The reproduction number and its measurement. A critique of the Robert Koch Institute

Quaas, Georg

30 May 2020

The reproduction number and its measurement. A critique of the Robert Koch Institute

Georg Quaas (May 2020)

Affiliated to the University of Leipzig

JEL-Classification:

C32, C61, I12, I18, J11

Keywords:

Classic epidemic model, reproduction number, contact rate, COVID-19, mathematics of highly infectious diseases, public health, economic cost of health policy

Abstract:

The Robert Koch Institute (RKI) has the goal of „protecting the population from disease and improving their state of health“ (RKI 2017). To this end, it develops research-based concrete recommendations for policy and makes data available to the expert public. Since the April 3, 2020, it has been publishing daily the numbers of corona infections reported by the health authorities, since the April 9, 2020, also the number of deaths from this infection and since April 25, 2020, the estimated number of convalescents. The so-called reproduction number reported since April 7, 2020, have largely superseded all other criteria by which the public health policy is guided. This article shows that the calculation of this figure by the RKI is neither theory-based nor particularly reliable. Nevertheless, there is a simple way to determine this number in the framework of the classic epidemic model (CEM). This study makes explicit important parts of the theoretical background of the CEM with the goal to underline that the method of determining the reproduction number empirically is a theoretically defined matter and cannot be replaced by a phenomenological method.

Acknowledgement:

I would like to thank Dr. Michael Wolff (Berlin) for critical discussions which led to improvements of formulas (18) and (19).

1. Exponential growth

Every process in which the increase or decrease dx/dt of a quantum $x \geq 0$ of objects over time depends on this number can be represented by an exponential function. Let $k \neq 0$ be a proportionality factor and

(1) $\frac{dx}{dt} = k \cdot x$,

it is easy to check that

(2) $x = x_0 e^{kt}$.

Supposedly, it is $x = x_0$ for two periods of time; in this case, the objects under consideration have simply reproduced themselves, their quantity is constant, and their growth rate is $k = 0$. It makes sense to assign the reproduction number 1 to this state. Accordingly, one could come up with the idea to define the reproduction number as follows:

(3) $R = x/x_0$

The definition (3) would be inappropriate. R would change with time according to e^{kt}, even if the rate of growth k is constant.

The model of exponential growth was applied to the growth of a population and of a capital stock (Solow 1957), a forest stock (Faustmann 1849), a nuclear chain reaction and many other problems, including the development of epidemics by the spread of virus infection. "For example, the transfer rate γI" – the transfer from the group of infectious people of a population with the amount I to the group of non-infectious people – "corresponds to $P(t) = e^{-\gamma t}$ as the fraction that is still in the infective class t units after entering this class and to $1/\gamma$ as the mean waiting time." (Hethcote 2000: 603)

2. Stock-flow model

A classic example of the relationship between flows and stocks "is a bath with water flowing in from a tap. The amount of water in the bath at any moment is a stock variable… The rate at which water enters the bath is a flow variable…" (Abel et al. 1998: 45-46) If there is an inflow (in) and an outflow (out) like a bathtub has, and the amount of water is x, the change of this amount is:

(4) $dx = x_t - x_{t-1} = in - out$

Of course, the change of x refers to the same time period in which both inflow and outflow happen.

The equation (4) applies to and solves many problems: The change of a capital stock equals investment minus depreciation (revaluation and losses ignored), the number of new officially registered unemployed people equals the number of new people out of work minus former unemployed people newly engaged, the money available for a household equals the income

minus expenditure plus former savings. Equation (4) is a basic element of stock-flow consistent models.

Let us suppose that $x_t \geq 0$ and x_t is constant over time, then $\Delta x = 0$. This implies $in = out$. The equilibrium of inflow and outflow is the characteristic of a *simple reproduction* of x.

It should no surprise that equations (1-4) can be specified in such a way that it is a reconstruction of the classic epidemic model (CEM).

3. The empirical domain of the CEM

The minimum of data required for an application of the classic epidemic model to an infectious disease like COVID-19 are:

(i) The number of new infected people on a daily basis ΔA, which can be summed up to the total number of infected people A according to the stock-flow equation (4) applied to discrete data.

(ii) The population size N_0; if a very deadly disease is expected the population size should be regarded as a variable dependent from the number of deaths $D(t)$:

(5) $N(t) = N_0 - D(t)$.

For accuracy, the number of births also should be added. Both birth and deaths are estimated in the framework of endemic models with the help of the corresponding average rates and N_0. (Hethcote 2000: 606-607).

(iii) Even though the number R from the disease–recovered people is often given as a statistical estimate made by health authorities, it is empirically exactly measurable, at least in principle.

(iv) The same can be said of the number of deaths D. To make the CEM as simple as possible, the number of deaths can be included in the number of recovered people R (an der Heiden & Buchholz 2020: 1)

In the framework of an exponential growth model the average time T of infected persons being infectious defines the transition probability

(6) $\gamma = 1/T$

of leaving the group of invectives.

The knowledge of T is not necessary for the *application* of the CEM, but would be useful (see below eq. 18). The same can be said of the proportion of people that are immune to an infection from the beginning. If the number of people with passive immunity is known, it will be treated as a given and constant part of the "recovered people." It turns out that the group of recovered people comprises convalescent, dead and a priori immune people. The specific feature of the group of recovered people is this: their members are not susceptible to the virus.

4. Theoretical variables: A classification scheme

(i) Let the variable I be the number of infectives, i.e., the number of infected people that are infectious and, therefore, a danger to people that are not immune.

(ii) The number of persons that are susceptible to a virus is named *S*.

There is a group in every population comprised of people that are infected but not infectious. For instance, a person just being infected becomes infectious after a few days only. Because we do not know the exact time span, this person is treated as being infectious from the beginning – belonging to the group I. Of course, a person who has recovered from the disease belongs both to the group R and to the group of infected people A, but not to the infectives I. In the case of COVID-19, it is supposed that these people are immune – at least for a while. For the difference between infected and infectious people, see Hethcote (2000: 601)

To summarize the classification scheme of people that makes up a population, we get the equation:

(7) $S(t)+I(t)+R(t)=N_0$

The CEM focuses on the stock I of infectives and its changes. In the case of COVID-19, any new infected very soon becomes a member of the pool of infectives. Oblivious to this small but exactly unknown time-lag, we can say that ΔA is not only the inflow to A, but also to I. The outflow comprises the recovered and deceased people. Put the last two groups together, the change of the unobservable group of infectives I is given by:

(8) $\Delta I=\Delta A-\Delta R$.

5. Exponential approach to the outflow

When the duration T of the average infectious period, which should be a constant for every disease, is over, infected persons leave the pool of infectives and become immune (disease-acquired immunity), they are the core of the recovered people *R*, or they belong to the group of dead people *D* which is also a part of R in the CEM-framework.

All crucial effects of a disease come from the pool of invectives. There is a probability γ to be recovered or dead after an uncertain number of days:

(9) $\Delta R=\gamma I$

6. Hamer's probability and frequency

The population shares of the infectives *I* and of the susceptibles *S* are

(10) $i=I/N$ und $s=S/N$.

These proportions can be interpreted as the probability of accidentally encountering one infectious or one susceptible person in the population. The probability that both persons meet each other has the value $i\cdot s$ (Hamer 1906) – treating the state of being infected and

susceptible as independent events (Larson 1974: 59). The frequency of infections is assumed to increase with the size of the population, resulting in a term we abbreviate to *H*:

$$(11) \qquad isN = \frac{ISN}{N^2} = \frac{IS}{N} = H$$

All variables in (11) are time dependent. From a practical point of view, *H* can be interpreted as the number of possibilities to become infected in a population of the changing extent $N(t)$ according to equation (5), so to speak the "abstract risk situation."

The degree with which this probability is effective is measured by another parameter, β, in the following way:

$$(12) \quad \Delta A = \beta IS/N = \beta H$$

The parameter β, which is called "contact rate," reflects the actual infection process, with regard to how many people are infected by an infectious agent per unit of time (here: per day) on average (Hethcote 2000: 602). This depends on a variety of factors that are not explicitly included in the model: population density, the number of daily interactions, common behaviors (hygiene, shaking hands, etc.). Some of these factors can be influenced pragmatically so that policymakers have access. However, the effect on the parameter β is recorded with a time delay. In the case of Corona, the media initially assumed that each infected person infects three more people during his or her infectious phase. The German Robert Koch Institute (RKI) reports a basic reproduction figure R_0 between 2 and 2.5 in its "Corona profile." Initial experience suggests that the infectious phase lasts for about ten days, so an average can be set to $\beta = 0,3$ – as long as no more accurate estimates are available. Equation (12) can be used to estimate the contact rate β empirically, even when ignoring the rest of the CEM.

The number of new infected persons reduces the number of susceptible people:

$$(13) \quad \Delta S = -\Delta A$$

To summarize and put together the causal hypotheses and the equation with the Hamer–probability, we get:

$$(14) \quad \Delta I = \beta IS/N - \gamma I$$

Equations (12), (13) and (14) form the core of the CEM (Hethcote 2000: 604) applied to discrete data. It can easily be seen that the last equation is an application of equation (1) to the discrete variable I.

A cause is often connected with a time lag before the corresponding effect takes place. Time lags have to be added to the second term on the right-hand side of equation (14) before the parameters β and γ are estimated empirically. In spite of the promising headline "Two SIS epidemiologic model with delays," Hethcote and van den Driessche (2000) are not very helpful to solve that econometric problem.

The main concern with this study is the correct estimation of the reproduction number.

7. The reproduction number

According to Hethcote (2000: 603-604), the reproduction number should be better called "replacement number" because of the danger of confusing it with the basic reproduction number. Conceptual clarity would afford the name "replacement number," as can be seen in the following paragraph. It is regrettable, but no serious harm for the attentive reader that the replacement number *R*, defined by the CEM, uses the same symbol as the number of recovered people. To be as unambiguous as possible, we name the reproduction number as R_{CEM}.

"The replacement number R is defined to be the average number of secondary infections produced by a typical infective during the entire period of infectiousness…" (Hethcote 2000: 603-604)

If the pool of infectives does not change, every person that leaves the pool has infected exactly one other person and is replaced (!) by this infected person. Of course, this is valid on average only. This special situation defines the replacement number 1 (one). There is a balance between inflow and outflow according to equation (4) that defines $R_{CEM} = 1$:

(15) $\beta IS/N = \gamma I$

Let us assume that one person has not exactly one, but $x \geq 0$ other people infected during his or her infectious period; the number of persons leaving the group is still γI. In this case the number of new infected people ΔA is x times higher than the number of persons leaving the group I to the group of recovered people *R*:

(16) $\beta IS/N = x\gamma I$

According to the conceptual definition, x is identical to the replacement number, if x is the number of persons infected by a typical infective during his or her infectious period *on average*. From this and (16), we derive the daily measurable reproduction or replacement number:

(17) $R_{CEM} = \frac{\beta IS/N}{\gamma I} = \frac{\beta s}{\gamma} = \frac{\Delta A}{\Delta R}$

As can be seen in the third term, R_{CEM} does not change with e^{kt} as was the case in the wrong equation (3).

8. The infectious time

If we remember the relationship (6) between γ and the average infectious period, T, we get the number of persons that are on average infected by one infectious person – a number that is identical with R_{CEM} per definition:

(18) $R_{CEM} = \beta Ts$

From (17), (18) and (12) follows:

(19) $T = \frac{I}{\Delta R}$

This equation can easily be used to determine the infectious time. With the data delivered by the RKI, I found $T \approx 11$.

9. The "reproduction number" made by RKI

The RKI started to report the value for the following variable in its daily situation reports on April 07, 2020:

> The reproduction number is the number of persons in average infected by a case. This number can only be estimated and not directly extracted from the notification system. The current estimation is R = 1.3 (1.0-1.6). This is based on the number of cases with disease onset between 31/03/2020-03/04/2020 and 27/03/2020-30/03/2020 and an average generation time of 4 days. Cases with more recent disease onset are not included because their low number would lead to an unstable estimation. (Daily Situation Report 2020, April 07)

The estimate of the RKI reflects a state of affairs that dates back at least 3 days. Due to late notifications by the health authorities, the actual new infections are only correct after at least three days. These figures are reported on the "dashboard" recommended by the RKI. After aggregation, they should correspond to the current figures in the daily situation reports, but this is not the case.

The methodological principles are explained in the management report dated April 13. The preparation of the data – the so-called nowcasting – is described in detail by an der Heiden and Hamouda (2020: 10 pp.). With the help of the average delay in the reporting system, a more exact date of the onset of the disease is determined (in most cases, the date was determined completely new). Based on this, the time-dependent reproduction rate of the RKI is determined by assuming that it takes an average of 4 days for one infected person to infect the next (= "generation time"). In another publication, the presumed duration of the infectious period is given in days $T = 10$ (an der Heiden M, Buchholz 2020). According to the RKI profile, the virus could still be detected in some infected persons 8 days after the outbreak of the disease. Together with the average of 2 days of an infectious prelude during the incubation period, one must therefore expect an average of ten days of infectivity at least.

The use of a generation time of 4 days coupled with the assumption of a much longer infectious period casts serious doubts about the theoretical rationale, because this would mean that in four days 1 + 1 = 2, in eight days 2 + 2 = 4, and in ten days about six people are infected. The reproduction number is six or more. When you already know the value of the reproduction number, you need not compute one.

For the adjusted time series of infections given, the authors an der Heiden and Hamouda (2020: 13) describe what they use as a method for computing the reproduction number based on the adjusted time series:

> With a constant generation time of 4 days, R is the quotient of the number of new cases in two consecutive periods of 4 days each. If the number of new cases has increased in the second time period, R is above 1. If the number of new cases is the same in both time periods, the re-production number is 1. This then corresponds to a linear increase in the

> number of cases. If, on the other hand, only every second case infects another person, i.e. R = 0.5, then the number of new infections is halved within the generational period. (an der Heiden und Hamouda 2020: 13)

The special "reproduction number" computed and published by the RKI is nothing else than the change of the number of new infected people averaged twice over 4 days. The underlying idea of this method is the inappropriate definition (3) above.

Meanwhile the RKI has changed its method two times. The first change happened on April 30 and included the elimination of the word "generation time" and a change of the averaging:

> The number of new cases estimated during the nowcasting process was previously presented as a moving 3-day average to compensate for random effects of individual days. Since April 29, 2020, the RKI has been using a 4-day average, which smooths the course of the bar chart to a certain extent. … The result of the R-estimate does not change thereby. Due to the smoothed course of the nowcasting, the calculation of the point estimator of R can be performed in fewer steps. For a given day, this value is now calculated as a simple quotient of the number of new cases for this day divided by the number of new cases 4 days before. (Daily Situation Report, 2020, April 30)

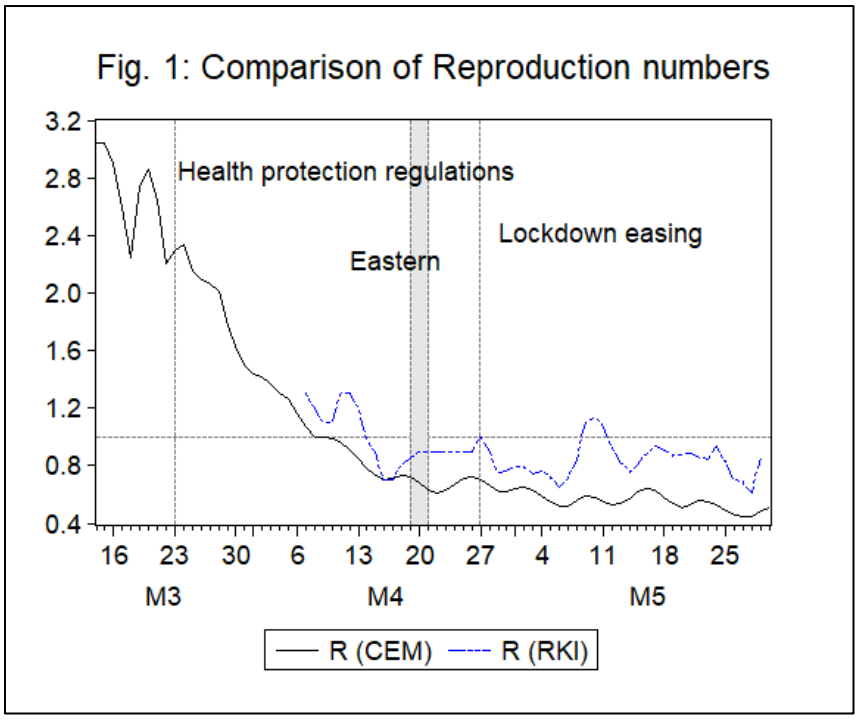
The next correction was achieved on May 14:

> The R-value reported to date reflects the trend in the number of new cases and can indicate possible changes in trend. However, this value is sensitive to short-term changes in the number of cases – such as those caused by individual outbreaks – which can lead to relatively large fluctuations, especially in the case of a small number of new cases. In addition to this sensitive R-value, the RKI therefore now provides a second more stable 7-day R-value, which refers to a longer period of time and is therefore subject to less short-term fluctuations. (Daily Situation Report, 2020, May 14)

The RKI changed the averaging of its nowcasting data, but never changed its method for computing the reproduction number, in spite of the fact that it was informed about its applying the wrong method several times since April 24 by the author.

10. The foregone chance for an alternative health policy

After some public discussions, the RKI accepted the easing of contact restriction under the condition of a reproduction number around one (but not more than 1.2 for a short time). If we take this benchmark for granted, the German economy could have had an opening on the base of the correct reproduction number ten days before Eastern (see Fig. 1).



If the damage of the lockdown can sometimes be estimated by real numbers, it will be possible to determine how many €-billions the wrong computation of the reproduction number has cost.

11. Bibliography

Abel, A., Bernanke, B., McNabb, R. (1998), Macroeconomics. European Edition. Munich.

an der Heiden, M., Buchholz, U. (2020), Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland, DOI 10.25646/6571.2

an der Heiden, M., Hamouda, O. (2020), Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting, Epid. Bull. 2020, Nr. 17, S. 10–15. DOI 10.25646/6692

Faustmann, M. (1849), Berechnung des Werthes, welchen Waldboden, sowie noch nicht haubare Holzbestände für die Waldwirthschaft besitzen. Allgemeine Forst- und Jagd-Zeitung, Dezember, 441–451.

Hamer, W. H. (1906), Epidemic Disease in England. Lancet, 1, 733–739.

Hethcote, H. W. (2000), The Mathematics of Infectious Diseases. SIAM Review, Vol. 42, No. 4. (Dec.), 599–653.

Hethcote, H. W., van den Driessche, P. (2000), Two SIS epidemiologic models with delays. Journal of Mathematical Biology, 2000 (40), 3–26.

Larson, H. J. (1974), Introduction to Probability Theory and Statistical Inference. Second Edition. New York.

Robert Koch Institute (2017), Leitbild, URL:
www.rki.de/DE/Content/Institut/Leitbild/Leitbild_node.html

RKI, SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19), URL:
www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html

RKI, Tägliche Situationsberichte, URL:
www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Archiv.html

Solow, R. M. (1956), A Contribution to the Theory of Economic Growth. The Quaterly Journal of Economics, Vol 70, No 1 (Febr.), 65–94.