 doi:10.20944/preprints202005.0372.v1

Modeling and Analysis of The Early-Growth Dynamics of COVID-19 Transmission

Abul Kalam al Azad[1], Asif Mostafa Hussain

Department of Computer Science and Engineering,

University of Liberal Arts Bangladesh (ULAB).

[1] abul.azad@ulab.edu.bd

Abstract – As an on-going pandemic caused by the out-break of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or simply COVID-19 sweeps through the globe at an unprecedented rate leaving behind trails of high infection and mortality, it is crucial to understand the propagation dynamics of the virus in a host population in order to take urgent and effective remedial and mitigating steps to save life. It is already observed in many countries and communities that accurate and timely testing, tracing, and tracking of the infection lead to better containment and slowing down of the spread. In this exploratory research, the early growth dynamics of infection within a population is pursued based on real data. The study posits that the early growth in a homogenous population follows an exponential pattern motivating further rigorous quantitative treatment based on a number of analytical models such as logistic model, Richard's model, and Gompertz model– the acceleration pattern of the outbreak is ascertained from the daily inflection data, and regression analysis against population models yields dynamic growth indices which allow very accurate prediction of the successive outbreak size when calibrated continually with updated data. The performance of the various models is evaluated with the real dataset. More, the basic reproduction number of the COVID-19 virus propagation in the community is estimated based on the on-set phase dataset using multi-compartmental epidemiological model. Also, the maximum infection size, infection doubling time and the scope of the herd immunity are also inferred for COVID-19 in a population.

Key Words – COVID-19, SARS-CoV-2, Coronavirus, Pandemic, Epidemiological Analysis, Exponential Growth, Herd Immunity, Doubling Period.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

1 Introduction

The World Health Organization (WHO) upgraded the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak to a pandemic on March 11, 2020, and chose COVID-19 (Coronavirus Disease 2019) as the suitable nomenclature for this virus-mediated disease [1, 2]. The Chinese authority first reported the outbreak in December 2019, and since then it spread throughout the globe relentlessly claiming lives of 258,326 people and infected 3,727,301 more in as many as 212 countries as of May 2020 [1]. The COVID-19 outbreak dynamics showed a critical nature of progressing to a large-scale transmission within a population from a rather mild local outbreak phase as observed in the first reported infection in the seafood market in Wuhan, China, via a community transmission phase where the contagion spread through interpersonal and cluster transmission. The acceleration of the outbreak is further facilitated by asymptomatic nature of contagion, i.e., an infected person may not show any symptoms of the COVID-19 upto nearly 15 days while may still keep on infecting others in his or her contact [2, 3, 4]. This renders the tracking and tracing of the infected individual a very demanding task. The virus targets the lungs by entering the body typically through mouth and nose as respiratory droplets [5, 6]. As a result, the transmissibility of the virus is rather high among communities with close proximity and high contact rate. This puts the countries with higher spatial population density at an extraordinary risk; specially, the countries in South Asia are apparently more vulnerable to the contagion due to a combination of factors ranging from population density, societal structure, culture to constrained health-care facilities. Bangladesh is among the south Asian countries with a significant population size drawing an utmost urgency to analyze and understand the COVID-19 outbreak dynamics.

Globally ranked 8th for the population size, Bangladesh is home to approximately 161 million people with a staggering population density of nearly 2864 per square miles, which is 7th highest in the world [7, 8]. Given this mammoth sized population packed in this rather limited area makes Bangladesh uniquely vulnerable to large outbreak and potential human catastrophe. Besides, the anticipated size of the infected population will

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

soon overwhelm any healthcare services and would make it difficult for any contingencies or interventions. The first infection broke with three cases on 7th March, 2020, in Bangladesh, and the Government imposed an emergency 'lockdown' from 26th March, by declaring full-fledged closure of religious and educational institutions, offices, businesses, shops, malls, and a ban on all social gatherings, making an exception to some emergency services. As of 10th May, almost 62 days since the first reported cases, the total number of infected people in Bangladesh, according to the official report, stands at 20,995 with a total 314 dead as of 15th May, and the numbers are mounting daily [9, 10]. It cannot be emphasized more, the importance and urgency to analyze the infection dynamics in this spreading phase of COVID-19 in Bangladesh to help in the decision of executing any exit plan to come out of the outbreak with minimal loss of life and economic hardship.

In recent times, for very obvious necessity and urgency, a huge effort is undertaken to understand the infection of COVID-19 pandemic based on the data in different spatial contexts globally, which is reflected by a growing number of research in this area [11, 12, 13]. Infectious diseases are, typically, modeled as a diffusion process involving a pathogen that disseminates among a given homogeneous or heterogeneous population with a characteristic transmission strength governed by a system of coupled differential equations with variables indicating different compartments the populations are subdivided; e.g., the SIR epidemiological model assumes a homogeneous population to undergo transitions from susceptible (S) to Infectious (I) to removed (R) states as a virus sweeps through; various model variants to this are widely used for understanding the outbreak and for exploring different scenarios to forecast future development of the outbreak, essentially to understand the 'flattening' behavior of the infection in order to impose and modulate emergency public measures like lockdown by gauging the healthcare facilities available to the population [14, 15, 16, 17, 18]. Moreover, another theoretical approach, to capture the inherent randomness of infection, is to recourse to stochastic epidemiological approaches based on (generalized) logistic models, and (sub) exponential models [19, 20, 21].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

In this study, a phenomenological approach, based on both stochastic and multi-compartment models, have been adopted to model the real dataset of COVID-19 infection in Bangladesh. The study shows that the growth stage follows a very random nature in Bangladesh with typical exponential growth feature. Motivated by the growth trend in the real data, exponential model techniques along with logistic model have been applied to extract trending parameters of the outbreak. Further regression analysis has been conducted to extract high precision fitting of the data with the models, and the fitting parameters have produced a very commendable accuracy in predicting future infection values in the region. Comparative analysis have shown Richard's model and Gompertz model have performed remarkably better than any other parametric approaches in fitting and predicting. The study, further, estimated the basic reproduction number, R_0, for infection in Bangladesh, using multi-compartmental model against the real dataset. The estimation posits $R_0 \sim 3.0$ for the infection which accounts for the observed exponential growth trend in the country. The temporal variation of reproduction is also noted. Further numerical analyses show the scope and rate of the growth of the pandemic in this region is fortunately lower than earlier grave projections. The doubling time of the infection, the maximum infection population size as well as the herd immunity threshold for the current growth are also estimated and presented in this study.

2 Materials and Methods

2.1 Phenomenological Growth Model

Phenomenological models are data-driven empirical approaches without any specific basis on first principle or any physical mechanisms [20, 22]. These models provide insight in the patterns in the observed data, and so are good candidates to use as the initial probe of investigation, which eventually motivate reproducibility of empirical observation using more rigorous mathematical models. As data suggests, Bangladesh is still in the early stage of epidemic growth with exponential features, so an exponential

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

model such as the following is a good special power-law based method to study its initial growth dynamics.

$$I(t) = ae^{bt} \qquad (1)$$

Here, $I(t)$ implies the cumulative incidences over time t, and a and b are parameters indicating the growth pattern – these can be estimated by fixing or fitting with data. The general generating rate form of equation 1 is $I'(t) \sim cI^p(t)$, where $I'(t)$ is the rate of the cumulative incidence; parameter c indicates the growth rate, with $p = 0$, the growth rate becomes constant resulting a linear cumulative infection growth. In semi-logarithmic scale, the exponential growth pattern is visually evident with straight-line fit of the data after consecutive generations in the growth pattern onto the equation $log_e I(t) = bt + log_e a$, from equation 1, whereas, a concave-down shape of the curve would indicate a sub-exponential growth pattern.

2.2 Logistic Model

The logistic model is a basic epidemiological model with empirical touch, which is characterized by a sigmoid curve signifying an initial exponential growth until reaching an inflection point where the growth gradually slows down to saturation, resulting in the S-shaped profile of the dynamics [23]. It can roughly predict the development and transmission behavior of the outbreak through logistic regression analysis. The rate of cumulative infection, $I(t)$, according to logistic model, may be written as

$$I'(t) = \frac{k}{N} I(t)\big(N - I(t)\big) \qquad (2)$$

Here k represents the exponential growth rate, and N the total population size with $\lim_{t\to\infty} I(t) = N$. The solution of (2), with the boundary condition $I(t = 0) = I_0$ indicating the initial infection incidence, the equation (2) reduces to

$$I(t) = \frac{N}{1 + \left(\frac{N - I_0}{I_0}\right) e^{-kt}} \qquad (3)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

The model parameters in (3) are N, I_0 and k, and these can be ascertained with fitting and regression analysis against the dataset.

2.3 Gompertz Model

Traditionally employed for population growth, The Gompertz model is characterized by a sigmoid function describing growth as being slowest at the onset and at the end of a given time period, and also the asymmetric nature of the resulting S-shaped curve is in contrast to that of the logistic curve [24]. The slow onset behavior as described by the model has rendered it very suitable for modeling the infection growth pattern in Bangladesh. The mathematical form of the Gompertz model may be written as

$$I(t) = ae^{-be^{-ct}} \qquad (4)$$

Here, the parameters are a, b and c, with a being an asymptote, $\lim_{t\to\infty} I(t) = a = N$, the size of the population; b modulates the translation of the curve along the time axis; c is the growth rate of the incidence.

2.4 Richards Model

The logistic and Gompertz models expound the growth dynamics in profile of a S-shaped sigmoid curve with fixed growth rate due to the rigid model parameters. Richards model, on the other hand, offers flexibility in the growth profile incorporating additional modulating parameters in the form of exponent, resulting in a power-law formulation of the original logistic model [25]. This renders the model very effective for application in the current COVID-19 outbreak, as the infection incidences in the onset period in Bangladesh is observed to be very random and irregular. The Richards differential equation model for epidemiological application is by

$$I'(t) = rI(t)\left(1 - \left(\frac{I(t)}{N}\right)^{s}\right) \qquad (5)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

Here, r represents the intrinsic growth rate of the epidemic, and s is the modulating parameter offering more freedom in the bending of the S-shaped sigmoid curve. The solution to equation (5) with boundary condition, $I(t=0)=I_0$, and asymptote, $\lim_{t\to\infty} I(t)=N$, leads to the following equation

$$I(t)=\frac{N}{(1+Re^{-srt})^{\frac{1}{s}}} \qquad (6)$$

Here, $R=-1+\left(\frac{N}{I_0}\right)^s$; s and r are the fitting parameters. The model captures the observed exponential growth behavior at the early phase of the outbreak, when $I(t)\ll N$. Note the model with value $s=0$ reduces to its traditional logistic model.

2.5 Compartmental Model

The counterparts to the phenomenological models are the mechanistic models describing the diffusion process of the transmission of the epidemic within a population as a spatial-temporal dynamical mechanism. Typically the population, in such models, is sub-grouped as compartments consisting of individuals who are susceptible, infectious, or removed, because of recovery or death, along the temporal progression of the infection [26]. Each of these compartments are expressed by an ordinary differential equation with coefficients indicating the virulence of the pathogen, physical contact chances, symptomatic as well as asymptomatic interactions among population, and also the probability of transference of individual among different sub-groups giving rise to nonlinear coupled dynamical system features in the model. Thus, compartmental models are useful for forecasting short-term as well as long-term development of the epidemic, and for assessing various interventions, such as lockdown, to control the speed and trajectory of the outbreak among a population. Compartmental models also help by being a crucial theoretical tool to interrogate the available public health resilience with respect to the predicted infection profile, and the interventions may be modulated along the temporal phases to better respond to infection, and thus the severity of the outbreak can be mitigated. One of the underlying assumptions of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

model is the homogeneity of the population with infection prevalence. The high spatial population density of this study case renders compartmental model especially suitable to explore the onset epidemiological states of the population and follow the development. The mechanistic model variant used in this study is the traditional SIR (susceptible-infectious-removed) compartmental model [27, 28] governed by the following system of differential equations

$$\frac{\partial S}{\partial t} = -\beta SI \qquad (7)$$

$$\frac{\partial I}{\partial t} = \beta SI - \gamma I \qquad (8)$$

$$\frac{\partial R}{\partial t} = \gamma I \qquad (9)$$

Here, the three-compartment variables, denoted by (S, I, R), evolves spatial-temporally by the dynamic systems (7) – (9) preserving the constraint for the total host population, $N = S(t) + I(t) + R(t)$, and β and γ are the non-negative parameters: β indicates the transmission rate per infectious individual, while γ is the recovery rate, so the infectious period is exponentially distributed with the mean $\frac{1}{\gamma}$. At the onset of the endemic growth phase, for a completely susceptible population indicated by $S(0) \sim N$, the equation (8) can be solved analytically to incur the exponential form $I(t) \sim I(0)e^{(\beta-\gamma)t}$. The product of the transmission rate and the recovery period gives the average number of secondary incidences generated by an individual primary infection, which is referred to as the basic reproduction number, R_0 (*R*- naught). In the late part of study, the reproduction number is estimated from the initial infection growth data of Bangladesh. Basic reproduction number can also be used to estimate another useful observable known as the herd immunity threshold (HIT), indicating the fraction of the susceptible population required to undergo immunization for the infection to die away statistically either by vaccine application or natural antibody generation. This value is also calculated for the case nation in later part.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

2.6. Model Performance Metric

The data obtained for the study is in the form of a time series infection incidence with a temporal resolution of 1 day. The observations are collected in a longitudinal vector *(I_1, I_2, I_3, … .. .., I_t, … ..)*, where *t* is the outbreak generation or simply time and I_t's are the related cumulative infection values. In the numerical computation the successive incidences are updated as $I_{t+1} = I_t + \Delta I_t \Delta t$, where ΔI_t is the increment rate in infection, and Δt is the time step. Likewise, the models, after fitting with the dataset and calibration of the respective fitting parameters, yield a vector comprised of predicted cumulative infection incidences, may be denoted as $\widehat{I_t}$. Following the regression analysis of the model against the data, the sum of squares of the residual values, comprised of the difference between the observed and predicted values, is minimized to measure the goodness-fit of the model. As the performance metric, R^2 value [29] or the regression coefficient is evaluated based on the following mathematical form

$$R^2 = 1 - \frac{\Sigma\left(I_t - \widehat{I_t}\right)^2}{\Sigma(I_t - \bar{I})^2} \qquad (10)$$

Here, $\bar{I}$ is the average of the actual cumulative confirmed incidences. It is obvious from the relation that a near unity value for the metric would imply an accurate prediction and so higher performance of the evaluating model. Also, the 95% confidence interval (CI) is estimated based on the t-values for the evaluation of the model fit with the reported data as well as to ascertain the model projection performances.

2.7 Data and Tools

Data exploited for analyses and predictions were extracted from the Institute of Epidemiology, Disease Control and Research (IEDCR) [9], a government research institute, and the Directorate General of Health Services (DGHS) [10], an agency of the government, both under the Ministry of Health and Family Welfare, Bangladesh. The World Health Organization's situation reports on COVID-19 [30, 31] were also followed as the secondary source, and this was used as a cross-check for any irregularities. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

primary fields of the data as resourced from IEDCR and DGHS, containing the daily infection number along with total test counts conducted, daily mortality, and recovery numbers.

The data configuration and management were done in MS Excel 2016, and the model simulation and regression analyses were conducted in Python 3 [32, 33] on Windows 10 Pro. The hardware for the study was Intel(R) Core(TM) i3 processor with clock-rate 2.70 GHz.

3 Results and Discussion

3.1 Empirical Growth Trend

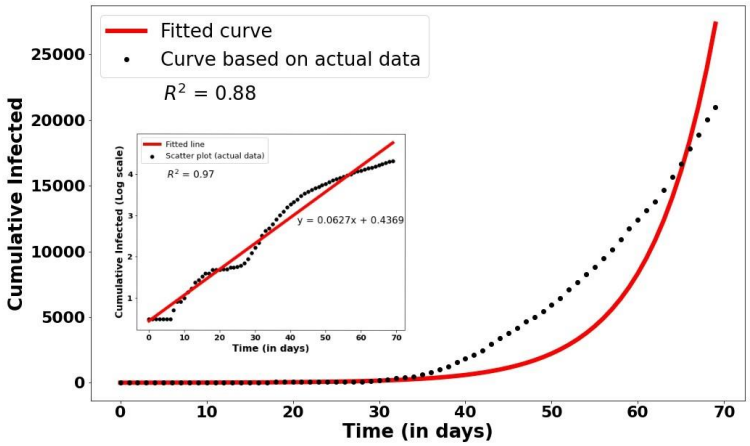

Figure 1: The COVID-19 cumulative infection incidences (back dots) of Bangladesh show exponential growth at the onset period and fitted to a phenomenological exponential function (equation1) with parameters: are $a = 3.0$, $b = 0.13214$. The semi-logarithmic plot, shown in the inset, posits an exponential growth feature of the data closely following linear trend (red).

For analysis, we use reported incidences data from 8th March to 15th May from the source as mentioned earlier. The onset transmission phase for a pandemic typically trends both sub-exponential and exponential pattern [11, 18, 22]. The exponential growth may alter to sub-exponential behavior over time indicating shift in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

transmission strength and development in a population. COVID-19 virus transmission in Bangladesh is observed to follow a sharp exponential feature. The main panel in Figure 1 shows both the data and exponential fitted graph, and the inset hosts the semi-logarithmic recasting of the cumulative infection resulting a straight line fit with $R^2 \sim 0.88$ estimated from equation 10. The observed deviation about the fitted curves and straight line is due to the sheer random nature of the dataset, though in the long progression to the outbreak the data appear to follow the generic exponential trend. According to the reported data, the COVID-19 outbreak in Bangladesh set off on 7th March with three incidences, and not until mid-April, the incidences did not really kick off. The reason behind the low number of incidences is due to the insufficient number of tests conducted in the country. However, the test number afterwards soared substantially on a daily basis, shown in Figure 2, along with the coronavirus positive incidences. As of the record on the 15th May, the cumulative figure stands at 20,065 individuals with positive incidence out of a total of 166,994 tests conducted, which measures to about 8.3% of coronavirus positive cases. The cumulative growth factor of the incidence in population, determined by the ratio of I_t to I_{t-1} where t is transmission generation period expressed in days, is computed to be approximately 1.05 as of reported data of 15th May.

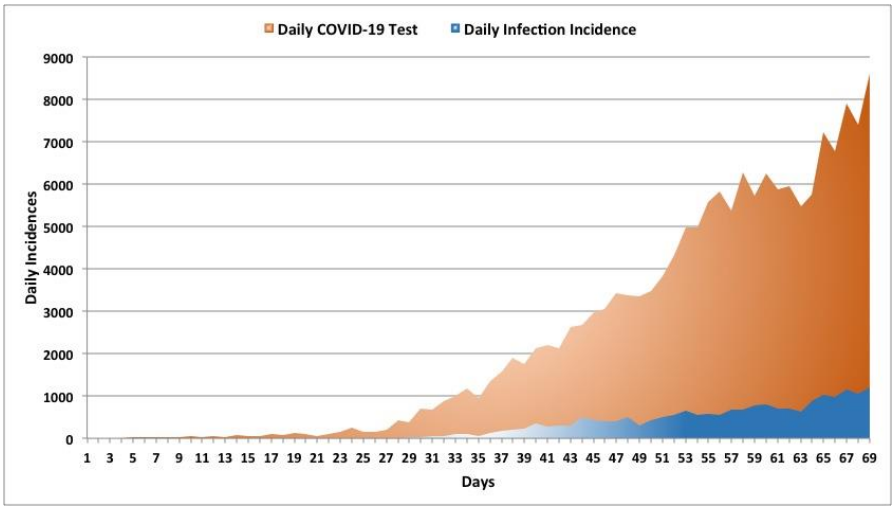

Figure 2: Comparative illustration of the daily positive COVID-19 incidences with respect to corresponding daily growing number of tests conducted throughout the country.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

3.2 Fitting and Analysis: Logistic Model

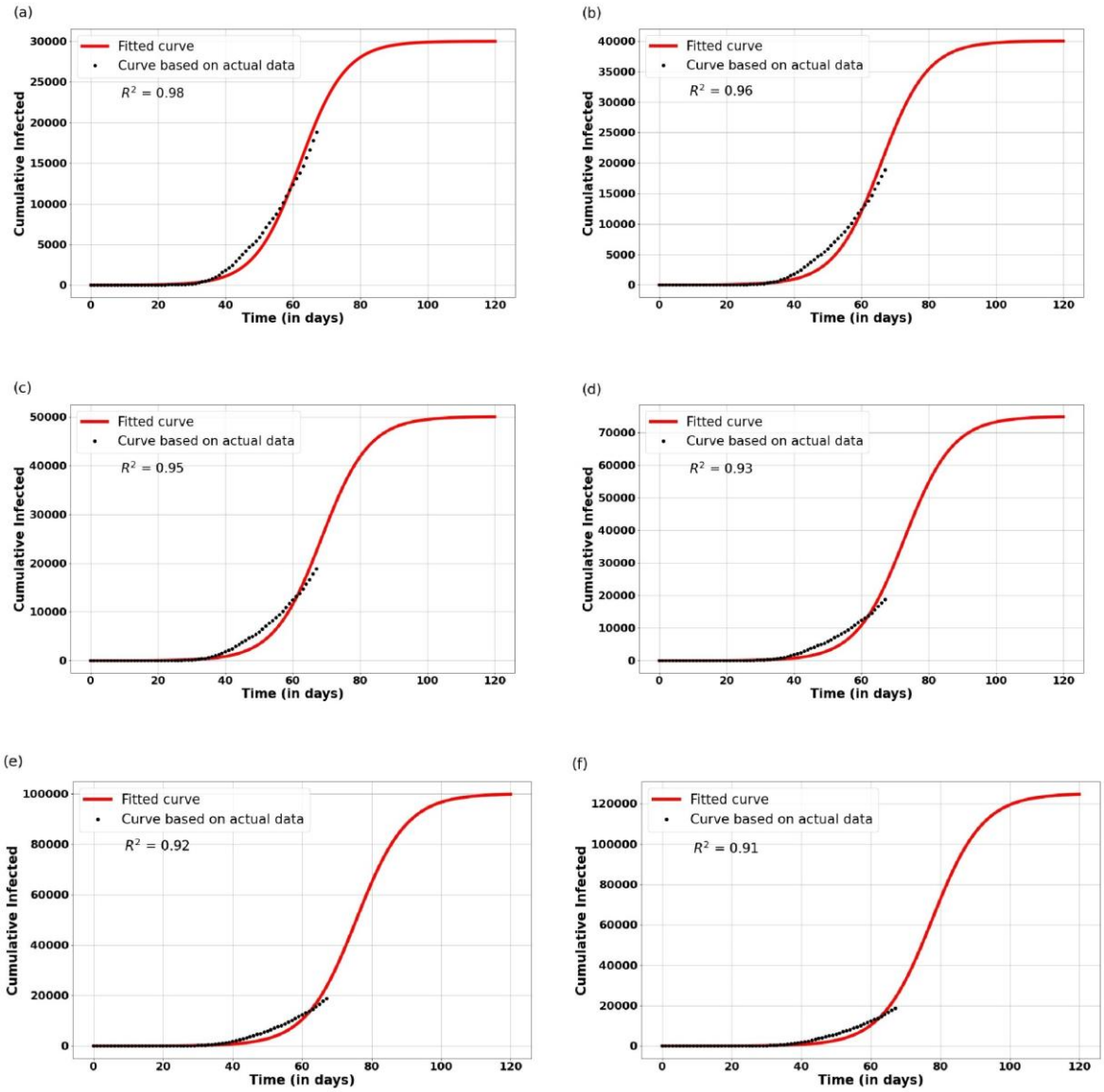

Figure 3: The realistic cumulative infection data has been fitted to the logistic model (equation 3) in a variety of variations involving carrying capacities, N: (a) 30K, (b) 40K, (c) 50K, (d) 75K, (e) 100K, and (f) 120K. The inflection point is observed to move rightward with the increasing carrying capacity, N, of the infection in the population implying increasingly longer endemic period, until infection dynamics see a slowing down. The corresponding regression coefficient, R^2, values after equation 10 represents the degree of goodness-of-fit.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

One characteristic feature of the logistic model described by equation 3 is the inflection point, positioned in the middle of the S-shaped solution profile, which indicates a functional alternation of the dynamics – following for the rising part, the portion of the curve preceding the inflection point indicates an accelerating exponential growth while the behavior past the inflection is of slowing to a saturation level, in population dynamics, is the carrying capacity, i.e., the size of the target population for the infection. Figure 3 panels demonstrate the fitting outcome of the reported data and the corresponding S-shaped growth pattern of the infection, for varying carrying capacities. The fits indicate the data accord with the accelerating growth part of the model indicating the early stage of the infection. The corresponding goodness-of-fit measurements, in Figure 3, indicate that it is possible to fit the data very efficiently in variety of configurations of the model parameters. Such degeneracy in the fitting implies that the logistic model is a good candidate to explore the onset transmission dynamics of a pandemic in contrast to its ability for the long-term projection of the transmission development. We note from the simulation results that Bangladesh is still to catch on with the inflection, which would result in the declining incidence rate. Figure 4 is the rate model generated by the solution of equation 2 with carrying capacity for 100K and other model parameters are found from fitting as $I(0) = 3$ and growth efficient parameter $k = 0.139065$. The bars are the corresponding daily infection incidence values as per the recorded data. The inset of the figure indicates that the inflection point moves farther with larger target population, implying that it will take longer to reach to a slowing down phase if more people are put into risk of infection. Also, the simulation result from the dynamical form of the logistic model (equation 2) offers an outlook to the future development of the pandemic by predicting the duration of the outbreak, which is approximately 140 days, given the fitted growth pattern and carrying capacity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

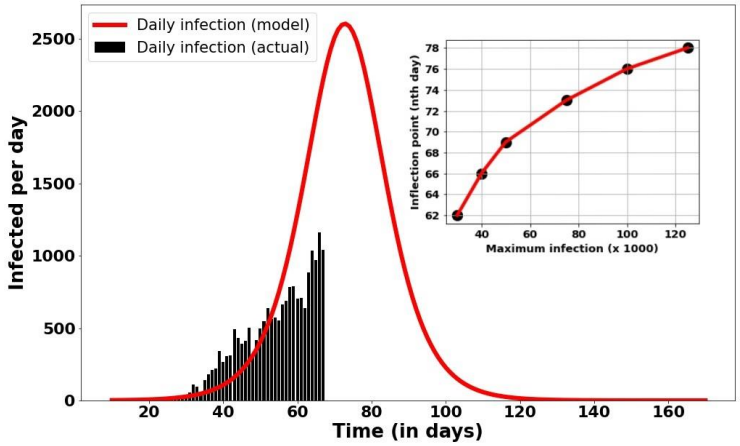

Figure 4: Overlay of the daily infection incidence (black bars) on the logistic derivative curve generated by equation 2 (red profile) for $N = 100K$ and the fitted parameters of Figure 3e. The maximum of the derivative profile, also known as the inflection point, indicates the daily maximum the outbreak project and at the corresponding day since the first report of the incidence; here it occurs approximately at day 75 of the pandemic. The inset illustrates the shifting of the inflection points with the increasing maximum infection size or the outbreak carrying capacity. The red lines joining the observables (black dots) are artificially added to guide the eyes.

3.3 Fitting and Analysis: Gompertz Model

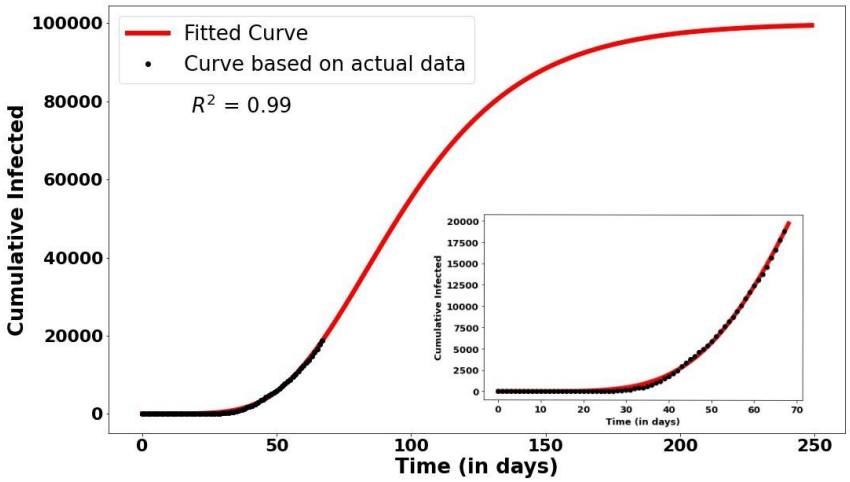

Figure 5: The real dataset fits the Gompertz model (equation 4) with high fidelity. The inset is the close up of the onset region reflecting the high correlation, $R^2 \sim 0.99$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

(equation 10) between the data (black dots) and the model fit; the parameters are $a = 100000$, $b = 13.96191598$, and $c = 0.03171862$.

The Gompertz model, borne out of population dynamical family of models, offers more flexibility in the S-shape with modulating parameters a, b, and c (equation 4), endowing more degrees of freedom in the fitting and simulation. Figure 5 shows the outcome of the model fit with the Bangladesh COVID-19 virus transmission data shown in black dots. The goodness-of-fit indicator, R^2, returns an almost perfect score of the fit. Unlike the logistic model, Gompertz model is not dependent on the carrying capacity and also the asymmetric profile of quick rise and longer period to slow down past the inflection is akin to what the observed pandemic transmission dynamics suggests.

3.4 Fitting and Analysis: Richards Model

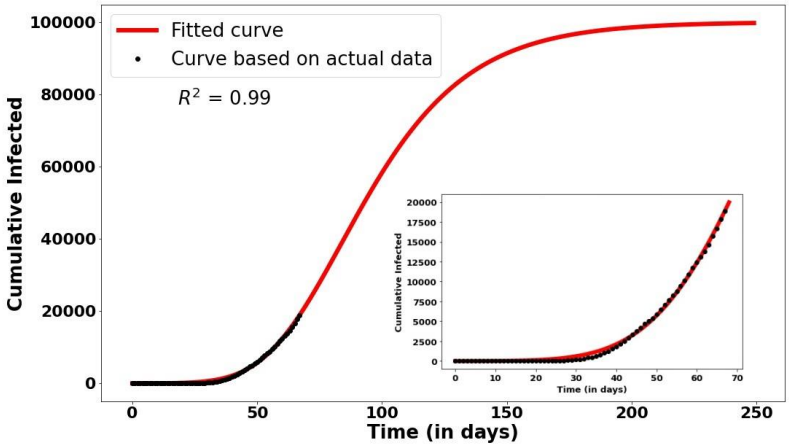

Figure 6: The Richards model reproduces the cumulative infection at the onset with respect to the read data (black dots) with an astonishing regression coefficient value 0.99 (after equation 10). The model parameters: $N = 100000$, $r = 0.29493425$, and $s = 0.12345197$. The large panel shows the solution to equation 5 yielding the characteristic S-shaped profile (in red) governed by equation 6. The inset zooms in the realistic data overlay at the onset to visualize the fit clearly.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

We observe a very similar high performance of the Richards model when trained by transmission incidence data of Bangladesh. Richards model, which is a generalized formulation of the logistic model (equation 2 and 3), with a distinguishing exponent s, which enables the model to be more flexible, like Gompertz model (equation 4), and allows reflecting the realistic data with trained model parameters. Our simulation estimates the value of the exponent parameter, $s = 0.12345$, for the early growth phase of the pandemic in Bangladesh. Also, like its logistic counterpart, Richards model incorporates the size of the host population.

3.5 Compartmental Model Fitting

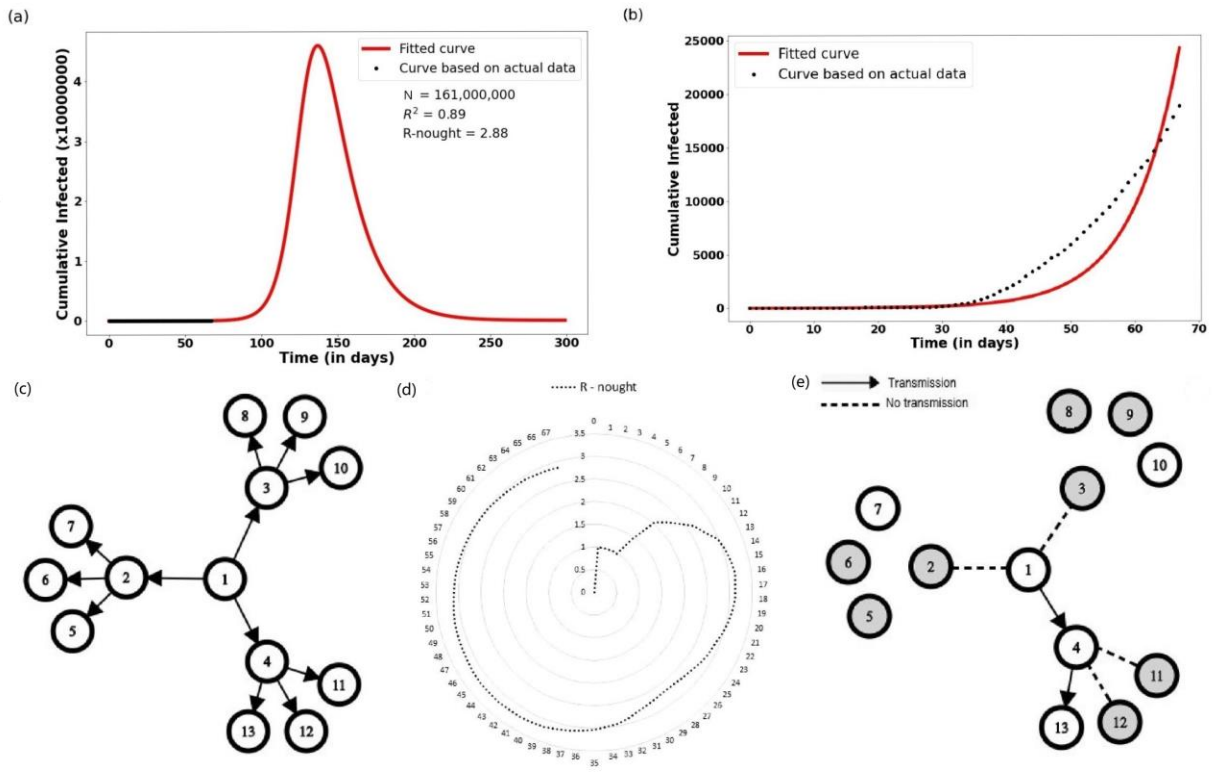

Figure 7: (a) The cumulative infection incidences $I(t)$, governed by SIR model (equation 8), are fitted to the observation for an entire susceptible population of 161 million, the model parameters thus obtained are the transmission rate, $\beta = 0.2057$, and the infection period $\frac{1}{\gamma} = 14$ days. The corresponding basic reproduction number is estimated as 2.88. The model offers a rather conservative projection to the future

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

development of the infection by indicating a die-down phase after 250 days from inception on March 7. The panel (b) is the blow-out of the onset with overlay of data (dotted) and the model fit (line), the regression coefficient for the fit yields $R^2 \sim 0.89$. The panel (c) illustrates in the cartoon how the secondary infection propagate (arrows) for a case with $R_0 = 3.0$, where there are three emergent secondary incidences (circles) from single primary infected individual in a homogenous population in successive generation of the infection. The panel (d) demonstrates the temporal progression of the estimated R_0 of Bangladesh; the estimated values are shown in the dynamic spiral, and evidently the current estimate has decreased to 2.8 from a little over 3.0 at the very onset of the infection. Panel (e) portrays the concept of Herd immunity Index (HIT), estimated as 2/3 for a case of $R_0 = 3.0$; the cartoon illustrates that the infection eventually dies out if two-third of the population grow an immunity to COVID-19, so the contagious transmission (solid arrow) stay within a small portion of the population and fail to affect most others (broken arrows).

Compartmental models are traditionally applied to understand and analyze various aspects ranging from infection pattern, recovery, and latency to effectiveness of mitigating intervention measures and future course of propagation of viral diffusion in a host population. In our study, we use classic formalism of mechanically intertwined susceptible, infectious and removed compartments to model the infection progression based on the incidence data in Bangladesh. High population density of Bangladesh satisfies the homogeneity condition underlying the formulation of the SIR model (equations 7–9). Figure 7(a) and 7(b) show the result of fitting the reported infection incidences in Bangladesh with the model both visually and numerically. In the simulation, we consider the entire population as the susceptible host. In the onset phase, the underlying condition, $S(0) \sim N$, leads to the prevalent exponential growth pattern: $I(t) \sim I(0) e^{(\beta - \gamma)t}$. According to data-trained estimation, the current transmission rate (β) of COVID-19 virus in Bangladesh is 0.21, and the infection period is set to 14 days as per the global observation [3, 4]. Though the simulation is tailored to capture the early-growth dynamics of the transmission, the simulations offer tentative insight into the future projection of the pandemic, given the current rates subsist. Figure 7(a) implies a 250-day duration window for the pandemic in the country, though it is observed that the parameters may change due applied interventions and also due to the intrinsic changes in the variables. Figure 8 depicts the degeneracy in the SIR model, where the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

reported data may train the same model and fit quite efficiently in every scenario with different host population sizes yielding different parameter values. So, depending on the size of the target population hosting the infection, the initial growth trend may lead to different projected outcomes.

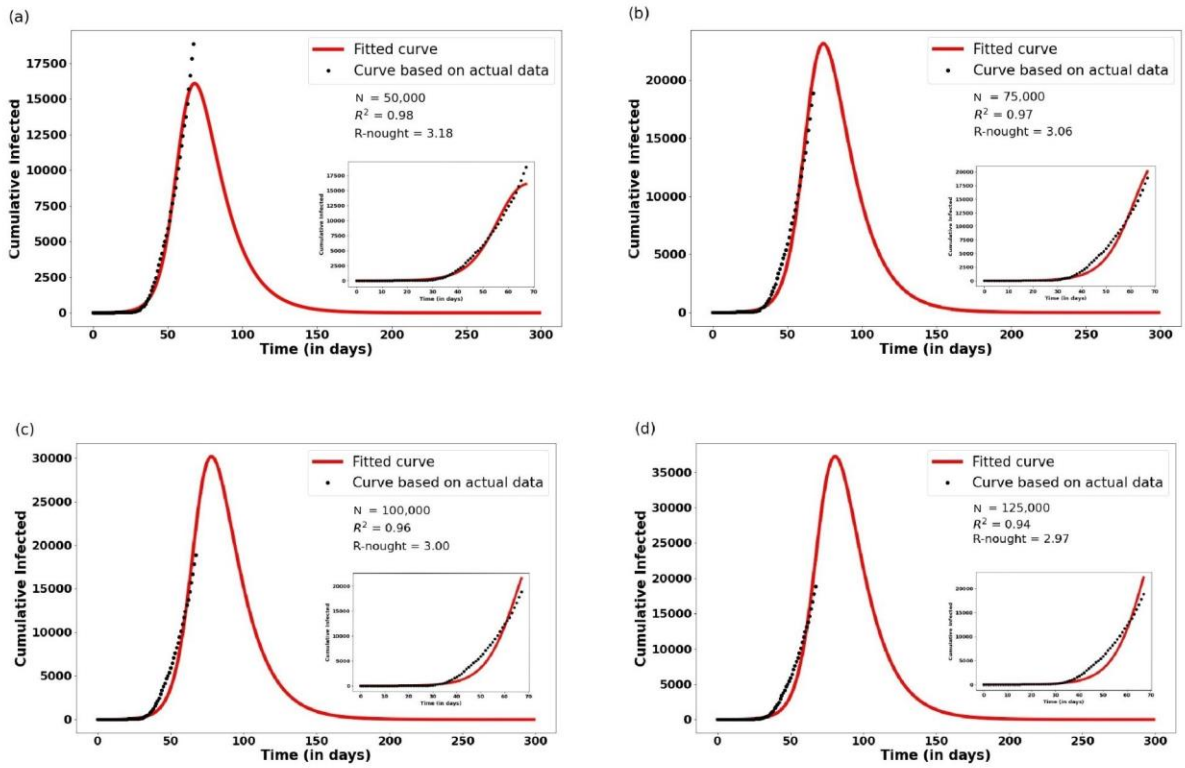

Figure 8: This panel cluster illustrates the degeneracy involving modeling the early exponential phase of the infection, where the same data may be fitted to a high degree of accuracy to a multiple scenario governed by the SIR-Model across different points in parameter space. Across the panels (a)–(d) the vulnerable population size has been varied from 50K, 75K, 100K and 125K, respectively. Notably in each case the fit of the data (black dots) to the model generation (red profile) score high on evaluation scale (equation 10) and also yield varying degree of secondary infection proliferation (R_0). The insets are the close-ups of the corresponding data regions. This observation implies that the model projection is constrained to short temporal scale.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

3.5.1 Basic Reproduction Number (R_0)

Compartmental model offers a quantitative way to compute the basic reproduction number or R_0 (R-naught) of the prevalent COVID-10 pandemic in a host population. Basic reproduction number is an important gauge to ascertain the virulence of an infection in a population and a risk assessment tool; it is expressed as the average secondary infection size from the one primary infected individual. Figure 7(c) illustrates the growing propagation envelope for a case $R_0 = 3.0$, where one infected individual infect three others, and each one of them infects three more and so on in the successive generation of infection. This leads to the ubiquitous exponential growth observed during an outbreak. In the early growth phase, when initial number of infected individual ($I(0)$) is far small than the total susceptible population leading to an approximation $S(0) \sim N$, the basic reproduction number can be computed simply by multiplying the transmission rate (β) with the recovery period ($\frac{1}{\gamma}$). From simulation (Figure 7(a, b)), we compute the basic reproduction number for Bangladesh is 2.88, which is well within the reported range of 1.4 to 6.5 for COVID-19 globally [34, 35]. We also track the change of R_0 with time in the country, the spiral in Figure 7(d) indicates the value is shifting slowly towards desired unit value, starting from over 3.0 initially, which would imply a cease to the exponential growth phase to constant propagation, and a value less than 1 would mean an eventual dying out of the pandemic. In the course development of the pandemic, as the number of infected individual increases and likewise the susceptible individual count declines, effective reproduction number, R_e, computed by $R_e = \frac{S(t)}{N} R_0$, is employed.

3.5.2 Herd Immunity Threshold

Herd immunity threshold (HIT) represents the fraction of host population that needs to undergo immunization against the virus to potentially neutralize the infection [36, 37, 38]. The fraction may be calculated in terms of basic reproduction number given by $1 - \frac{1}{R_0}$. Figure 7(e) illustrates the functional concept of HIT for a case with $R_0 = 3.0$; the infection eventually eradicated as $\frac{2}{3}$ of the population is immunized, i.e., the infection

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

fails to affect 2 individuals out of 3 incidences per infection generation, resulting in the decline of the pandemic. According to estimation from current reported data as of mid-May at the time of writing, Bangladesh, to reach immunity, 65.3% of the population must be vaccinated or treated with antiviral drugs to eradicate the current pandemic in the country. It should be noted that the HIT value would change with temporal progression of infection due to its dependency on varying basic reproduction number.

3.5.3 Doubling Period

Doubling period is a useful long-term metric to probe the temporal developmental pattern of a pandemic. Doubling period of a pandemic, as the name suggests, is the time period taken for the infection to get double with respect to the baseline, mathematically, it can be computed as: for the early-growth, i.e., $S(0) \sim N$, then doubling the infection would imply that $I(t_d) = I(0)e^{(\beta-\gamma)t_d} = 2I(0)$ following from equation 8, where t_d indicates the doubling period of the infection; solving this results the mathematical equation to compute doubling time as $t_d = \frac{\ln 2}{(\beta-\gamma)} = \frac{\ln 2}{\gamma(R_0-1)}$. For COVID-19 in Bangladesh, we compute the latest doubling period for the study period to be 5.16 days, which is comparable to the early stage doubling periods in global cases [39].

3.5.4 The Maximum Size of the Infection

In this section, we estimate the projected maximum size of infection, I_{max}, in the host population based on the data and trained parameters from model fitting. At the maximum infection, $\frac{dI}{dt} = 0$, in the equation 8, which gives $S = \frac{\gamma}{\beta} = \frac{1}{R_0}$, also dividing equation 8 and 9 yields $\frac{dS}{dI} = \frac{-\beta SI}{\beta SI - \gamma I}$, which may be recast as $I(t) + S(t) - \frac{\gamma}{\beta}\ln(S(t)) = I(0) + S(0) - \frac{\gamma}{\beta}\ln(S(0))$. Now, considering the entire host population as susceptible, $S(0) \sim N$, these equations may be manipulated to yield an expression for maximum infection: $\frac{I_{max}}{N} = 1 - \frac{1}{R_0}(1 + \ln(R_0))$. We use this expression to calculate the maximum infection size for Bangladesh, and we report, based on the calculated basic

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

reproduction number estimated earlier, 28.5% of the total population or 46 million people are at the risk of the infection if the current rate of transmission propagated unabated.

3.6 Comparative Analysis of Model Performance

In this study, we use both phenomenological and mechanical models to train and test for the on-going COVID-19 infection trend in Bangladesh. In our simulation, we find the Gompertz and Richards models perform the best in capturing the early growth trend as observed in the reported data. Table 1 summarizes the key comparative features of the models. And Figure 9 depicts a closer look at the fitting and projection capabilities of the best performing Gompertz and Richards models found in this study. The simulations for the two models have been projected to the 90th day beyond modeled 62 days' observation period. The 95% confidence interval shown by the upper and lower bounds in the results imply the accuracy of the models both in the fitting area and the projection, along the future generation of infection in the host population.

Table 1: Performance summary of the early-growth models

Model Type	Models	Governing Growth Profile	Carrying Capacity (N)	Goodness of Fit (R^2)
Phenomenological	Simple Exponential	$I(t) = ae^{bt}$	–	0.88
	Logistic	$I(t) = \frac{N}{1 + \left(\frac{N - I_0}{I_0}\right) e^{-kt}}$	100K	0.92
	Gompertz	$I(t) = ae^{-be^{-ct}}$	–	0.99
	Richards	$I(t) = \frac{N}{(1 + Re^{-srt})^{\frac{1}{s}}}$	100K	0.99
Mechanical	Compartmental	$I(t) \sim I(0)e^{(\beta - \gamma)t}$	161 Million	0.89

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

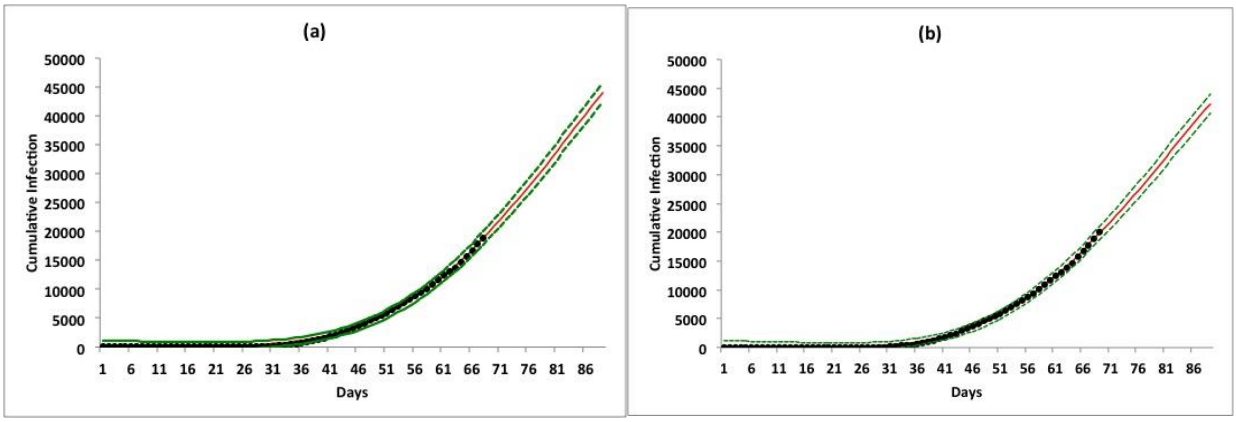

Figure 9: Richards (a) and Gompertz (b) model Performance Panels (a) and (b) are the visual demonstration of the model fit accuracy and projection performance of Richards and Gompertz models, respectively. The black dots are the real data seen closely superimposed on the red line generated by the calibrated Richard (equation 5 with fitted parameters as in Figure 6) and Gompertz (equation 4 with the model parameters as in Figure 5) models. The bounding contour around the red model generation indicates the 95% confidence interval (CI) of the future projection by the model down the infection progression. The narrower CI along the dataset on the uplifting stage indicates the high degree accuracy in the model fitting, and also in the projection until 90-day period shown in the computation, holds to a high accuracy.

4 Conclusions and Outlook

Bangladesh continues to see an unabated exponential growth of the COVID-19 infection, even during the preparation of the paper, after 60 days since the reported case in the country, though the suppressive intervention measures are in place. So, dynamic epidemiological modeling approach to understand the underlying nature and peculiarities of the infection diffusion dynamics in the country is imperative to ascertain and to interrogate the effectiveness and defectiveness of interventions [40, 41]. Our study explores the available real data in this early phase through the pandemic, despite the inevitable constraints involving the size and scope of the data, and attempts to depict a scale up assessment of the pandemic throughout the country, employing both phenomenological and mechanistic models. The spatial-temporal dynamics explored in this study can be generalized beyond the premises considered, and may be used to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2020 doi:10.20944/preprints202005.0372.v1

understand the pandemic in other countries, especially those sharing similar demography. We may earmark few key conclusions of the study as following:

1. The pandemic growth in the country paves a strong exponential growth pattern.
2. The study finds population dynamical models based on the power-law featured growth, such as Gompertz model and the generalized logistic model counterpart, Richards model, perform very well in the onset phase of the pandemic, and are very successful in capturing the current infection pattern as well as projecting the future development of the pandemic in short temporal range.
3. The study posits the basic reproduction number, R_0, of current the COVID-19 outbreak in Bangladesh stands at 2.88, at the time of the write up, and the study shows the deceleration trend in the growth reflected in the observed decrease in the R_0 over time.
4. The Herd Immunity Threshold (HIT) estimated in the study asserts that 65.3% of the population needs to achieve immunity in order to get out of the pandemic.
5. The study finds that the mechanical model, based on the current available data, estimates the doubling period for the infection in Bangladesh is 5.16 days.
6. The study estimates the upper bound of the total infection size in Bangladesh; according to the estimation, the current growth trend projects about 45.9 million people may be infected by coronavirus.

The epidemiological modeling performance study conducted here offers a crucial insight into the dynamical features and numerical measures of the outbreak size in Bangladesh, which can be used as guiding tools to assess the responses and outcome by continuous monitoring of the situation. The basic reproduction number is a crucial indicator of any pandemic, and it must be computed and monitored on a regular basis to undertake and assess the suppressive and mitigating phases during the pandemic, more so at the onset phase, as the response interventions to safe-guard both life and economy of the country. The theoretical modeling and simulation as pursued in this study is an important step towards development of automation-based tools and machine learning techniques (e.g., [42]) to further the understanding of dynamical aspects of the outbreak and gauge various response protocols.

References

1. World Health Organization's Official COVI-19 Dashboard: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

2. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance

3. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill. 25, 2000180 (2020). doi:10.2807/1560-7917.ES.2020.25.10.2000180pmid:32183930

4. YU, I., & SUNG, J. (2004). The epidemiology of the outbreak of severe acute respiratory syndrome (SARS) in Hong Kong – what we do know and what we don't. *Epidemiology and Infection, 132*(5), 781-786. doi:10.1017/S0950268804002614

5. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020; 579: 265–269. doi: 10.1038/s41586-020-2008-3. pmid:32015508

6. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro Surveill. 2020;25(4):2000058

7. "World Population prospects – Population division" - *population.un.org.* United Nations Department of Economic and Social Affairs*, Population Division*. Retrieved 9 November 2019*.*

8. "Overall total population" – World Population Prospects: The 2019 Revision"*.* population.un.org. United Nations Department of Economic and Social Affairs*,* Population Division*.* Retrieved 9 November 2019*.*

9. Institute of Epidemiology, Disease Control and Research (IEDCR), official website: http://iedcr.gov.bd/index.php/component/content/article/11-others/227-pressrelease

10. Directorate General of Health Services (DGHS), Ministry of Heath and Family Welfare, Government of People's Republic of Bangladesh: https://dghs.gov.bd/index.php/en/home/5373-novel-coronavirus-covid-19-press-release

11. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. 2020;382(13):1199–207.

12. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. https://doi.org/10.1016/S1473-3099(20)30243-7.

13. Tuite A, Fisman DN, Greer AL. Mathematical modeling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. *medRxiv* 2020. https://doi.org/10.1503/cmaj.200476 .

14.Yuan D F, Ying L Y, Dong C Z. Research Progress on Epidemic Early Warning Model [J]. Zhejiang Preventive Medicine, 2012(08): 20-24+27.

15. Zhang F, Li L, Xuan H Y. Overview of infectious disease transmission models [J]. Theory and Practice of Systems Engineering, 2011, 31(9):1736-1744.

16. Grassly N C,Fraser C.Mathematical models of infectious disease transmission[J].Nature Reviews Microbiology 2008,6(6):477 487

17 Li Z L, Zhang LM. Mathematical Model of SARS Prediction and Its Research Progress [J]. Journal of Mathematical Medicine, 2004, 17(6):481-484.

18 Ma, J., et al. (2014). Estimating initial epidemic growth rates. Bulletin of Mathematical Biology, 76(1), 245e260.

19. Chowell, G., et al. (2016a). Mathematical models to characterize early epidemic growth: A review. Physics of Life Reviews, (16), 30064e30071. pii: S1571-0645.

20. Chowell, G., et al. (2016b). Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics. PLOS Currents Outbreaks. http://dx.doi.org/10.1371/currents.outbreaks.f14b2217c902f453d9320a43a35b9583.

21. Chowell, G., et al. (2016c). Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics. Journal of The Royal Society Interface, 13(123). p. pii: 20160659.

22. Chowell, G., & Viboud, C. (2016). Is it growing exponentially fast? Impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics. Infectious Disease Modelling, 1(1), 71e78.

23. Sun Y, Koh V, Marimuthu K, et al. Epidemiological and Clinical Predictors of COVID-19 [Preprint, 2020 Mar 25]. *Clin Infect Dis*. 2020;ciaa322. doi:10.1093/cid/ciaa322

24. Gompertz, B. (1825). "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies". Philosophical Transactions of the Royal Society of London. 115: 513–583. doi:10.1098/rstl.1825.0026. JSTOR 107756.

25. Richards, F. J. (1959). "A Flexible Growth Function for Empirical Use". Journal of Experimental Botany. 10 (2): 290–300. doi:10.1093/jxb/10.2.290.

26. Blackwood JC, Childs LM. An introduction to compartmental modeling for the budding infectious disease modeler. Lett Biomath. 2018;5(1):195–221.

27. Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society A. 115 (772): 700–721.

28. Harko, Tiberiu; Lobo, Francisco S. N.; Mak, M. K. (2014). "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates". Applied Mathematics and Computation. 236: 184–194.

29. Colin Cameron, A.; Windmeijer, Frank A.G. (1997). "An R-squared measure of goodness of fit for some common nonlinear regression models". Journal of Econometrics. **77** (2): 1790–2. *doi:10.1016/S0304-4076(96)01818-0*

30. World Health Organization, Situation Reports on Coronavirus Disease 2019 (COVID-19): https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

31. The Worldometer. COVID-19 CORONAVIRUS PANDEMIC. 2020. https://www.worldometers.info/coronavirus/.

32. Oliphant TE. Python for scientific computing. Comput Sci Eng. 2007;9(3):10–20.

33. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng. 2007;9(3):90–5.

34. S. Zhao et. al., Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak, International Journal of Infectious Diseases 92, P214-217 (2020), https : //doi.org/10.1016/j.ijid.2020.01.050.

35 A. Lai et. al., Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2, Journal of Medical Virology, 25 February 2020, https : //doi.org/10.1002/jmv.25723.

36. Fine P, Eames K, Heymann DL. "Herd immunity": a rough guide. *Clin Infect Dis*. 2011;52(7):911-916. doi:10.1093/cid/cir007

37. Fox JP, Elveback L, Scott W, et al. Herd immunity: basic concept and relevance to public health immunization practices. Am J Epidemiol1971; 94:179–89.

38. Anderson RM, May RM. Vaccination and herd immunity to infectious diseases. Nature1985; 318:323–9.

39. K. Muniz-Rodriguez et. al., Epidemic doubling time of the COVID-19 epidemic by Chinese province, medRxiv 2020, DOI : https : //doi.org/10.1101/2020.02.05.20020750.

40. N. M. Ferguson et. al., Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, Imperial College COVID-19 Response Team, 16 March 2020, DOI:10.25561/77482.

41. Chowdhury, R., Heng, K., Shawon, M.S.R. *et al.* Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. *Eur J Epidemiol* (2020). https://doi.org/10.1007/s10654-020-00649-w

42. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. 2020; 368(6491): eabb6936. doi:10.1126/science.abb6936