Monitoring the spread of COVID-19 by estimating reproduction numbers over time

Thomas Hotz[1], Matthias Glock[1], Stefan Heyder[1], Sebastian Semper[1],
Anne Böhle[2], Alexander Krämer[2]

[1] Institut für Mathematik, Technische Universität Ilmenau
{thomas.hotz,matthias.glock,stefan.heyder,sebastian.semper}@tu-ilmenau.de

[2] School of Public Health, Bielefeld University
{anne.boehle,alexander.kraemer}@uni-bielefeld.de

17/04/20 – 15:04 GMT

Abstract

To control the current outbreak of the Coronavirus Disease 2019, constant monitoring of the epidemic is required since, as of today, no vaccines or antiviral drugs against it are known. We provide daily updated estimates of the reproduction number over time at https://stochastik-tu-ilmenau.github.io/COVID-19/. In this document, we describe the estimator, developed in (Fraser; 2007), we are using, and we give details on its implementation. Furthermore, we validate the estimator on simulated data, demonstrate that estimates on real data lead to plausible results, and perform a sensitivity analysis. Finally, we discuss why the estimates obtained need to be interpreted with care.

Contents

1 Introduction

As the Coronavirus Disease 2019 (COVID-19) threatens humanity, unprecedented measures to stop its spread have been adopted around the globe. In many countries, schools have closed and curfews have been imposed. Given the enormous burden these measures place on the economy, sooner or later they have to be relaxed. This raises important questions for policymakers and public health specialists. How large is the effect of these measures? Do they effectively stop the spread of COVID-19? What will happen if restrictions get relaxed? And in the future, how can we see whether the epidemic is getting out of hands again?

To answer these questions, one needs to know how fast the epidemic is growing. In infectious disease epidemiology, this is measured by the **reproduction number**, i.e. the mean number of people someone who got infected will infect in the course of time. Its **critical value** clearly is 1: for larger values the epidemic will grow, for smaller values it will diminish.

Since conditions may change in the future, e.g. when countermeasures are introduced or lifted, the reproduction number may also change. We therefore follow Fraser (2007) and consider what he calls the *instantaneous* reproduction number $R(t)$ at time t, and for which he suggests the estimator

$$\hat{R}(t) = \frac{I(t)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)} \tag{1}$$

where $I(t)$ is the number of incident cases at time t and w specifies the so-called **infectivity profile**, i.e. the distribution of the **generation time**, which is assumed to be known. To the best of our knowledge, this estimator has first been published by Fraser and others in (Grassly et al.; 2006). An overview of other estimators may be found in (Obadia et al.; 2012).

We explain the probabilistic model behind this estimator following (Cori et al.; 2013, Web Appendix 1) in Section 2. In addition, we analytically derive asymptotic confidence intervals (with details given in Appendix A) which are simple to compute. Here, we differ from Grassly et al. (2006) who use computationally more elaborate resampling techniques, namely the bootstrap, to obtain confidence intervals; Cori et al. (2013) on the other hand take a Bayesian approach, assuming a certain gamma prior distribution for $R(t)$.

In Section 3, some epidemiologically relevant properties of COVID-19 are discussed, and the infectivity profile is modelled. The estimator and corresponding confidence intervals are validated on simulated data in Section 4. Then, we apply this methodology to real data for Germany in Section 5, followed by a sensitivity analysis in Section 6. Finally, the results are summarised in Section 7, also discussing difficulties with this approach.

In order to continuously monitor the spread of COVID-19, a designated website has been created where the results of our analysis are shown and updated daily. It is available at `https://stochastik-tu-ilmenau.github.io/COVID-19/` in English for all affected countries based on the data from (Johns Hopkins University Center for Systems Science and Engineering; 2020) as well as in German for Germany and its federal states based on the data from (Robert Koch-Institut; 2020) at `https://stochastik-tu-ilmenau.github.io/COVID-19/germany`. The source code for that website as well as for this report may be found at `https://github.com/Stochastik-TU-Ilmenau/COVID-19/tree/gh-pages`, rendering this fully reproducible research. We note that a similar analysis using the Bayesian approach of (Cori et al.; 2013) was presented by Abbott et al. (2020) with updates at `https://epiforecasts.io/covid/posts/global/`.

2 Derivation of the estimator

The following is an adaptation of the modelling in (Fraser; 2007) and (Cori et al.; 2013, Web Appendix 1).

Time is taken to be discrete, i.e. we consider days $t \in \mathbb{Z}$, since the spread of the epidemic shows a strong intraday variability (e.g., there are fewer infections during the night when people are at sleep), and the time scales of incubation and infectious period are on the order of days. Also, cases are reported on a daily basis.

The number of **incidences**, i.e. newly infected cases, at day t will be given as $I(t)$. The **infection**

age of an infected person in days, i.e. the number of day elapsed since the infection, is denoted by $\tau \in \mathbb{N}_0$.

The spread of the epidemic depends strongly on the time-dependent **transmissibility** $\beta(t,\tau) \geq 0$ specifying the expected number of susceptible individuals an infectious person at infection age τ, a so-called **primary case**, will infect at time t. The transmissibility is in particular affected by the **contact rate**, i.e. the mean number of people an infected person meets per day, and the **infectiousness** of the primary case. The former is addressed by **non-pharmaceutical interventions** such as school closures and curfews, the latter is a virological feature of the disease. Therefore we make a crucial **structural assumption**, namely that they separate:

$$\beta(t,\tau) = R(t)\,w(\tau), \tag{2}$$

where $R(t) \geq 0$ denotes the (instantaneous) **reproduction number** at time t of transmission, i.e. when the **secondary case** gets infected by the primary case, and $w(\tau) \in [0,1]$ specifies the **infectivity profile** at infection age τ. This models the belief that contact rates change over time but the infectiousness of the primary case depends only on τ which is debatable, however: when rules for isolation or quarantine change are loosened, e.g. because hospital capacities are exhausted, β will change differently for different values of τ; we will reiterate this point in Section 7. It is also reflected in the fact that any constant factor may be alternatively incorporated into R or w. The latter is therefore standardised such that

$$\sum_{\tau=0}^{\infty} w(\tau) = 1, \tag{3}$$

i.e. w is a **probability distribution** which can be interpreted as follows: for a fixed time t randomly pick a pair of individuals where the first one is a primary case that got infected at time t, in turn infecting the second one later; $w(\tau)$ is the probability that the second case got infected at time $t+\tau$, i.e. at infection age τ of the primary case. w thus specifies the distribution of the **generation time**. It is assumed to be **known**; see Section 3 on how we model it for COVID-19.

In a **stochastic model** for the dynamics of the epidemic, $I(t)$ is given as the number of successful transmissions from an infectious person to someone who is susceptible to the disease. Assuming that each possible transmission succeeds independently (thus ignoring the possibility of multiple infections) with a probability corresponding to β, and if there are many possible transmissions, $I(t)$ is – by the law of small numbers – approximately **Poisson distributed** conditional on the past. The intensity of this Poisson distribution is equal to

$$\mathbf{E}(I(t)\,|\,I(t-1),\dots) = \sum_{\tau=1}^{\infty} \beta(t,\tau)I(t-\tau) = R(t)\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)\,. \tag{4}$$

Here, transmissions **on the same day** are ruled out, i.e. $w(0) = 0$, which is a realistic assumption since the incubation period will be at least one day.

The last equation suggests the **estimator** $\hat{R}(t)$ for $R(t)$ given in (Fraser; 2007, Equation (9)),

$$\hat{R}(t) = \frac{I(t)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)}\,. \tag{5}$$

Note that the **case reproduction number** $R_c(t)$, i.e. the expected number of people a primary case infected at time t will infect, is given by, cf. (Fraser; 2007, Equations (2) and (8)),

$$R_c(t) = \sum_{\tau=1}^{\infty} \beta(t+\tau,\tau) = \sum_{\tau=1}^{\infty} R(t+\tau)\,w(\tau)\,. \tag{6}$$

This is of course difficult or even impossible to estimate as it depends on future contact rates, i.e. on countermeasures that will be imposed. However, assuming that conditions remain the same in the future, i.e. $R(s) = R(t)$ for $s > t$, we obtain $R(t)$ again, cf. (Fraser; 2007, Equation (3)),

$$\sum_{\tau=1}^{\infty} R(t+\tau)w(\tau) = R(t)\sum_{\tau=1}^{\infty} w(\tau) = R(t)\,. \tag{7}$$

This explains why $R(t)$ is called *reproduction number*.

3

For large intensities, i.e. if the conditional expectation in Equation (4) is large, the distribution of $\hat{R}(t)$ can be well approximated by a Gaussian distribution, with small standard errors. From this, **asymptotic confidence intervals** can be derived, see Section A. If q denotes the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution then

$$\left[\hat{R}(t) - q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)}}, \hat{R}(t) + q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)}} \right] \tag{8}$$

is an (asymptotic) $(1-\alpha)$-confidence interval for $R(t)$. Note that in practice ten or more incident cases should suffice for the asymptotics to be reliable.

3 Specifics of COVID-19

As COVID-19 is a new disease, first being described at the end of 2019, its virological features have not yet been conclusively determined. Nonetheless, we tried to choose parameters in agreement with the current state of research.

For comparisons, we note that in a population without any countermeasures, the **basic reproduction number** R_0 is believed to be given by some value between 2.4 and 4.1 (Read et al.; 2020).

The **incubation time**, i.e. the time from infection until symptom onset, ranges from 1 to 14 days with a mean of 5 to 6 days; the virus can be detected from 1 to 2 days before symptom onset for up to 7 to 12 days in moderate cases, and even up to two weeks in severe cases (World Health Organisation; 2020). We therefore may indeed assume that $w(0) = 0$.

For modelling the infectivity profile w, it is important to realise that it is not proportional to the amount of viral specimens that can be detected in an infected person's sputum, say. Indeed, since severe cases are very likely to be hospitalised and thus strictly isolated, the probability of infecting someone more than 12 days after infection is very low. Similarly, before symptom onset the probability for transmission might be very low since no sputum is spread.

The infectivity profile is therefore set to start with 0 on the first day after infection with a linear increase up to day 4, remaining constant up to day 6 and decaying linearly again until being 0 at day 11; see Figure 1. In Section 6 we discuss the effect this choice has on the analysis.

4 Validation on simulated data

To validate the estimator, we simulate a **stochastic SEIR** (a.k.a. Kermack-McKendrick) **model**. To be more precise, we consider a discrete-time Markov chain describing a population of $n = 1$ million people with each individual being in one of four states: **susceptible**, i.e. not yet infected; **exposed**, more aptly called **latent**, i.e. infected but not yet infectious; **infectious**; or **recovered** and thus immune. We start at time 0 with 100 latent individuals, all others initially being susceptible. At each time step, a susceptible person becomes infected if the virus is transmitted through contact with an infectious person; such contacts happen independently with probability p_E. A latent person becomes infectious with probability p_I, and an infectious person recovers with probability p_R; otherwise an individual remains in its state.

This results in incubation times, i.e. times spent in the latent state, which are geometrically distributed with mean $1/p_I$; for this to be 3, we set $p_I = 1/3$. Similarly, the infectious period is geometrically distributed with mean $1/p_R$ which we would like to be 4, so we set $p_R = 1/4$. The corresponding infectivity profile w is then given by the convolution of these two geometric convolutions. It can be calculated analytically, see Appendix B for details; the result is shown in Figure 2. Note that $w(1) = w(0) = 0$ since it takes at least one day to become latent and another one to become infectious in this model.

The basic reproduction rate is then given by $R_0 = np_E/p_R$ since an infected person on average infects np_E individuals per day (if all were susceptible) for $1/p_R$ days on average. In order to simulate an epidemic with $R_0 = 2.5$, we set $p_E = R_0 p_R/n$ accordingly.

Over time, the reproduction number changes naturally because more people recover and become immune: $R(t)$ is R_0 times the proportion of susceptible individuals at that time. In addition, we
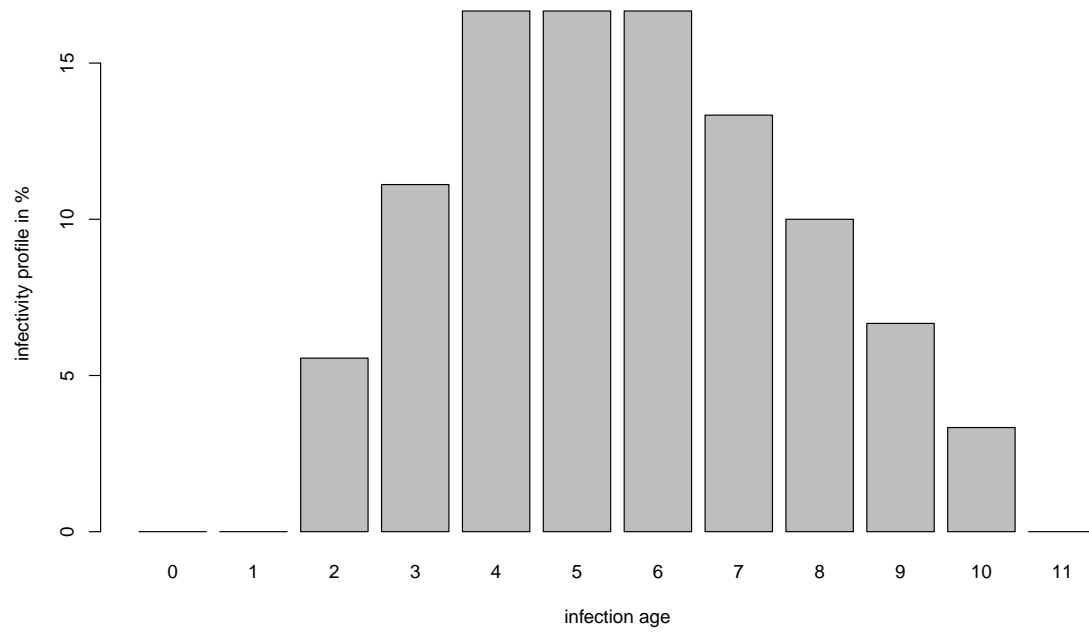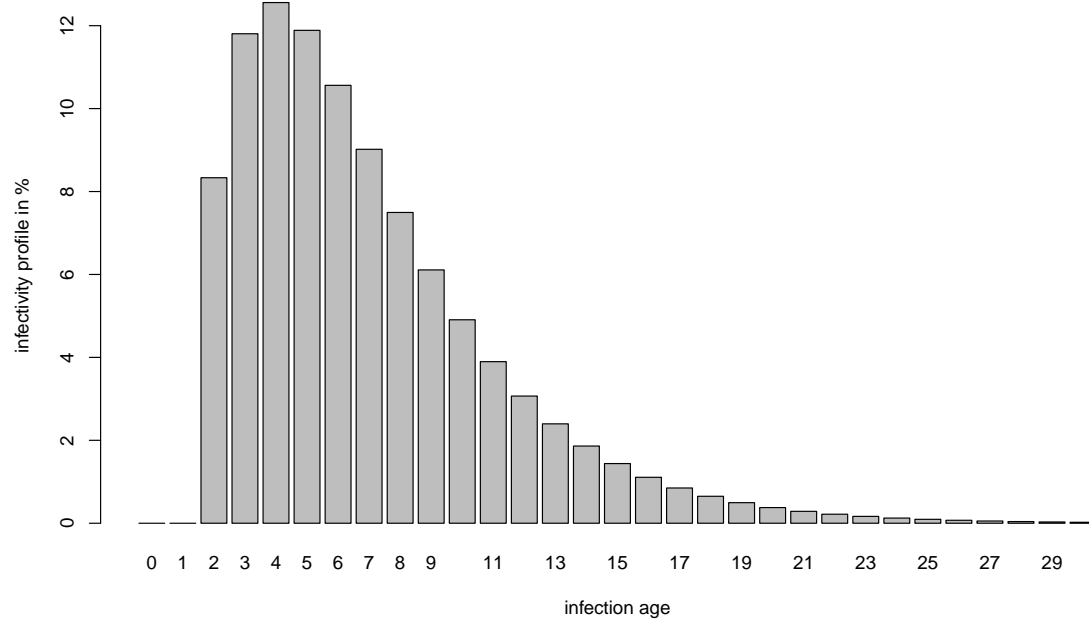
Figure 1: Modelled infectivity profile w.

Figure 2: Computed infectivity profile w corresponding to the simulation.

date of implementation	measure
13–18/03/2020 (mostly 16/03/2020)	school closures
14–22/03/2020 (mostly 16–22/03/2020)	closure of institutions, restaurants etc.
20–25/03/2020 (mostly 22/03/2020)	contact restrictions

Table 1: Summary of starting dates for non-pharmaceutical interventions introduced by federal states in Germany.

assume that countermeasures have been imposed at time 30, resulting in $R(t)$ being 0.7 times the proportion of susceptibles afterwards, and that measures have been relaxed at time 50, resulting in $R(t)$ being 1.3 times the proportion of susceptibles thereafter.

Figure 3 shows one simulation run. The resulting estimates $\hat{R}(t)$ and pointwise 95%-confidence intervals ($\alpha = 5\%$ as usual) can be compared with the true reproduction rate $R(t)$ in Figure 4.

The simulation has been repeated 100 times, and for each time point the proportion of confidence intervals containing the true reproduction number has been determined, see Figure 5. They appear not quite to have the desired nominal coverage but given that they are only asymptotic confidence intervals, and modelling errors are typically much larger, we consider them acceptable in practice.

These simulations demonstrate how well the estimator is able to detect changes in the reproduction number. From a practical viewpoint, this is an overly optimistic result. In fact, Equation (4) and consequently the estimator $\hat{R}(t)$ in Equation (5) are based on the number of newly infected cases. But **infection dates** are rarely known. Instead, cases are reported when they are tested with a positive test result. In our simple simulation, one should therefore consider the **newly infectious cases** at day t as input data $I(t)$ for the estimator. Note that their increase lags behind the one of the **newly infected cases**, i.e. the **newly latent cases**, by the incubation time, see Figure 3 where they lag behind by about 1 day, the mode of the incubation time distribution.

We use a naïve approach to deal with this which we call **infection-to-observation period**: we shift the estimator back by the observed lag, i.e. by 1 day. The result is shown in Figure 6 where the jump in $R(t)$ leads only to a rapid change of $\hat{R}(t)$, approaching the true value $R(t)$ exponentially fast, though. For real data, the infection-to-observation period is even larger, since symptomatic cases are usually not reported immediately. This will be taken into account in the following section.

5 Application to real data

As an example, we consider data for Germany and its federal states (Bundesländer) provided by the Robert Koch-Institut (2020), see Figure 7 for the total daily reported cases. Each case in this dataset is labelled with a **reporting date**, i.e. the day when the local health authority (Gesundheitsamt) has been notified about the case. Of course, this is not the day of symptom onset, let alone the day of infection which is needed for the estimator in Equation (5). We therefore set an **infection-to-observation period** by which we backdate the cases. It is pragmatically chosen as 5 days of incubation time (cf. Section 3) plus 2 more days **reporting delay** for testing etc., i.e. the infection-to-observation period is set to 7 days.

Since cases are reported to local health authorities, then collected at the level of states who in turn report them to the federal Robert Koch-Institut, they appear in the dataset a few days later, although with the date of reporting to the local health authority. Therefore, we exclude data from yesterday and the two days before.

Based on the backdated data and the infectivity profile from Section 3 (see Figure 1), we estimated the reproduction numbers for Germany over time, see Figure 8. Note that there are no estimates for the last 7 days for which new cases are shown in Figure 7 due to the infection-to-observation period.

Starting with Bremen on 13/03/2020, more and more restrictive non-pharmaceutical countermeasures have been adopted by the federal states; see Table 1 for a short overview. Their effect on the reproduction number is clearly visible in Figure 7, resulting in a reproduction number of less than 1 with all measures in place.
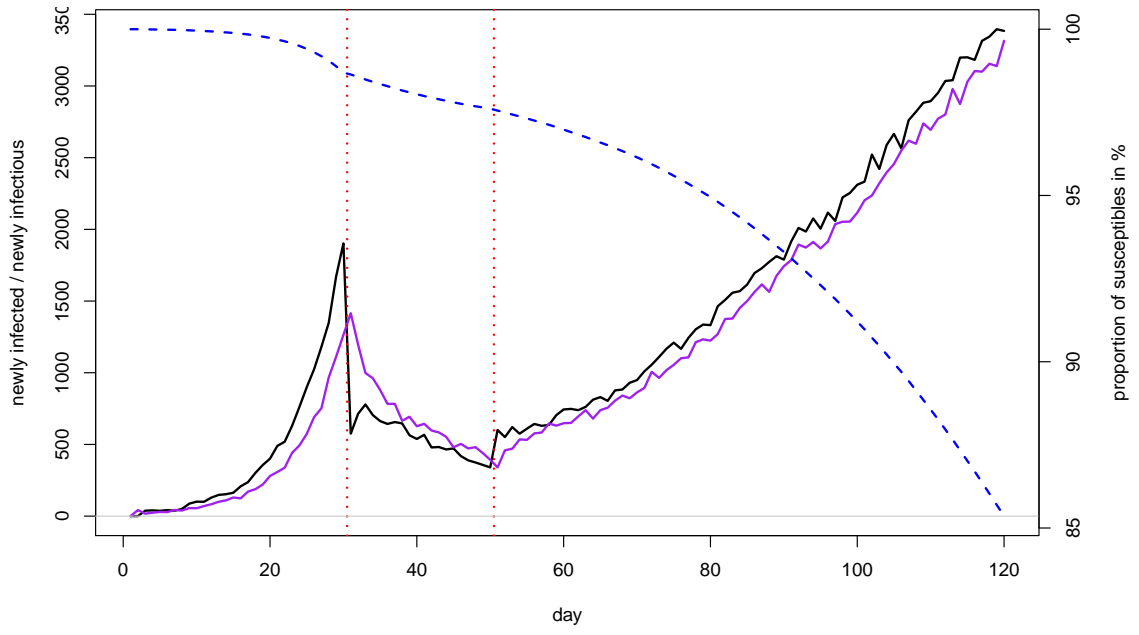
Figure 3: One simulation of the SEIR model; black solid line (left axis): newly infected; purple solid line (left axis): newly infectious; dashed blue line (right axis): proportion of susceptibles; vertical red dotted lines: intervention times.
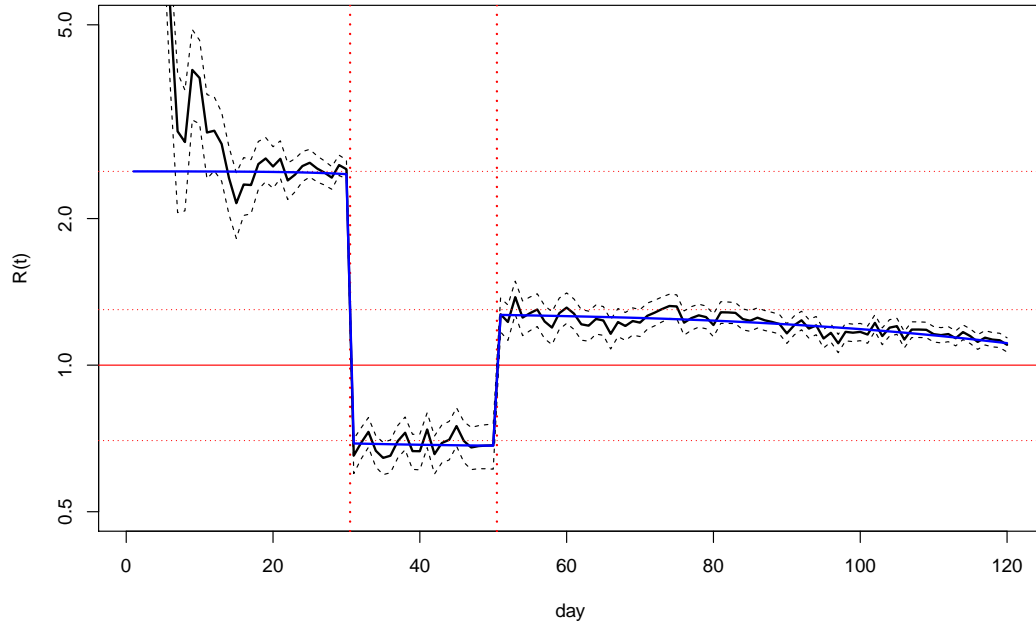

Figure 4: One simulation of the SEIR model; black solid line: $\hat{R}(t)$; black dashed lines: pointwise 95%-confidence intervals; blue solid line: $R(t)$; vertical red dotted lines: intervention times; horizontal red dotted lines: corresponding reproduction numbers without decrease in susceptibles taken into account.
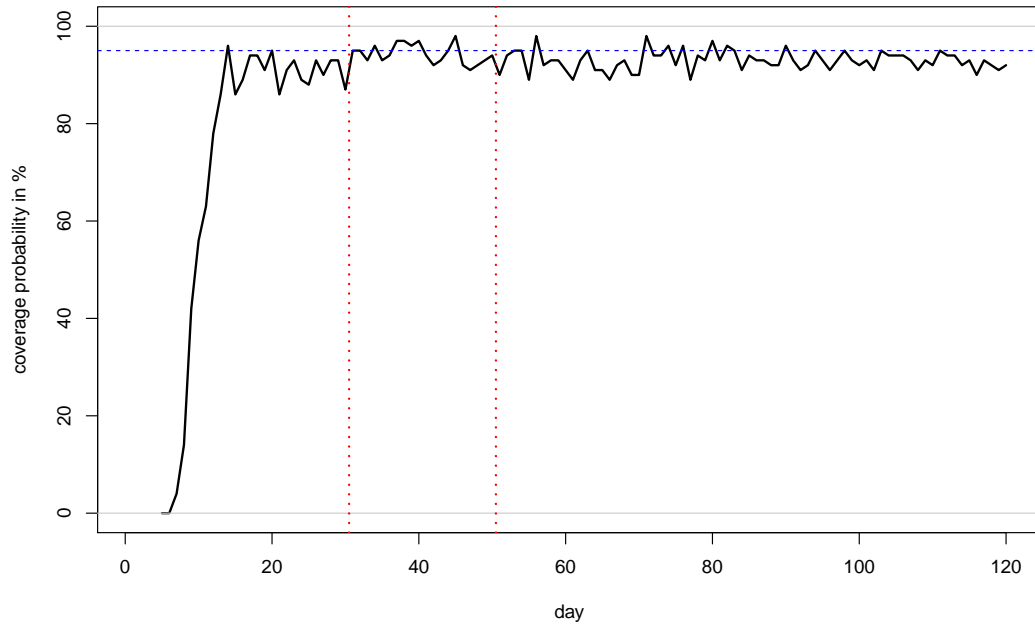
Figure 5: Estimated coverage probability based on 100 simulations (black solid line); horizontal blue dahed line: nominal coverage (95%); vertical red dotted lines: intervention times.
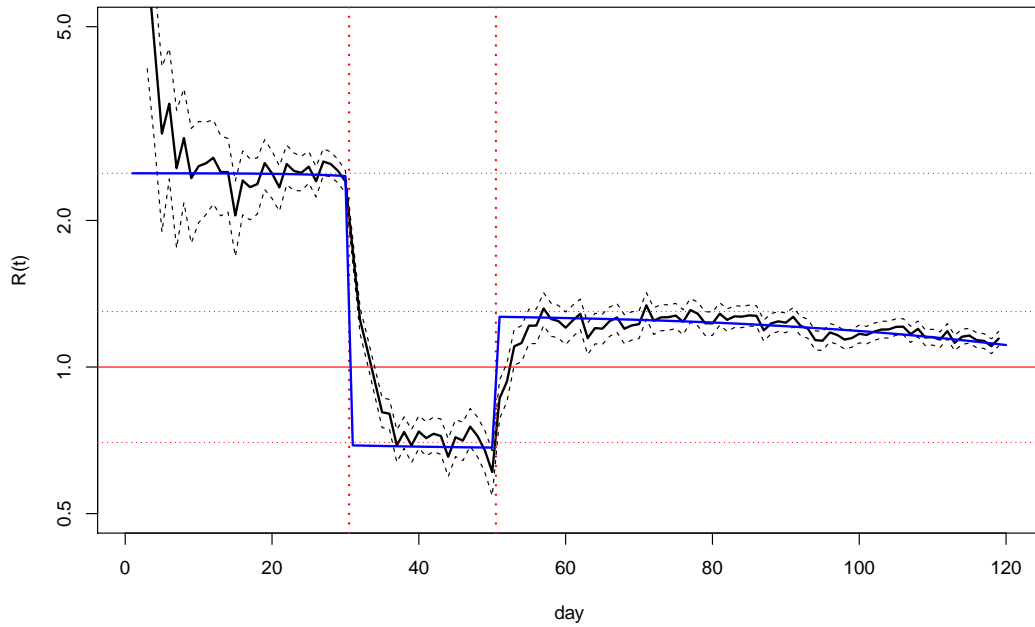

Figure 6: Estimator based on the newly infectious of one simulation of the SEIR model shifted by 1 day; black solid line: $\hat{R}(t)$; black dashed lines: pointwise 95%-confidence intervals; blue solid line: $R(t)$; vertical red dotted lines: intervention times; horizontal red lines: corresponding reproduction numbers without decrease in susceptibles taken into account. This is to be compared with Figure 4 where the estimator is based on the newly infected cases.
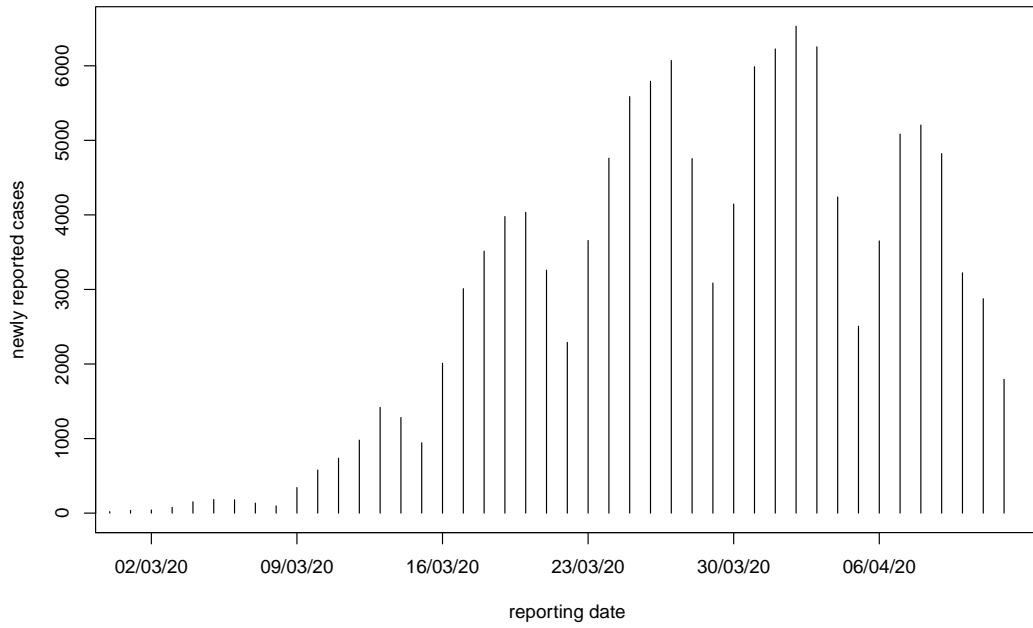
8

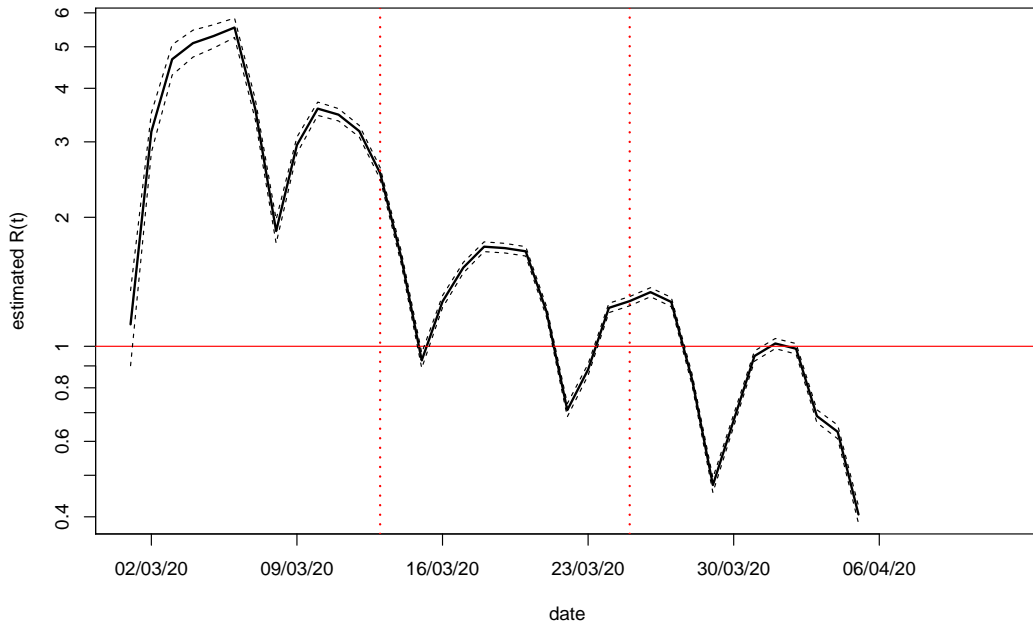Figure 7: Newly reported cases for Germany over time, based on data from the Robert Koch-Institut (2020).

Figure 8: Estimated reproduction numbers for Germany over time (solid line) with pointwise 95%-confidence intervals (dashed lines); vertical red dashed lines indicate the time period over which countermeasures have been implemented, cf. Table 1.

The strong weekly pattern in the estimates is due to the fact that less cases are reported around weekends, cf. Figure 7 where Mondays are marked on the horizontal axis. We do not compute an average over a sliding window of seven days so the viewer immediately recognizes the size of such artefacts, warning her to be overly confident in the results. In fact, these artefacts are much larger than the statistical uncertainty due to the stochastic nature of the epidemic which is reflected in the confidence intervals.

6 Sensitivity analysis

The estimator depends on two ingredients, the data, of course, and the infectivity profile. For Germany, there is a second dataset provided by Johns Hopkins University Center for Systems Science and Engineering (2020) whose source are mainly official data, too, but collected at the local level once they are available. In particular, as soon as data become available, they are marked as reported on that very day. They therefore show a far less pronounced weekday effect than the data from the Robert Koch-Institut (2020), see Figure 9 and compare with Figure 7. Moreover, data are not backedited, so even yesterday's data are final and can be used.

The estimates based on the data from the Johns Hopkins University Center for Systems Science and Engineering (2020) differ quantitatively but not qualitatively from the ones using the data of the (Robert Koch-Institut; 2020), see Figure 10.

To understand the effect the infectivity profile exerts on the estimates, we consider the infectivity profile we computed for the stochastic SEIR-model used for simulations in Section 4 (see Figure 2), employing it to estimate the reproduction numbers using the data from the Robert Koch-Institut (2020) again. When comparing the results with the ones obtained using the infectivity profile modelled in Section 3 (see Figure 1), one observes that the former profile has a longer tail, so the estimator takes values from further in the past into account, which for increasing case numbers reduces the denominator in Equation (5), and hence somewhat increases the estimates. Once case numbers stabilise, this effect obviously vanishes. As with the data source, the influence of the infectivity profile appears to be small enough not to matter qualitatively.

7 Discussion

The results for simulated data in Section 4 demonstrate the validity of the estimator, and of the asymptotic confidence intervals we derived. This is substantiated further by the fact that the estimated reproduction numbers' decrease for Germany correlate strongly with enforcement of non-pharmaceutical countermeasures there.

Let us stress the advantage of this estimator over approaches which determine growth rates or doubling times by fitting **exponential growth models** to numbers of either new cases or total cases in the initial phase of the epidemic where the proportion of susceptibles is close to 100%, cf. (Obadia et al.; 2012). In fact, the latter models are implicitly based on the assumption that conditions do not change such that the epidemic spreads with a constant growth rate, and thus with a constant reproduction number. But here we aim to determine a varying reproduction number. Fitting exponential growth models to total case numbers therefore is not conducive, even when localising the procedure by considering short time windows. Indeed, the case numbers from the past which occurred under different conditions will always affect the estimates. This problem is alleviated when exponential growth models are fitted locally to the numbers of new cases. Still, one needs to assume that conditions change slowly – which is debatable for the drastic measures which have been implemented quickly. In any case, even if one could observe new infections directly, the resulting estimates would be (additionally) smoothened, as opposed to the unbiased, sharp results obtained for the estimator we consider here (cf. Figure 4).

Nonetheless, the estimates have to be cautiously interpreted. For one, the calculated confidence intervals quantify only a rather small part of the uncertainty, namely the one which stems from the stochastic nature of the epidemic's evolution over time. Other uncertainties may affect the estimates much more, in particular when case numbers are large. In the following, we discuss those which we believe to be most important.

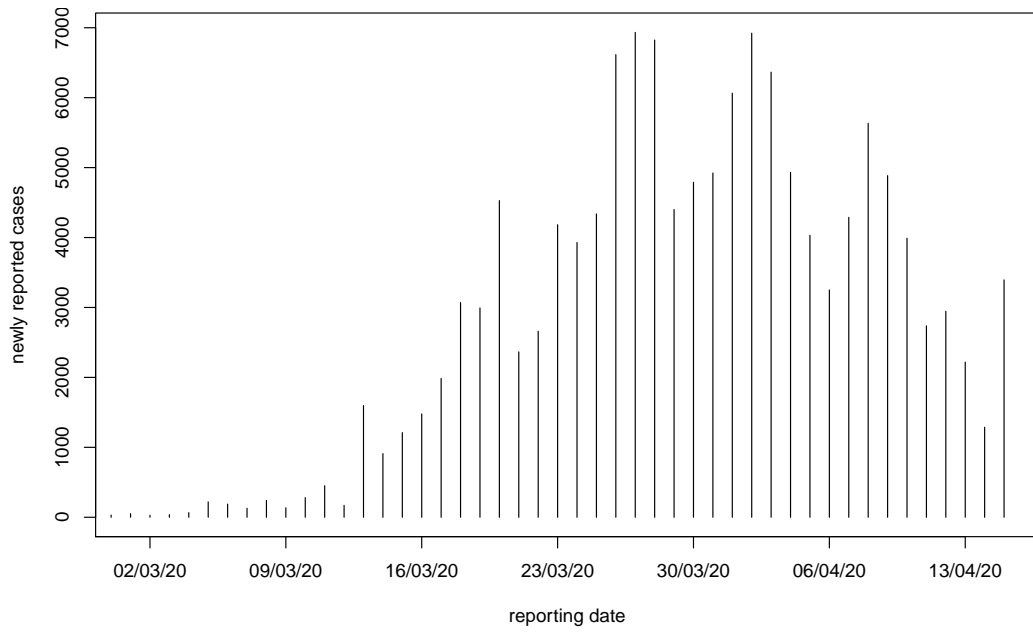The first set of difficulties concerns the **quality of the data**.

Figure 9: Newly reported cases for Germany over time, based on data from the Johns Hopkins University Center for Systems Science and Engineering (2020); compare with Figure 7.
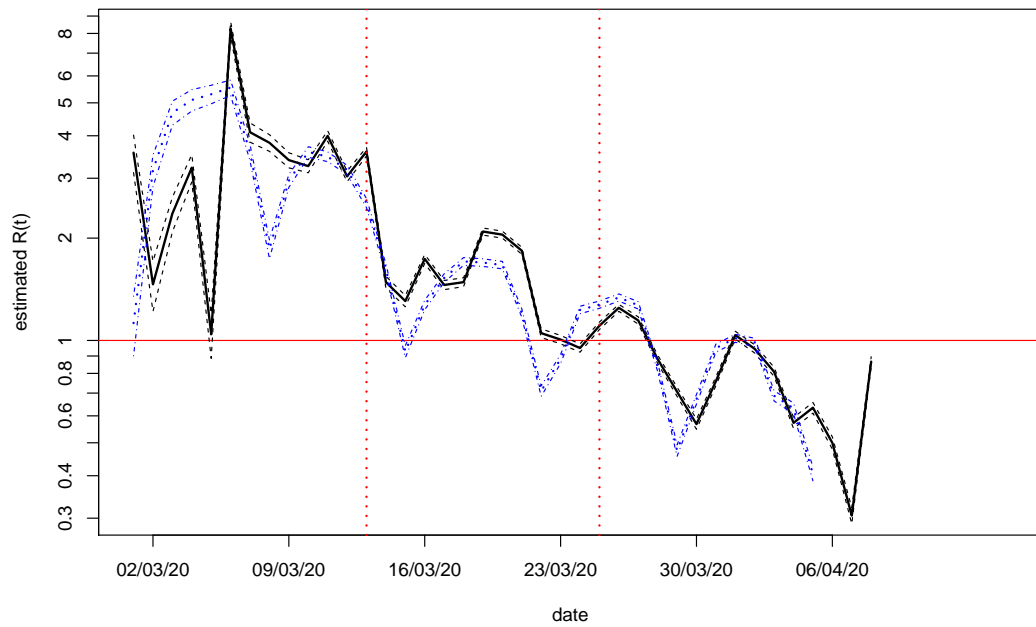

Figure 10: Estimated reproduction numbers for Germany over time (solid/dotted lines) with pointwise 95%-confidence intervals (dashed/dash-dotted lines) based on data from Johns Hopkins University Center for Systems Science and Engineering (2020), shown in black (solid), and Robert Koch-Institut (2020), shown in blue (dotted), respectively; vertical red dashed lines indicate the time period over which countermeasures have been implemented, cf. Table 1.

(a) Not all infections are reported, for example because cases remain asymptomatic, or because infected persons die without having been tested. If the proportion of infections which get reported stays constant over time (or at least varies slowly), both numerator and denominator in Equation (5) are multiplied by the same factor, so they cancel and the estimates are not affected. Changes in the reporting or testing methodology, however, will affect the estimates, as they will be indistinguishable from a true increase or decrease in the number of infections. This will for example happen if more people are tested due to higher capacities in testing facilities, or if lower case numbers allow more extensive tests of potential contacts, or if deaths are attributed to the disease without testing when such capacities are exhausted. Potential remedies include to not only consider reported infections but take fatalities, test rates etc. into account.

(b) The reporting date is not the date of infection: when the patient becomes symptomatic, he has to visit a physician, samples have to be tested, the test results need to be interpreted, and finally reported to the authorities. The strength of this effect is visible from the periodic pattern related to the days of the week in Figures 7 and 8. For some of the data provided by the Robert Koch-Institut (2020), both the reporting date, and the day of symptom onset are known, which in principle allows to infer dates of symptom onset for the entire dataset, thus getting rid of the weekday's influence. But the difficulty that the estimator is based on knowing the date of infection remains, cf. Section 4 with Figure 6. To treat this properly, one would need to know the distribution of the incubation times, and compute a deconvolution.

(c) Imported cases, i.e. travellers who became infected abroad and got reported after returning home, should not be counted as secondary cases because the corresponding primary case has not been accounted for. However, the location of infection is often unknown; such cases will then unduly increase the numerator in Equation (5), and hence also the estimated reproduction number. This might explain the surprisingly large estimated values – larger than 4, cf. Section 3 – at the beginning of March in Germany, see Figure 8, when many infections were acquired during holidays abroad.

Other problems originate from the modelling approach.

(d) In Equation (2), a structural assumption was made: the infectivity profile does not change over time. If changing conditions affect cases at different infection ages differently, e.g. because the health system is overwhelmed and no longer can provide for high quality isolation of severe cases (with higher infection ages), or because better medical treatment for such cases becomes available, then the change of the transmissibility β depends on the infection age. As a result, the estimates for the reproduction number will combine the changes for the different infection ages into a certain average.

(e) Similarly, the method does not distinguish individuals in different strata of the population, e.g. age groups or regions. So changes which affect certain strata more and others less, e.g. school closures, will again be averaged over the population.

(f) Finally, the infectivity profile requires modelling. We stress that this needs to be distinguished from a virological assessment of a case's level of infectiousness, as it rather describes the potential to successfully transmit the virus. For example, the probability at a late stage of the infection may be assumed to be very low: such a person is most likely to be well isolated, either at home (where either all other members of the household have already been infected or apparently are immune) or at a hospital (where isolation measures are strict), so even though from a virological point of view the person may be highly infectious, she probably will not cause a secondary infection at that stage. From data on chains of infection, it may be possible to directly estimate the infectivity profile from the observed generation times.

All these issues render comparisons between countries particularly difficult. In any case, they sound a note of caution when looking at the proposed estimates. Nonetheless, encouraged by the results in Section 5, we believe that qualitatively correct conclusions may be drawn from them, and that they may prove useful to continuously monitor the spread of COVID-19.

A   Derivation of confidence intervals

Starting from Equation (4), the conditional expectation of $\hat{R}(t)$ given the past is

$$\mathbf{E}(\hat{R}(t)\,|\,I(t-1),\dots) = \frac{\mathbb{E}(I(t)|I(t-1),\dots)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)} = R(t)\,. \tag{9}$$

Therefore, $\hat{R}(t)$ is **unbiased**,

$$\mathbf{E}\,\hat{R}(t) = R(t)\,, \tag{10}$$

and the **conditional variance** of $\hat{R}(t)$ is given by

$$\mathbf{Var}(\hat{R}(t)\,|\,I(t-1),\dots) = \frac{R(t)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)}\,. \tag{11}$$

An application of Slutsky's lemma gives an asymptotic $(1-\alpha)$-confidence interval for $R(t)$: if q denotes the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution it is given by

$$\left[\hat{R}(t) - q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)}},\; \hat{R}(t) + q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau)I(t-\tau)}}\right]. \tag{12}$$

Note that (approximate) coverage is always guaranteed conditionally on the past, and hence also without conditioning.

B   Derivation of the infectivity profile for the SEIR-model

Both latent period and infectious period are geometrically distributed with parameters p_I and p_R, respectively. We need to compute the convolution (summing over time s of getting infectious). For $\tau > 1$ and assuming $p_I > p_R$ (the other cases are similar), we obtain

$$\begin{aligned}
w(\tau) &= \sum_{s=1}^{\tau-1} p_I(1-p_I)^{s-1}(1-p_R)^{\tau-1-s} = p_I(1-p_R)^{\tau-2}\sum_{s=0}^{\tau-2}\left(\frac{1-p_I}{1-p_R}\right)^s \\
&= p_I(1-p_R)^{\tau-2}\frac{1-\left(\frac{1-p_I}{1-p_R}\right)^{\tau-1}}{\frac{p_I-p_R}{1-p_R}} = p_I(1-p_R)^{\tau-1}\frac{1-\left(\frac{1-p_I}{1-p_R}\right)^{\tau-1}}{p_I-p_R} \\
&= p_I\frac{(1-p_R)^{\tau-1}-(1-p_I)^{\tau-1}}{p_I-p_R}\,.
\end{aligned} \tag{13}$$

References

Abbott, S., Hellewell, J., Munday, J. D., Chun, J. Y., Thompson, R. N., Bosse, N. I., Chan, Y.-W. D., Russell, T. W., Jarvis, C. I., CMMID nCov working group, Flasche, S., Kucharski, A. J., Eggo, R. and Funk, S. (2020). Temporal variation in transmission during the COVID-19 outbreak. https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html.

Cori, A., Ferguson, N. M., Fraser, C. and Cauchemez, S. (2013). A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, *American Journal of Epidemiology* **178**(9): 1505–1512. https://doi.org/10.1093/aje/kwt133.

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic, *PLOS ONE* **2**(8): 1–12. https://doi.org/10.1371/journal.pone.0000758.

Grassly, N. C., Fraser, C., Wenger, J., Deshpande, J. M., Sutter, R. W., Heymann, D. L. and Aylward, R. B. (2006). New strategies for the elimination of polio from India, *Science* **314**(5802): 1150–1153. https://science.sciencemag.org/content/314/5802/1150.

Johns Hopkins University Center for Systems Science and Engineering (2020). 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. https://github.com/CSSEGISandData/COVID-19.

Obadia, T., Haneef, R. and Boëlle, P.-Y. (2012). The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks, *BMC Medical Informatics and Decision Making* **12**(147). https://doi.org/10.1186/1472-6947-12-147.

Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A. and Jewell, C. P. (2020). Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. medRxiv, https://doi.org/10.1101/2020.01.23.20018549.

Robert Koch-Institut (2020). RKI COVID19. dl-de/by-2-0, https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0.

World Health Organisation (2020). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19).
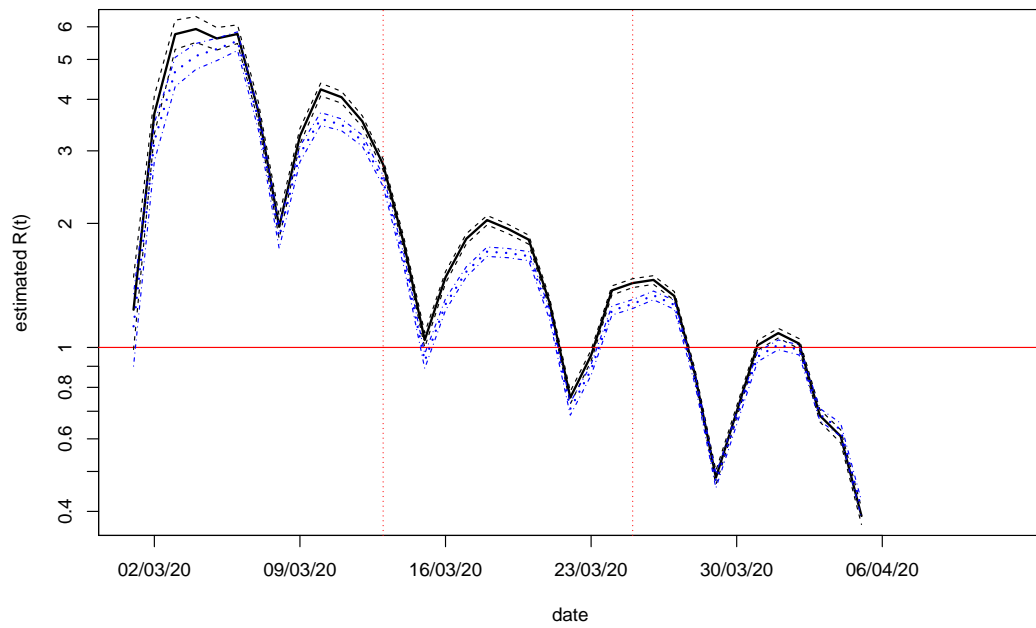
Figure 11: Estimated reproduction numbers for Germany over time (solid/dotted lines) with point-wise 95%-confidence intervals (dashed/dash-dotted lines) based on data from Robert Koch-Institut (2020) using the infectivity profile of the SEIR-model in Section 4 (see Figure 2), shown in black (solid), and using the infectivity profile modelled in Section 3 (see Figure 1), shown in blue (dotted), respectively; vertical red dashed lines indicate the time period over which countermeasures have been implemented, cf. Table 1.