Estimation of the current development of the SARS-CoV-2 epidemic in Germany - nowcasting

Explanation of the data used

There is great interest in presenting and understanding the current infection and the timely development of SARS-CoV-2 infections and Covid-19 cases in Germany. Naturally, no one can know or accurately determine the actual number of infections that occurred today or in the past week. Only when the data subjects have tested positive can their number be recorded and analyzed in a survey system. In general, however, it is true that not all infected people develop symptoms, not all who develop symptoms go to a doctor's office, not all who go to the doctor are tested and not all who test positive are also recorded in a survey system. In addition, a certain amount of time passes between all of these individual steps,

In Germany, doctors and laboratories report infections with SARS-CoV-2 to the responsible health authorities in accordance with the notification obligation under the Infection Protection Act (IfSG) and transmit them to the Robert Koch Institute (RKI) via the responsible state authorities.

There were 103,140 SARS-CoV-2 cases at the current data status (8.4.2020, 00:00 a.m.). Among them were 50,761 (49.8%) men and 51,988 (50.0%) women. In 391 (0.2%) other cases, the sex was either diverse, not assessed or not known. The median age was 50 years (interquartile range [IQR]: 33 - 61 years), 207 cases had no age information. If one plots the cases after the date of receipt at the RKI, the curves in Figure 1 result. Up to 6,000 cases per day have been transmitted to the RKI by the health authorities, somewhat less in the past few days. A separate presentation of this development by gender and age group shows above all a significant increase in the number of new cases in the age group (80+).

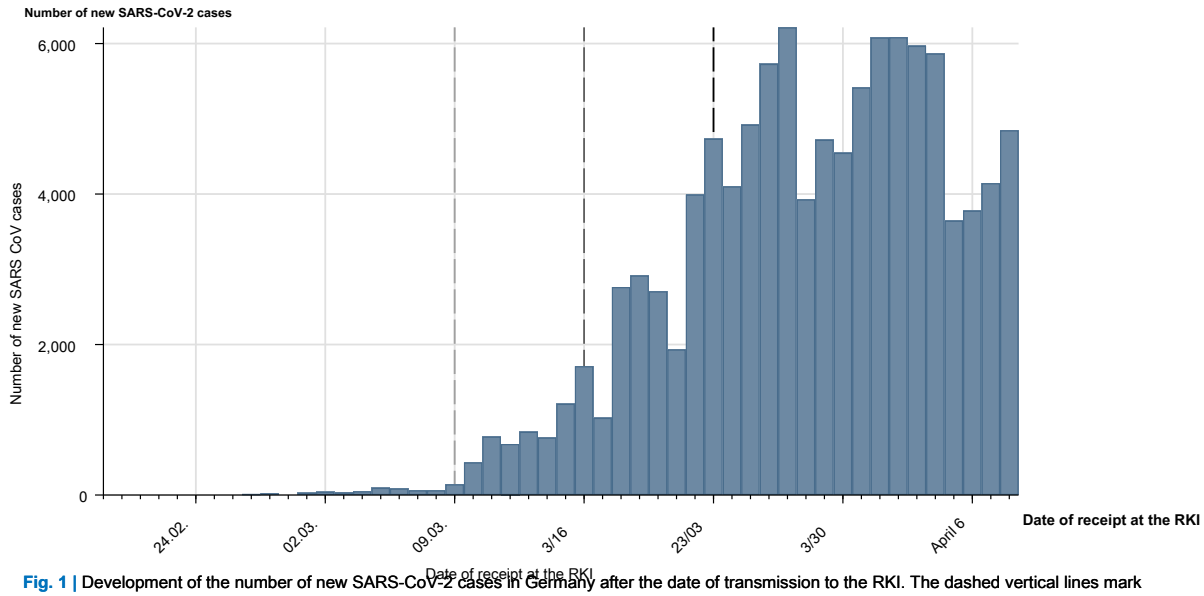

Fig. 1 | Development of the number of new SARS-CoV-2 cases in Germany after the date of transmission to the RKI. The dashed vertical lines mark the start of the measures listed in Tab. 1 (p. 15) on March 9, March 16 and March 23, 2020.

The cases reported to the RKI by the health authorities on a daily basis do not directly reflect the current infection, as stated above. The chronological course of the current infection would best be represented by the number of infections that occurred daily. In most cases, however, the exact time of infection is not known or cannot be determined.

Imputation of missing values at the beginning of the illness

In order to present the current development of the SARS-CoV-2 epidemic, the date of illness (date of symptom onset) is the most suitable and available parameter from the reporting data. The beginning of the disease was indicated by the health authorities in 62,909 (61%) cases. In some cases of confirmed SARS-CoV-2 infection, an asymptomatic course develops, so that the onset of the disease never occurs. Nevertheless, these cases are also assigned an artificial onset of disease in the context of our analysis; they are treated as if the onset of the disease was not specified. In 511 cases, the time interval between the date of transmission to the RKI and the onset of the disease was negative or was more than 30 days *.

A so-called method was used to replace this missing information *multiple imputation* carried out,[1] where the missing data values are estimated based on the statistical relationships of the known data. The most important information to determine the missing onset of illness was the date of receipt of the case report at the RKI, the missing values were estimated separately according to gender and age group.

The distribution of the duration between the onset of the disease and the date of submission of the notification

The report to the RKI shows that 50% of the cases were transmitted after 7 days. In most cases, this duration is between 5 and 10 days. An analysis of the temporal dynamics of this distribution after the day of receipt at the RKI showed: The average rose between March 12 and 21 from 5.3 days to 6.6 days. Between March 22 and 28 it was about 8 days, between March 29 and 31 it was about 9 days. Since April 1, the duration from the onset of illness to the submission of the report to the RKI has been decreasing again and was recently around 8 days. These shifts are taken into account in the imputation of the missing values at the onset of the disease.

To carry out the *multiple imputation* 200 realizations from the empirical distribution of the duration between the onset of illness and the date of transmission were assigned to the cases without onset (separated by gender and age group). The difference between the transmission date and this distance then gives the different realizations of the simulated onset of the disease. This gives us an estimate of the onset of the disease in the cases already transmitted (see Fig. 2 "Onset of the disease imputed", p. 12).

Explanation of the nowcasting

Nowcasting [2nd] creates an estimate of the course of the number of SARS-CoV-2 cases in Germany that have already occurred, taking into account the delay in diagnosis, reporting and transmission. To do this, we determine the proportion of cases that were reported after a certain number of days, x, after the onset of the disease. This percentage is used to correct the number of reported reports with the onset of illness x days before the status of the analysis. It must be borne in mind that only recently diagnosed, reporting and transmission intervals have been recorded for recently ill cases. [3rd]

In the current procedure we make the simplifying assumption that the correction is constant over time. An evaluation of this assumption shows that the estimated distribution of the share over time leads to a more complete transmission

* Note: A duration of 0 days or even a small negative duration can be explained by cases that were tested as part of a contact person follow-up of a confirmed case and only developed symptoms after the positive test.

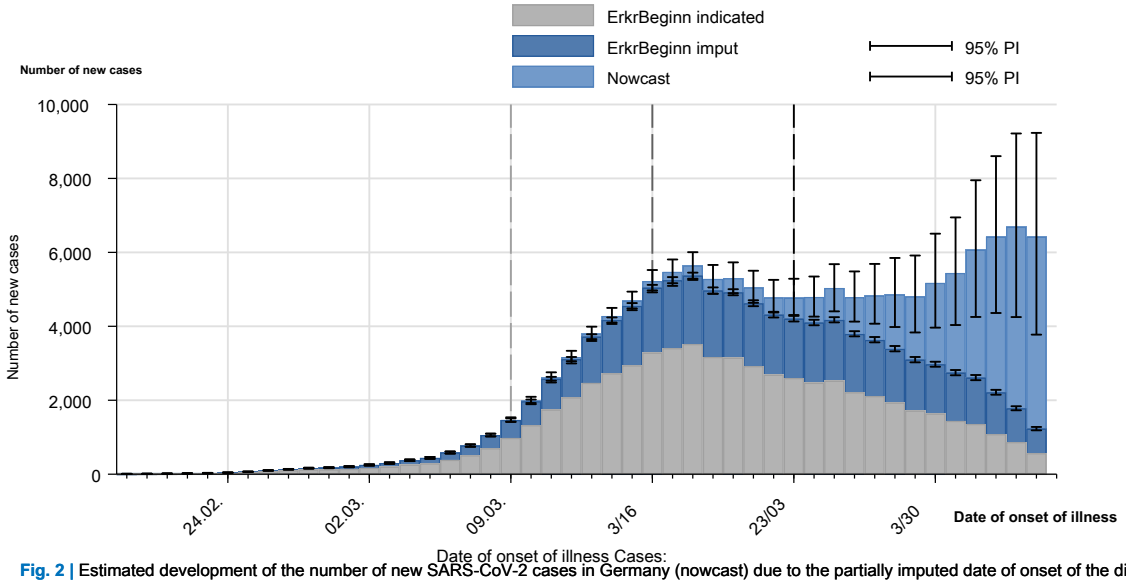

Fig. 2 | Estimated development of the number of new SARS-CoV-2 cases in Germany (nowcast) due to the partially imputed date of onset of the disease and adjusted for diagnosis and reporting delay with 95% prediction intervals (95% PI). The dashed vertical lines mark the start of certain measures on March 9, March 16 and March 23, see Tab. 1 (p. 15). Shown until April 4th; 3 days before the analysis.

developed after a given number of days. Therefore, nowcasting tends to slightly overestimate the expected number of cases.

Nowcasting results

The course of the curve of the cases transmitted to the RKI (gray and dark blue), especially in the last 10 days, is significantly below the estimated course of the already symptomatic cases (light blue) in Germany, which had an onset of disease a few days ago and only a small one Partly diagnosed, reported and transmitted (see Fig. 2). The course of the expected new cases, corrected for the delay in reporting and transmission, initially rises steadily, but stabilizes at a level of around 16 March 16

5,000 new cases a day. Since March 30, this number may have increased to over 6,000 new cases per day (see Fig. 2).

The 95% prediction intervals show the uncertainty due to the adjustment after the diagnosis and reporting delay as well as due to the partially missing information on the onset of the disease. Nowcasting is unstable for cases with an onset of disease 3 days or less before the status of the analysis, since within 3 days one to

small part of these cases is recorded. Overall, nowcasting tends to be relatively sensitive to fluctuations in the number of new cases close to the status of the analysis, but stabilizes after a few days if the increase or decrease is not confirmed (see Fig. 3, p. 13). .

A look at the development by gender and age groups (0 - 19, 20 - 39, 40 - 59, 60 - 79 and 80+) shows that the forecast number of cases per 100,000 inhabitants in the age group (80+) increases particularly sharply. This will presumably also be seen in a stronger increase in the number of hospitalized cases and cases requiring intensive care. In absolute numbers, the adults dominate between the ages of 20 and 79, due to the high proportion of the total population, especially the 40 to 59 year olds.

Estimation of the reproduction number R

Based on the nowcasting, an estimate of the time-dependent reproduction number R can be carried out. The number of reproductions is the number of people who are infected by an index case on average. Based on the current state of knowledge, we assume that an average of 5 days pass between the infection and the onset of the first symptoms. Are probably infected

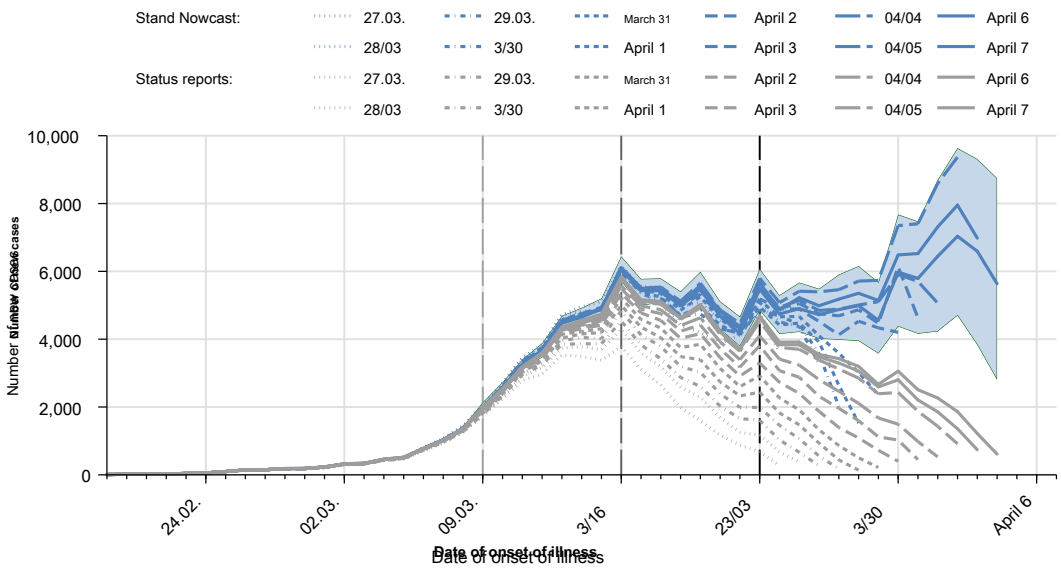

Fig. 3 | Comparison of the estimated development of the number of SARS-CoV-2 cases in Germany (nowcast) to different data levels, the current estimate with prediction interval. The dashed vertical lines mark the start of certain measures on March 9, March 16 and March 23, see Tab. 1 (p. 15).

However, people become infectious about 2 days before the onset of symptoms and can infect other people as early as 3 days after their own exposure. The generation time describes the mean period of time from the infection of a person to the infection of the subsequent cases infected by him. We estimate this period of time to be about 4 days because the infectivity at the beginning of the infection is particularly high and the infected person is not aware before the onset of symptoms that he can already infect others. The generation time is not a stable property of the pathogen, but, like the number of reproductions, depends on various factors and can change over time.

If each case infects 2 subsequent cases on average (R = 2), then the number of new infections doubles after each generation time. In contrast, the number of new infections is halved with a reproductive number R = 0.5. It is precisely this dynamic that can be used in reverse to determine the effective number of reproductions from the data confirmed to appreciate new SARS CoV-2 diseases.

With a constant generation time of 4 days, R is the quotient of the number of new cases in two successive periods of 4 days each. If the number of new cases increased in the second period, the R is above 1. If the number of new cases is the same in both periods, the number of reproductions is 1. This 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 halves within the generation period.

The R estimate for the beginning of March shows values in the range of R = 3, which then decrease and have stabilized around R = 1 since March 22 (see Fig. 4, p. 14). R has risen slightly since March 30 and is 1.2 (April 4: 95% PI:

0.9-1.6). One possible reason for the slight increase is that the virus is now spreading more widely among older people as we are increasingly seeing outbreaks in nursing homes and hospitals. Another aspect is that the test capacities in Germany have increased significantly and that a greater proportion of the infections can be seen as a result of stronger testing

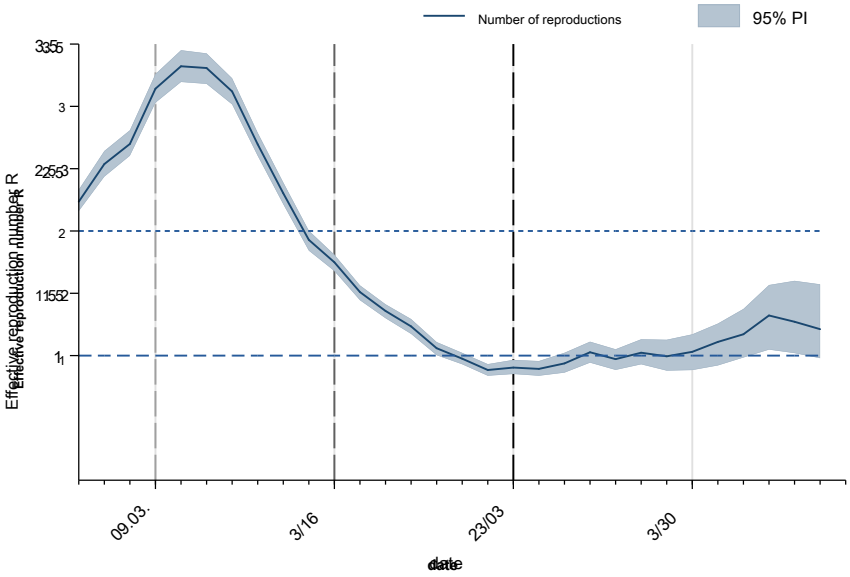

Fig. 4 | Estimation of the effective reproduction number R for an assumed generation time of 4 days. The dashed vertical lines indicate the start of the measures listed in Table 1 (p. 15) on March 9, March 16 and March 23, 2020.

becomes. This structural effect and the resulting increase in the number of reports can lead to the current R-value somewhat overestimating real events. Adjustment for the higher test rates is not possible without further ado because there is no sufficiently differentiated test data.

A stability analysis of the R estimate shows that the R value is more stable overall than the nowcasting itself (see Fig. 5). Still find yourself

Here, too, individual small outliers that disappear after a few days.

The figures for the nowcast and the R estimate show the date of the start of important measures to contain the SARSCoV-2 epidemic in Germany for orientation. However, these times are not included in the estimate of the nowcast itself. In addition to the testing of suspected cases, the isolation of confirmed cases

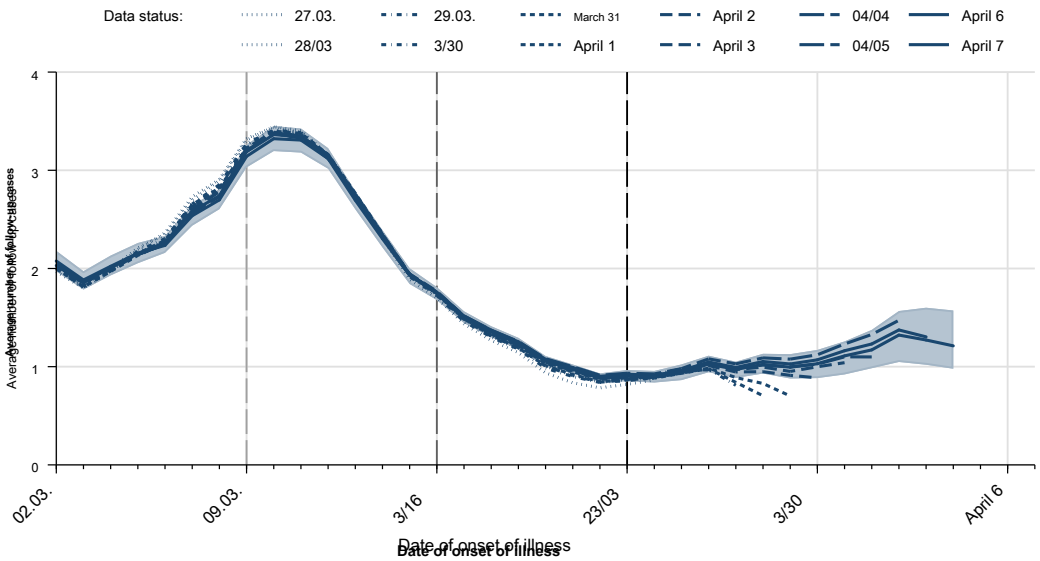

Fig. 5 | Comparison of the estimate of the effective reproduction number R for an assumed generation time of 4 days for different data levels. The dashed vertical lines indicate the start of the measures listed in Table 1 (p. 15) on March 9, March 16 and March 23, 2020.

Start of the measure	measure
9th March	Cancellation of large events in different federal states (with over 1,000 participants)
March 16	Federal-state agreement on guidelines against the spread of the coronavirus
March 23	Nationwide extensive ban on contacts

Tab. 1 | Start of measures to contain the SARS-CoV-2 epidemic in Germany, 2020.

and the quarantine of close contacts in confirmed cases, these are general contact-reducing measures to reduce the spread of the virus (see Table 1).

An R estimate is also possible based on the course of the new cases, for example after the reporting date, and this is often the only option for international data. In principle, this should lead to similar results, but should be somewhat more susceptible to reporting artifacts.

literature

1 Little RJ, Rubin DB: Statistical analysis with missing data: John Wiley & Sons; 2019

2nd Höhle M, an der Heiden M: Bayesian nowcasting during the STEC O104: H4 outbreak in Germany, 2011. Biometrics 2014; 70 (4): 993 - 1002

3rd Lawless J: Adjustments for reporting delays and the prediction of occurred but not reported events. Canadian Journal of Statistics 1994; 22 (1): 15-31

author

a) Dr. Matthias an der Heiden | b) Dr. Osamah Hamouda

a) Robert Koch Institute | Department 3 | FG 34 HIV / AIDS and other sexually or blood-borne infections

b) Robert Koch Institute | Department 3

Correspondence: anderHeidenM@rki.de

Conflict of interest

The authors state that there is no conflict of interest.

Suggested citation

an der Heiden M, Hamouda O: Estimate of the current development of the SARS-CoV-2 epidemic in Germany - nowcasting.

Epid Bull 2020; 17: 10 - 15 | DOI 10.25646 / 6692