Elemente der Mathematik (Online first)
DOI 10.4171/EM/514

Published by EMS Press

Relationship between the evolutionary development of a virus and logistic growth

Sebastian Bauer and Lukas Donner

Sebastian Bauer is professor of mathematics didactics at KIT, Karlsruhe. He received his PhD in mathematics from the University of Duisburg-Essen and taught mathematics and physics in higher secondary schools for several years. His interests lie in the development of innovative teaching content for mathematics education.

Lukas Donner finished his PhD in mathematics in February 2019 in Graz (Austria) and is now a postdoctoral researcher of mathematics didactics at the University of Duisburg-Essen (Germany). His research interests lie in various fields of mathematics education, especially in cognitive-psychological aspects of mathematical problem solving and of mathematics competitions, subject-matter didactics and interdisciplinary mathematics and ethics education.

1 Introduction

Starting from the original variant of the SARS-CoV-2 virus, the so-called wild type, several further virus variants (like alpha, delta or omicron, to name just a few) have developed in the past years.

The following phenomenon was observed in Germany (and at several further places worldwide) at the beginning of 2021 and in the middle of 2021: one variant of the SARS-CoV-2 virus predominant at that time was displaced by another variant of the virus, as depicted in Figure 1.

This context seems to us to be accessible and interesting as a cause for mathematical modeling in upper secondary education and for teacher training at universities. In particular, it offers the possibility to relate data and models. In this article, we want to prepare

Während der Covid-Epidemie 2019 bis 2022 haben mehrmals neuartige Virusvarianten die vormals vorherrschenden verdrängt. Dabei scheinen sich die Anteile der neuen Variante grob nach einer logistischen Gesetzmässigkeit verhalten zu haben. Die vorliegende Arbeit untersucht dieses Verhalten anhand von Daten, die von staatlichen Stellen veröffentlicht wurden, und mit Hilfe theoretischer Überlegungen vor dem Hintergrund verschiedener Modelle. Dadurch kann der Beitrag auch als fachlicher Hintergrund für die schulische Behandlung logistischer Wachstumsphänomene in epidemiologischen Kontexten dienen.

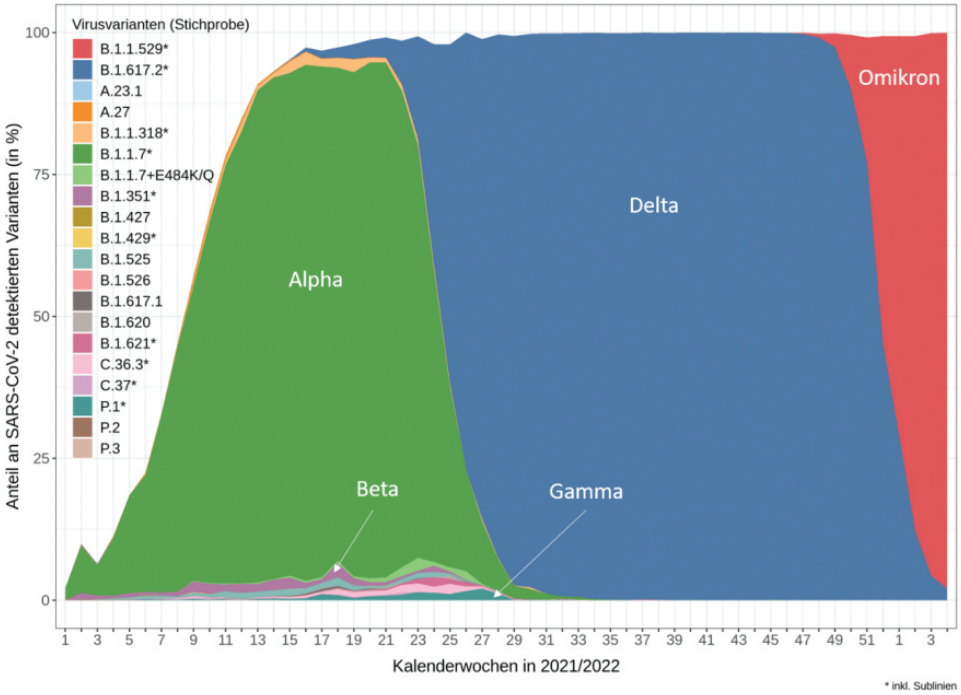

Figure 1. Ratio of virus variants in 2021 [8, p. 36]

the subject background for this. We approach this phenomenon from two perspectives. In the medical epidemiology literature, such transitions between different viral variants are often modeled with a logistic growth function, see e.g. [3, 4, 9]. We investigate how well the measured ratio of cases can be represented by a logistic growth model. Second, we deductively derive a logistic relationship from models of epidemiology, namely, an exponential and an SIR model. In addition, we discuss the question of whether it is theoretically possible that one of the repressed viral variants could re-emerge as a dominant variant.

2 Analysis of surveillance data

The phenomenon of changing dominant viral variants described in the previous section is addressed using real infection data from Germany as an example. In Germany, the Robert Koch Institute (RKI) is responsible for detecting new infections with the SARS-CoV-2 virus and differentiating variants by genome sequencing. This section refers to data published in two of their surveys of the so-called *Variants of Concern*, i.e., those variants that are under special scrutiny because of their potentially significant impact on the future course of the SARS-CoV-2 epidemic.

Exemplarily, based on data published by the RKI, we take a look at the development of the ratios of the respective Variants of Concern at the beginning of the year 2021 and

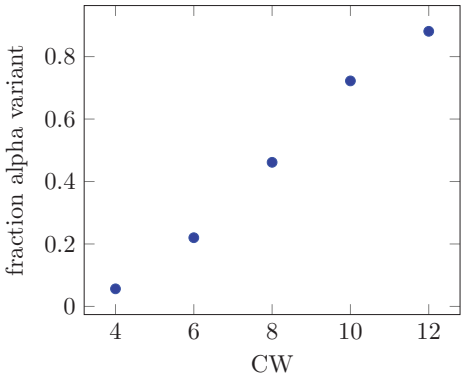

Figure 2. Fraction of alpha variant among all infections, CWs 4–12 in Germany in 2021. Data retrieved from RKI [7, Table 1 (p. 6)]

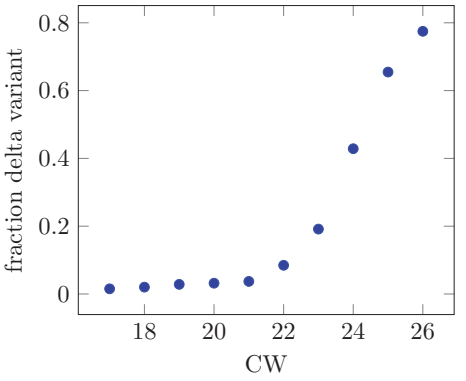

Figure 3. Fraction of delta variant among all infections, CWs 18–26 in Germany in 2021. Data retrieved from RKI [6, Table 3 (p. 9)]

CW	04	06	08	10	12
Fraction α	0.056	0.220	0.461	0.722	0.881
CW	17	18	19	20	21
Fraction δ	0.014	0.018	0.026	0.030	0.035
CW	22	23	24	25	26
Fraction δ	0.078	0.175	0.393	0.601	0.740

Table 1. Fractions of new virus variants, data retrieved from RKI [7, Table 1 (p. 6)] and [6, Table 3 (p. 9)]

in the middle of the year 2021 in Germany. In calendar weeks (CW) 4 to 12, the ratio of the so-called alpha variant among all new infections increased, while in turn the ratio of the rest (e.g. the wild type as well as several other variants that only made up a very small part of the total at that time) steadily decreased (see Figure 2). Between CW 18 and 26, the fraction of the delta variant increased rapidly and displaced the rest and in particular the alpha variant (see Figure 3). The data is also given in Table 1.

These specific trajectories of the data points suggest that a logistic growth function might be appropriate to represent the respective ratios. This is quite remarkable. Usually, logistic differential equations arise from structural modeling considerations in the context of an infection model in the following way: of a population of size N, a part of size X is infected and the complementary part of size $Y = N - X$ is susceptible (i.e., uninfected and infectible). Assuming that infection is permanent and comes about through contacts between susceptibles and infected and that the number of susceptibles and infected cannot change in any other way, the rate of change of X is proportional to the product of X and Y, i.e., $X' = rX(N - X) = rNX(1 - \frac{X}{N})$. The situation at hand differs from this: the quantity X measured here is the proportion of new infections of a given virus type. The complementary quantity $1 - X$ is the proportion of newly infected persons of the other virus type. In this case, it is not immediately obvious why the product XY determines

the rate of change of X: after all, the number of infections do not change due to contacts between carriers of the different virus variants, but due to contacts with uninfected persons.

For the case of a logistic growing fraction, the logistic differential equation is given by $X' = r \cdot X \cdot (1 - X)$ with logistic growth rate r and is solved by

$$X(t) = \frac{1}{1 + e^{-r \cdot (t - t_0)}}$$

for $0 < X < 1$, with

$$\lim_{t \to \infty} X(t) = 1, \quad \lim_{t \to -\infty} X(t) = 0, \quad \text{and} \quad X(t_0) = \frac{1}{2}.$$

Using the transformation to a linear model

$$Z := \ln\Big(\frac{X}{1 - X}\Big) = r(t - t_0),$$

we can quickly check for logistic growth behavior in a qualitative way. For the first transition from wild type to alpha variant, the transformed data points seem to lie well on a straight line. At the transition from alpha to delta, there is a significant change in slope from CW 20 to 21, so a logistic growth model should only be used for CW 21 to 26. (In the first attempt, however, we wanted to take all data published in [6] into account and are therefore only now limiting ourselves to the period mentioned. We have chosen weeks 21–26 rather than 17–21 because this is when the transition from the alpha to the delta variant took place.) In order to fit the data points with this type of function, we have to properly estimate the values of the parameters r and t_0 for each case. We estimate the parameters based on the given data points by using the least square method. For the second transition, we only use CWs 21–26. With the aid of the numerical method of the generalized reduced gradient, we get the optimized functions

$$X_{w,\alpha}(t) = \frac{1}{1 + e^{-0.5682 \cdot (t - 8.3123)}} \quad \text{and} \quad X_{\alpha,\delta}(t) = \frac{1}{1 + e^{-0.8740 \cdot (t - 24.62)}}$$

for the first and second case respectively (t in weeks, with $t = 0$ denoting the beginning of 2021). The optimal functions are depicted in Figures 6 and 7.

While the meaning of the parameter t_0 is obvious – the point in time when the fraction is just 0.5 – the interpretation of the meaning of the logistic growth rate r is not so clear. So, after having determined an apparently good fit, the question arises whether the logistic growth behavior can be explained in a theoretical way, eventually giving more interpretation to the logistic growth parameter r.

3 Theoretical justification of logistic growth from different epidemiological model assumptions

In this section, we derive the logistic growth behavior from two theoretical perspectives. In the first one, we start with an exponential model of infection, and in the second, we rely on the well established susceptible-infected-removed compartment model (SIR) first introduced by Kermack and McKendrick [5]. For the second case, see also [1, 10].

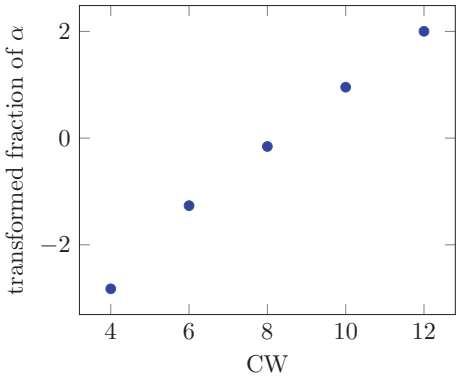

Figure 4. Transformed fraction of alpha variant in CWs 4–12

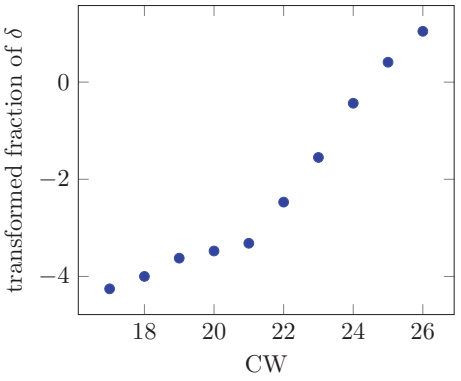

Figure 5. Transformed fraction of delta variant in CWs 17–26

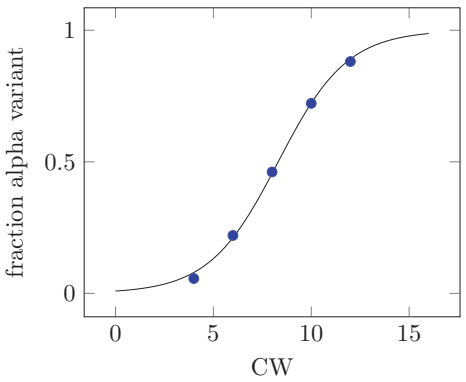

Figure 6. Optimal logistic growth function for the fraction of alpha variant in CWs 1–16

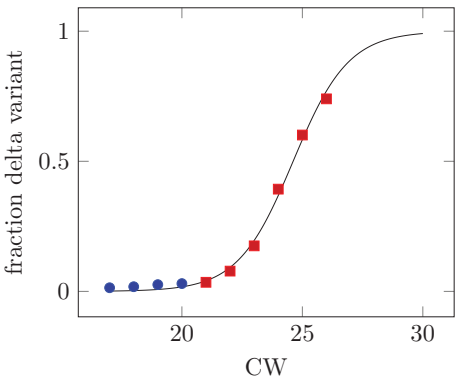

Figure 7. Optimal logistic growth function for the fraction of delta variant fitted to data from CWs 21–26 (red squares); the data of CWs 17–20 (blue circles) were not considered for the fit.

3.1 Exponential model

We call a contact between two people effective if, in the case where one is infectious and the other is susceptible, infection of the susceptible person would occur. We assume that the contact behavior of the population is homogeneous, i.e., all individuals have effective contacts at the same rate. We denote this rate by k. In the literature, this parameter is known as *effective* or *infectious contact rate* (in the sense that the contact is effective as it leads to an infection), or as *transmission rate* (see e.g. [10]). We further assume that every single person is either infected or susceptible. The number of infected individuals at time t is denoted by $I(t)$. The next crucial assumption is that the proportion of susceptibles in the total population does not change significantly over the period of time under consideration. If we denote the number of susceptibles by S and the size of the total population by N, it is assumed that $\frac{S}{N}$ remains constant. If we now consider the contacts of an infected person,

only the proportion $\frac{S}{N}$ leads to a new infection using a Laplace probability approach. Thus, infections change the number of infected individuals at a rate $k\frac{S}{N}I(t)$.

On the other hand, infected persons also recover after a certain time. By g, we denote the mean recovery rate and thus obtain the differential equation

$$I' = k\frac{S}{N}I - gI = \Big(k\frac{S}{N} - g\Big)I = aI \tag{1}$$

with $a = k\frac{S}{N} - g$. We call (1) the exponential model of epidemics. Usually, the exponential model is formulated only for the beginning of an epidemic with $\frac{S}{N} = 1$. In contrast, we already use the concept of probability of contact with a susceptible person, $\frac{S}{N}$, which is a well established ingredient of SIR models; see e.g. [10]. In contrast to the full SIR model, here, this fraction is assumed to be constant, thereby enhancing the applicability of the exponential model from the beginning of an epidemic to situation in which this ratio remains nearly constant.

In general, we consider situations with two different virus variants, say variant 1 and variant 2. The variants may differ in terms of infectivity and recovery time. Thus, we obtain specific contact rates k_1 and k_2 and specific recovery rates g_1 and g_2 for the variants. With $a_1 = k_1\frac{S}{N} - g_1$ and $a_2 = k_2\frac{S}{N} - g_2$, this leads to the differential equations

$$I_1' = a_1 I_1 \quad \text{and} \quad I_2' = a_2 I_2 \tag{2}$$

for the case numbers of the first and second variant respectively. We are now interested in how the ratio X of the case numbers of the second variant to the total case numbers evolves. In the statistics used in Figures 3 and 2, case numbers are related to a time interval of one or two weeks respectively. In a precise way, we should model the case numbers of week $n+1$ as integral over the infection rates

$$N_1 := k_1\frac{S}{N}I_1 \quad \text{and} \quad N_2 := k_2\frac{S}{N}I_2 \tag{3}$$

from $t = n$ to $t = n+1$. We replace this quantity by the infection rate times the time step. By building the fraction, the time step drops out. Therefore, we define

$$X = \frac{N_2}{N_2 + N_1} \quad \text{and} \quad Y = \frac{N_1}{N_2 + N_1}. \tag{4}$$

Thus, of course, $X + Y = 1$, and using (2), one obtains $N_2' = a_2 N_2$ as well as $N_1' = a_1 N_1$. We now calculate which differential equation X solves,

$$\begin{aligned} X' &= \frac{N_2'(N_2 + N_1) - N_2(N_2' + N_1')}{(N_2 + N_1)^2} = \frac{N_2' N_1 - N_2 N_1'}{(N_2 + N_1)^2} \\ &= (a_2 - a_1)\frac{N_2 N_1}{(N_2 + N_1)^2} = (a_2 - a_1)XY = (a_2 - a_1)X(1 - X). \end{aligned}$$

So X satisfies a logistic differential with logistic growth rate $r = a_2 - a_1$. If $r > 0$, i.e., $a_2 > a_1$, X converges to 1 for t at infinity and the first variant no longer plays a role after a certain time. If $r < 0$, the proportion of the second variant could not increase at all, but the variant would disappear.

With this theoretical background at hand, we now can interpret the logistic growth parameter r estimated in the last section. In the transition from the wild type to the alpha variant, we found

$$r = (k_\alpha - k_w)\frac{S}{N} - (g_\alpha - g_w) = 0.5682 \text{ (per week)}.$$

In the transition from alpha to delta, this parameter is found to be even larger, $r = 0.8740$, indicating that the fitness advantage of delta over alpha is even larger than the advantage of alpha over the wild type. The fitness advantage stems from two ingredients, higher infectivity and longer duration of infectivity. Since only the differences of $k\frac{S}{N}$ and g enter, but not the parameters themselves, it is not possible to disentangle both effects by the given data. It also could be the case that one variant has an advantage in higher infectivity and the other in longer duration.

3.2 SIR model

The exponential model can only describe an epidemic well for a certain period of time since, in general, the proportion of susceptibles in the total population $\frac{S}{N}$ will change over time. This effect is accounted for in the so-called SIR model, introduced for the first time by Kermack and McKendrick for one virus variant; see [5]. In order to consider two virus variants, the model is extended in an obvious way; see also e.g. [1, 10]. In the SIR model, it is assumed that the recovered receive everlasting immunity and no longer participate in the infection event. We denote the time-varying number of individuals in the new classes at time t by $S(t)$ for the susceptibles and $R(t)$ for the recovered, which are usually also referred to as the removed. The size of the total population is then given by $N = S + I_1 + I_2 + R$. According to these considerations, differential equations for the time-dependent variables S and R must be added,

$$\begin{aligned} I_1' &= k_1 I_1 \frac{S}{N} - g_1 I_1 = I_1 \cdot \left(k_1 \frac{S}{N} - g_1\right), \\ I_2' &= k_2 I_2 \frac{S}{N} - g_2 I = I_2 \cdot \left(k_2 \frac{S}{N} - g_2\right), \\ S' &= -(k_1 I_1 + k_2 I_2)\frac{S}{N}, \\ R' &= g_1 I_1 + g_2 I_2. \end{aligned} \tag{5}$$

As a result of these equations, the size of the total population N is conserved. It should be noted that the infection rates $k \cdot \frac{S}{N} \cdot I$ are often also given in the form pSI with $p = \frac{k}{N}$ in the literature. Since we are dealing with a closed system here, i.e., the size of the total population N is constant, both versions are equivalent. In our notation, interpretations seem to be simpler: $\frac{S}{N}$ is the probability that a given contact is a contact with a susceptible and k is the effective contact rate. The parameter p, on the other hand, has the meaning of the pairwise effective contact rate, which is the rate of infection per susceptible and per infective; see e.g. [2].

As in the exponential model, we now compute according to which differential equation the ratio of the infection rate of the second variant to the total infection rate evolves. The

infection rates are given again by formula (3); compare also with (5), but note that the quantity S is time dependent here. By the SIR equations (5) and the product rule, they solve the following differential equations:

$$N_1' = N_1\Big(k_1\frac{S}{N} - g_1 - \frac{k_1 I_1 + k_2 I_2}{N}\Big), \quad N_2' = N_2\Big(k_2\frac{S}{N} - g_2 - \frac{k_1 I_1 + k_2 I_2}{N}\Big). \tag{6}$$

We define the fraction X of the infection rate of the first variant as in (4). Using (6), we conclude again

$$X' = \Big((k_2 - k_1)\frac{S(t)}{N} - (g_2 - g_1)\Big)X(1 - X) \tag{7}$$

since the extra terms $\frac{k_1 I_1 + k_2 I_2}{N}$ in (6) are symmetric in 1 and 2 and thus drop out. The logistic growth rate $r(t)$ is now time dependent.

During the revision of this paper, we have learned that logistic growth behavior is also theoretically derived in [1,10] from slightly different variants of the SIR model. In contrast to this contribution, [1] consider the fraction $\tilde{X} = \frac{I_2}{I_1 + I_2}$. However, the data do not indicate the ratio of those currently infected, but the ratio of those newly infected. For this reason, the approach chosen here, (4), seems more appropriate to us. Due to the structure of the SIR equations, both approaches ultimately come to the same result. In [10], the authors partly work with explicit solutions instead of the differential equations and replace $S(t)$ by $S(\tau)$ with an arbitrarily chosen but fixed time τ. For this reason, our representation seems clearer and more general to us.

The fraction X is not governed by an exact logistic law, but the coefficient

$$r(t) = (k_2 - k_1)\frac{S(t)}{N} - (g_2 - g_1)$$

is time dependent and monotonously decreasing since, by (5), the term $\frac{S(t)}{N}$ is monotonously decreasing. With regard to the transition from the wild type to alpha and from alpha to delta, it could be that this effect is reflected in a slight decrease in the slope in the plot of the transformed data points in Figures 4 and 5. The question arises whether a better model for the transitions results if the time dependence of the susceptibles is taken into account. In order to do this, the parameters in the SIR model cannot be estimated from the given data: there is too little data and only the differences, but not the parameters themselves, are included in the model equation (7). Because we cannot use a structure-theoretical model for S for the reason just mentioned, we choose the mathematically simplest non-constant model that depends on two parameters, namely, a linear model, for the number of susceptibles: $S(t) = S(t_0) + m(t - t_0)$. With this approach, we get

$$r(t) = (k_2 - k_1)\frac{S(t_0)}{N} - (g_2 - g_1) + s(t - t_0) = \tilde{r} + s(t - t_0) \tag{8}$$

with $\tilde{r} := (k_2 - k_1)\frac{S(t_0)}{N} - (g_2 - g_1)$ and $s := (k_2 - k_1)m$. Using separation of variables, solutions of (8) are given by

$$X(t) = \Big(1 + \exp\Big(-\tilde{r}(t - t_0) - \frac{1}{2}s(t - t_0)^2\Big)\Big)^{-1}.$$

Transition	r	t_0	$\tilde{r}$	s	t_0
w to α	0.5684	8.3141	0.5683	−0.02348	8.241
α to δ	0.8740	24.6161	0.8365	−0.15233	24.5186

Table 2. Estimated parameters for models with two and three parameters

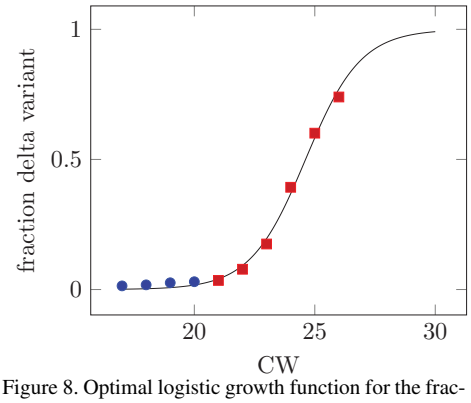

Figure 8. Optimal logistic growth function for the fraction of delta variant fitted to data from CWs 21–26 (red squares); the data of CWs 17–20 (blue circles) were not considered for the fit.

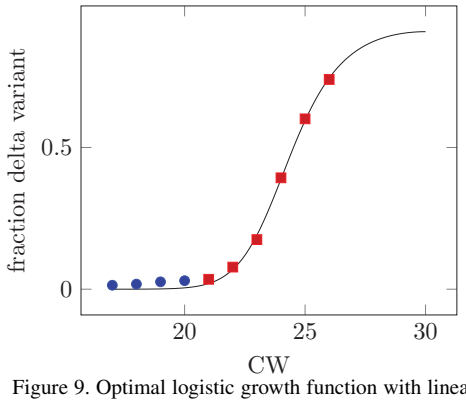

Figure 9. Optimal logistic growth function with linear $S(t)$ for the fraction of delta variant fitted to data from CWs 21–26 (red squares); the data of CWs 17–20 (blue circles) were not considered for the fit.

Analogous to Section 2, we estimate the three parameters $\tilde{r}, s, t_0$ using least squares. The resulting parameters are recorded in Table 2. There is hardly any visible difference in the plots of the transition from the wild type to alpha. The two plots for the transition from alpha to delta are compared in Figures 8 and 9.

Is the new model with three parameters now better than the old model with only two parameters? Does the decrease in susceptibles in the transitions play a role that should be taken into account? Visually, it seems that, at least in the transition from alpha to delta, the fit with three parameters provides a recognizably better fit than the fit with two parameters and that the decrease in susceptibles should be taken into account. The advantage of fitness $r = 0.8740$ decreases by approximately 4.3 % to $\tilde{r} = 0.8365$ due to the consideration of the decrease of $\frac{S}{N}$.

In order to clarify the question statistically whether the second model is better or not, we would need an error estimate for the data, which is not given by the RKI.

We end this article with one more theoretical consideration. Using the SIR equations to model the transition from one virus variant to the next, there could be situations in which a variant suppressed at the beginning of an epidemic could come back. Consider the following hypothetical scenario: the second variant is more infectious than the first, but persons infected by the first variant are infectious over a longer period of time than persons infected by the second variant, and the difference in infectiousness is greater than the difference in recovery rates, i.e., $k_2 > k_1$, $g_2 > g_1$ and $k_2 - k_1 - (g_2 - g_1) > 0$. In this case, at the beginning of the epidemic, when $\frac{S}{N}$ is approximately 1, the fraction of the second variant increases with an approximate logistic growth behavior, in which,

however, the rate parameter $r(t) = (k_2 - k_1)\frac{S(t)}{N} - (g_2 - g_1)$ becomes smaller and smaller as the ratio $\frac{S(t)}{N}$ decreases. However, if at some point

$$\frac{S(t)}{N} < \frac{g_2 - g_1}{k_2 - k_1},$$

then r actually becomes negative and the proportion of the first variant begins to grow again and will dominate again by the end of the epidemic.

Whether this is the case for the concretely considered transitions cannot be answered with the help of the available data since the necessary parameter differences cannot be derived from the estimated parameters $\tilde{r}$ and s.

References

[1] L. Boyle, S. Hletko, J. Huang, J. Lee, G. Pallod, H. R. Tung, and R. Durrett, Selective sweeps in SARS-CoV-2 variant competition, *Proc. Natl. Acad. Sci. USA* **119** (2022), no. 47, article no. e2213879119.

[2] N. F. Britton, *Essential mathematical biology*, Springer, London, 2003.

[3] R. Earnest et al., Comparative transmissibility of SARS-CoV-2 variants delta and alpha in New England, USA, *Cell Rep.* **3** (2022), no. 4, article no. 100583.

[4] V. A. Gushchin et al., Dynamics of SARS-CoV-2 major genetic lineages in Moscow in the context of vaccine prophylaxis, *Int. J. Mol. Sci.* **23** (2022), no. 22, article no. 14670.

[5] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc.* **115** (1927), no. 772, 700–721.

[6] Robert-Koch-Institut, *Bericht zu Virusvarianten von SARS-CoV-2 in Deutschland. 14.07.2021*, www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/DESH/Bericht_VOC_2021-07-14.pdf, accessed 7 February 2023.

[7] Robert-Koch-Institut, *Bericht zu Virusvarianten von SARS-CoV-2 in Deutschland, insbesondere zur Variant of Concern (VOC) B.1.1.7. 31.03.2021*, www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/DESH/Bericht_VOC_2021-03-31.pdf, accessed 7 February 2023.

[8] Robert-Koch-Institut, *Wöchentlicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19). 10.02.2022 – Aktualisierter Stand für Deutschland*, www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Wochenbericht/Wochenbericht_2022-02-10.pdf, accessed 7 February 2023.

[9] R. S. Paton, C. E. Overton, and T. Ward, The rapid replacement of the SARS-CoV-2 delta variant by Omicron (B. 1.1. 529) in England, *Sci. Transl. Med.* **14** (2022), no. 652, article no. eabo5395.

[10] S. Zhao, I. Hu, J. Lou, M. K. C. Chong, L. Cao, D. He, B. C. Y. Zee, and M. H. Wang, The mechanism shaping the logistic growth of mutation proportion in epidemics at population scale, *Infectious Disease Modelling* **8** (2023), no. 3, 107–121.

Sebastian Bauer
Institut für Algebra und Geometrie
Karlsruher Institut für Technik (KIT)
Englerstr. 2, 76131 Karlsruhe, Germany
sebastian.bauer2@kit.edu

Lukas Donner
Fakultät für Mathematik
Universität Duisburg-Essen
Thea-Leymann-Straße 9, 45127 Essen, Germany
lukas.donner@uni-due.de