

Modified SIR-model applied to covid-19, similarity solutions and projections to further development

E. Rebhan*

Institut für Theoretische Physik,
Heinrich–Heine–Universität,
D–40225 Düsseldorf, Germany

July 30, 2020

Abstract

The SIR-model is adapted to the covid-19 pandemic through a modification that consists in making the basic reproduction number variable. Independent of it, another reproduction number is introduced, which is defined similarly to the usual net reproduction number. Due to its simple analytic form, it enables a clear interpretation for all values. A further parameter, provisionally called acceleration parameter, is introduced and applied, which enables a more differentiated characterization of the infection number dynamics. By a variable transformation the 3 equations of the modified SIR-model can be reduced to 2. The latter are solved up to ordinary integrations. The solutions are evaluated for current situations, yielding a pretty good match with the data reported. Encouraged by this, a variety of possible future developments is examined, including linear and exponential growth of the infection numbers as well as sub- and super-exponential growth. In particular, the behavior of the two reproduction numbers and the acceleration parameter is studied, which in some cases leads to surprising results. With regard to the number of unreported infections it is shown, that from the solution for a special one solutions for others can be derived by similarity transformations.

*e-mail: rebhan@thphy.uni-duesseldorf.de

1 Introduction

In this paper it is shown that the SIR-model [1] can pretty well be adjusted to the data reported for the covid-19 pandemic by means of a modification, which consists in allowing a dynamic variability of the usually constant basis reproduction number R_0 . An additional reproduction number is introduced which is similarly defined as the sensitive R-number of the Robert Koch Institute (RKI) or the net reproduction rate. Another number turns out to be particularly useful, the values of which allow even more differentiated informations about the progression of the infection numbers. For it, the term acceleration parameter is proposed.

First, by a suitable transformation of variables the system of 3 coupled nonlinear differential equations of the SIR-model is converted into a system with only 2 equations. The determination of the general solution of the latter is then carried on so far that only ordinary integrations remain. This makes it unnecessary to develop a numerical program for the solutions. Rather, the simplicity and clarity of the solutions obtained make it possible to quickly achieve results for concrete problems, especially when programs like MATHEMATICA are employed. With regard to the number of unreported infections it turns out, that despite the non-linearity of the equations, from the solution for a special one the solutions for other numbers can be derived by similarity transformations under rather unobtrusive assumptions.

In applying the solutions to the covid-19 pandemic, the reported data for one variable are replaced by a best least-squares fit function and thus integrated in the expanded SIR-model. Thereupon, from this the associated solution for a further variable is determined and compared with the corresponding reported data. It turns out that at least for more advanced states there is fairly good agreement. This encourages solutions to be explored, that continue preceding solutions to some extent into the future. Various assumptions about the further course of the infection or reproduction numbers are thus examined and compared. Also considered are the conditions under which the pandemic comes to an end. In the concluding section, some more qualitative aspects of the pandemic are discussed.

In order to keep the health damage and death cases caused by the pandemic as low as possible, significant restrictions and unfamiliar actions are expected of the population. So that people understand why these are necessary, it is important that the reasons for them are communicated as precise and intelligible as possible. This in turn presupposes that the knowledge about the various aspects of the pandemic is as extensive as possible, while being precise and thorough. This article was written with the intention and hope of making a small contribution to this. It is mainly based on the system of SIR-equations which were taken from Wikipedia [2]. Therefore, only a few references are given.

2 Definitions, basic equations and solutions

2.1 Definitions and basic equations

Neglecting birth and non-covid death rates, the basic equations of the SIR-model are

$$\dot{S}(t) = -\beta SI/N \quad (1)$$

$$\dot{I}(t) = \beta SI/N - \gamma I \quad (2)$$

$$\dot{R}(t) = \gamma I, \quad (3)$$

where the following notations are used: S =susceptible individuals, I =infectious individuals, R =removed individuals (recovered with acquired immunity to the disease or deceased), N =(invariable) total number of individuals, t =time with the unit day, and $\dot{S}(t)=dS/dt$ etc. Furthermore, γ is the daily recovery rate of infected individuals, and β is the daily rate of new infections caused by an infected individual. The SIR-model is based on the assumptions that individuals can be infected only once, become contagious immediately after infection and remain so until they gain immunity or die. The time-dependent variables of the SIR-equations satisfy the relation

$$S + I + R = N. \quad (4)$$

We also use the basic reproduction number

$$\mathcal{R} = \beta/\gamma, \quad (5)$$

thereby omitting the subscript 0 of the usual notation R_0 in view of a modification specified further on.

Clearly, the numerical solution of the above equations is not a problem. However, considerable simplifications are possible that make it easier to answer specific questions and enable a better understanding of solution properties. In a first step, the system of 3 equations (1)-(3) can be reduced to a system of 2 equations by introducing the variable

$$Z(t) = I(t) + R(t) \quad (6)$$

which is also used in the infection diagrams of institutions determining the infection numbers like the Robert Koch Institute (RKI) or the Johns Hopkins University (JHU). Eliminating S and R by use of $S=N-Z$ and $R=Z-I$ (which follows from Eqs. (4) and (6)), and using $\beta=\gamma\mathcal{R}$ (which follows from Eq. (5)), Eqs. (2)-(3) become

$$\dot{I}(t) = \gamma I[\mathcal{R}(N-Z)/N - 1] \quad (7)$$

$$\dot{Z}(t) = \gamma I\mathcal{R}(N-Z)/N. \quad (8)$$

With the definitions

$$a = I/N \quad \text{and} \quad z = Z/N, \quad (9)$$

we switch from the extensive variables I and Z to the intensive variables a and z , thus finally ending up with the equations

$$\dot{a}(t) = \gamma a [\mathcal{R}(1-z) - 1] \quad (10)$$

$$\dot{z}(t) = \gamma a \mathcal{R}(1-z). \quad (11)$$

With the help of Eqs. (4), (6) and (9), the solutions $a(t)$ and $z(t)$ of these equations immediately lead to $S(t) = N(1-z(t))$ and $R(t) = N(z(t) - a(t))$.

2.2 Variable basic reproduction number and solutions

Dividing Eq. (10) by Eq. (11), with $a'(z) = da/dz = \dot{a}(t)/\dot{z}(t)$ we obtain

$$a'(z) = 1 - \frac{1}{\mathcal{R}(1-z)}. \quad (12)$$

For fixed \mathcal{R} this is a differential equation for a with the solution

$$a(z) = a_0 + z - z_0 + \frac{1}{\mathcal{R}} \ln \frac{1-z}{1-z_0} \quad (13)$$

to the initial condition $a(z_0) = a_0$.

We also want to consider situations in which $a(t)$ and/or $z(t)$ are predetermined by empirical data on the progress of the pandemic. In this case Eq. (13) cannot be expected to be a suitable solution. However, it turns out that the SIR-model can be adapted to the covid-19 pandemic by allowing the basic reproduction number to be variable. In practice, such changes are brought about by ordering measures to protect the population, such as quarantine, social distancing, hygienic washing of the hands or mandatory wearing of respirators, but also through sensible and responsible behavior of the population. Only experience can teach how this quantitatively affects \mathcal{R} or the reproduction number R_z introduced further down. Specifically, we introduce the replacement $\mathcal{R} \rightarrow \mathcal{R}(z)$, where $\mathcal{R}(z)$ is obtained by resolving Eq. (12) with respect to \mathcal{R} ,

$$\mathcal{R}(z) = \frac{1}{(1-z)(1-a'(z))}. \quad (14)$$

Inserting this in Eq. (11) yields

$$\dot{z}(t) = \frac{\gamma a(z)}{1-a'(z)} \quad \text{or} \quad t'(z) = \frac{1}{\gamma} \left(\frac{1}{a(z)} - \frac{d \ln a(z)}{dz} \right), \quad (15)$$

where $t(z)$ is the inverse function of $z(t)$. $\dot{z}(t)$ is the temporal increase of $z(t)$, which can be looked up in the daily records of the RKI [3] or JHU [4]. Integration of Eq. (15b)¹ yields

$$t(z) = t_0 + \frac{1}{\gamma} \left(\int_{z_0}^z \frac{dx}{a(x)} - \ln \frac{a(z)}{a(z_0)} \right) \quad (16)$$

with $t(z)=t_0$ for $z=z_0$. (Eqs. (15)-(16) also apply to the case of fixed \mathcal{R} .)

Eq. (16) is the solution for given $a(z)$. We also consider the case where $t(z)$ is given and a solution for $a(z)$ is sought. The differential equation from which $a(z)$ must be determined is Eq. (15a), which with $\dot{z}(t)=1/t'(z)$ can be converted into

$$a'(z) + \gamma a(z) t'(z) = 1.$$

The solution to this equation is readily obtained by using the method of variation of constants and is

$$a(z) = \left(a_* e^{\gamma t(z_*)} + \int_{z_*}^z e^{\gamma t(x)} dx \right) e^{-\gamma t(z)}, \quad (17)$$

where a_* and z_* are arbitrary fixed values in the range of possible a - and z -values.

Annotation: From a mathematical point of view, the modification $\mathcal{R} = \text{const}$ to $\mathcal{R} = \mathcal{R}(z)$ in the SIR-model means that the system (10)-(11) of 2 equations is under-determined because it contains 3 unknown functions, $z(t)$, $a(t)$ and $\mathcal{R}(z)$. However, this is not a problem because it would not even be desirable for $\mathcal{R}(z)$ to be determined by the equations. In practical applications, there is the possibility of specifying any of the 3 functions, either by adaption to empirical data or by making assumptions about the future development

2.3 Additional reproduction number

The variable reduction number introduced in Eq. (14) does not allow a clear interpretation of its values, even though we will see later on that it has values and time variations similar to usual ones. For better comparison with the latter and to enable better interpretability, we introduce a second reproduction number that is similarly defined as the sensitive R-number of the RKI or the net reproduction rate.

For this purpose we assume that a list of successive daily values Z_n for the total number of infected individuals is given. First, a best least-squares fit function $z(t)$ [5] to the corresponding data $\{t_n, z_n\}$, $n=0, 1, 2, \dots$ [6] is constructed, which

¹If there are 2 equations in one line, the first is labeled a and the second b.

in many cases can be done using a power series $z(t) = \sum_0^N a_n t^n$. (In this way, the data is smoothed, as is also done in a similar way in the usual calculation of R values by averaging). The current increase in $z(t)$ is $\dot{z}(t) \Delta t$, and furthermore

$$\dot{z}(t - \Delta t) = \dot{z}(t) - \ddot{z}(t) \Delta t$$

holds. With this, we define our new reproduction factor as

$$R_z(t) = \frac{\dot{z}(t) \Delta t}{\dot{z}(t - \Delta t) \Delta t} = \frac{\dot{z}(t)}{\dot{z}(t) - \ddot{z}(t) \Delta t}.$$

We still relate it to the time unit *day* by putting $\Delta t = 1$, and after a slight reshaping we finally get

$$R_z(t) = \frac{1}{1 - \ddot{z}(t)/\dot{z}(t)} = \frac{1}{1 - \dot{v}_z(t)/v_z(t)} \quad \text{with} \quad v_z(t) = \dot{z}(t). \quad (18)$$

(v_z is the growth velocity of $z(t)$.) On cursory inspection, it looks like $R_z(t)$ is an instantaneous value related only to the moment t . Contrary to this, in the calculation of the sensitive R -value of the RKI, due to a time delay in the reported infection data past values are included, and the data are smoothed by averaging over several days to compensate for statistical fluctuations. On closer inspection, however, this also applies to our $R_z(t)$ because to determine the function $z(t)$ via a best least-squares fit to a data set $\{t_n, z_n\}$, past data are included as well as smoothing takes place.

For the R -number of RKI and the usual net R -number the value 1 is easy to understand, because on average every infected individual infects another one during its contagious stage, whence the number of newly infected individuals remains constant. For values $\neq 1$ a similarly simple explanation is not possible, except that a deviation downwards or upwards is more favorable or unfavorable the larger it is. For our R_z -number the situation is slightly better, because Eq. (18) can be used to quantify the conditions under which a deviation increases or decreases: the greater the relative acceleration \dot{v}_z/v_z , the greater the deviation from 1 where $R_z(t) \equiv 1$ for constant $\dot{z}(t)$ as above. It would even be possible to use the number

$$Q(t) = \frac{\ddot{z}(t)}{\dot{z}(t)} = \frac{\dot{v}_z(t)}{v_z(t)} \quad (19)$$

as another number for characterizing the state of the pandemic. However, in the next section it is shown that a modification of $Q(t)$ is a much better choice.

We would like to apply our reproduction number also where not $z(t)$ is specified but its inverse $t(z)$. For this purpose, we differentiate the identity $t'(z) \dot{z}(t) =$

$(dt/dz)/(dz/dt) \equiv 1$ with respect to t , obtaining $t''(z)\dot{z}^2(t) + t'(z)\ddot{z}(t) = 0$ or $\ddot{z}(t)/\dot{z}(t) = -t''(z)\dot{z}(t)/t'(z) = -t''(z)/(t'(z))^2$, and by insertion into Eq. (18) get

$$R_z(z) = -\frac{1}{1 - t''(z)/(t'(z))^2}. \quad (20)$$

For the case of given $a(z)$ addressed in Section 3.1.1, we still determine how $\mathcal{R}(z)$ can be expressed in terms of $a(z)$. From Eq. (11) with $\mathcal{R} \rightarrow \mathcal{R}(z)$ we get

$$\ddot{z}(t) = \gamma [\mathcal{R}'(z)(1-z)a(z) - \mathcal{R}a(z) + \mathcal{R}(z)(1-z)a'(z)] \dot{z}(t).$$

Dividing by $\dot{z}(t)$ and using

$$\mathcal{R}(z)(1-z)a'(z) = \mathcal{R}(z)(1-z) - 1,$$

which follows from Eq. (14), we get

$$R_z(z) = \frac{1}{1-Q(z)} \quad \text{with} \quad Q(z) = \gamma [\mathcal{R}'(z)(1-z) - \mathcal{R}(z)] a(z) - \mathcal{R}(z)(1-z) - 1. \quad (21)$$

γ must be determined so that $R_z(z_0)$ coincides with the $R_z(t_0)$ of Eq. (18).

2.4 Acceleration parameter

The terms linear and exponential growth are often used to characterize the increase in the number of infections, the net reproduction number being used to distinguish them. As we will see, this is a rather vague and possibly misleading characterization. In the following, another parameter is proposed that enables a more differentiated assessment.

Combining Eq. (18) and Eq. (19) yields

$$R_z(t) = \frac{1}{1-Q(t)} \quad \text{or} \quad R_z(z) = \frac{1}{1-Q(z)} \quad (22)$$

from which follows that R_z and Q offer the same information content. The acceleration parameter proposed by us consists in a modification of $Q = \ddot{z}/\dot{z}$ and is

$$A(t) = \frac{z(t)\ddot{z}(t)}{\dot{z}^2(t)} = \frac{z(t)\dot{v}_z(t)}{v_z^2(t)} \quad \text{or} \quad A(t) = \frac{z(t)Q(t)}{\dot{z}(t)}. \quad (23)$$

The relationship between $A(t)$ and $Q(t)$ shows that $A(t)$ contains more information than $Q(t)$.

Now, we first calculate $A(t)$ for the case of exponential growth, $z(t)=z_0e^{\alpha t}$. With $\dot{z}(t)=\alpha z(t)$ and $\ddot{z}(t)=\alpha^2 z(t)$ we get $A(t)\equiv 1$. For other growth patterns we expand $z(t)$ around an arbitrary $z_0=z(t_0)$ according to

$$z(t) = z_0 + v_0(t-t_0) + A_0(t-t_0)^2/2$$

and get with $\dot{z}(t)=v_0+A_0(t-t_0)$ and $\ddot{z}(t)=A_0$ for $t \rightarrow t_0$ the value

$$A(t_0) = \frac{z_0 A_0}{v_0^2}. \quad (24)$$

Let us now consider different values of A_0 for the same values of z_0 and v_0 . In the case of exponential growth we have $v_0=\alpha z_0$, $A_0=\alpha^2 z_0$ and from this $A_0=v_0^2/z_0$ since $A(t_0)=1$. For $A_0>v_0^2/z_0$ (i.e. $A(t_0)>1$), the curve $z(t)$ is curved stronger upward than an exponential curve of the same slope through the same point z_0 and therefore (locally) shows a super-exponential growth. For $A_0<v_0^2/z_0$ (or $A(t_0)<1$), it is less curved and shows a sub-exponential growth. For $A_0=0$ it is not curved and (locally) linear, and for $A_0<0$ its growth is slower than linear and stops growing at all for $v_0 \rightarrow 0$ with the consequence $A(t_0) \rightarrow -\infty$. These results are summarized in the formula

$$A(t) \begin{cases} > 1 & \text{super-exponential} \\ = 1 & \text{exponential} \\ < 1 & \text{sub-exponential} \\ = 0 & \text{linear} \\ < 0 & \text{sub-linear} \\ \rightarrow -\infty & \text{no growth.} \end{cases} \quad (25)$$

Applications are presented in later sections.

We still want to find out how $A(t)$ and $R_z(t)$ are related. According to Eq. (23b) we have $A=Qz/\dot{z}$, and from Eq. (22) follows $Q=(R_z-1)/R_z$. Combining these results yields

$$A(t) = \frac{z(t) (R_z(t) - 1)}{\dot{z}(t) R_z(t)} = \frac{R_z(t) - 1}{R_z(t) d \ln z(t) / dt}. \quad (26)$$

The function $z(t)$, required to calculate $A(t)$, is determined as discussed in the last section. If, as in the practical determination of the net reproduction number, only the data from the last 8 days are used, a best fit in the class of polynomials $z(t)=\sum_0^N a_n t^n$ should provide suitable results. Daily updating is easy to do by omitting the oldest from the data and adding the newest. Although the parameter $A(t)$ provides more information than the reproduction number $R_z(t)$, it is also not sufficient for a complete assessment of the stage of the pandemic. For decisions on measures to its containment or about the relaxation of such measures, still other quantities (like $\dot{z}(t)$ etc.) and results of statistical investigations must be consulted.

2.5 Various growth possibilities of $z(t)$

2.5.1 Linear growth:

The general solution for linear growth is

$$z(t) = z_0 + \alpha t \quad (27)$$

with constant z_0 and α . Inserting the resulting relations $\dot{z}(t)=\alpha$ and $\ddot{z}(t)=0$ in Eq. (18) yields

$$R_z = 1, \quad (28)$$

and vice versa Eq. (27) follows from Eq. (28).

Inserting $\dot{z}(t)=\alpha$ in Eq. (15a) leads to $\alpha=\gamma a(z)/(1-a'(z))$ or

$$a'(z) + \delta a(z) = 1 \quad \text{with} \quad \delta = \gamma/\alpha.$$

The solution to this equation is $a(z)=1/\delta$ so that $a'(z)=0$, and with this Eq. (14) delivers

$$\mathcal{R}(z) = \frac{1}{(1-z)}. \quad (29)$$

R_z and \mathcal{R} agree only for $z=0$. In order to keep the number of infectious individuals (I or a resp.) constant, R_z must be held fixed at the value 1, while \mathcal{R} increases continuously according to Eq. (29) with Eq. (27).

2.5.2 Exponential growth:

The general solution for exponential growth is

$$z(t) = z_0 e^{\alpha t} \quad \text{with} \quad z_0 = z(0). \quad (30)$$

With $\dot{z}(t)=\alpha z(t)$ and $\ddot{z}(t)=\alpha^2 z(t)$ Eq. (18) yields

$$R_z = 1/(1-\alpha), \quad (31)$$

and from Eq. (11) we get

$$\mathcal{R}(z) = \frac{\alpha z}{\gamma a(z)(1-z)}. \quad (32)$$

In the last result $a(z)$ still has to be calculated what Eq. (17) can be used for. The inverse function $t(z)$ to $z(t)$ from Eq. (30) required for this is

$$t(z) = \frac{1}{\alpha} \ln \frac{z}{z_0} \quad (33)$$

so that

$$e^{\gamma t(z)} = \left(\frac{z}{z_0}\right)^\delta \quad \text{with} \quad \delta = \gamma/\alpha. \quad (34)$$

According to this $t(z_0)=0$. Choosing $z_*=z_0$ and $a_*=a_0$ in Eq. (17), inserting in it the result (32), and using

$$\int_{z_0}^z x^\delta dx = (z^{\delta+1} - z_0^{\delta+1})/(\delta + 1),$$

we finally get

$$a(z) = \frac{z}{\delta + 1} + \left(a_0 - \frac{z_0}{\delta + 1}\right) \left(\frac{z_0}{z}\right)^\delta. \quad (35)$$

When this is inserted in Eq. (32), the solution $\mathcal{R}(z)$ is complete. The time behavior of \mathcal{R} is obtained by substituting $z(t)$ from Eq. (30) for z in $\mathcal{R}(z)$.

2.5.3 Accelerated growth:

We are also interested in how constant acceleration or deceleration $\ddot{z}(t)=\dot{v}_z(t)=\alpha$ of $z(t)$ affects the reproduction number R_z . In this case we have

$$\dot{z}(t) = v_z(t) = v_0 + \alpha t, \quad z(t) = z_0 + v_0 t + \alpha t^2/2$$

and

$$R_z(t) = \frac{1}{1 - (\delta + t)^{-1}} \quad \text{with} \quad \delta = \frac{v_0}{\alpha}. \quad (36)$$

For $\alpha > 0$ we have $\delta > 0$, $R_z(0) = 1/(1 - 1/\delta) > 1$ and $R_z(0) > R_z(t) > 1$. Asymptotically we get

$$R_z(t) \rightarrow \frac{1}{1 - 1/t} \rightarrow 1 \quad \text{for} \quad t \rightarrow \infty. \quad (37)$$

In spite of an accelerated increase of $z(t)$ the reproduction number R_z decreases. This illustrates that the specification of a reproduction number, R_z in our case, is not sufficient to characterize the pandemic.

For $\alpha < 0$ or $\delta < 0$ resp. we can write

$$R_z(t) = \frac{1}{1 + (|\delta| - t)^{-1}}$$

whence $R_z(t) < R_z(0) < 1$, and asymptotically we get

$$R_z(t) \rightarrow 0 \quad \text{for} \quad t \rightarrow |\delta|. \quad (38)$$

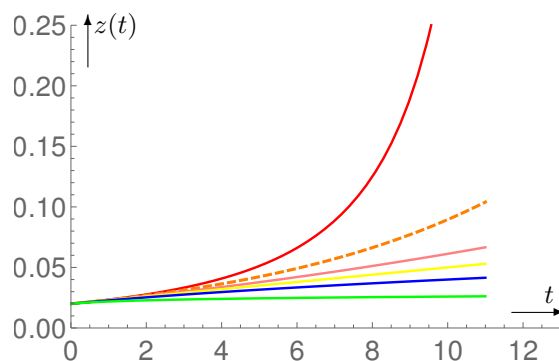


Figure 1: $z(t)$ for $A = 1.5, 1, 0.5, 0, -10$ in the order from top to bottom, the curve for exponential growth being dashed.

2.5.4 Growth with fixed $A(t)$

For fixed $A(t) = \text{const} = A$, a conversion of Eq. (23) with use of $v_z = \dot{z}(t)$ yields

$$\frac{\dot{v}_z}{v_z} = \frac{d}{dt} \ln \frac{v_z}{v_0} = \frac{A \dot{z}(t)}{z} = \frac{d}{dt} \ln \frac{v_z}{v_0}$$

and after integration

$$\ln \frac{v_z}{v_0} = A \ln \frac{z}{z_0} \quad \text{or} \quad v_z = \dot{z}(t) = v_0 \left(\frac{z}{z_0} \right)^A.$$

Further integration leads to the result

$$z(t) = \left(z_0^{1-A} + v_0(1-A)(t-t_0)/z_0^A \right)^{1/(1-A)}. \quad (39)$$

Fig. 1 shows $z(t)$ for 5 different values of A as an illustration, the initial values z_0 and v_0 being arbitrarily chosen but equal for all curves $z(t)$.

Annotation: Linear, exponential and super-exponential growth are all possible, but due to $z(t) \leq 1$ they can only occur temporarily.

2.6 Linear growth of $R_z(t)$

After the surprising result that $R_z(t)$ decreases when $z(t)$ is accelerated, we are interested under which circumstances R_z increases, and examine which consequences result from the ansatz

$$R_z(t) = R_* + \alpha t.$$

With this we get from Eq. (18)

$$\frac{\dot{v}_z(t)}{v_z(t)} = \frac{d(\ln(v_z(t)/v_0))}{dt} = 1 - \frac{1}{R_* + \alpha t}.$$

Integration with respect to t results in

$$\ln(v_z(t)/v_0) = t - \ln(1 + \alpha t/R_*)$$

or

$$v_z(t) = \frac{v_0 e^t}{(1 + \alpha t/R_*)}. \quad (40)$$

$v_z(t)$ growth faster than exponentially (super-exponential growth) for increasing $R_z(t)$ (positive α) and slower than exponentially (sub-exponential growth) for decreasing $R_z(t)$ (negative α).

2.7 Similarity solutions

The data published by the RKI or JHU and updated daily refer to the reported cases. Since many individuals are infected and do not know because they have no symptoms of illness, there is a high number of unreported cases. This can be taken into account by multiplying the reported data by factors. In the equations of the modified SIR-model then the substitutions $Z \rightarrow \tilde{Z} = \delta Z$ with constant δ must be made. Eq. (6) must be satisfied by both sets of variables, which means that

$$\tilde{Z} = \delta I + \delta R = \tilde{I} + \tilde{R}$$

must apply. We now make the (not compulsory) assumption that $\tilde{I} = \delta I$ holds, from which $\tilde{R} = \delta R$ follows. (\tilde{R} is the sum of those who have recovered and those who have died. Since the latter are generally well known, practically none of them are unreported. For the validity of $\tilde{R} = \delta R$, this would have to be compensated by multiplying the recovered ones by a somewhat higher factor.) To the specific variables a and z the same relationships apply as for I and Z , i.e.

$$\tilde{a} = \delta a, \quad \tilde{z} = \delta z. \quad (41)$$

The multiplying factor δ is not known, but can be estimated² and be quite high (e.g. 10 or even higher). It would be an unpleasant surprise if, after completion of a calculation, it turned out that the factor δ used for it was wrong and the

²The number of unreported cases can be roughly estimated from the ratio of death cases to the total number of officially registered infected individuals. On July 4, 2020, the apparent death rate caused by covid-19 was 4.7 % for the world and 4.6 % for Germany. According to preliminary informations provided by the virologists, some of them based on statistical surveys but with some uncertainty due to the low number of cases, the true death rate (lethal rate) is only about 0.5 %. In order for this percentage to come out for the populations mentioned, the number of reported cases e.g. in the world must be multiplied by the factor $\delta = 4.7/0.5 = 9.4$.

whole calculation had to be repeated with another δ . The following shows that this can be avoided under the above conditions.

Let $a(z)$ be the function resulting from a concrete reported data set $\{a_n, z_n\}$ via a best least-squares fit. The unreported cases are taken into account through the replacements $a_n \rightarrow \tilde{a}_n = \delta a_n$ and $z_n \rightarrow \tilde{z}_n = \delta z_n$, and for the associated fit function applies $a(z) \rightarrow \tilde{a}(\tilde{z}) = \delta a(z)$ with $\tilde{z} = \delta z$, in short

$$\tilde{a}(\tilde{z}) = \delta a(\tilde{z}/\delta). \quad (42)$$

(From this follows $\tilde{a}(\tilde{z}_n) = \tilde{a}_n = \delta a_n = \delta a(z_n)$ as required.) Using $dz/d\tilde{z} = 1/\delta$, the derivation of this with respect to \tilde{z} yields

$$\tilde{a}'(\tilde{z}) = \delta a'(z) \Big|_{z=\tilde{z}/\delta} dz/d\tilde{z} = a'(z) \Big|_{z=\tilde{z}/\delta}. \quad (43)$$

Eq. (14) must also apply to \tilde{a} and \tilde{z} , why with use of Eqs. (41) and (43) we get

$$\tilde{\mathcal{R}} = \mathcal{R}(\tilde{z}) = \frac{1}{(1 - \tilde{z})(1 - a'(z) \Big|_{z=\tilde{z}/\delta})} \quad (44)$$

and

$$\tilde{\mathcal{R}}(t) = \mathcal{R}(\tilde{z}(t)) = \frac{1}{(1 - \delta z(t))(1 - a'(z) \Big|_{z(t)})}. \quad (45)$$

With $\tilde{z}(t) = \delta z(t)$, from Eqs. (18) and (19) we obtain

$$\tilde{R}_z(t) = \frac{1}{1 - \tilde{z}(t)/\dot{z}(t)} \quad \text{and} \quad \tilde{Q} = \frac{\dot{z}(t)}{\dot{z}(t)}. \quad (46)$$

Finally we want to find out how the result (16) is influenced by δ . From Eq. (42) follows

$$\frac{\tilde{a}(\tilde{z})}{\tilde{a}(\tilde{z}_0)} = \frac{a(z)}{a(z_0)} \Big|_{z=\tilde{z}/\delta}$$

and

$$\int_{\tilde{z}_0}^{\tilde{z}} \frac{d\tilde{x}}{\tilde{a}(\tilde{x})} = \int_{\delta z_0}^{\delta z} \frac{\delta dx}{\delta a(\tilde{x}/\delta)} = \int_{\delta z_0}^{\delta z} \frac{dx}{a(x)} = \int_{z_0}^z \frac{dx}{a(x)} \Big|_{z=\tilde{z}/\delta},$$

which eventually leads to the result

$$t(\tilde{z}) = t(z) \Big|_{z=\tilde{z}/\delta}. \quad (47)$$

3 Investigation of concrete situations

3.1 Modified SIR-model adapted to specific cases

The SIR-model describes the average behavior of the statistical variables S, I and R or z and a in our case. Good agreement with the data supplied by the RKI or JHU can therefore only be expected when the number of cases is sufficiently high. Our approach to checking the applicability of the modified SIR-model consists in either specifying the function $a(z)$ and calculating the function $z(t)$ or $t(z)$ resp. from the SIR-equations, or vice versa. To be more precise, we first look for an analytic function $a(z)$ (or $t(z)$), whose values $a(z_i)$ (or $t(z_i)$) at the positions z_i agree as well as possible with the given data pairs $\{a_i, z_i\}, i=0, 1, 2, \dots$ (or $\{t_i, z_i\}, i=0, 1, 2, \dots$). The function $t(z)$ calculated from Eq. (16) (or $a(z)$ calculated from Eq. (17)) is then compared with the corresponding reported data. The parameter γ , which has not yet been determined, is used as a fit parameter.

There are several problems with the described procedure regarding accuracy. In our calculations, we arbitrarily determined the number of unreported cases so that all reported data a_i and z_i are multiplied by the same factor $\delta = 10$. As more and more people get tested for Covid-19, that number gets smaller, which we didn't take into account. Furthermore, the data a_i of the infectious individuals are less reliable than the total number z_i of infections, which may be the cause why they are not specified in the official dashboards (see e.g. [3] or [4]). Also, the determination of the functions $a(z)$ or $z(t)$ adapted to the given data is not straight forward. The path we have chosen avoids too much effort and may not be the best. However, in view of the other error sources, it is not worth investing too much effort into it.

3.1.1 Given $a(z)$

We examine the case of given data for the function $a(z)$, using the data reported for the world as an example and multiplying them by $\delta=10$ to account for unreported cases. Corresponding data for the time interval from 3.1.2020 to 7.9.2020 (128 days) are taken from Ref. [7], and the function $a(z)$ is the best least-squares fit in the class of polynomials $\sum_{v=0}^n c_v z^v$, both plotted in Fig. 2.³ The function $a(z)$ was then used to calculate $t(z)$ from Eq. (16), γ being determined such that $t(z_i)=t_i$ for a specific pair $\{z_i, t_i\}$ of the data set from Ref. [7]. The comparison of the solution $t(z)$ with this data is also shown in Fig. 2. The correspondence is not perfect, but good enough that one can infer the usefulness of the modified SIR-model for certain questions, at least in terms of quality. Furthermore, the

³It suffices to consider the function $a(z)$ because it runs through the same values as $a(t)$ and increases monotonically with t because of $\dot{z}(t) \geq 0$.

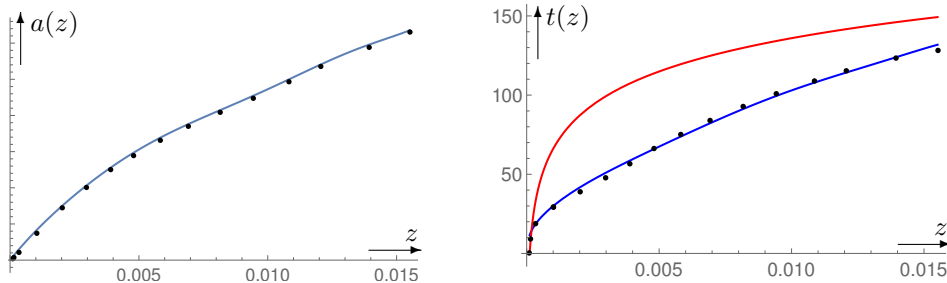


Figure 2: Left figure: reported data $\{z_n, a_n\}$ (black dots) for the world and associated best fit curve $a(z)$. Right figure: curve $t(z)$ calculated from $a(z)$ with Eq. (16) together with the corresponding reported data $\{z_n, t_n\}$. An exponential curve with the same initial values is entered above it, showing, that $z(t)$ is initially super-exponential.

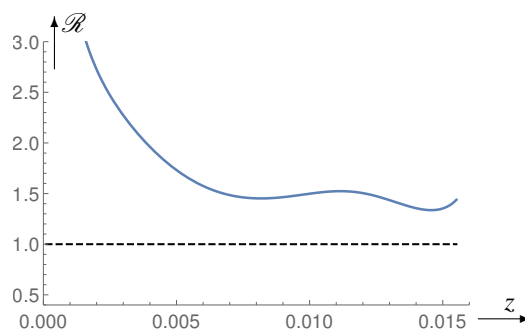


Figure 3: $\mathcal{R}(z)$ and for the world, same time interval as in Fig. 2.

reproduction number $\mathcal{R}(z)$ is shown in Fig. 3. (R_z is shown as a function of t in Fig. 5.)

3.1.2 Given $t(z)$

The reverse case, in which the data for $t(z)$ are given and $a(z)$ is calculated, is examined by the example of Germany. Corresponding data for the time interval from 4.4.2020 to 7.13.2020 are taken from Ref. [8]. Instead of $z(t)$, we determine the inverse function $t(z)$ as a fit to the reported data, but this time not in the class of polynomials, because this resulted in a function with too many fluctuations. (For similar reasons, only the range with decreasing values of the target function $a(z)$ was examined.) Instead, the approach

$$t(z) = (d + \tan(z/a))/b$$

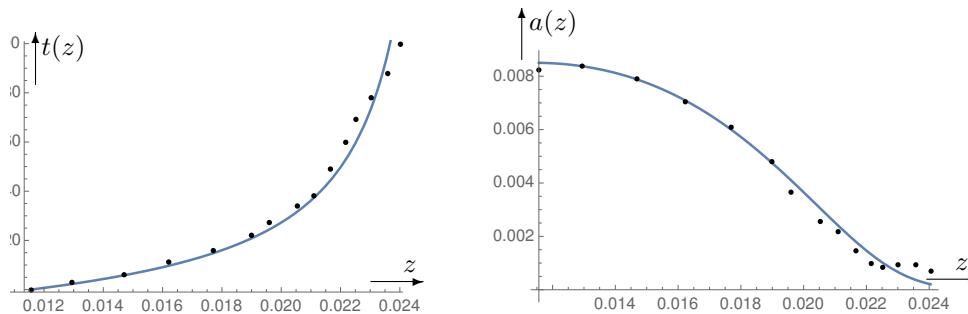


Figure 4: Left figure: reported data $\{z_n, t_n\}$ for Germany and associated fit curve $t(z)$. Right figure: curve $a(z)$ calculated from $t(z)$ with Eq. (17) together with the reported data $\{z_n, a_n\}$.

was used, the parameters a, b , and d being chosen to optimize the fit to the data (see Fig. 4). $a(z)$ was then calculated from Eq. (17) and γ determined such that $a(z_i) = a_i$ for a specific pair $\{z_i, a_i\}$ of the data. The comparison of the solution $a(z)$ with the data is also shown in Fig. 4.

3.2 Application of the acceleration parameter $A(t)$

The acceleration parameter $A(t)$ can be calculated directly from reported data without recourse to the SIR-model, the function $z(t)$ again being determined by a (best) fit. Fig. 5 illustrates this process, using again the example of the world. For $A(t) > 1$, the growth becomes super-exponential, and almost linear around day 60.

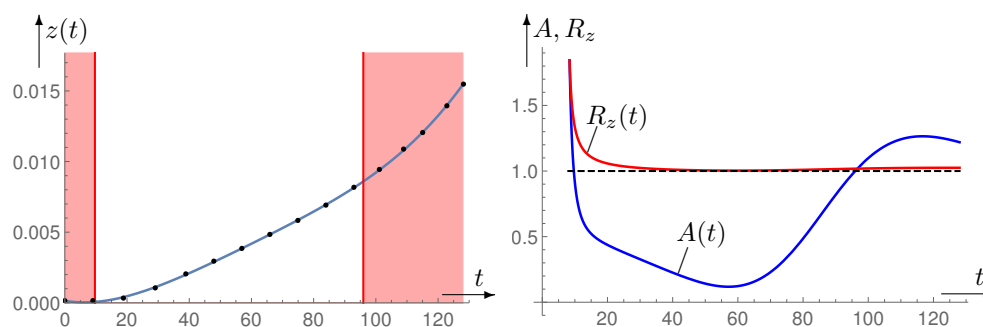


Figure 5: Left figure: reported data $\{t_n, z_n\}$ for the world over a time interval of 128 days, and associated best fit curve $z(t)$. In the shaded areas, $z(t)$ is super-exponential. Right figure: Associated parameters $A(t)$ and $R_z(t)$ calculated from $z(t)$ using Eqs. (18) and (23).

4 Future projections

We have seen that the modified SIR-model is well applicable to the pandemic, although in some cases rather only to its more advanced stage. We therefore assume that it can also be applied to future developments to a certain extent. Of particular interest is how $a(t)$ and $z(t)$ develop when certain assumptions about the further course of $\mathcal{R}(t)$, $R_z(t)$ or $A(t)$ are made.

4.1 Forecast solutions

Predefined $\mathcal{R}(z)$: In this case, from Eq. (10) results the equation

$$a'(z) = 1 - \frac{1}{(1-z)\mathcal{R}(z)},$$

which is solved by

$$a(z) = a_* + z - z_* - \int_{z_*}^z \frac{dx}{(1-x)\mathcal{R}(x)}. \quad (48)$$

The associated solution $z(t)$ is given by its inverse function $t(z)$ to be calculated from Eq. (16). For γ , the value of the preceding solution can be used, to which the forecast solution is linked.

Predefined $R_z(t)$: For given $R_z(t)$, from Eq. (18) results the equation

$$\dot{v}_z(t)/v_z(t) = 1 - 1/R_z(t),$$

which is solved by

$$v_z(t) = v_0 e^{t - \int_0^t (1/R_z(\tau)) d\tau} \quad \text{for} \quad v_z(0) = v_0, \quad (49)$$

and integration of $\dot{z}(t) = v_z(t)$ yields

$$z(t) = z_0 + \int_0^t v_z(\tau) d\tau \quad (50)$$

with $v_z(\tau)$ given by Eq. (49). For determining the associated function $a(t)$, we draw on Eq. (15a) in the form

$$a'(z) + \gamma a/\dot{z}(t) = a'(z) + \gamma a/v_z(t) = 1.$$

Using $\dot{a}(t) = a'(z)\dot{z}(t) = a'(z)v_z(t)$ or $a'(z) = \dot{a}(t)/v_z(t)$, from this we get the equation

$$\dot{a}(t) + \gamma a(t) = v_z(t),$$

which is solved by

$$a(t) = \left(a_0 + \int_0^t v_z(\tau) e^{\gamma\tau} d\tau \right) e^{-\gamma t}. \quad (51)$$

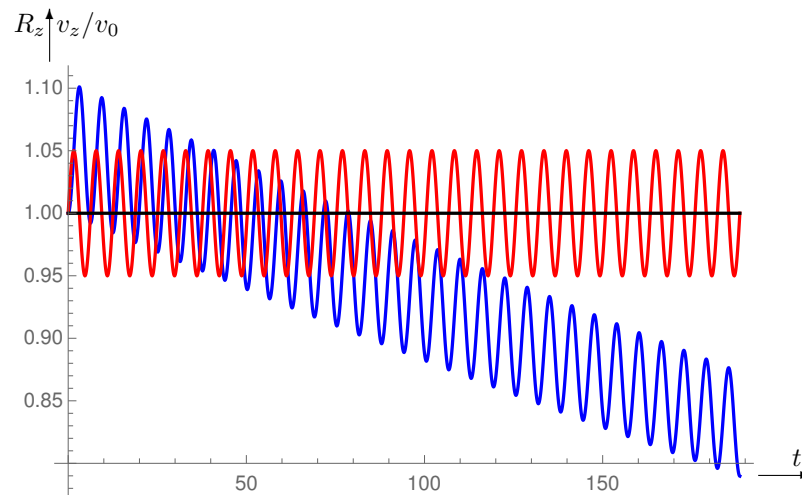


Figure 6: The horizontal curve is $R_z(t)$, the top-down curve $v_z(t)/v_0$.

$R_z(t)$ oscillations around $R_z=1$: In practice, after the number of newly infected individuals has dropped to a lower level, it is variously observed that the net reproduction number is close to 1 and alternately exceeds this value and falls below it. We simulate this situation by making the ansatz

$$R_z(t) = 1 + \varepsilon \sin(\omega t). \quad (52)$$

According to this and Eq. (49) the relative rate of new infections is

$$\dot{z}(t) = v_z(t) = v_0 e^{t - \int_0^t (1/(1 + \varepsilon \sin(\omega \tau))) d\tau}. \quad (53)$$

The integral in the exponent can indeed be expressed through analytic functions,

$$\int_0^t \frac{d\tau}{1 + \varepsilon \sin(\omega \tau)} = \frac{2 \arctan\left(\frac{\varepsilon + \tan(\omega t/2)}{\sqrt{1 - \varepsilon^2}}\right)}{\omega \sqrt{1 - \varepsilon^2}}.$$

It is, however, rather tedious to get the various branches of the trigonometric functions aligned correctly adjacent to each other. We have therefore evaluated Eq. (53) numerically. The result is shown in Fig. 5. The surprising outcome is that $v_z(t)$ gradually decreases. The upward exceedings of limit 1 by v_z are therefore harmless if they are compensated for by undershoots of the same duration and strength.

4.2 End of the pandemic

A pandemic ends when there are no more infectious people, i.e. for $I=0$ or $a=0$.

Case of constant \mathcal{R} : In this case Eq. (13) applies, and for $a=0$ we obtain

$$\ln \frac{1-z}{1-z_0} = -\mathcal{R}(a_0 + z - z_0).$$

With the definitions

$$W = -\mathcal{R}(z-1), \quad W_0 = -\mathcal{R}(z_0-1) \quad (54)$$

and by transition from the logarithm to the exponential function this results in

$$W e^W = V \quad \text{with} \quad V = W_0 e^{W_0 - \mathcal{R} a_0}. \quad (55)$$

The resolution of this equation with respect to W is called Lambert-W-function, omega function or product logarithm. It has several branches, of which the one we need is called the principal solution $W_0(V)$. Thus, in terms of W and V the solution of our problem is $W=W_0(V)$. Returning to our original variables by use of Eqs. (54) and (55b), for the value $z=z_f$, at which the pandemic ends, we get

$$z_f = 1 + W_0(V)/\mathcal{R} \quad \text{with} \quad V = -\mathcal{R}(1-z_0) e^{-(1+a_0-z_0)}. \quad (56)$$

Case of constant R_z : According to Eq. (18) for constant R_z we get

$$\dot{v}_z(t) = \alpha v_z(t) \quad \text{with} \quad \alpha = (R_z - 1)/R_z \quad (57)$$

and from this

$$v_z(t) = v_0 e^{\alpha t} \quad (58)$$

with $v_0 \geq 0$, because $z(t)$ cannot decrease. This means that the constancy of R_z implies exponential growth of $v_z(t)$. According to Eq. (11) the condition $a=0$ for ending the pandemic calls for $\dot{z}(t)=v_z=0$ which is not possible with $\alpha \geq 0$, and for $\alpha < 0$ this is only achieved after an infinitely long time. With $v_z=\dot{z}(t)$ from Eq. (58) results the differential equation $\dot{z}(t)=v_0 e^{\alpha t}$ which is solved by

$$z(t) = z_0 + (v_z(t) - v_0)/\alpha. \quad (59)$$

According to the termination condition $v_z=0$, from this and Eq. (58) we obtain

$$\alpha \leq 0 \quad \text{and} \quad z_f = z_0 + v_0 |\alpha|. \quad (60)$$

The final value z_f that $z(t)$ assumes at the end of the pandemic remains below its maximum 1 only for $|\alpha| > v_0/(1-z_0)$.

5 Concluding remarks

By many, the exponential growth in infection numbers is considered as the worst of all possibilities, and it is assumed to occur immediately when the net reproduction number exceeds 1. The first assumption overlooks the fact that the term exponential growth does not represent a complete and adequate characterization, because there is slow and fast exponential growth ($e^{\alpha t}$ with a small or large α) as there can be slow and fast linear growth ($\dot{z}(t)=\alpha$ with a small or large α). Accordingly, rapid linear growth can be worse than slow exponential growth over a long period of time. So the growth of $Z(t)$ in the world was almost linear from the beginning of April to the middle of May, but unfortunately at very high speed. That the second assumption is wrong can be seen from Eqs. (25)-(26). According to them $A(t)=0$ for $R_z(t)=1$. If $R_z(t)$ were to become a little larger from this value, then $A(t)$ would have to jump to 1 immediately for exponential growth, which is impossible according to Eq. (26).⁴ As we have seen, the modified SIR-model allows for all growth possibilities, from sub-linear up to super-exponential. How dangerous the current growth is depends, however, not only on its type as characterized by $A(t)$, but also on its current growth rate v_z , and on how long this kind of growth will last.

A question, that has been asked many times and the answer to which is important for the acceptance of restrictions that must be endured to contain the pandemic, is: Why do the latter still have to be maintained when the rate of new infections (i.e. v_z) has become sufficiently low. The answer to this is not easy, but there is an analogy familiar to all of us which maybe contributes to some understanding. Imagine driving a car and pressing the accelerator halfway to maintain a certain speed. If you want to drive faster, you have to push it further, e.g. three quarters. If you would release the pedal now, you would fall back to lower speed. In a similar way, you either have to maintain the measures undertaken for keeping low the number of new infections, or to replace them with equally effective, but more targeted and less restrictive measures. Note that this analogy is of a more symbolic kind, because neither frictional forces nor inertia and energy supply can be attributed to the pandemic. At most, in the case $R_z < 1$ or $A(t) < 0$ a constant negative acceleration $\dot{v}_z(t) = \alpha < 0$ (treated at the end of Section 2.5) can to a limited extent be compared with the acceleration of a car. However, according to Eq. (5), independently of α the final velocity v_z is always zero due to the lack of friction. The fact that it is so difficult to bring and keep R_z well below 1 or $A(t)$ far below zero can possibly be explained as follows: Either the effort to maintain a negative

⁴This does not conflict with the fact that according to Eqs. (58)-(59) $v_z(t)$ and $z(t)$ grow exponentially for constant $R_z > 1$, because also $A \geq 1$ in this case. The consequence of this is that R_z may not be constant. An example for $A(t) < 1$ with simultaneous $R_z(t) > 1$ is provided by $z(t) = \alpha t^2 / 2$, for which $A(t) = 1/2$ and $R_z(t) = (1 - 1/t)^{-1}$.

acceleration increases with decreasing $v_z(t)$, or else, there are processes that work like negative friction. A third mechanism is discussed in the next paragraph.

When looking at the infection rates which the JHU releases for the different countries of the world, one notices that for those who have weathered the crisis rather well (e.g. Germany, Austria, Japan, South Korea or New Zealand), $\dot{z}(t)$ does not decrease completely down to zero, but only to a low and almost constant level. One reason for this could also be, that the different countries do not form closed systems, so that due to the pandemic nature of the covid-19 crisis infections are always brought in from outside. Since the SIR-model only applies to closed systems, this factor cannot be treated with it. That would be possible for the world as a whole, but the latter is still far from low infection rates.

Acknowledgements: For calculations and plots MATHEMATICA 10 was used.

References

- [1] Kermack W.O., McKendrick A.G., A contribution to the mathematical theory of epidemics, Proc. Roy. Soc. A, **115** (1927), p. 700721
- [2] <https://de.wikipedia.org/wiki/SIR-Modell>
- [3] <https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4>
- [4] <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>
- [5] Wolfram S., Mathematica: A System for Doing Mathematics by Computer, Addison-Wesley Publishing Company, Redwood City Cal., (1988) p. 461
- [6] Wolfram S., Mathematica: A System for Doing Mathematics by Computer, Addison-Wesley Publishing Company, Redwood City Cal., (1988) p. 542
- [7] <https://interaktiv.tagesspiegel.de/lab/sars-cov-2-das-virus-in-echtzeit/>
- [8] <https://www.welt.de/vermishtes/article206504969/Corona-Wieder-unter-400-neue-Faelle-in-Deutschland-gemeldet.html>