



American Journal of Epidemiology & Public Health

Research Article

The Dynamic Reproduction Index: Accurate Determination from Incidence and Application for an Early Warning System -

Robert NJ Conradt and Stephan Herminghaus*

Max Planck Institute for Dynamics and Self-Organization (MPI-DS), Am Faßberg 17, D-37077 Göttingen, Germany

***Address for Correspondence:** Stephan Herminghaus, Max Planck Institute for Dynamics and Self-Organization (MPI-DS), Am Faßberg 17, D-37077 Göttingen, Germany, E-mail: stephan.herminghaus@ds.mpg.de

Submitted: 29 April 2022; **Approved:** 18 May 2022; **Published:** 20 May 2022

Cite this article: Conradt RNJ, Herminghaus S. The Dynamic Reproduction Index: Accurate Determination from Incidence and Application for an Early Warning System. American J Epidemiol Public Health. 2022 May 20;6(2): 030-037. doi: 10.37871/ajeph.id56

Copyright: © 2022 Conradt RNJ, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Two methods of calculating the reproduction index from daily new infection data (incidence) are considered, one by using the generation time t_G as a shift (R_G), and an incidence-based method directly derived from the differential equation system of an SIR epidemic dynamics model (R_I). While the former (which is commonly used) is shown to be at variance with the true reproduction index, we find that the latter provides a sensitive detection device for intervention effects and other events affecting the epidemic, making it well-suited for diagnostic purposes in policy making. Furthermore, we introduce a similar quantity, R_I^{calc} , which can be calculated directly from R_G . It shows largely the same behaviour as R_I , with less fine structure. However, it is accurate in particular in the vicinity of $R = 1$, where accuracy is important for the correct prediction of epidemic dynamics. We introduce an entirely new, self-consistent method to derive, an improved R_I^{corr} which is both accurate and contains the details of the epidemic spreading dynamics. Hence we obtain R accurately from incidence data alone. Moreover, plotting R versus incidence reveals the orbital structure of epidemic waves, whose fine structure features clearly correlate with public events and interventions, thus providing a sensitive diagnostic tool for policy making. It is demonstrated that the widespread use of only incidence as a diagnostic tool is clearly inappropriate.

Keywords: Dynamic reproduction index; Epidemic dynamics model; COVID-19 pandemic; SARS-CoV-2

INTRODUCTION

As the impact of the human species on the biosphere intensifies and biodiversity is in rapid decline, zoonoses are expected to become more frequent, posing an increasing hazard on global human health [1-3]. The outbreak of the COVID-19 pandemic, caused by the SARS-CoV-2 virus, had unprecedented impact on societies all over Earth. Mitigation measures included complete lockdowns of societal life, with severe social, economic, and individual consequences [4,5]. The dramatically varying success [6-9] of the interventions owed in part to cultural differences [6], but also to only limited understanding of infection spreading dynamics and a severe lack of established methods in epidemic state diagnosis and prediction. Improvement of this general situation is imperative, in particular as similar events are expected to strike more often in the future.

In search for optimized strategies, a two-fold view must be adopted. On the one hand, one needs to understand, in retrospect, which interventions have had what effect on the epidemic spreading dynamics, in order to properly design future interventions. This requires sensitive diagnosis tools for assessing the state of the epidemic on a (if possible) daily basis. On the other hand, tools are needed for predicting the future of epidemic dynamics as reliably as possible. Aside from extensive simulation, this requires careful analysis of data, such as the number of infected citizens [5,10].

MATERIALS AND METHODS

Here we discuss the system in terms of an SIR model [11,12], referring to the number of Susceptible (S), Infected (I), and Recovered (R) individuals, respectively, in a population of N citizens. Here we identify with R all those who are neither susceptible nor infected ($R = N - S - I$), which includes those who are deceased. While the ratio of deceased vs. recovered individuals is undoubtedly of utmost societal concern, it can be disregarded here, as we will solely discuss prevention measures addressing the spreading of the disease. We define I as the number of individuals who carry sufficient viral load to be contagious. They are assumed to remain in this state for an average duration τ . Although the viral load changes with time during the illness, contagion can be sufficiently well described by this simple picture for our purposes [13].

The spreading dynamics of an epidemic can then be described by a set of two equations [12],

$$\begin{aligned} \partial_t s &= -csi, \\ \partial_t i &= csi - \frac{i}{\tau}, \end{aligned} \quad (1)$$

where ∂_t is the derivative with respect to time, while $s(t) = S/N$ and $i(t) = I/N$ are the fractions of susceptible and infected individuals in the population, respectively. C is the average number of new infections a single infected individual would cause per unit time in an otherwise infection-free (but susceptible) population. It is accessible to interventions such as closing schools, wearing facial masks etc., but this shall not concern us here, as we focus solely on methods to detect the current stage of an epidemic and to predict its near future development, at given C .

The reproduction index

The base reproduction index, R_0 , is related to C via

$$R_0 = c\tau. \quad (2)$$

It denotes the total average number of individuals newly infected by a single infected one under the above conditions. Since the probability of infection is directly proportional to the fraction of susceptibles, we have

$$R(t) = c\tau s(t). \quad (3)$$

for the dynamic reproduction index, $R(t)$ [12]. The latter is of tremendous importance for assessing the current status of an epidemic. If $R = 1$, I stays constant, but when $R > 1$, each infected individual causes more than one new infection on average, such that $I(t)$ increases exponentially. It is thus of major interest to determine R from epidemiological data as accurately as possible, in particular in the vicinity of unity.

By combining eq. (3) with the first eq. (1), we obtain

$$R(t) = \frac{\tau |\partial_t s|}{i}. \quad (4)$$

This relies only on quantities which can be derived from data usually available from the health care system. In particular, C does not appear in eq. (4). $\partial_t S$ represents the number of new infections per unit time (daily incidence) and can be considered known accurately. τ is known from clinical experience with the disease, and $I(t)$ (and hence i) can be estimated once τ is known. We will discuss further below how to treat uncertainties and possible temporal variations in τ . It is in particular the dynamic variations of R which we will show to yield considerable insight into the infection process.

Since infection data are discrete data collected on a daily basis [14], we will now write down a discrete version of eq. (4). The daily incidence will be called $\dot{S} = |\partial_t S|$. τ as well as t will henceforth be expressed in units of days, and treated as discrete variables. We then may be tempted to simply write $R(t) = \tau \dot{S} / I$. However,



we must be aware that for rather general infrastructural reasons, the reporting efficiency of infection numbers varies characteristically, e.g., on weekends. We therefore should provide for suitable averaging. Hence we write

$$R_I^{lk}(t) = \frac{\frac{1}{2l+1} \sum_{j=-l}^l -\dot{S}(t-j)}{\frac{1}{k} \sum_{j=l+1}^{l+k} I(t-j)} \tau, \tag{5}$$

where l and k are parameters determining the intervals over which $-\dot{S}$ and I are being averaged, respectively. Since the typical variability of data reflects the sequence of seven weekdays, it appears reasonable to average $-\dot{S}$ over seven consecutive days. If we furthermore average the (less variable) number of infected individuals over a period $k = \tau$, we obtain

$$R_I(t) = R_I^{3\tau}(t) = \frac{\frac{1}{7} \sum_{j=-3}^{+3} \dot{S}(t-j)}{\sum_{j=4}^{\tau+3} I(t-j)} \tau^2 \tag{6}$$

for the incidence-based dynamic reproduction index. While \dot{S} is known precisely, I can only be estimated based, among others, on τ . However, this has only minor effects on the accuracy of R_I and its dynamic variations, as I enters only as an average over the duration τ .

The merits of R_I show up clearly when compared to other definitions of reproduction indices which are currently used in epidemic data based diagnostics of the infection dynamics. Since we will later use data from Germany in our analysis, we refer to what is issued by the Robert Koch Institute (RKI) in Germany as the “reproduction index”. It is based on the idea of calculating the ratio of incidence data, taken on two successive instants, separated by a delay time t_G [15]. The latter is called the generation time and represents the average time interval between an infection and a subsequent “successful” transmission of the infection to a third person. The RKI uses $t_G = 4$ days. The definition of this generation-time based reproduction index is then [16,17]

$$R_G(t) = \frac{\sum_{j=0}^3 \dot{S}(t-j)}{\sum_{j=0}^3 \dot{S}(t-j-t_G)}. \tag{7}$$

Its mathematical meaning becomes clearer in a continuous formulation,

$$R_G^{cont}(t) = \frac{\partial_t s(t)}{\partial_t s(t-t_G)}. \tag{8}$$

This can be written as

$$R_G^+(t) = R_G^{cont}(t+t_G) = 1 + t_G \partial_t \ln |\partial_t s(t)|, \tag{9}$$

where we have truncated the Taylor expansion after the first term. This reveals that R_G is directly related to the logarithmic derivative

of the daily incidence, $|\partial_t s(t)|$, with some delay equal to t_G . Hence whenever the daily incidence happens to vary exponentially, $\partial_t s(t) \propto \exp(R-1)t$, R_G can indeed be interpreted as a reproduction index. At any other time, however, when this is not the case, the use of R_G as a reproduction index lacks mathematical foundation. Note furthermore that while c enters directly in R (as given by eqs. (3) and (4)), it cancels out in all expressions for R_G . Hence public measures affecting c will readily show up in R_I , which has been defined according to eq. (4), but not in R_G .

Application to epidemic data

From data obtained in Germany during the SARS-CoV2 pandemic in 2020 and 2021 [18], let us now calculate $R_I(t)$ and $R_G(t)$ by means of eqs. (6) and (7), respectively. The result is displayed in the top panel of figure 1, exposing the remarkable differences between the two quantities. There is a strong tendency of $R_G(t)$ to stay closer to unity than $R_I(t)$, which reflects the dynamics more pronouncedly. Events like the strong increase of $R_I(t)$ up to a value of 2.3 at the end of October 2020 (around day 245) hardly show up in $R_G(t)$.

For a more detailed discussion, a number of important events are listed in table 1. After day 63, which corresponds to May 5, 2020, we see a sharp increase of R_I . At the end of the first epidemic wave, German officials had decided to relax public life to almost normal conditions. Hence stores, restaurants, cultural institutions, and museums were opened at the beginning of May. Because there was only little change in incidence (and prevalence, grey shade at the bottom of the panels) during the rest of spring and summer, it went unnoticed that the reproduction index was undergoing strong changes. That these were much less pronounced in R_G (which was used by officials) than in R_I made their detection particularly difficult.

A few days after a strong rise in R_G was noticed, a major disease outbreak was reported on day 109 at the Tönnies slaughterhouse site near Gütersloh, among the large number of loan workers living at the site. But since prevalence remained low after that, the further rise of R was not noticed, or taken seriously, certainly in part due to the noisiness of the data. After the incidence then rose very sharply during October 2020, the German government decided on what became known as a “lockdown light”, starting from November 2nd (day 245, labelled LL in figure 1). This was associated with a hope of possibly easing restrictions for Christmas. Note that the maximum of $R_I(t)$ lies slightly but distinctly before day 245. We will come back to this aspect further below.

Because the pronounced decrease of the reproduction index is not reflected by R_G , and incidence did not seem to decrease, it was not noticed that the situation was actually relaxing during November. Consequently, a hard lockdown (including widespread closures of shops, businesses, schools, etc.) was announced on December 1st (day 275), which would start on December 16th (day 290). In particular, the relaxations for Christmas that had been previously promised were withdrawn. This announcement led to a short-lived but sharp increase in infections, as many people squeezed through the two-weeks bottleneck into (consequently crowded) shops for their Christmas shopping. The subsequent decline (presumably due to the

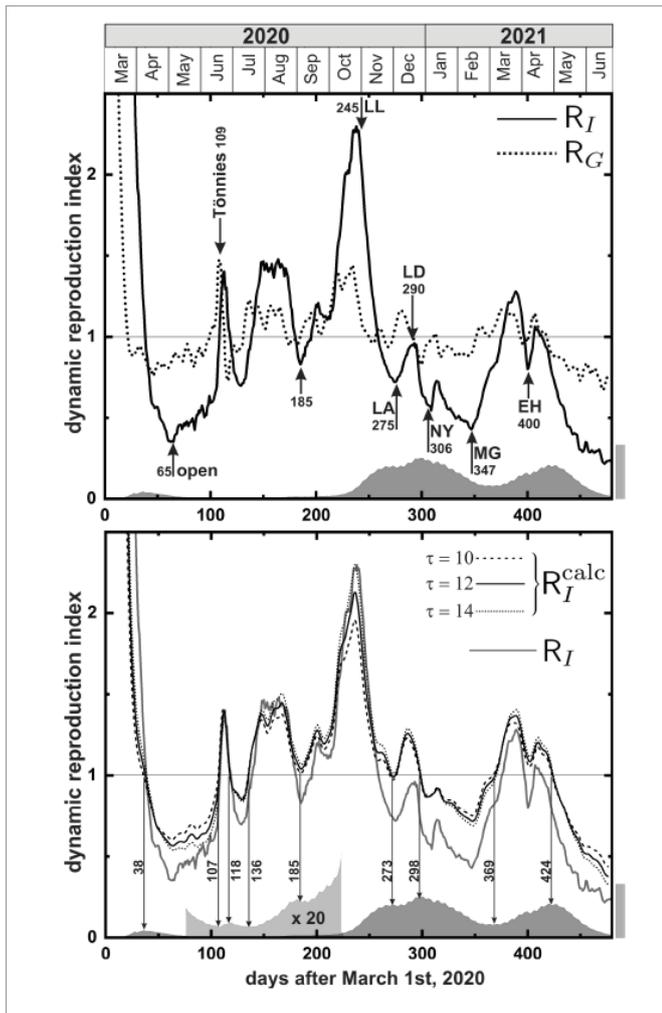


Figure 1: Top panel: the two reproduction indices $R_G(t)$ (dotted) and $R_I(t)$ (solid) as obtained from infection data in Germany during the SARS-CoV2 pandemic until mid 2021. The arrows indicate certain events and interventions (Table 1) as discussed in the text. The grey shaded curve at the bottom indicates the prevalence (number of infected). The corresponding grey scale bar to the right corresponds to one half million people. Bottom panel: $R_I(t)$ as obtained from $R_G(t)$ through eq. (14) for different values of τ (black curves) and numerically from infection data (grey, same as black curve in top panel). The vertical arrows are located where the black curves reach unity, and coincide well with the extrema of the prevalence. In the interval $t \in [75, 225]$, the prevalence curve has been scaled by a factor of 20 for visibility.

Table 1: A number of events marked in the top panel of figure 1.

t	Date	Label	Event
[d]	(d/m/y)	(Figure 1)	
65	5/5/2020	open	Reopening stores and restaurants
109	18/6/2020	Tönnies	Outbreak at Tönnies company site (Gütersloh) is noticed
245	1/11/2020	LL	Lockdown light
275	1/12/2020	LA	December lockdown announced
290	16/12/2020	LD	December lockdown starts
306	1/1/2021	NY	New year's day
347	11/2/2021	MG	Mardi Gras
400	5/4/2021	EH	Easter holidays

start of the Christmas school holidays) abruptly terminated on new year's day (day 347), when many people had visited relatives and friends. A similar feature appears on day 400 at the Easter holidays, for similar reasons. Clearly, most (if not all) of these features are very prominent in $R_I(t)$, but only poorly (if at all) discernible in R_G .

Nevertheless, it may seem that some features in $R_G(t)$ appear a little earlier than corresponding features in $R_I(t)$. This is apparent most clearly from the points where unity is crossed, which for $R_G(t)$ lie significantly before those for $R_I(t)$. This could be interpreted as R_G being better suited for forecast purposes than R_I , as it discloses the same information at an earlier time. However, this turns out to be a delusive mathematical artifact. As we see from eq. (9), with help from eq. (1), R_G^+ can be written as

$$R_G^+ = 1 + t_G \frac{\partial_t |\partial_t s|}{cs i} \tag{10}$$

By means of eqs. (4) and (1), the numerator of the second term can be expressed as

$$\partial_t |\partial_t s| = \frac{i}{\tau} \partial R + |\partial_t s| (cs - \frac{1}{\tau}) \tag{11}$$

Consequently, we have

$$\frac{1}{t_G} (R_G^+ - 1) = \frac{1}{\tau} (R - 1) + \partial_t \ln R \tag{12}$$

In other words, the deviation of $R_G^+(t)$ from unity is composed of a term proportional to the deviation of $R(t)$ from unity and the logarithmic time derivative of $R(t)$. Hence R_G has a much more complex structure than R , exhibiting additional features (and additional noisiness) from the time derivative of R . In particular, the shift towards earlier times of the transitions through unity is merely due to the derivative term in eq. (12), and has no significance concerning epidemic spreading dynamics. Clearly, R_G cannot be interpreted in terms of a true reproduction index.

We have seen above that R_I is a powerful diagnosis tool, as it reacts sensitively to events and interventions in society. Nevertheless, it comes with its downsides. From the second wave (around day 300) we see that the absolute magnitude of R_I cannot be accurate, as the number of infections rises considerably shortly before day 300, while R_I is clearly smaller than unity. This may be attributed to the fact that R_I depends on τ (cf. eqs. (6)), which cancels out for R_G , as it is obvious from eq. (7). The inherent problem is that τ is known with only poor accuracy and can undergo gradual changes during the epidemic. An obvious cause may be the appearance of new mutations, which often result in shifts in the clinical picture, possibly including changes in the duration of the illness, hence in τ [?]. It will only rarely be possible to receive regular reliable data on τ .

RESULTS AND DISCUSSION

It turns out that the independence of R_G from τ may be exploited here by inverting eq. (12), in order to calculate R_I from R_G . If we denote by $D_G = R_G^+ - 1$ the deviation of R_G^+ from unity, we can rewrite eq. (12) (by multiplying with τ and dividing by R) into

$$\partial_t \left(\frac{\tau}{R(t)} \right) + \left(\frac{D_G(t)}{t_G} + \frac{1}{\tau} \right) \frac{\tau}{R(t)} - 1 = 0 \tag{13}$$



This is a linear differential equation in τ / R and can be solved by means of the method of variation of parameters. The result is

$$R_I^{calc}(t) = \frac{\tau e^{\int p(t)dt}}{\int e^{\int p(t)dt} dt}, \tag{14}$$

where

$$p(t) = \frac{D_G(t)}{t_G} + \frac{1}{\tau}. \tag{15}$$

One finds that R_I^{calc} depends on τ only weakly, because its role as a prefactor in eq. (14) and its appearance in $p(t)$ cancel each other to a large extent. In the bottom panel of figure 1, $R_I^{calc}(t)$ thus obtained is plotted for three different values of τ as the black curves. The grey curve is R_I as in the top panel. Clearly, all of the more prominent features of R_I are reproduced.

There are two main differences between $R_I^{calc}(t)$ and $R_I(t)$. First, there is a vertical shift which varies with time only very slowly. Second, much finer details are visible in R_I . The striking feature of R_I^{calc} is that close to its transitions through unity, there is almost no sensitivity to τ . Hence should τ vary over the course of the epidemic by, e.g., as much as 40 percent (as between the dashed and the dotted curve), the shape of $R_I^{calc}(t)$ would not change much. In particular, the points where it hits unity do not change their position appreciably. As the vertical arrows show, these points are very close to the extrema of the number of infected people, as one would correctly expect for the reproduction index.

Hence what we display in the bottom panel of figure 1 may well be called the best of both worlds. In $R_I(t)$ we see very fine details which allow to identify the effects of social events, and to assess the effectiveness of public interventions in retrospect. In $R_I^{calc}(t)$, which we derived from $R_G(t)$ by means of eq. (14), we see less detail, but obtain a more accurate estimate of the reproduction index. This allows for more reliable predictions of near-future epidemic dynamics, as R_I^{calc} is particularly accurate close to unity. The difference between R_I and R_I^{calc} is presumably due to a (slowly) varying τ .

Self-consistent correction of R

In the next step, we seek to exploit the separation of time scales between variations in τ and R , i.e., the fact that variations in τ can be expected to be slow as compared to the rapid variations seen in R . As we will show below, this allows to determine $\tau(t)$ from R_I and R_G in a self-consistent manner. We can then write

$$\tau(t) = q(t)\tau_0, \tag{16}$$

where $q(t)$ is a slowly varying function of order unity and τ_0 is the (assumed constant) value of τ we have initially used to obtain R_I . Once we will have determined $q(t)$, we can write down a corrected reproduction index,

$$R_I^{corr} = q(t)R_I, \tag{17}$$

since R_I proportional to τ according to eq. (4).

In order to obtain $q(t)$, we refer to the only minor impact variations in τ have on $R_I^{calc}(t)$, as visible in the bottom panel of figure 1. In contrast, R_I^{corr} is directly proportional to (and hence strongly dependent on) q (cf. eq. (17)). Hence we seek a slowly varying function $q(t)$ which matches $R_I^{corr}(t)$ to $R_I^{calc}(t)$ as closely as possible. The remaining deviations are then solely due to the truncation of the Taylor expansion we had to introduce to derive eq. (9) (and hence eq. (14)). Formally, this procedure can be expressed as minimizing the integrated deviation,

$$\Delta\{q(t)\} = \int (R_I^{calc} - qR_I)^2 dt, \tag{18}$$

where R_I^{calc} is to be determined using eqs. (14) through (16). A straightforward possibility is then to use a truncated (i.e., low frequency) Fourier series or low order polynomial for $q(t)$ and find the minimum of Δ with respect to the coefficients of its terms.

Alternatively, one can use an iterative approach, which we present here. Since it only uses standard fitting procedures, it easily lends itself to application. First we calculate the ratio $r(t) = R_I^{calc} / R_I$ and fit a slowly varying $q(t)$ to it (cf. eq. (17)), thereby cutting off high-frequency components. Then $\tau_1(t) = \tau_0 q(t)$ yields a new $p_1(t) = \frac{1}{t_G} D_G + \frac{1}{\tau_1(t)}$, which we use, together with $\tau_1(t)$, to recalculate R_I^{calc} from eq. (14). Then R_I is replaced with $q(t)R_I$ and $r(t)$ is recalculated. This is repeated until $q(t)$ (i.e., the result of the polynomial fit) has become stable.

The result we obtained using fourth-order polynomials for $q(t)$ is presented in figure 2. After five iterations, the result for $q(t)$, did not change anymore. As one can clearly see, R_I^{calc} has been rather well matched to R_I^{corr} , aside from some high-frequency variations. One may choose to admit faster variations to τ in order to achieve an even better match between R_I^{corr} and R_I^{calc} , but this shall not concern us here. $q(t)$ is indicated by the dotted curve in the figure, to be read off the left scale. The scale to the right shows the corresponding values of τ . Note that the pronounced peak in R around day 240 reaches a very high value, deviating from unity about four times more strongly than R_G (dotted curve in figure 1). This highlights the importance of our method to calculate R_I for predictions of near-future epidemic spreading dynamics. Around day 240, using R_G would have (or actually has) under-estimated the doubling rate of infections by as much as a factor of 4.5.

Revealing the orbital structure of epidemic waves

Finally, it is instructive to elaborate on some additional aspects of data presentation and analysis. In figure 3a we plot the number of new infections during a period τ against the prevalence, i.e., the total number of currently ill individuals. The ordinate is calculated from the seven-day averaged incidence by multiplying with $\tau / 7$

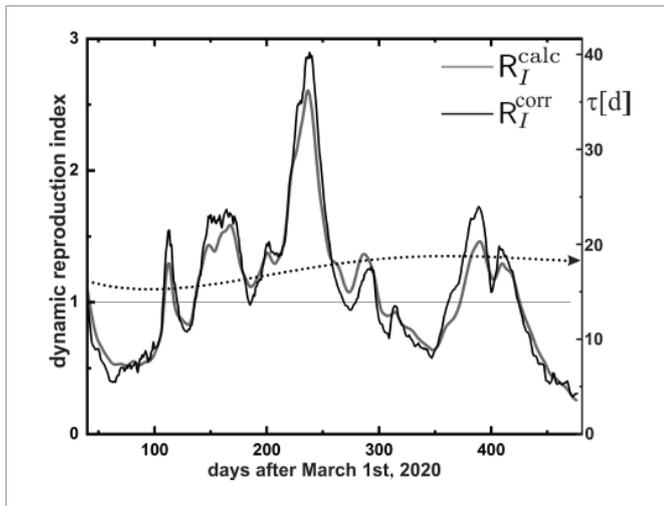


Figure 2: Self-consistent matching of the low-frequency components of $R_G(t)$ (black) and $R_I(t)$ (grey). Matching has been achieved by setting $\tau = q(t)\tau_0$, where $R_G(t)$ has been used for the initial calculation of $R_I(t)$, and $q(t)$ is a fourth-order polynomial. It is indicated by the dotted curve. The corresponding values of $\tau(t)$ can be read off the scale to the right.

Each of the small circles represents one day, with the symbol style representing the three epidemic waves (open, first wave. full grey, second wave. full black, third wave). The data are gathering into an elongated cloud along the first diagonal (dashed line). If we assume that about 3% of individuals infected with SARS-CoV2 need intensive care, we can estimate the maximum prevalence the society could bear. Since there are 16734 intensive care beds in Germany [19], we conclude that the displayed range of the abscissa represents about the maximum “acceptable” range of prevalence (one half million infected individuals).

In order to analyse the internal structure of the data cloud, we compare with a numerically simulated sample trajectory of eq. (1), which is shown as the solid curve. It forms a lobe, starting off at the origin with a slope equal to R_0 , proceeds clockwise (arrows) and re-enters the origin at an inferior slope of $R_1 = 0.14$ (lower dash-dotted line) for $t \rightarrow \infty$. The initial slope, which is indicated by the upper dash-dotted line, follows eq. (4), since the ordinate and abscissa just represent $\tau |\partial_t S|$ and i , respectively. For the simulation, we have set $R_0 = 3.3$ in order to match typical values assumed for SARS-CoV2. The size of the lobe corresponds to the number of people affected by the epidemic. As a consequence of the structure of the solutions to eq. (1), R_1 is a function of R_0 , and $R_1(3.3) = 0.14$.

In fact, the data representing the first wave (open circles close to the origin) exhibit just the same lobe shape, initially following the dash-dotted line, and as time proceeds is traversed in clockwise direction. A closer look at the second (grey) and third (black) wave reveals that their trajectories tend to form clockwise orbits as well, with smaller sub-orbits, thus exposing additional fine structure of the infection dynamics.

Fine structure of orbits and their relation to events and interventions

This orbit structure is revealed more clearly in a different

presentation, when we plot $R(t)$ versus incidence, as shown in figure 3b. Foremost, we note that since these orbits are being passed through in clockwise sense, R is generally a forerunner to incidence. From the number of data points on the orbits, one may appreciate that forecasting epidemic dynamics on the basis of R can be easily two or even three weeks ahead of forecasting on the basis of incidence data.

In order to assess the diagnostic power of this presentation of data, we first consider the vicinity of day 65 (lower left, first entry of table 1). While the seven-days incidence remained on the decline, R abruptly switched from the decline to a sharp rise when stores and restaurants were re-opened. This underpins that R is a much more sensitive diagnostic tool than the (widely used) incidence. Furthermore, no comparable feature can be discerned in the variations of R_G displayed in figure 1.

It is similarly illuminating to discuss the vicinity of day 245, marked as LL. We had already seen from figure 1 that the decline of R during November 2021 cannot have been due to the measures having taken effect on that day, because they came when the decline had already started. Figure 3b shows even more clearly that the situation was already on a relaxing path. When on day 275 (LA) the hard December lockdown was announced, which would take effect on day 290, the orbit was in fact almost finished and had already curved back towards the origin. It is interesting to note that, as it is obvious from the undisturbed shape of the orbit, that the (rather mild) measures which had been taken on day 245 had no discernible effect on the epidemic dynamics.

After Mardi Gras (day 347), incidence stayed calm, and R_G remained featureless within noise level, as figure 1 shows. Hence no measures of caution were taken. The trajectory in figure 3b, however, clearly shows that this was when a new orbit had formed, launching what became the third epidemic wave. Most importantly, we see from figure 3b that the widespread exclusive use of the incidence (abscissa) for assessing the state of the epidemic is void of any sound basis, as it is insensitive to some of the dynamic features of the epidemic and fails to reflect the orbital structure of epidemic waves. In particular, the use of threshold values for incidence in legislation on public mitigation interventions is clearly inappropriate.

At first glance, it may appear remarkable that quantities involving averages over several days in their derivations, as apparent in eqs. (5) and (6), exhibit such sharply local features in time as visible in Fig. 3b for Mardi Gras (day 347) or at the end of the summer holidays in some larger provinces (day 185). A key here is the averaging over seven consecutive days, thereby eliminating all spurious contributions from intra-week variations in social activity. Although the impact of a temporally local event is thereby reduced by a factor of seven, it clearly stands out against data noise, because the latter is strongly reduced due to the averaging over a full week.

CONCLUSION

We have presented analysis tools for different aspects of epidemic mitigation and management interventions. First, we showed that from data of daily incidence (and prevalence derived therefrom) one can derive, *via* eq. (6), a quantity $R_I(t)$ which resembles the dynamic reproduction index and provides a very sensitive seismograph of the current state of the epidemic and its dynamics.

Second, we have shown that a similar quantity, $R_I^{calc}(t)$, can be derived *via* eq. (14) from the generation-based reproduction

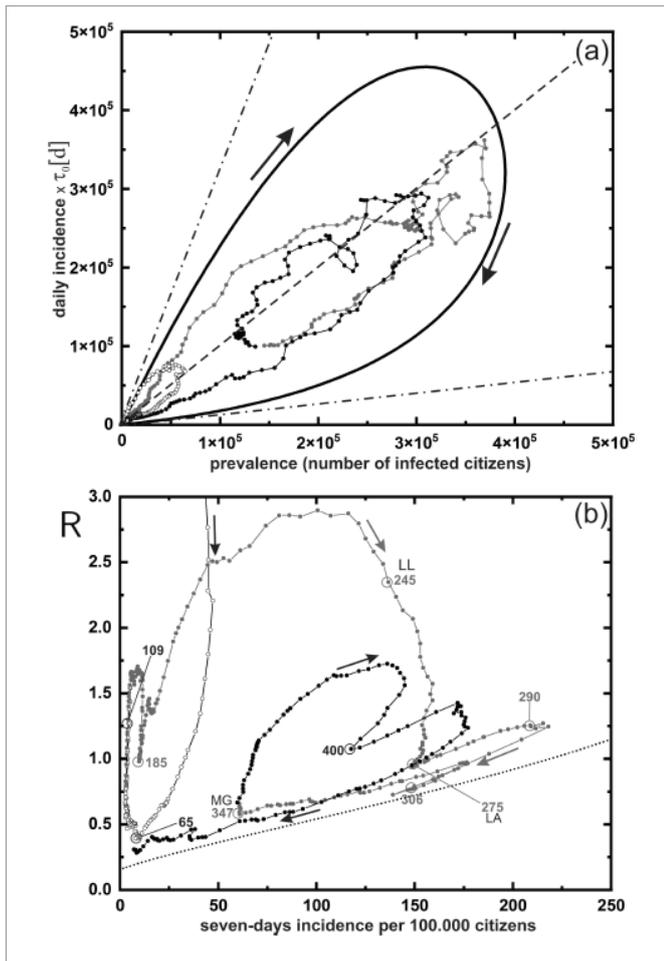


Figure 2: Day-by-day epidemic trajectories (circles with polygons). First wave (open), $t < 130$. Second wave (filled grey), $t \in [130, 351]$. Third wave (filled black), $t > 351$. **a)** Total number of new infections within a time interval $R_0(t)$ versus total number of acutely infected. Each wave appears as a clockwise orbit, with smaller sub-orbits. Data points tend to group along the first diagonal (dashed line). Solid curve: a sample simulation of an epidemic trajectory, showing the generic clockwise orbit structure, very similar to the data from the first wave. Both are asymptotic to the dash-dotted line with slope $R_0 = 3.3$. The second (lower) asymptote corresponds to the terminal value of $t \in [75, 225]$. The horizontal axis spans the maximum range of SARS-CoV2 infections acceptable to the German health system (about one half million). **b)** Reproduction index $R = R_i,corr(t)$ versus seven-days incidence. The orbit structure is more clearly revealed. This presentation allows to assess in which phase of an epidemic wave the system currently is. All data stay just above the asymptotic (dotted curve) solution of eq. (1), which enters the vertical axis at the terminal value of $R_1 = 0.14$. Various incidents mentioned in table 1 are marked along the trajectory. Note that R is generally a forerunner to incidence, and a new orbit is marked as a sharp increase of R . Hence one may miss important developments when monitoring incidence alone.

index, R_G . It has the same general behaviour as $R_I(t)$, but has the particular merit of very accurately representing the true reproduction index, $R(t)$, whenever it is close to unity. This is of great interest for forecasting the epidemic development, which change abruptly whenever R crosses unity. As it can be derived directly from available data on R_G , a renewed analysis of infection data is not needed for its calculation.

Third, we have shown how $R_I(t)$ and $R_G(t)$ can be combined to accurately determine the reproduction index, $R(t) = R_I^{corr}$, in a self-consistent manner. This method also yields the average infection time parameter, $\tau(t)$, which may slowly vary in time and is difficult to obtain reliably from clinical data [13].

Moreover, we have demonstrated that a presentation of standard epidemic data in the plane spanned by the reproduction index and the incidence displays the internal orbit structure of epidemic waves in a way beneficial for assessing the current epidemic state of affairs. Using accurate data for the reproduction index, which can be obtained with the methods outlined above, it proves to be very sensitive diagnostic tool to correlate mitigation measures with epidemic dynamics. Finally, we have shown that epidemic dynamics forecasting on the basis of R is potentially very powerful as it reacts immediately on events and interventions, despite the involved averaging procedure. Furthermore, it may be weeks ahead of forecasting on the basis of incidence, as revealed by the orbital character of epidemic waves. Hence we believe that the tools presented here should be very useful for policy makers during epidemics, such as the recent outbreak of COVID-19.

ACKNOWLEDGEMENT

The authors are grateful to Paul Leiderer, Christoph Klein, and Armin Lambacher for inspiring discussions and critical reading of the manuscript. This paper is dedicated to all those who died in the COVID-19 pandemic, especially to those who passed away in loneliness.

REFERENCES

- Cascio A, Bosilkovski M, Rodriguez-Morales AJ, Pappas G. The socio-ecology of zoonotic infections. *Clin Microbiol Infect.* 2011 Mar;17(3):336-42. doi: 10.1111/j.1469-0691.2010.03451.x. PMID: 21175957.
- Morse SS, Mazet JA, Woolhouse M, Parrish CR, Carroll D, Karesh WB, Zambrana-Torrel C, Lipkin WI, Daszak P. Prediction and prevention of the next pandemic zoonosis. *Lancet.* 2012 Dec 1;380(9857):1956-65. doi: 10.1016/S0140-6736(12)61684-5. PMID: 23200504; PMCID: PMC3712877.
- Wang LF, Cramer G. Emerging zoonotic viral diseases. *Rev Sci Tech.* 2014 Aug;33(2):569-81. doi: 10.20506/rst.33.2.2311. PMID: 25707184.
- World Health Organization (WHO). Tech Rep. 2020.
- Enserink M, Kupferschmidt K. With COVID-19, modeling takes on life and death importance. *Science.* 2020 Mar 27;367(6485):1414-1415. doi: 10.1126/science.367.6485.1414-b. PMID: 32217707.
- Baniamin HM, Mizanur R, Hasan MT. The COVID-19 pandemic: Why are some countries coping more successfully than others? The COVID-19 pandemic: why are some countries coping more successfully than others? *Asia Pacific Journal of Public Administration.* 2020;42:153169. doi: 10.1080/23276665.2020.1784769.
- Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, Pinior B, Thurner S, Klimek P. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav.* 2020 Dec;4(12):1303-1312. doi: 10.1038/s41562-020-01009-0. Epub 2020 Nov 16. PMID: 33199859.
- Ng TC, Cheng HY, Chang HH, Liu CC, Yang CC, Jian SW, Liu DP, Cohen T, Lin HH. Comparison of Estimated Effectiveness of Case-Based and Population-Based Interventions on COVID-19 Containment in Taiwan. *JAMA Intern Med.* 2021 Jul 1;181(7):913-921. doi: 10.1001/jamainternmed.2021.1644. PMID: 33821922; PMCID: PMC8025126.
- Lu G, Razum O, Jahn A, Zhang Y, Sutton B, Sridhar D, Ariyoshi K, von Seidlein L, Müller O. COVID-19 in Germany and China: mitigation versus elimination strategy. *Glob Health Action.* 2021 Jan 1;14(1):1875601. doi:



- 10.1080/16549716.2021.1875601. PMID: 33472568; PMCID: PMC7833051.
10. Dehning J, Spitzner FP, Linden MC, Mohr SB, Neto JP, Zierenberg J, Wibral M, Wilczek M, Priesemann V. Model-based and model-free characterization of epidemic outbreaks. medRxiv preprint. 2020. doi: 10.1101/2020.09.16.20187484.
11. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Review. 2000;42:599-653. <https://bit.ly/3yO2g5v>
12. Godara P, Herminghaus S, Heidemann KM. A control theory approach to optimal pandemic mitigation. PLoS One. 2021 Feb 19;16(2):e0247445. doi: 10.1371/journal.pone.0247445. PMID: 33606802; PMCID: PMC7894916.
13. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013 Nov 1;178(9):1505-12. doi: 10.1093/aje/kwt133. Epub 2013 Sep 15. PMID: 24043437; PMCID: PMC3816335.
14. Robert Koch Institut. Aktueller Lage-/Situationsbericht des RKI zu COVID-19. 2020.
15. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, Priesemann V. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020 Jul 10;369(6500):eabb9789. doi: 10.1126/science.abb9789. Epub 2020 May 15. PMID: 32414780; PMCID: PMC7239331.
16. RKI. Tech Rep. Robert Koch Institut, Berlin. 2020. <https://bit.ly/3PAo6iY>
17. An der Heiden M, Hamouda O. Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland-Nowcasting. Epidemiol Bull. 2020;10-16. <https://bit.ly/3sMuz0x>
18. Robert Koch Institut. Nowcasting und R-Schätzung: Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie. <https://bit.ly/3G13egq>
19. Stang A, Stang M, Jöckel KH. Estimated Use of Intensive Care Beds Due to COVID-19 in Germany Over Time. Dtsch Arztebl Int. 2020 May 8;117(19):329-335. doi: 10.3238/arztebl.2020.0329. PMID: 32527378; PMCID: PMC7207202.