Prediction of CoVid-19 infection, transmission and recovery rates: A new analysis and global societal comparisons

Romney Duffey, Enrico Zio

Version of Record: https://www.sciencedirect.com/science/article/pii/S0925753520302514
Manuscript_d01604e18cce174f3a7144cad025fa07

Prediction of CoVid-19 infection, transmission and recovery rates : a new analysis and global societal comparisons

Romney B Duffey[1] and Enrico Zio[2]

[1]Idaho Falls, Idaho, USA
[2]Centre de Recherche sur les Risques et les Crises (CRC), MINES ParisTech/PSL Université Paris, Sophia Antipolis, France.
Department of Energy, Politecnico di Milano, Milan, Italy.
Eminent Scholar at the Department of Nuclear Engineering, Kyung Hee University, Seoul, South Korea.

Abstract

We analyze the process of infection rate growth and decline for the recent global pandemic, applying a new method to the available global data. We describe and utilize an original approach based on statistical physics to predict the societal transmission timescale and the universal recovery trajectory resulting from the countermeasures implemented in entire societies. We compare the whole-society infection growth rates for many countries and local regions, to illustrate the common physical and mathematical basis for the viral spread and infection rate reduction, and validate the theory and resulting correlations. We show that methods traditionally considered for the numerical analysis and the control of individual virus transmission (e.g. R_0 scaling) represent one special case of the theory, and also compare our results to the available IHME computer model outcomes. We proceed to illustrate several interesting features of the different approaches to the mitigation of the pandemic, related to social isolation and "lockdown" tactics. Finally, we use presently available data from many countries to make actual predictions of the time needed for securing minimum infection rates in the future, highlighting the differences that emerge between isolated "islands" and mobile cities, and identifying the desired overall recovery trajectory.

1. Introduction

Traditional pandemic modeling and epidemiological methods are based on the concept of viral spread due to person-to-person contact, via an empirical transmission parameter, R_O, which is adjusted to fit dynamic (time-varying) infection data. We develop a new way to predict the recovery rate of infections following a pandemic outbreak, using the basic postulates of statistical physics and learning theory. We differ completely from the prior published work on the topic (see e.g. Kucharski et al., 2020) by using infection rates, statistical physics theory and comparisons between multiple countries and regions, to make universal predictions of transmission/incubation timescales and recovery trajectories. Our objective is to improve predictive accuracy based on data, by providing an alternative basis for the physical modeling of pandemics in entire societies.

The fundamental postulate is that humans learn from experience in correcting their mistakes and errors (sometimes even just by trial and error), to reduce undesirable outcomes, as they gain knowledge on the problem and skills for addressing it. Hence, in the specific coronavirus case, infection growth represents failure to learn to control or manage the outbreak, whereas infection decrease reflects the effectiveness of the countermeasures – whatever they may be – as devised and implemented following the learning process and the resulting knowledge acquired on the virus and its mechanisms of diffusion. The theory is consistent with the models and data in cognitive psychology of how humans behave and the brain operates (Ohlsson, 1996; Fiondella and Duffey, 2015; Anderson, 1990; Duffey, 2017). The learning process is reflected in the overall observed societal behavior where human errors and incorrect decisions are the dominant contributors to accidents, crashes, system failures, errors, and operational incidents, just as they are in failing (forgetting) or succeeding (learning) in containing and managing a viral pandemic (Duffey and Zio, 2020).

By mid-2020, the world has counted over 4,000,000 cases of CoVid-19 infection (and still growing at the time of writing) and over several 100 thousands of deaths, with the infection having spread quickly across borders, imported from nation-to-nation mainly via travelers, visitors and tourists, and spread internally from just social and day-to-day human contact. In this regard, the US Centers for Disease Control correctly states the simple obvious: “Risk depends on characteristics of the virus, including how well it spreads between people” (Source: www.cdc.gov/coronavirus/2019-ncov/summary), which is what we examine in this paper.

The early onset of the CoVid-19 pandemic led to the postulation of many gloomy scenarios, generally assuming no effective countermeasures to the spread of infections. The infection numbers grew quickly at first, before countermeasures, such as isolation, distancing, mobility and business restrictions, and curfews, were implemented to reduce infection rates and “flatten the curve” of count numbers versus time. The characteristic shape is well demonstrated, for example, by the data from Italy for 70 days after the number of infections reached an initial value of 100 infection cases (see Figure 1).

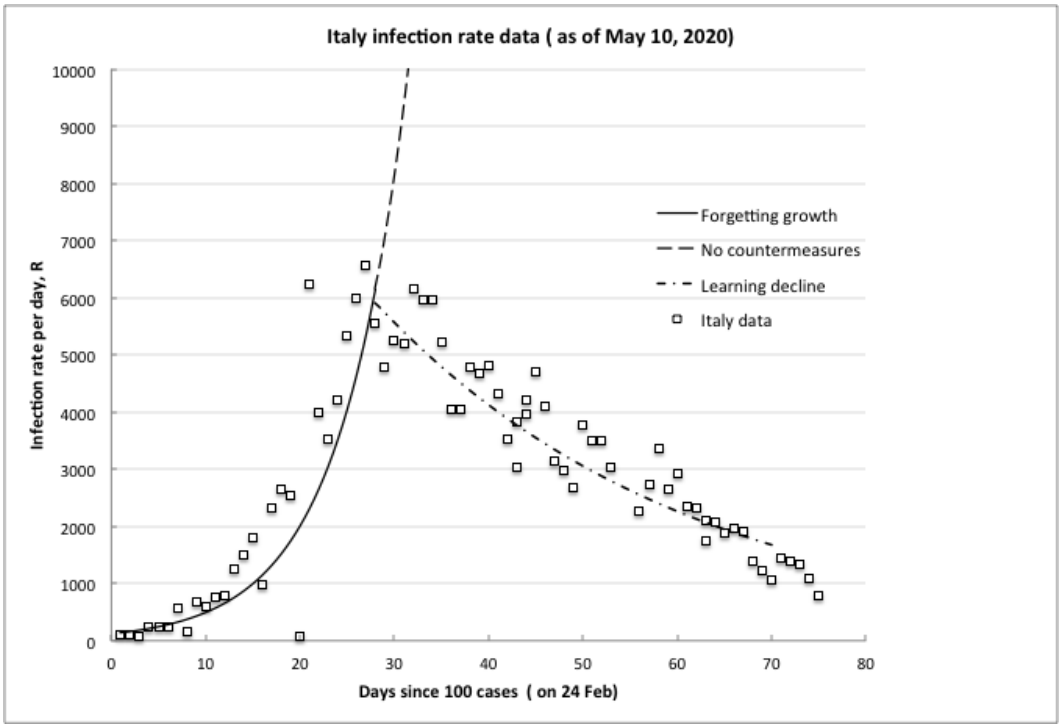

Figure 1 Typical observed trends in infection rate growth and decline, from the day of 100 recorded infection cases: the Italian experience

In any accident (e.g. a bridge collapse, a car crash, a toxic gas spill from a processing industry, etc.) or natural disaster (e.g. an earthquake, a flood, a forest fire), the numbers of persons killed or injured are highly variable, depending on who and how many risk-exposed individuals happen to be present in the area at the moment of the accident and on the accident physical dynamics. This depends on so many factors that the occurrence and outcome of the accident or disaster can be considered random. In the case of the virus spread also, the number of deaths depends on too many uncontrolled variables and factors (age, pre-existing health conditions, health care system capacity, individual propensity, socio-economic status etc.), so that the average death percentage per infection also varies widely in magnitude, location and time (as the data clearly show). Deaths are also heavily weighted by the older population and in some countries dominated by fatalities in "elder care" or "assisted living" facilities that have been left without adequate countermeasures. Sad to say, deaths (distressing as they are) are then *not* the right measure to look at for the analysis and control of the dynamics of the virus. Rather, infections are the measure useful to look at for the spread and control of infectious diseases, as recognized in traditional epidemiological models, where a characteristic basic reproduction parameter, R_0 , is introduced to reflect the individual or person-to-person transmission ratio (see e.g. Heesterbeck, 2002; Jones, 2019). True, there may be inaccuracies in infection counts due to the testing types, protocols and extent in and for any given population or sample, and some other minor inconsistencies. But the data indicate that these have relatively small effect on the trends, and possibly introduce only a small systematic error, that we address and minimize by using non-dimensional country-to-country and relative rate inter-comparisons.

Therefore, for our analysis we calculate the reported *infection rates*, not the numbers or rates of deaths or hospitalizations, because we want to determine the overall social scale of infection and use a *leading* indicator, whereas deaths are a *lagging* indicator.

The infections numbers also depend on which country/region they refer to and on the stage reached by the infection (early onset, spread extent, countermeasures employed etc.). We already know that if uncontrolled, the increase in infections rises exponentially if the rate of infections is assumed to be proportional to the number of infected individuals, (Heesterbeek, 2002; Duffey and Zio, 2020). After some "learning" leads to the definition and introduction of countermeasures, the rate falls, precisely in reflection to the society's learning on how to reduce the infection rate or "flatten the curve" of the accumulated number of deaths, infections and/or hospitalizations. The characterization of this trend can be, then, used for guiding the societal response to the pandemic and control the effectiveness of the countermeasures introduced. For example, the USA official guidelines for CoVid-19 pandemic state that before a "comeback" phase can be initiated, in addition to some social distancing and other measures of reduction of the virus diffusion, any given State should have a 14-day period (White House, 2020, www.whitehouse.gov/openingamerica) with:

a) downward trajectory of reported influenza-like illnesses (ILI)
b) downward trajectory of reported CoVid-19-like syndromic cases.

No quantitative measure of statistical significance is given or required, no desirable or minimum rates are specified, nor any means or method of prediction. However, everywhere in the World, the public health policies considered in the aftermath of the virus outbreak are "data-driven" periodically adjusted and computer model projections are updated, as death and infection rates change. For effective control and decision-making, it is, therefore, fundamental to establish the onset and significance of the downward trajectory by numerical scientific criteria that can be used by policy makers in their region-by-region and country-by-country analysis, to predict the future trend expectation and to discern significant systematic deviations (learning by cross-fertilization).

A logical public question is: does the infection recovery data show any sign that we are learning how to reduce risk and which countermeasures were or are most effective and should be kept (and which were or are not, and can be eliminated)?

The global and local data now available from the CoVid-19 pandemic provides an opportunity for new thinking and new methods to predict recovery rate and evaluate countermeasures effectiveness. Specifically, decision makers want to know when and how to 're-open" or relax stringent social distancing, "lockdown", self-isolation and medical emergency decrees, while medical and epidemiological analyses focus more clinically on deaths and hospital workloads or anticipated "worst case" or upper bound scenarios. With an initial focus of our analysis on Italy-the leading and worst hit region in Europe- we proceed to a systematic examination of the infection rate data for many countries to determine :

a) Overall societal transmission timescales to reach defined rate peaks determined by introduction of countermeasures (e.g. in California, Belgium, Brazil, Italy, Spain, UK, USA)
b) Trends for "island states" exercising complete border and access control, where importation of cases by travelers could be completely controlled or minimized (e.g. New Zealand, Hong Kong and Iceland)
c) Results of imposing only limited "societal" restrictions that rely in large part on more voluntary or lax countermeasures, confiding in guided public awareness and mindfulness, and in ethics of resilience (Rajaonah and Zio, 2020) (Sweden, Wyoming, Nebraska and Iowa)
d) Comparisons between USA regions with completely different environmental and living milieu but all with low infection rates (Idaho and St Louis, Missouri)
e) Demonstration of recovery rates due to effective internal countermeasures that established a well-defined trajectory (e.g. China, Hong Kong, Italy, Spain, New Zealand and South Korea);
f) Extent of less defined peaks where an extended "plateau" occurs (Canada, USA and UK) and rates with potential low testing or under-reporting of cases (India, Ecuador and others)

In the following Section 2 of the paper, we first develop the fundamental statistical theory and compare it to more traditional methods. To the best of the authors' knowledge, this method and theory is not known to and has not been previously published in the epidemiological literature. Then, we make a series of comparisons and predictions for the overall growth and decline of infection rates, using the available data[1]. From the analyses and comparisons, we draw a number of conclusions in Section 3, relevant to the use of the proposed framework for dealing with the future phases of recovery from the pandemic in a risk-informed way that can guide policy decisions for resilience.

2. Theoretical model for social infection rate evolution

In any observation interval considered (e.g. successive days), new infections appear distributed as outcomes from prior or past exposure of some portion of the entire infection risk-exposed population. We adopt a dynamic infection model that is based on well-known statistical physics and is fundamentally different from classic infectious disease theory, which is based on the (sequential) person-to-person rate multiplier, R_0 (Heesterbeek, 2002; Jones, 2007).

For infectious diseases, in any society the number of observed infections, n, is randomly distributed in the exposed population, since anyone can get it by interacting with someone else. Independent of the individual or overall transmission mechanism(s), for a whole society the probability of cross-infection given contact, $p=n/N$, then depends solely on the total (fixed) *number*, N, of the equally *infection risk-exposed* individuals of the recipient

[1] Data sources are all on-line, and include global lists from WHO and world-in-data.org, and individual local state (e.g. Idaho, Iowa, California, Nebraska), and country internet "dashboards" and city file records (Italy, Hong Kong, New York, St Louis, USA (CDC), etc.), which are all updated periodically.

population. Using the established physics and mathematical principles of classical statistical mechanics (Rushbrooke, 1949; Greiner et al, 1997; Duffey and Saull, 2008) and standard reliability theory (Lewis, 1994), the rate and distribution of disease infections that are observed among a sample or group of any size emerge from applying the following conventional and physically reasonable assumptions:

- infections appear as outcomes and are counted in number, *n*, during some prior and present observed infection risk exposure intervals (here measured in days, *d*);
- outcomes (infections) occur and are observed randomly, but are a systematic function of the risk exposure by person-to-person contact, or by any and all transmission spread processes, with some average or overall societal characteristic timescale, G, of the transmission and incubation growth process;
- a multitude of possible distributions of infections, $N!/\Pi n!$, are equally possible and enable the use of the well-known Stirling's formula for large numbers (Rushbrooke, 1949);
- the distribution of the number of occurred infections, *n*, as a function of the incremental or actual infection risk exposure interval, Δd, is the most likely, because it is the one that has actually occurred and been observed;
- the most likely infection distribution is that which gives a maximum of the likelihood among all the possible distributions for a given number of observed infections;
- the total number of all possible infections observed in the prior infection risk exposure interval in days, *d*, is finite and constant (the exposed or sample population does not change);
- the rate of outcomes (infections) , $\lambda \equiv R(d)$, per incremental risk exposed interval, Δd, is proportional to the incremental change, Δn, in the number of infections during the infection risk exposure interval of observation (taken equal to a step of d= 1 day, in our case).

As usual, the validity of the above assumptions is justified by data comparisons and predictions. The number of actual infections, *n*, observed in the total risk-exposure interval *d*, *n(d)*, is (hopefully) much less than the total number of possible infections and the sample size , so $n<<N$. As usual for any sample $n<<N$, the *social rate*, *R(d),* of observed infections is (since *n* are removed from the population by already becoming infected),

$$\lambda \equiv R(d) = \frac{1}{(N-n)} \frac{\Delta n}{\Delta d} \approx \frac{1}{N} \frac{dn}{dd} \qquad (1)$$

With the above introduced assumptions and conditions, the most likely distribution of the number of possible observed infections, *n*, is, then, *always* systematically exponential in form (Rushbrooke, 1949). In any incremental observation interval of infection risk exposure, the observed outcome rate*, R(d)*, is derived as (Duffey and Saull, 2008),

$$R(d) = R_m + (R_0 - R_m) e^{\pm k(d-d_0)} \qquad (2)$$

independent of the total number of possible infections, N, where in our present notation, R_0 is the initial rate value at $d = d_0$, the time of the onset of infection observation.

The positive sign describes the "forgetting" phenomenon resulting in rate growth due to (initially) uncontained and/or still continuing infection spread until the time, d_M, at which a peak or *maximum* rate value, R_M, is reached. The negative sign gives rise to a rate decline due to (successively, day-by-day) "learning" by society and individuals deploying and obeying effective countermeasures until the *minimum* attainable rate value, R_m, is reached. The constant, k, is therefore the characteristic e-folding timescale of the rate, dependent on the presence (learning) or absence (forgetting) of effective countermeasures (societal and medical, in the case of CoVid-19).
Equation (2) is actually the solution of the *second-order* differential equation in the observed infection number, n, as a function of risk exposure days,

$$\frac{dR}{dd} = \pm k(R - R_m) \qquad \text{or} \qquad \frac{d^2n}{dd^2} = \pm k\left(\frac{dn}{dd} - \frac{dn_m}{dd}\right) \qquad (3)$$

This relation describes quantitatively the Learning Hypothesis that *the rate of change of the rate (reported infection cases per day, in this paper) is proportional to the rate itself*. The single physical parameter, k, thus represents the process of acquisition of knowledge and learning, as human societal behavior affects the rate of unwanted outcomes.

This purely theoretical result should be contrasted to the traditional multi-parameter epidemiological model for growth due to transmission by individuals through one-by-one contact(s). In its simplest form with some contact removal rate[2], ν, the infection rate is given by the straightforward assumption that the individual cross-infection rate of growth is proportional to the number of individuals already infected. Conventionally and empirically, this is governed by the *first-order* differential equation (e.g. Heesterbeek, 2002; Jones, 2007),

$$\frac{dn}{dd} \propto n = \nu n\,(R_0 - 1) \qquad (4)$$

Thus, the successive one-to-another-one infection rate grows *if and only if* $R_0 > 1$. This formula is fundamentally different, in that this traditional individual R_0 arises as a special case of the more general societal Learning Theory previously introduced. Hence, by equating Equation (3) with the differential of Equation (4) straightforwardly we obtain the equivalent societal reproduction number as,

$$R_0 = \frac{\{1 \pm k(R - R_m)\}}{\nu R} \qquad (5)$$

Only in the limits of high rates, $kR >> 1$, *with* $R >> R_m$, , and taking $\nu \sim O(1)$, is, then, $R_0 \sim \pm k$.

[2] Note that in the theory underpinning the analysis presented in this work, removal of already infected cases from the population is captured by using the correct formula for the rate (Equation (1)) not by introducing the empirical adjustable factor, ν.

Being intrinsically dependent on the overall societal rate, in general $R_0 = f(R)$, and is *not* constant but varies as R changes, which explains why any such parametric transmission model requires periodic adjustment to fit growth data, and often yields a wide spread of apparent values (see e.g. Holme and Masuda, 2015; Wikipaedia "Basic Reproduction Number"; and IHME, 2020). So, this new relation in Equation (5) also implies that we must know or determine the dynamic societal learning or forgetting rate , k, and importantly its sign.

For the initial positive exponential process of infection rate growth, for clarity we denote *positive* $k=G$, and consider its onset and trend in infection risk-exposure days from the initial "zero" day of spread increase corresponding to the day, d_0, of first observing or detecting an infection threshold number, say, 100 cases. Then, Equation (2) becomes, for $d_0<d<d_M$,

$$R(d) = R_m + (R_0 - R_m)e^{G(d-d_0)} \qquad (6)$$

Similarly, for the successive *negative* exponential process of infection rate decline, we use directly the symbol, k, for the exponential parameter, and consider its onset and trend in infection risk-exposure days from the initial "zero" day of spread decline, $d=d_M$, corresponding to the day of reaching the infection rate peak, R_M. Then, Equation (2) becomes, for $d_M<d<d_T$,

$$R(d) = R_m + (R_M - R_m)e^{-k(d-d_M)} \qquad (7)$$

These growth and decline curves (negative and positive exponential, with parameters G and k, respectively) intersect at the peak day, d_M, when the rate achieves its maximum value, R_M. This can sometimes generate almost a "plateau" in the actual data, depending on the absolute rate value and the actual intensity of the force of the spread and of the counterforces of the measures deployed to control it and contain it (see actual data below). From the Equations (6) and (7) evaluated at their peak intersection, and considering a sufficiently long span of evolution after the onset in Equation (6), $d_M>>d_0$ at which $R_M >> R_m$, we obtain,

$$d_M = \left(\frac{1}{G-k}\right) ln\left(\frac{R_M}{R_0-R_m}\right) \qquad (8)$$

The peak rate day, d_M, is, then, found to depend on the ratio of the peak rate value, R_M , to the value of observed rate onset, R_0, and on the difference between the e-folding parameters, G *and* k, governing the rate growth (forgetting) and decline (learning) processes, whose values can be readily determined from fitting the recorded data. We have already shown that this intersection occurs at the correct peak day by fitting to the data for Italy shown in Figure 1 above (Duffey and Zio, 2020).

For further data inter-comparisons, it is useful to adopt the non-dimensional form of these trends, obtaining the so-called Universal Learning Curve (ULC) (Duffey and Saull, 2008). For example, with reference to the infection rate decline phase of the learning part of the

process, taking the normalized infection risk-exposure interval measure $d^* = (d-d_0)/(d_T-d_0)$, where $d_0=d_M$ and $d_T >> d_0$ is the total infection risk-exposure interval of observation when learning is completed and the attainable minimum rate, R_m, is reached, Equation (7) can be written as (Duffey and Zio, 2020),

$$E^* = \frac{R(d)-R_m}{R_M-R_m} = e^{-kd^*} \qquad (9)$$

Interestingly, a value $k \sim 3$ fits "universally" the decline trend (due to "learning") of many failures/errors/accidents time series rate data in industrial, surgical, transportation, mining, manufacturing, chemical, maintenance, software and a multitude of other societal systems (see Duffey and Saull, 2008). This same learning rate trend and constant value emerges also in cognitive psychological testing from skill acquisition tasks performed and learnt by individuals, in which case one talks of the "Universal Law of Practice" (ULP) thus linking individual to collective learning.

Similarly, for the forgetting growth part of the process, taking the normalized infection risk-exposure interval measure $d^* = (d_0-d)/(d_M-d_0)$, where $d_M >> d_0$, Equation (4) becomes,

$$E^* = \frac{R(d)-R_M}{R_0-R_M} = e^{Gd^*} \qquad (8)$$

We have provided here the fundamentally human-societal basis for the observed trends, and now show the universal applicability of the theory on a wide range of growing and declining infection rate data.

2.1 Predicting the infection rate growth trajectory to the peak

The trajectory of growth to the peak of the infection rate (e.g. the solid line in Figure 1 for the Italian case) has been examined for many countries and the best fit exponential has been determined (Duffey and Zio, 2020). The application to 14 countries/regions has yielded 295 data points for over 1,000,000 total infections worldwide[3]. The comparisons demonstrate an overall similarity of the characteristic e-folding timescale, $1/G$, independent of the absolute peak infection rate value reached, which ranges over $200 < R_M < 50000$ per day in the different regions considered. The average global growth characteristic timescale value is 5.9 ±1.4 days with a coefficient of determination R^2 =0.84, or a one standard deviation interval of 4.5- 7.3 days.

[3] China, Belgium, Brazil, California USA , Canada, Germany, Idaho USA, Italy, Spain, South Korea, Sweden, Turkey, UK, and USA

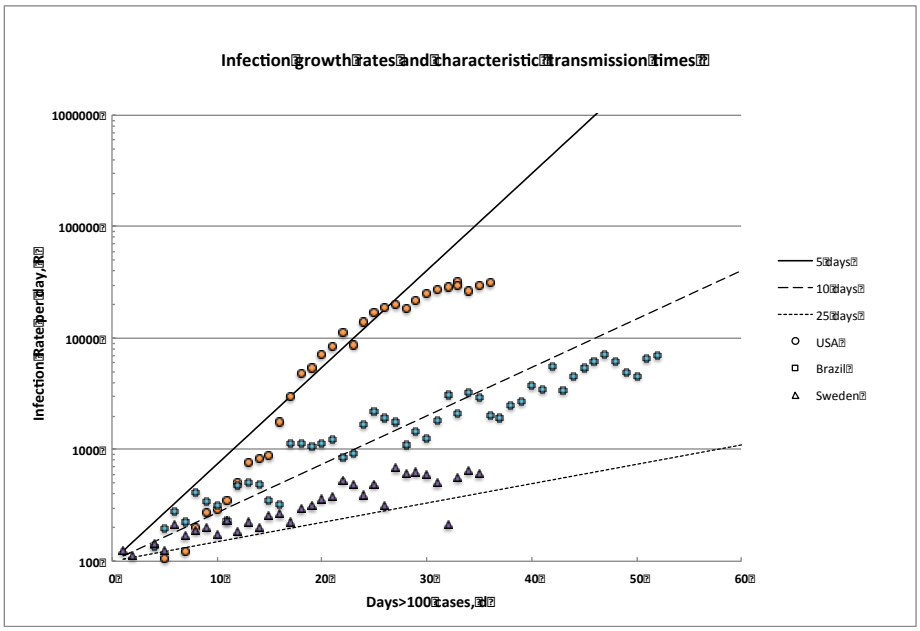

Figure 2 Typical characteristic growth trajectories of societal infection rates (semi- logarithmic plot)

Differences of up to a factor of five in growth timescales of $0.04<G<0.2$ are apparent between countries, as shown by the extreme examples in Figure 2, suggesting important impacts of environmental, cultural and social factors on transmission rates. Only Sweden adopted limited voluntary 'social distancing' countermeasures without industry and service closures, but initially had a low rate of infections anyway, and is relatively geographically isolated with limited international travel, and largely mono-cultural. Brazil did not impose restrictions immediately and is a highly diverse society with heavily populated urban developments, major cities and extensive rural/agricultural spaces. The USA contains all these different factors and characteristics, the highest rate and rapid spread being in cities which are diverse, highly mobile, very densely populated and with unavoidably close community contacts (dominated by New York) .

From our data analysis, averaging the individual and world G-values gives a characteristic incubation timescale of 5.5 days, which is in agreement with the likely theoretical incubation timescale of about 5 days and the results of other studies. Indeed, the characteristic transmission/incubation timescale value for entire societies that we have found can be regarded as reasonable, comparing quite well to the *individual incubation* period following infection in the USA indicated in (Ghinai et al, 2020). In the paper, an average incubation time of 4.4±1.7 days[4], or between 2.7- 6.1 days, was found to result from tracking cross-infections for just 15 individual cases; these values are well within the one standard deviation interval of 4.5-7.3 days determined in our study for the global population, and match well also the 1/G value of approximately 5 days found for the USA. Furthermore confirmation comes from a study (Becker et al, 2020) that used standard statistical distribution fitting on 88 individual cases in Wuhan, China to find a mean of 6.4

[4] This average and standard deviation were not reported in the paper, so these were calculated using the original data given in Figure 2 of the referenced paper.

days, with standard deviation interval of 5.6- 7.7 days, again values perfectly fitting the range of uncertainties estimated in our paper. As another demonstrative example, Casella (2020) analyzed the growth and decay of infections using an empirical "control" model that is commonly used to tune electronic system response behavior utilizing variable feedback loops. As stated clearly by Casella, it has no theoretical foundation but was fitted using (limited) infection data from Hubei, China and Lazio, Italy, obtaining a growth exponent of 3.8 days and 3.1 days, respectively.

Furthermore, we can compare the present theory to the early growth estimates obtained by applying SEIR models for the very first cases in Wuhan (Kucharski et al, 2020), looking at Equation (5) with positive $k=G$. At $d=d_0$, the true initial rate is R_0 and we can assume, reasonably, that the threshold rate for detection/reporting is at best, say, ψ -times this low initial value above background infectious diseases or similar (influenza-like) infections: then, $R \sim \psi R_0$. Numerically, with testing and early diagnosis, the initial rates are of the order of a few $\sim O$(1-3) per day (as the real data initially show), and infections are detected and "removed" so that the contact removal rate $\nu \sim O(1)$, yielding from Equation (5) the equivalent *initial* reproduction number, $R_0 \sim 1 + \psi G$. Since the world societal growth data indicate $0.04<G<0.2$, with a detection threshold only $\psi \sim 2$ times the value R_0, we can estimate from Equation (5) that $1.08 < R_0 < 1.4$, well within the mean early onset values of $1.05 < R_0 < 2.5$ and ±95% of $0.41 < R_0 < 4.77$ reported in (Kucharski et al,2020). This shows that our new societal learning theory presented in this paper, indeed, predicts global R_0 values consistent with the individual initial transmission estimates, but utilizing a much broader global database than that of initial individual infections.

We have further verified the applicability of our analysis at the international level by comparing Sweden (total population of around 10M) to rural States in the USA (Iowa, Nebraska and Wyoming) far away from large conurbations and with a total combined population of about 5M, and that have announced and adopted relatively mild countermeasures short of a full "lockdown". As shown in Figure 3, for 40 days the infection rate growth is essentially the same and $1/G \sim 25$ days for these apparently very disparate regions, reflecting the environmental conditions related to ample personal space and decreased opportunity for transmission, with incubation timescale 1/G increasing by a factor of five when outside heavily urbanized and highly mobile communities. This effect is confirmed also in the comparisons between regions and cities with limited total populations of less than circa 300,000 (e.g. Idaho and St Louis), which result in similar and very low rates (<100 per day).

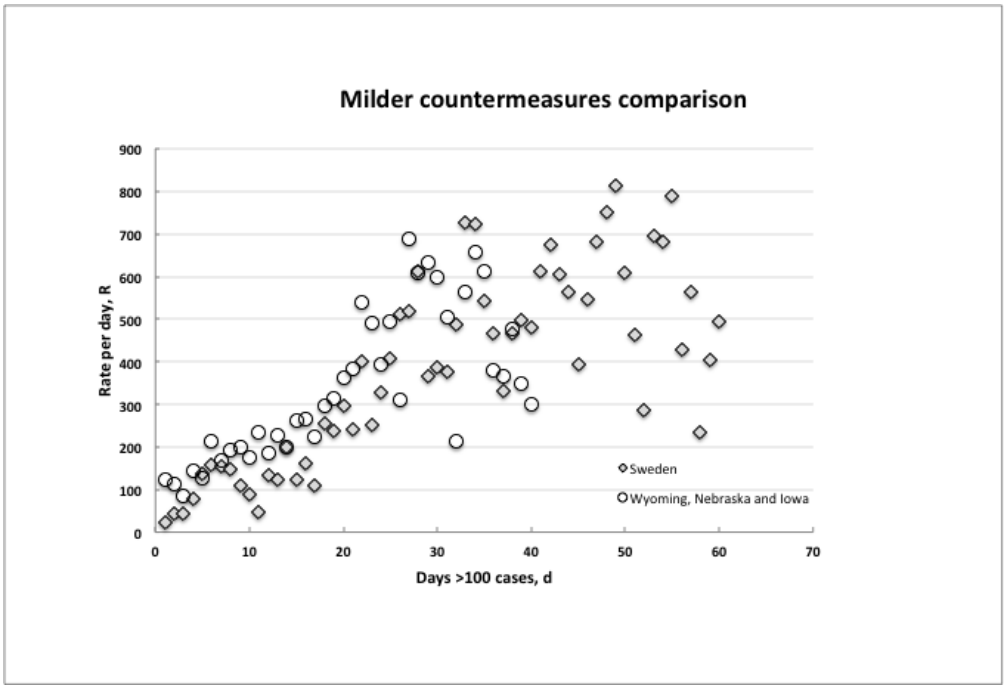

Figure 3 Comparison of trajectories of infection rate growth in relatively little densely populated regions of two continents where mild countermeasures have been applied

One can confirm that, as expected, the underlying cause of the exponentially increasing infection growth rate is indeed random person-to-person societal transmission followed by incubation. Our analysis demonstrates not only that a similar trend of growth is seen globally everywhere, but also it provides a timeframe for readily implementing resilience countermeasures to contain the infection spread and, as we shall see in the next section, for judging their effectiveness in reducing it.

2.2 Predicting the infection rate reduction trajectory from the peak

The present analysis illustrates the fact that humans learn how to control the CoVid-19 outbreak spread and reduce infection rates by using countermeasures (treatments, isolation, "social distancing" etc.). Therefore, if these are effective the infection *rate* must reach a peak and, then, decline. In Italy, after some delay in implementing countermeasures, the infection rate peaked at about 6000 per day after about 30 or so days, as can be seen in the graph of Figure 1 and, then, declined. In our previous paper (Duffey and Zio, 2020), using Equation (7), and initially using data from just four countries (China, South Korea, Spain and Italy), we demonstrated the finding that the recovery curve is universal in shape.

We can now update and further verify this prediction by comparing to other countries that have already achieved recovery by adopting effective countermeasures. In particular, "island" nations (which include Austria, Australia, New Zealand, Hong Kong and Iceland) have been able to control their pandemic and almost completely block entry of any imported cases while limiting internal infection spread. The comparison also includes the recovery prediction by the elaborate IHME model (IHME, 2020, see also Duffey and Zio,

2020), a detailed epidemiological computer model used to estimate the evolution of the virus diffusion. As shown in Figure 4, these seemingly ideal countermeasures have worked effectively and the outcomes in terms of infection rate evolution follow the statistical learning theory predictions, providing thus a benchmark for countries where differences in trend are a potential cause for concern regarding the effectiveness of their countermeasures, and the management and control of the pandemic.

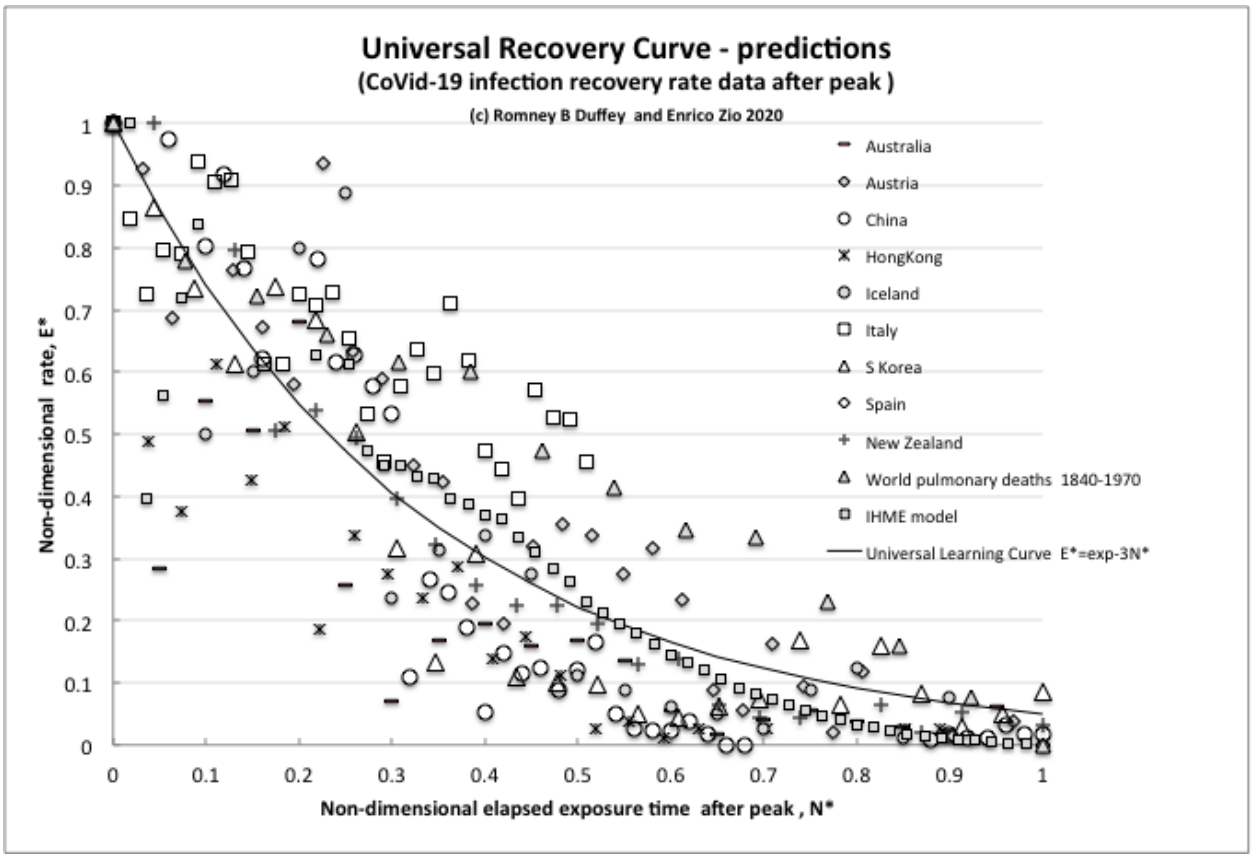

Figure 4 Predictions of recovery in agreement with the analytical exponential trend of learning theory

The key point is that these successful recovery data all follow similar decreasing trajectories with, furthermore, the learning curve being nearly the same (k~3), as previously found for any learning process on undesired outcomes, accidents, events of other modern technological systems operated by humans: we call such a Universal Learning Curve the Universal Recovery Curve (URC).

2.3 Ineffective countermeasures and consequent departures from recovery

Ideally, all countries should show or demonstrate some trend of recovery after peaking as reported in Figure 4, which plots the evolution of E^*, the non-dimensional infection rate normalized to the initial peak value, versus N^*, the non-dimensional elapsed time of experience/knowledge or risk exposure after the rate has peaked (number of days after peak/day of peak). However, sometimes the peak is delayed, and may not be well defined, so both rate peaking and recovery do not perfectly follow the ideal theory everywhere.

There are several examples of "imperfect" learning that can be observed:

1) after the initial rate growth, infection rate values persist somewhat stably, despite countermeasures being in effect: this shows in a lack of a steady decline and, rather, in the presence of a plateau, as residual cases of infections still contribute to balancing the reducing effects of the countermeasures (see Figure 5);
2) infection rates do not peak but continue to slowly climb, due to the fact that countermeasures and/or reporting are not being completely effective;
3) infection rates are irregular, showing no clear growth or decline trend and/or exhibit occasional or persistent gaps in data reporting.

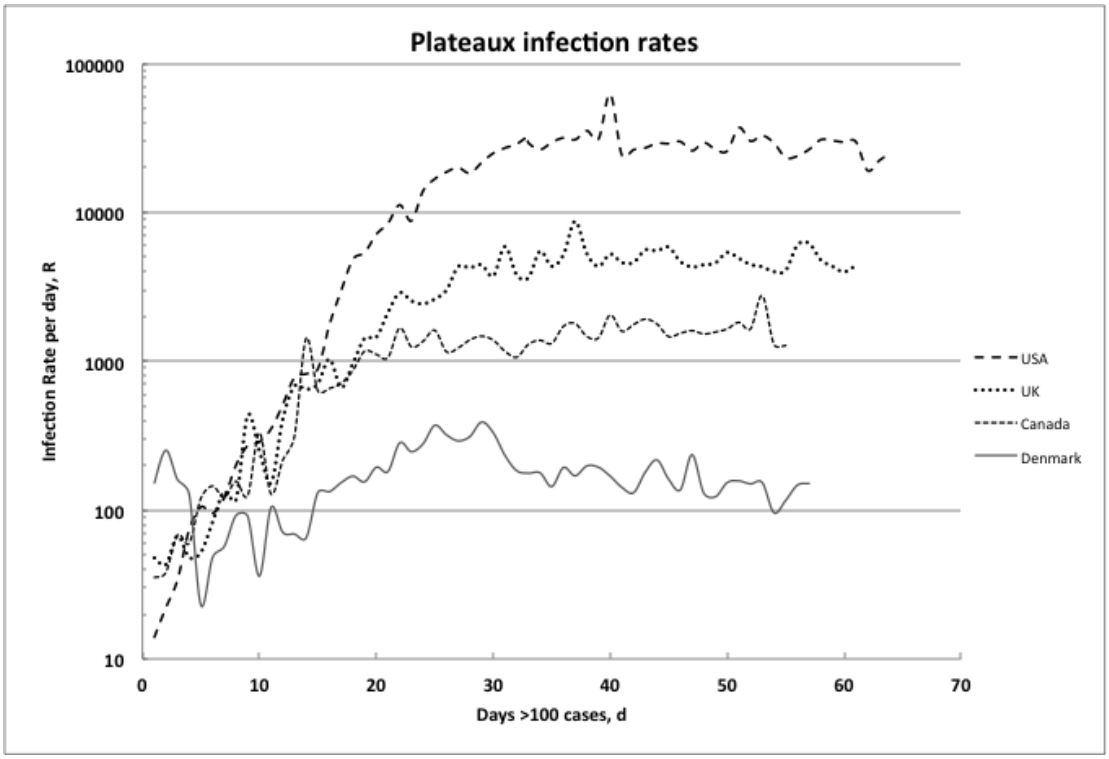

Figure 5 Examples of plateaux in infection rates trajectories

Formally, the plateau rate, R_M, occurs when the countermeasure effectiveness in growth reduction precisely balances the increases in growth rate, which from Equation (8) , since $R_M >>> R_m$, results when

$$R_M \approx R_0 e^{(G-k)d_M}$$

From Figures 1 and 5, the peak is typically reached by $d_M \sim 30$ days, and from the observed initial growth rates $0.1<G<0.2$ per day. Numerically, without effective recovery countermeasures, the extreme case is when $k<<G$, where for the fastest transmission, $G \sim 0.2$: so, with the nominal onset threshold of $R_0 = 100$ per day, and no immediate effective reduction, $k \sim 0$, the plateau rate is of course given by, $R_M = 100\ e^{0.2x30} \sim 40000$ per day, which is close to the observed maximum USA rate in Figure 5, and it is what should be expected, whereas the minimum value of the plateau would be when k~G.

Finally, the above analysis of the trajectories of infection rate growth and decline are based on estimates from data and are, thus, inevitably subject to uncertainty due to the many endogenous factors related to the virus spreading in the different environmental, social and medical conditions, and to the actuation and respect of the countermeasures implemented (Rajaonah and Zio, 2020). Still, the findings can provide clear guidance for evaluating the situation and reflecting on the best approach to mitigation and control, and on which countermeasures to keep, take or eliminate.

3. Conclusions

In this paper, we have used statistical theory to predict the growth and reduction of pandemic infections using infection rate as a measure of observed outcomes of the infection process. We have showed that the traditional person-to-person R_0 model is a special case of this new societal learning theory. Based on publicly available data, the analyses for many countries and regions show that the CoVid-19 transmission and incubation rate is circa 5 days, independent of global location. After infection rate peaking, consistent with world outcome data, the recovery trajectories follow the Universal Learning Curve (Universal Recovery Curve), and different countermeasures result in differing growth and decline trajectories.
The theory and analysis here presented, indeed provide fundamental insights useful for risk handling during the development of and recovery from a pandemic.

Acknowledgement

The authors are grateful to the numerous websites and data collectors that made this work possible.

References

Anderson, J.R., 1990, *Cognitive Psychology and its Implications*, W.H. Freeman, 3rd edition. ISBN 0-7167-2085-X.
Becker, J A., D Klinkenberg and J Wallinga, 2020, Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020, Euro Surveill. 2020 ;25(5) pii=2000062 DOI;10.2807/1560-7917.ES.2020.25.5.2000062
Casella,F., 2020, Can the COVID-19 epidemic be controlled on the basis of daily test reports?,arXiv:2003.06967v2 [physics.soc-ph]
Du Z, Wang L, Chauchemez S, Xu X, Wang X, Cowling BJ, et al. , 2020, Risk for transportation of 2019 novel coronavirus disease from Wuhan to other cities in China. Emerging Infectious Diseases Journal, May . doi.org/10.3201/eid2605.200146.
Duffey,R.B., 2017, "Search theory: the probabilistic physics and psychology of memory and learning", J.Sci. Res. and Studies, Vol 4 , No 5,pp107-120.

Duffey, R.B. and Saull, J.W., 2008 , *Managing Risk: the human element* , J.Wiley and sons, UK, ISBN 978-0-470-69976-8.
Duffey, R.B. and Zio, E., 2020, Analysing recovery from pandemics by Learning Theory: the case of CoVid-19 medRxiv preprint doi: https://doi.org/10.1101/2020.04.10.20060319.
Fiondella, L. and Duffey, R.B., 2015, "Software and Human Reliability: error reduction and prediction", Proc. PSA 2015, paper 12221, Am. Nucl. Soc., April 21-26, Sun Valley, Idaho.
Ghinai, I., S Woods, K A Ritger, T D McPherson, S R Black, L Sparrow, M Fricchione, J L Kerins, M Pacilli, P S Ruestow, M A Arwady, S F Beavers, D C Payne, H L Kirking, and J E Layden 2020, Community Transmission of SARS-CoV-2 at Two Family Gatherings in Chicago, Illinois, February-March 2020, Morbidity and Mortality Weekly Report (MMWR) April 17, 2020, Vol. 69 , No. 15 pp446-450, US Department of Health and Human Services/Centers for Disease Control and Prevention (https://www.cdc.gov/mmwr/).
Greiner, W., L Neise and H Stocker, 1995*, Thermodynamics and Statistical Mechanics*, Springer, New York, ISBN :0-387-94299-8
Heesterbeek, J A P., 2002, A brief history of R_0 and a recipe for its calculation, Acta Biotheoretica , 01 January , 50(3):189-204, DOI:10.1023/A:1016599411804
Holme, P and N Masuda, 2015, The Basic Reproduction Number as a Predictor for Epidemic Outbreaks in Temporal Networks , March, PLoS ONE 10(3): e0120567. doi:10.1371/journal.pone.0120567
IHME COVID-19 health service utilization forecasting team**,** 2020**,** Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. *MedRxiv.* 26 March. doi:10.1101/2020.03.27.20043752.
Jones, J H, 2019, "Notes on R_0" , Dept of Earth System Science,, Stanford University, CA, USA April, accessed at web.stanford.edu/class/earthsys214/notes/Jones_R0_notes2019.pdf
Kucharski A J, T W Russell, C Diamond, Y Liu, J Edmunds, S Funk and R M Eggo,., 2020, Early dynamics of transmission and control of COVID-19: A mathematical modelling study. Lancet Inf. Dis., The Lancet, 20, pp553- 558, May, doi:10.1016/S1473-3099(20)30144-4
Lewis, E. E., 1994, *Introduction to Reliability Engineering*, John Wiley and Sons, New York, 2nd edition.
Ohlsson, 1996, "Learning from performance errors", Psychological Review, 103 (2):241-262.
Rajaonah B., Zio E.. Contributing to Disaster Management as an Individual Member of a Collectivity: Resilient Ethics and Ethics of Resilience. 2020. hal-02533290.
Rushbrooke, G. S., 1949*, Introduction to Statistical Mechanics*, Oxford University Press, London