{30}, and to other countries attempting to accelerate their progress towards herd immunity\textsuperscript{31}. The variety in these national prescriptions is a result of the different socioeconomic situations in individual countries which have to take into account not only the costs of these efforts both monetarily and in terms of lives, but also what can be reasonably achieved depending on the relationship between individual governments and their citizenry. Additionally, the spread of SARS-CoV-2 has been putatively linked to several inherent factors within a country, such as average population density\textsuperscript{32}, normal social behaviors\textsuperscript{22}, and even weather may have an effect\textsuperscript{33}. The efficacy of similar prescriptions can vary in pairs of neighboring countries (\textit{i.e.} the UK or Ireland). Therefore, it behooves every nation to review the results of its own policy prescriptions in order to make necessary course adjustments as quickly and accurately as possible.

While predictions about the future course of an epidemic, especially one as novel as COVID-19 are difficult under the best of circumstances, the severity of the pandemic has resulted in an unprecedented amount of epidemiological data being produced with daily frequency. Because logistic epidemic models have been in use for more than a century\textsuperscript{16} it is almost tautological that the rate of the spread of SARS-CoV-2 cannot be well-modeled by simple exponential growth equations, further demonstrated by statistical analysis of and the poor predictions made by the exponential model for the future number of cases as compared to the quadratic (parabolic) model\textsuperscript{15}. However, simple exponential models do not generate entirely terrible fit statistics (Table 1), and this may account for the conflation of the course of the SARS-CoV-2 pandemic with truly exponential growth. That the exponential growth constant term, $k$, is constantly decreasing after day 10 in 10 (68 \%) countries (SI Fig. 3) further
indicates the overall utility of logistic models, which were explicitly developed to model the a constantly decreasing rate of growth due to consumption of the available resource (i.e. the susceptible population pool of the SIR model)\textsuperscript{16}. But, while logistic models are implicitly the correct model, they are difficult to accurately fit during the early portion of an epidemic due to inherent uncertainties in the mathematical shape parameters (Equation 1) of the curve itself and the population carrying capacity for SARS-CoV-2, $N_M$, which still has a significant uncertainty as the virus has only recently moved into the human population. Herd immunity is defined as $1 - 1/R_0$, and since current estimates for $R_0$ vary from 1.5 – 6.5\textsuperscript{14}. This implies that 33 – 85% of the population will need to have contracted the disease and developed immunity in order to terminate the epidemic. A discrepancy of this size will significantly affect predictions based on logistic growth models.

Here we note the utility of the quadratic (parabolic) and simple square models in predicting the course of the pandemic more than a month in advance. The simple exponential model vastly overpredicts the number of cases (Fig. 2, Table 2). The Gompertz growth model, while often making largely correct predictions often generates wildly inaccurate estimates of the population carrying capacity $N_M$ (SI Table 2), and the generalized logistic model simply fails to produce a statistically reliable result with the currently available data. Overestimation of the future number of cases will cause problems because the failure of the number of predicted cases to materialize may be erroneously used as evidence that poorly implemented and ineffective policy prescriptions are reducing the spread of SARS-CoV-2, which may lead to political pressure for premature cessation of all prescriptive measures and inevitably an increase in the number of cases and excess, unnecessary morbidities. Fortunately, the quadratic model produces accurate, prospective predictions of the number of cases (Fig. 3, Table 2). Use of this model is simple as it is directly implemented in common spreadsheet programs and can be implemented without much difficulty or technical modeling expertise. In theory, this model can also be applied to smaller, sub-national populations, although the smaller number of total cases in these regions will undoubtedly give rise to larger statistical errors.

In no way does the empirical agreement between the quadratic model and empirical data negate the fact that the growth of the SARS-CoV-2 epidemic is logistic in nature in all 28 countries (Table 1, SI Table 2). We expect the suitability of these empirical quadratic fits is related to either the
fact that quadratic form of the slope of the generalized logistic function or the limitation of the virus to a physical radius of infectivity around infectious individuals, or that it is still early in the pandemic as no country has yet officially logged even 1% of its population as having been infected, or all three. Of course, the true number of COVID-19 cases is a matter of debate as there is speculation that a significant fraction of infections are not being identified\textsuperscript{14}. However, because this method is focused on the rate of case growth over time, the errors that lead to any undercounting within a given country are likely to remain largely unchanged over the short time periods observed here and still provide a reasonable estimate of the number of positively identified cases. While despite their similar predictive power we largely focus on the quadratic model rather than the simple square model for the aforementioned reasons, we must also note that quadratic curve fitting is natively implemented in most common spreadsheet software while the simple square model is not. By monitoring the $R^2$ values for the quadratic models, it is a simple task to identify when the epidemic is beginning to subside within a country (\textit{i.e.} “bending the curve”). Here we recommend the use of an $R^2$ value of 0.985 for identifying when the rate of infection is beginning to subside, but more conservative estimates can also be made by lowering this threshold.

Examination of the data collected here suggests that early, aggressive measures have been most effective at reducing disease burden within a country. Countries that initially adopted less stringent measures (such as the US, UK, Russia, and Brazil) are currently more heavily burdened than those countries that started with more intense prescriptions (such as China, South Korea, Australia, Denmark, and Vietnam)\textsuperscript{35}. Vietnam was not analyzed here as it does not currently have a large number of COVID-19 cases likely due to its early aggressive action against viral spread\textsuperscript{7,35}. The effectiveness of aggressive measures may also be due to the apparent quadratic rate of growth of total cases with time (Equation 2); while growth in proportion to the square of the number of days is fast, it is not as fast as exponential growth. Early reductions in the number of infected individuals and the number of interactions they have with susceptible individuals clearly pays compounded dividends in future case reductions as advantage can be taken of this slower spreading rate.

Quadratic modeling of the increases of COVID-19 cases within national boundaries currently gives a more accurate prospective prediction of the growth in the number of total cases. This allows
for near real time adjustments to policy prescriptions and obviates the need to estimate the effect of human behaviors on these predictions and instead focus on the available empirical data and fine tune the needed responses and efficiently guide the direction of health care resources. Until an effective vaccine or cure is available, social distancing, contact tracing, and other aggressive quarantine measures are the most effective tools to combat the spread of SARS-CoV-2 and it is imperative to monitor whether these measures are being effectively implemented. Accurate prospective modeling of the future number of cases within countries will help to minimize the social costs and financial burdens of these necessary mitigation measures.
**TABLES:**

**Table 1:** Statistical fit parameters for the quadratic, simple square, simple exponential, and Gompertz growth models for the early course of the epidemic in each country.

|                     | Mean  | Strd. Dev | Median     |
|---------------------|-------|-----------|------------|
| **Quadratic** *(N=28)* |       |           |            |
| R squared           | 0.9917| 0.0044    | 0.9930     |
| Sum of Squares      | 21916684140| 1.05689E+11| 87110317.5 |
| Sy.x                | 5612.9 | 15068.7   | 1250.0     |
| **Simple Square** *(N=28)* |       |           |            |
| R squared           | 0.9780| 0.0159    | 0.9821     |
| Sum of Squares      | 44700980001| 2.02519E+11| 174119853  |
| Sy.x                | 8572.1 | 21235.6   | 1591.5     |
| **Simple Exponential** *(N=28)* |       |           |            |
| R squared           | 0.9568| 0.0215    | 0.9548     |
| Sum of Squares      | 1.0317E+11| 5.23176E+11| 466907869.5|
| Sy.x                | 11257.0 | 32734.8   | 3067.0     |
| **Gompertz growth** *(N=26)* |       |           |            |
| R squared           | 0.9978| 0.0028    | 0.9991     |
| Sum of Squares      | 2048425776| 9767369361| 11600080   |
| Sy.x                | 1699.2 | 4598.3    | 459.7      |

*Because fit parameters were unable to be obtained for the Gompertz model for two countries, the statistical parameters were calculated from the remaining 26 countries (indicated by *N* = 26). The generalized logistic model failed to adequately determine fit parameters in every cases so is not included in this table.*
Table 2: Results of prospective predictions of total case load made using the various models.

|                      | Mean Predictive Length (Days) | Mean Minimum Normalized Prediction (Observed/Actual Total Cases) | Mean Maximum Normalized Prediction (Observed/Actual Total Cases) |
|----------------------|-------------------------------|----------------------------------------------------------------|----------------------------------------------------------------|
| Simple Exponential   | 14.6 ± 8.0 (11)               | 1.11 ± 0.09 (1.14)                                               | 8.9 x 10^{14} ± 4.7 x 10^{15} (5 x 10^{4})                     |
| Quadratic (Parabolic)| 35.6 ± 15.8 (31.5)            | 0.25 ± 0.29 (0.09)                                               | 1.55 ± 0.64 (1.33)                                              |
| Simple Square        | 36.9 ± 21.6 (31)              | 0.42 ± 0.34 (0.24)                                               | 1.26 ± 0.35 (1.12)                                              |
| Gompertz Growth      | 28.2 ± 17.9 (25.5)            | 0.14 ± 0.24 (0.02)                                               | 3.8 x 10^{9} ± 2.0 x 10^{9} (2 x 10^{2})                        |

^b All data given are mean results ± the sample standard error. Median values are given in parentheses.
**Figure 1:** Illustrative comparison of exponential, quadratic, generalized logistic and Gompertz growth curves. The Gompertz growth curve (Equation 4, solid black line) representing the progress of a theoretical epidemic for a disease with an arbitrarily chosen $R_0$ value of 3.4 ($r = 0.045, N_0 = 1, N_M = 70\%$, dotted line). The solid grey line is an equivalent logistic curve, note that while the midpoint for both logistic curves is the same, the Gompertz curve reaches the population carrying capacity more slowly, resulting in a long tailed epidemic. The initial part of the Gompertz curve (including time points until 5% of the population has been infected) was fit to the simple exponential (red dashes), quadratic (blue dashes) and simple square (green dashes) models. It is apparent from these curves how quickly the exponential curve overestimates the rate of growth for the epidemic as compared to the quadratic and simple square fit curves and how the quadratic model more closely follows the Gompertz growth curve, evidenced by the smaller $Sy.x$ value for the quadratic fit in Table 1.
Figure 2: The development of COVID-19 cases over time in 28 nations. The total number of cases as of June 1, 2020 is indicated by black circles while the early part of the curve is indicated by orange triangles. A quadratic fit curve based on the early part of the curve extrapolated into the future is shown as an orange dashed line. The black circles are obscured in those countries which had not begun to effectively reduce SARS-CoV-2 spread by June 1, 2020.
Figure 3: Comparison of the errors in prospective predictions for COVID-19 case numbers for different growth models for 28 countries for the simple exponential model (red triangles), the simple square model (green squares), the quadratic model (black circles), the Gompertz growth model (blue triangles), and the basic logistic growth model (purple diamonds). Note the log scale for the vertical axis which indicates the ratio of the predicted to observed number of cases. In each graph the fit values for each model using only data up to that day are used to predict the number of expected cases for the last day for which data is available (or the last day before significant curve deviation is observed, see figure 2). Days on which the fit was not statistically sound were omitted from the graph.
REFERENCES:

1. Organization WH. Coronaviruse disease 2019 (COVID-19) Situation Report - 51. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10; World Health Organization, 2020.

2. Sun J, He W-T, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou J, Veit M, Su S. COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. Trends in molecular medicine 2020;26(5):483-495.

3. Rothen HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of autoimmunity 2020;109:102433.

4. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 2020;25(10).

5. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung S-M, Hayashi K, Kinoshita R, Yang Y, Yuan B, Akhmetzhanov AR, Linton NM. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases 2020;94:154-155.

6. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. The Lancet Infectious diseases 2020.

7. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, Liao Q, Kwan MY, Lee SL, Chiu SS, Wu JT, Wu P, Leung GM. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. The Lancet Public health 2020;5(5):e279-e288.

8. Thanh Le T, Andreadakis Z, Kumar A, Gomez Roman R, Tollefsen S, Saville M, Mayhew S. The COVID-19 vaccine development landscape. Nature reviews Drug discovery 2020;19(5):305-306.

9. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, Tam C, Dickens BL. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet Infectious diseases 2020;20(6):678-688.

10. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling 2020;5:256-263.

11. Tsoularis A. Analysis of Logistic Growth Models. Res. Lett. Inf. Math. Sci. 2001;2:23-46.

12. Anderson RM, May RM. Herd-Immunity to Helminth Infection and Implications for Parasite Control. Nature 1985;315(6019):493-496.

13. Fine P, Eames K, Heymann DL. "Herd Immunity": A Rough Guide. Clinical Infectious Diseases 2011;52(7):911-916.

14. Liu Y, Gayle AA, Wilder-Smith A, Rocklov J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine 2020;27(2).

15. Brandenburg A. Piecewise quadratic growth during the 2019 novel coronavirus epidemic. arXiv:2002.03638 2020.

16. Tjorve KMC, Tjorve E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. PloS One 2017;12(6):e0178691.

17. Laird AK. Dynamics of Tumor Growth. Br J Cancer 1964;13:490-502.

18. Vora A, Burke DS, Cummings DAT. The impact of a physical geographic barrier on the dynamics of measles. Epidemiology and Infection 2008;136(5):713-720.

19. Kupferschmidt K, Cohen J. Can China's COVID-19 strategy work elsewhere? Science (New York, N Y ) 2020;367(6482):1061-1062.

20. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science (New York, N Y ) 2020;368(6492):742-746.
21. Covid-19 National Emergency Response Center E CMTK, Prevention. Contact Transmission of COVID-19 in South Korea: Novel Investigation Techniques for Tracing Contacts. Osong Public Health Res Perspect 2020;11(1):60-63.

22. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, Crockett MJ, Crum AJ, Douglas KM, Druckman JN, Drury J, Dube O, Ellemers N, Finkel EJ, Fowler JH, Gelfand M, Han S, Haslam SA, Jetten J, Kitayama S, Mobbs D, Napper LE, Packer DJ, Pennycook G, Peters E, Petty RE, Rand DG, Reicher SD, Schnall S, Shariff A, Skitka LJ, Smith SS, Sunstein CR, Tabri N, Tucker JA, Linden Svd, Lange Pv, Weeden KA, Wohl MA, Zaki J, Zion SR, Willer R. Using social and behavioural science to support COVID-19 pandemic response. Nature human behaviour 2020;4(5):460-471.

23. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. COVID-19 Image Data Collection: Prospective Predictions Are the Future. arXiv:2006.11988 2020.

24. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clinical practice 2009;113(3):c214-7.

25. Yang W, Zhang D, Peng L, Zhuge C, Hong L. Rational evaluation of various epidemic models based on the COVID-19 data of China. medRxiv 2020.

26. Control ECFPa. Accessed June 1, 2020, 2020.

27. Liu W, Tang S, Xiao Y. Model Selection and Evaluation Based on Emerging Infectious Disease Data Sets including A/H1N1 and Ebola. Computational and mathematical methods in medicine 2015;2015:207105.

28. Oh J, Lee J-K, Schwarz D, Ratcliffe HL, Markuns JF, Hirschhorn LR. National Response to COVID-19 in the Republic of Korea and Lessons Learned for Other Countries. Health systems and reform 2020;6(1):e1753464.

29. Moatti J-P. The French response to COVID-19: intrinsic difficulties at the interface of science, public health, and policy. The Lancet Public health 2020;5(5):e255.

30. Sjodin H, Johansson AF, Brannstrom A, Farooq Z, Kriit HK, Wilder-Smith A, Aström C, Thunberg J, Soderquist M, Rocklov J. COVID-19 healthcare demand and mortality in Sweden in response to non-pharmaceutical (NPIs) mitigation and suppression scenarios. medRxiv 2020.

31. Altmann DM, Douek DC, Boyton RJ. What policy makers need to know about COVID-19 protective immunity. Lancet (London, England) 2020;395(10236):1527-1529.

32. Rocklov J, Sjodin H. High population densities catalyse the spread of COVID-19. Journal of travel medicine 2020;27(3).

33. Wang J, Tang K, Feng K, Lin X, Weifeng L, Chen K, Wang F. High Temperature and High Humidity Reduce the Transmission of COVID-19. arxiv 2020.

34. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science (New York, N Y ) 2020;368(6490):489-493.

35. Gibney E. Whose coronavirus strategy worked best? Scientists hunt most effective policies. Nature 2020;581:15-16.
Supporting Information for “Prospective Prediction of Future SARS-COV-2 Infections Using Empirical Data on a National Level to Gauge Response Effectiveness” Blanco et al.
**SI Table 1:** The start date (day 1, in 2020) of the fit curves for the 28 analyzed countries.

| Country   | Date   | Country    | Date   |
|-----------|--------|------------|--------|
| Australia | Mar. 10| Italy      | Feb. 22 |
| Austria   | Mar. 5 | Malaysia   | Mar. 10 |
| Belgium   | Mar. 5 | Norway     | Mar. 5  |
| Brazil    | Mar. 12| Philippines| Mar. 14 |
| Canada    | Mar. 8 | Poland     | Mar. 13 |
| China     | Jan. 18| Portugal   | Mar. 12 |
| Denmark   | Mar. 13| Russia     | Mar. 14 |
| France    | Feb. 28| S. Korea   | Feb. 19 |
| Germany   | Feb. 28| Spain      | Feb. 28 |
| Holland   | Mar. 5 | Sweden     | Mar. 5  |
| India     | Mar. 17| Swiss      | Mar. 5  |
| Iran      | Feb. 22| Turkey     | Mar. 16 |
| Ireland   | Mar. 13| UK         | Mar. 4  |
| Israel    | Mar. 12| US         | Mar. 2  |
SI Table 2: The fit parameters for the development of the early portion of the SARS-CoV-2 epidemic in 28 countries for the exponential, quadratic, simple square, simple exponential, Gompertz growth models as calculated for each individual day during the early portion of the epidemic.\(^a\)

| Country         | Days | \( N_0 \times e^{kt} \) | Sum of Sq. | Sy.x | A     | \( B \) | R     | D     |
|-----------------|------|--------------------------|------------|------|-------|-------|-------|-------|
| Australia       | 20   | 1.0513                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Austria         | 27   | 1.0236                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Belgium         | 49   | 0.0617                   | 25096359   | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Brazil          | 80   | 0.0384                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Canada          | 78   | 0.0332                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| China           | 7268 | 0.1174                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Denmark         | 59.7 | 0.0794                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| France          | 44   | 0.0793                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Germany         | 40   | 0.0532                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| India           | 75   | 0.0511                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Israel          | 35   | 0.0002                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Italy           | 48   | 0.0028                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Norway          | 75   | 0.0004                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Philippines     | 375.5| 0.0081                  | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Portugal        | 2266 | 0.0542                   | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Russia          | 10551| 0.0458                  | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| South Africa    | 448.3| 0.0157                 | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| South Korea     | 4062 | 0.0283                | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Spain           | 4336 | 0.0582                | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Switzerland     | 65    | 0.0278                | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| Turkey          | 4154 | 0.0843                | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| UK              | 9411 | 0.0488                | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |
| USA             | 14532| 0.0288            | 246622     | 267  | 13.37 | -90.68 | 69.38 | 19    |

\(^a\) The fit equations for each are as follows:

**Simple exponential:**

\[ N(t) = N_0 e^{kt} \]

**Quadratic:**

\[ N(t) = At^2 + Bt + C \]

**Simple square:**

\[ N(t) = At^2 + C \]

---

23
Gompertz growth: 

\[ N(t) = N_0 e^{\left( \ln\left( \frac{N_M}{N_0} \right) (1 - e^{-rt}) \right)} \]

where \( N \) is the total number of cases, \( t \) is the time in days, \( N_0 \) is the initial seeding population of the epidemic, \( N_M \) is the population carrying capacity (the amount of the population that must be infected to achieve herd immunity), \( A, B \) & \( C \) are the standard quadratic terms (or for the simple square model equation). Additionally, the number of days of data used in the fitting, the \( R^2 \), sum of squares, and Sy.x statistical values are given. For the Gompertz growth model, an adequate fit could not be achieved for Brazil or Denmark and this is indicated by N/A.
**SI Table 3:** Summary of the statistical parameters of linear fits to the plots of $N$ and $\ln N$ with time.

| Square root of total cases | Mean      | Str. Dev. | Median |
|----------------------------|-----------|-----------|--------|
| Slope                      | 5.639464  | 3.648843  | 4.355  |
| Y-intercept                | -21.7129  | 35.10683  | -8.081 |
| X-intercept                | 1.442404  | 5.604858  | 1.8695 |
| 1/slope                    | 0.261796  | 0.166244  | 0.2302 |
| R squared                  | 0.977639  | 0.012277  | 0.98075|
| Sy.x                       | 14.90246  | 17.01295  | 7.7295 |

| Natural log of total cases | Mean      | Str. Dev. | Median |
|----------------------------|-----------|-----------|--------|
| Slope                      | 0.1353    | 0.057693  | 0.1251 |
| Y-intercept                | 5.492857  | 0.82118   | 5.435  |
| X-intercept                | -51.5525  | 30.04849  | -42.905|
| 1/slope                    | 8.937607  | 4.159834  | 8.014  |
| R squared                  | 0.891386  | 0.068346  | 0.9139 |
| Sy.x                       | 0.624629  | 0.278802  | 0.63595|
| SQRT Mean                  |           |           |        |
SI Table 4: The range of prediction results for the quadratic, simple square, simple exponential, and Gompertz growth models based on days of good predictions before the target, last day of observed data, inclusive.

| Days of Good Predictions | Minimum prediction for final day | Maximum prediction for final day |
|--------------------------|----------------------------------|----------------------------------|
| Australia                | Quad: 0.503115 | Exp: 0.447172 | Gompertz: 1.137994 | Quad: 0.580903 | Exp: 0.514972 | Gompertz: 1.216466 |
| Austria                  | Quad: 0.016875 | Exp: 0.215960 | Gompertz: 1.169537 | Quad: 0.021274 | Exp: 0.160253 | Gompertz: 1.220548 |
| Belgium                  | Quad: 0.006686 | Exp: 0.293578 | Gompertz: 1.166721 | Quad: 0.009560 | Exp: 0.120840 | Gompertz: 1.272823 |
| Brazil                   | Quad: 0.058007 | Exp: 0.895547 | Gompertz: 1.038249 | Quad: 0.094144 | Exp: 0.781347 | Gompertz: 1.185865 |
| Canada                   | Quad: 0.063630 | Exp: 0.155024 | Gompertz: 1.188932 | Quad: 0.065138 | Exp: 0.131318 | Gompertz: 1.295561 |
| China                    | Quad: 0.095759 | Exp: 0.199268 | Gompertz: 1.119746 | Quad: 0.095759 | Exp: 0.199268 | Gompertz: 1.119746 |
| Denmark                  | Quad: 0.352518 | Exp: 0.874563 | Gompertz: 1.180571 | Quad: 0.385402 | Exp: 1.140502 | Gompertz: 1.324723 |
| France                   | Quad: 0.049381 | Exp: 0.110337 | Gompertz: 1.151091 | Quad: 0.072456 | Exp: 0.100218 | Gompertz: 1.192701 |
| Germany                  | Quad: 0.001451 | Exp: 0.009676 | Gompertz: 1.050707 | Quad: 0.002736 | Exp: 0.000268 | Gompertz: 0.929757 |
| India                    | Quad: 0.120032 | Exp: 0.178258 | Gompertz: 1.040921 | Quad: 0.135907 | Exp: 0.124184 | Gompertz: 1.215347 |
| Iran                     | Quad: 0.243367 | Exp: 0.211323 | Gompertz: 1.060169 | Quad: 0.262009 | Exp: 0.190298 | Gompertz: 1.263993 |
| Ireland                  | Quad: 0.502011 | Exp: 0.655517 | Gompertz: 1.078376 | Quad: 0.515569 | Exp: 0.540988 | Gompertz: 1.280466 |
| Israel                   | Quad: 0.394938 | Exp: 0.583151 | Gompertz: 1.104638 | Quad: 0.411518 | Exp: 1.125427 | Gompertz: 1.336302 |
| Italy                    | Quad: 0.088719 | Exp: 0.184681 | Gompertz: 1.117078 | Quad: 0.102372 | Exp: 1.124866 | Gompertz: 1.221707 |
| Malaysia                 | Quad: 0.069649 | Exp: 0.087035 | Gompertz: 1.129023 | Quad: 0.076907 | Exp: 1.132076 | Gompertz: 1.245426 |
| Netherlands              | Quad: 0.200439 | Exp: 0.512291 | Gompertz: 1.150419 | Quad: 0.225327 | Exp: 1.172287 | Gompertz: 1.286714 |
| Norway                   | Quad: 0.027684 | Exp: 0.701163 | Gompertz: 1.109444 | Quad: 0.030184 | Exp: 1.097820 | Gompertz: 1.249701 |
| Philippines              | Quad: 0.132421 | Exp: 0.680459 | Gompertz: 1.037679 | Quad: 0.142517 | Exp: 1.117904 | Gompertz: 1.249701 |
| Poland                   | Quad: 0.777324 | Exp: 0.776356 | Gompertz: 1.325667 | Quad: 0.823958 | Exp: 0.875218 | Gompertz: 1.359904 |
| Portugal                 | Quad: 0.424212 | Exp: 0.698348 | Gompertz: 1.162532 | Quad: 0.478098 | Exp: 1.197017 | Gompertz: 1.245426 |
| Russia                   | Quad: 0.043712 | Exp: 0.045382 | Gompertz: 1.157029 | Quad: 0.053094 | Exp: 1.182701 | Gompertz: 1.327888 |
| South Korea              | Quad: 0.689575 | Exp: 0.866666 | Gompertz: 1.172747 | Quad: 0.717617 | Exp: 1.145092 | Gompertz: 1.361294 |
| Spain                    | Quad: 0.071436 | Exp: 0.078543 | Gompertz: 1.165177 | Quad: 0.086666 | Exp: 1.172864 | Gompertz: 1.392293 |
| Sweden                   | Quad: 0.040168 | Exp: 0.105126 | Gompertz: 1.142619 | Quad: 0.046093 | Exp: 1.137028 | Gompertz: 1.411293 |
| Switzerland              | Quad: 0.000212 | Exp: 0.466564 | Gompertz: 1.130035 | Quad: 0.003158 | Exp: 0.163152 | Gompertz: 1.454795 |
| Turkey                   | Quad: 0.302459 | Exp: 0.217373 | Gompertz: 1.122809 | Quad: 0.320809 | Exp: 1.130808 | Gompertz: 1.523934 |
| UK                       | Quad: 0.006558 | Exp: 0.109414 | Gompertz: 1.146397 | Quad: 0.009307 | Exp: 1.132734 | Gompertz: 1.572007 |
| USA                      | Quad: 0.034959 | Exp: 0.072806 | Gompertz: 1.180173 | Quad: 0.052715 | Exp: 0.208431 | Gompertz: 1.721519 |

b Good predictions are defined as the predicted result being within a factor of 2, predictions from 50 – 200% of the actual total number of cases). Thus, the quadratic model was able to predict the total number of cases in the United States in each of the 45 days before that day, while the exponential model was only within the defined good range for the 20 days preceding that day. The range of minimum predictions (under predictions) and maximum predictions (overpredictions) is also given
SI Figure 1: Plots of the square root of the total number of cases ($\sqrt{N}$) for the early portion of the COVID-19 epidemic in each of the 28 countries (black circles) along with a linear fit line.
SI Figure 2: Plots of the natural log of the total number of cases (lnN) for the early portion of the COVID-19 epidemic in each of the 28 countries (black circles) along with a linear fit line.
SI Figure 3: The change of the exponential rate term (k) over time for each of the 28 countries. It can be clearly seen that k is generally decreasing over time, often on each day but sometimes after an initial bit of increasing. This indicates that the exponential rate is regularly decreasing, as expected for a situation where growth resource is decreasing, as is expected for the logistic models family of models, including the generalized logistic and Gompertz growth models.