medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Extrapolation of Infection Data for the CoVid-19 Virus and Estimate of the Pandemic Time Scale.

Walter Langel, Institut für Biochemie, Universität Greifswald,

Felix-Hausdorff-Straße 4, 17489 Greifswald, Germany

Abstract

Predictions about the further development of the Corona pandemic are widely diverging. Here, a simple yet powerful algorithm is introduced for extrapolating infection rate and number of total infections from available data. The calculation predicts that under present conditions the infection rate in Germany will culminate in a few weeks and decrease to low values by mid-June 2020. Total number of infections will reach several 100000 though. A refinement of the calculation is presented in the supplemental material and shows that the lock down in Germany has reduced the total number of infections from a target value of 338 000 to 184 000, corresponding to a decrease of about 45%.

Introduction

News pages like [1] publish data on the CoVid19 infection for several countries as a function of time. The total infection numbers show a dramatic increase, and so far no decrease yet with the exception of China (P.R.C). The increase is in general called “exponential”, which implies that finally a significant part of the population will be infected. However, such an exponential function cannot explain the saturation of infection numbers as observed for China (**Figure 1**).

Obviously, it is of great interest to have some forecast on the extension of the corona crisis. Many measures taken by governments only extend for weeks so far, but other sources discuss two years of ongoing spreading of the virus and an infection from 10 million to about 2/3 of the total German population [2]. In this worst case, a death rate of e.g. 2% would result in 1.1 millions of victims.

For Germany, the maximum number of infections obviously is by far not yet reached (by march, 23rd, 2020). On the other hand a logarithmic plot shows a significant curvature (**Figure 2a**), suggesting that the number of confirmed infection cases already increases significantly less than strictly exponential.

In this communication I try to derive information on the length of the infection period from the deviation of the infection rate from a strict exponential increase. A fit for the total infection numbers as a function of time is presented. This calculation uses an analytical function and is based on three parameters only. The procedure is validated using the data from China, which has overcome the climax of the infection rate already, and then applied to data from Germany and Italy.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Method

A function for fitting the logarithm of the reported figures of total infection rates is defined by

$$ln\big(N(t)\big) = \frac{ln\big(N(t\to\infty)\big)}{1+\left(\frac{ln\big(N(t\to\infty)\big)}{ln\big(N(t=0)\big)}-1\right)\cdot exp(-t/\tau)} \quad (1)$$

For the fit of equation (1) to the data points, the standard solver in Microsoft Excel 2016 was applied to the logarithm of the number $N(t)$ of confirmed infections as taken in steps of one day from the website [1], and the least squares error with respect to equation (1) was minimized by varying the three parameters $ln\big(N(t=0)\big)$, $ln\big(N(t\to\infty)\big)$, and τ. Final values are quoted in **Table 1**. The data were weighted by $\sqrt{N(t)}$ to account for poor statistics of small numbers.

The formula correctly reproduces the limiting asymptotic cases:

- At $t = 0$ we obtain consistently

  $ln\big(N(t)\big) = ln\big(N(t=0)\big)$ (2).

  This starting point cannot be taken directly from some scattering initial data but is used as fit parameter. Its actual value depends on the time when counting of cases started.

- For short times, $t \ll \tau$, the formula reproduces a nearly exponential increase:

  $N(t) \approx \exp\,(k \cdot t)$ (3)

  and a time for increasing by a factor of two given as

  $\tau_2 = \frac{ln(2)}{k}$ (4).

  The derivation of equation (3) and of the relation of k and τ_2 to the parameters in equation (1) is lengthy and may be found in the appendix. (This main text is, however, fully comprehensible omitting the appendix). For reference purposes, the values of τ_2 are quoted in **Table 1**. The small values of only a few days for this parameter have largely contributed to the fear, which has raised the appearance of CoVid-19.

- The cited data (**Figures 1a, 2a, 3a**) show that already after a few days, significant deviation from this exponential increase has been found in practice. The transition range from initial exponential increase to final saturation is described by the third parameter τ, which is characteristic for the rise time of the number of infections.

- At long times, $t \gg \tau$, this function attains a maximum of $ln\big(N(t\to\infty)\big)$ for $t\to\infty$. This saturation behavior is generated by the exponential function in the denominator approaching zero for long times t.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results

The results for the fits are plotted in **Figures 1-3** for China, Germany and Italy, respectively. **Figure 1** shows that the expression (1) with appropriate parameters yields a very good fit of the data for China in the full range from beginning to the end of the pandemic.

For Germany (**Figure 2**), only the beginning of the curve is known so far, and the numbers of confirmed infections are still strongly increasing. Here, the aim of the fit is to predict the maximum number of infections $N(t \to \infty)$, and the time scale, on which this is reached. The orange line suggests that a maximum of more than 7000 new infections per day should be reached early in April, and that the rate should drop to small values by end of May.

The situation in Italy (**Figure 3**) is intermediate, as the fit suggests that the point of maximum infections per day is nearly reached. The maximum infection rate is predicted with more than 9000 per day in early April, and the total number of infections sums up to more than 400 000 in June.

All three curves suggest that only a small part of the respective population is infected by the spreading of the virus under present conditions and that between 80 000 and 400 000 people may be attained. The period with a high number of new infections should be a few months rather than two years and extend till about June 2020. After this time the further spread of the virus should slow down.

This fit obviously can only refer to the infection under the present conditions with precautions taken so far. It can neither be predicted if more severe political measures reduce the infection rate, or if other mechanisms of virus transfer, which did not yet contribute significantly to the present numbers, may gain importance later.

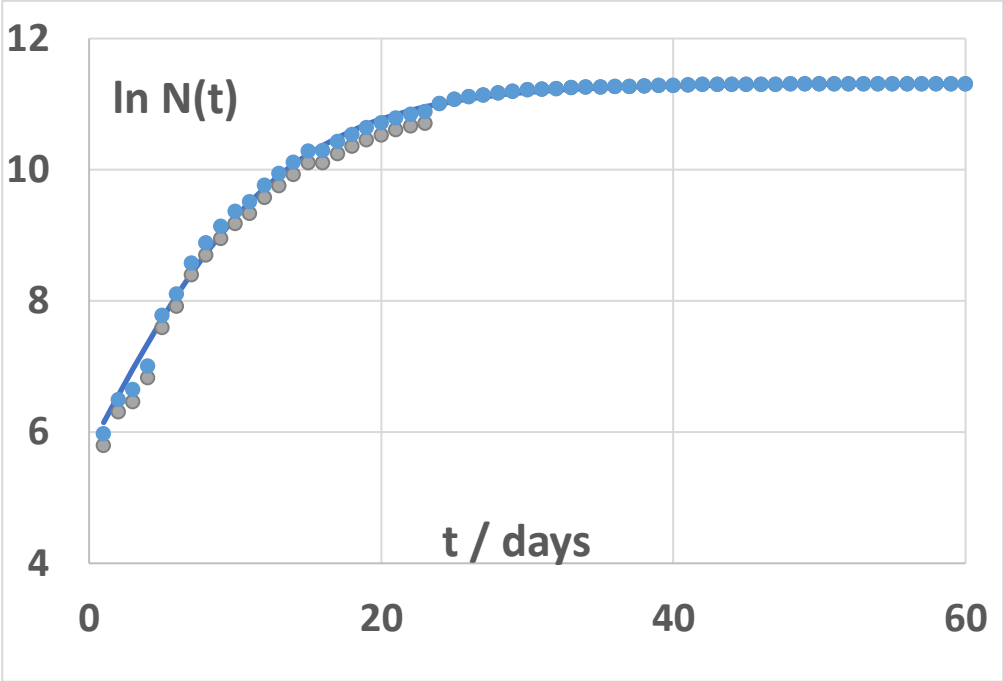
Table 1: Fitparameter for equation (3). The parameters result in the initial doubling times τ_2 for the total number of CorVid19 infections and the maximum numbers $N(t \to \infty)$ in three countries.

	China	Germany	Italy
$ln\big(N(t=0)\big)$	5.73	4.01	2.39
$ln\big(N(t \to \infty)\big)$	11.31	12.31	12.93
$N(t \to \infty)$	81 300	220 000	410 000
τ/d	6.72	10.55	15.39
τ_2/d	1.65	2.70	5.86

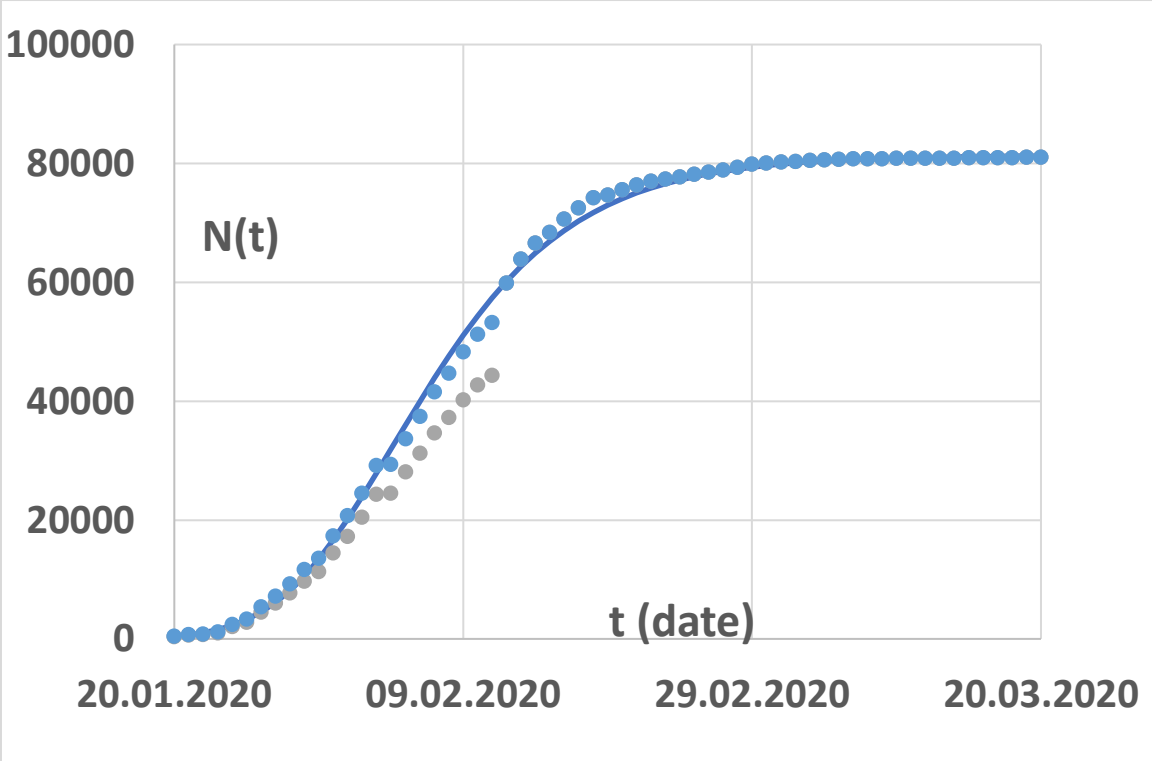
medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1: Confirmed infections in China.

a) Logarithmic plot of (filled circles, blue) and fit according to equation (3) (blue line). Some known inconsistency in the counting during the first days [3] was accounted for by multiplying the original data (gray) by 1.2 during the first 23 days till feb. 11th.



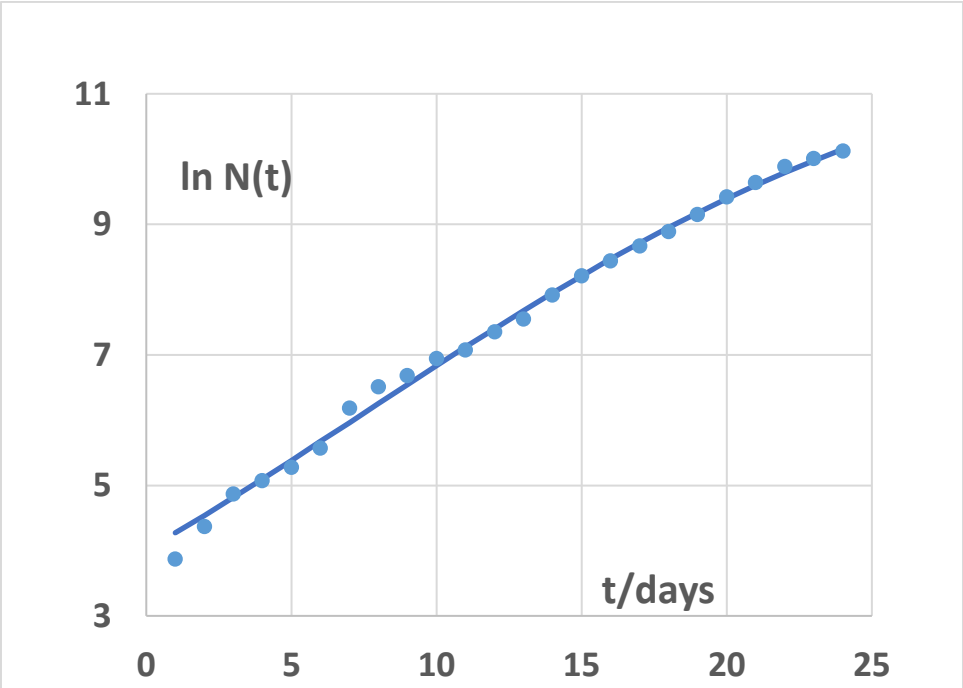
b) As **a)** but linear scaling of the y-axis.



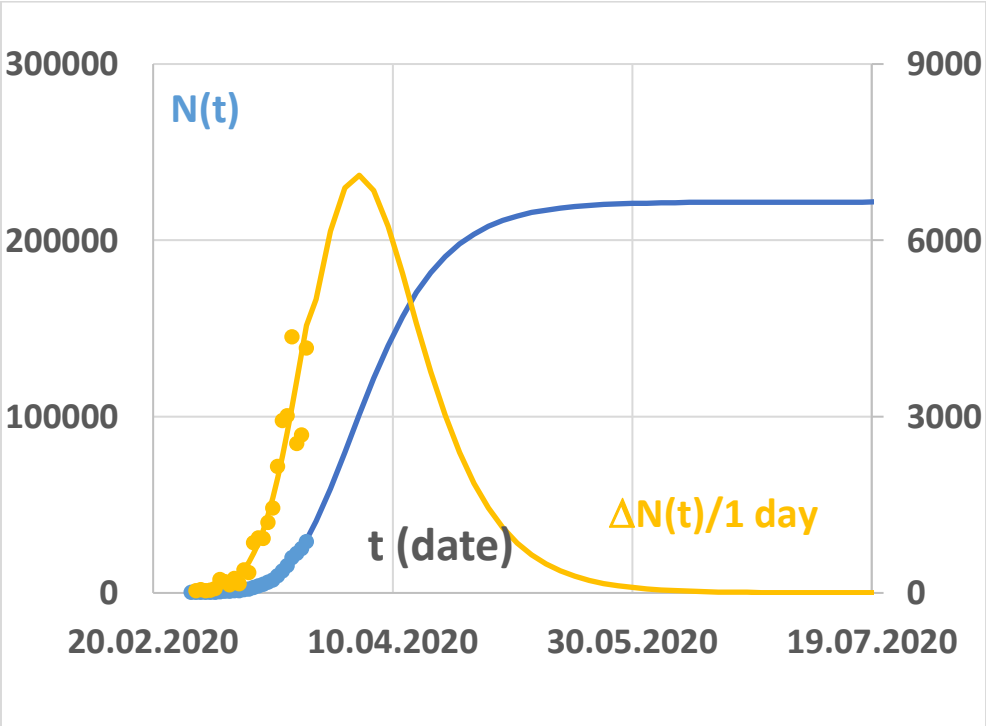
medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2: Data for Germany till march, 23rd 2020.

a) logarithmic plot as **Figure 1a)**,



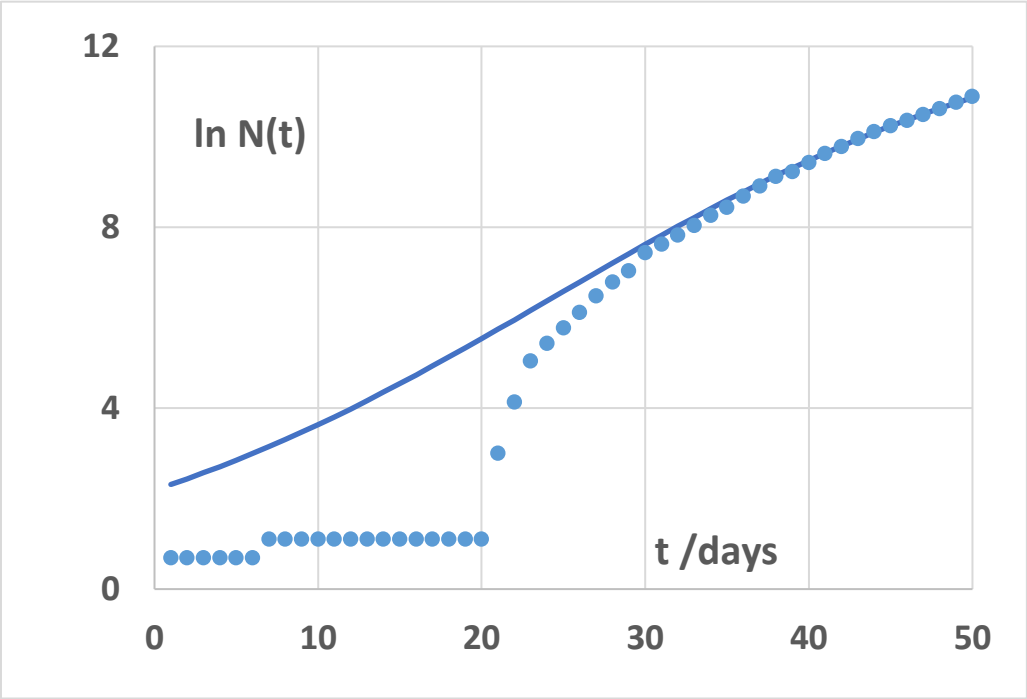
b) linear plot of total infection numbers, data (blue filled circles, left y-scale) and fit (blue line). From these numbers, the daily infection rate is obtained as difference between the total numbers from one day and the one before (orange filled circles, right y-scale). Such differential data show a larger scatter than the total numbers. Similarly the fit (orange line) is obtained from the original fit of equation (3). By applying equation (3) to dates in the future, for which no data are available yet, a forecast is obtained.



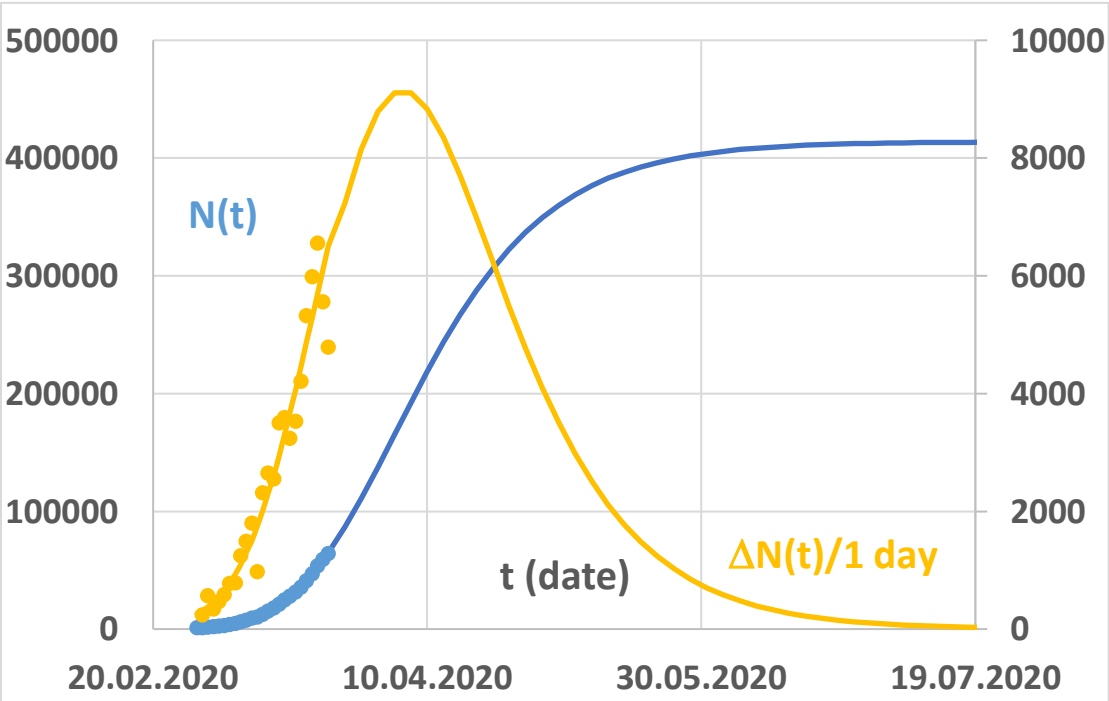
medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3: As **Figure 2** for Italy

a) The logarithmic plot of data before the 30th day shows an extreme curvature. Any extrapolation of these data would result in an unrealistically short time scale. Thus only data from 40th day on were fitted. At short times either the statistic error in the data was high or infection followed another mechanism than now.



b)



medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix:

1. Derivation of the time τ_2 of doubling $N(t)$ at small times from the fit parameters.

With the substitutions

$$f(t) = ln\big(N(t)\big), a = ln\big(N(t \to \infty)\big), b = \left(\frac{ln\big(N(t\to\infty)\big)}{ln\big(N(t=0)\big)} - 1\right), c = {}^{1}/_{\tau} \quad \text{(A1)}$$

the basic equation (1) is transferred into

$$f(t) = \frac{a}{1+b\cdot exp(-c\cdot t)}; \text{(A2)}$$

At small times, one may replace

$$exp(-c \cdot t) \approx 1 - c \cdot t \text{ (A3)}$$

and

$$f(t) = \frac{a}{1+b\cdot exp(-c\cdot t)} \approx \frac{a}{1+b\cdot(1-c\cdot t)} = \frac{a}{1+b} \cdot \frac{1}{1-\frac{b\cdot c\cdot t}{1+b}} \approx \frac{a}{1+b} \cdot \left(1 + \frac{b\cdot c\cdot t}{1+b}\right) = \frac{a}{1+b} + \frac{a\cdot b\cdot c}{(1+b)^2} \cdot t; \text{ (A4)}$$

Replacing the substitutions by the original parameters, equation (A5) is obtained:

$$ln\big(N(t)\big) = \frac{ln\big(N(t\to\infty)\big)}{\frac{ln\big(N(t\to\infty)\big)}{ln\big(N(t=0)\big)}} + \frac{ln\big(N(t\to\infty)\big)\cdot\left(\frac{ln\big(N(t\to\infty)\big)}{ln\big(N(t=0)\big)}-1\right)\cdot{}^{1}/_{\tau}}{\left(\frac{ln\big(N(t\to\infty)\big)}{ln\big(N(t=0)\big)}\right)^2} \cdot t =$$

$$ln\big(N(t=0)\big) + \frac{\big(ln\big(N(t\to\infty)\big)-ln\big(N(t=0)\big)\big)\cdot\big(ln\big(N(t=0)\big)\big)}{ln\big(N(t\to\infty)\big)} \cdot \frac{t}{\tau}; \text{ (A5)}$$

The approximation for small times t thus reads

$$ln\big(N(t)\big) = ln\big(N(t=0)\big) + k \cdot t \Leftrightarrow N(t) = N(t=0) \cdot exp(k \cdot t) \text{ (A6)}$$

with

$$k = \Big(ln\big(N(t \to \infty)\big) - ln\big(N(t=0)\big)\Big) \cdot \frac{ln\big(N(t=0)\big)}{ln\big(N(t\to\infty)\big)} \cdot \frac{1}{\tau} \text{ (A7)}$$

and the time for an increase of the number of infections in the initial state by a factor of two is

$$\tau_2 = \frac{ln(2)}{k} = \frac{ln(2)}{\big(ln\big(N(t\to\infty)\big)-ln\big(N(t=0)\big)\big)} \cdot \frac{ln\big(N(t\to\infty)\big)}{ln\big(N(t=0)\big)} \cdot \tau \text{ (A8).}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044081; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

[1] https://www.n-tv.de/infografik/Coronavirus-aktuelle-Zahlen-Daten-zur-Epidemie-in-Deutschland-Europa-und-der-Welt-article21604983.html

download March 24th, 2020

[2] https://www.sueddeutsche.de/gesundheit/krankheiten-experte-erwartet-60-bis-70-prozent-infizierte-in-deutschland-dpa.urn-newsml-dpa-com-20090101-200228-99-108884,

download March 24th, 2020

[3] https://www.tagesschau.de/ausland/corona-china-115.html

download March 24th, 2020