

## Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting

### Erläuterung zu den verwendeten Daten

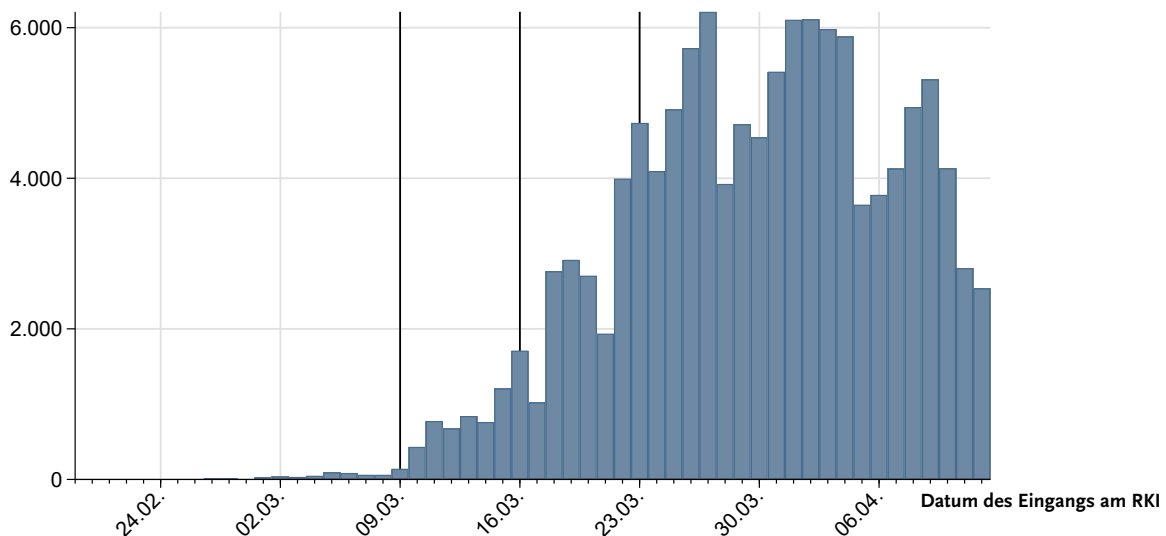
Es besteht ein großes Interesse daran das aktuelle Infektionsgeschehen und die zeitnahe Entwicklung von SARS-CoV-2-Infektionen und Covid-19-Erkrankungsfällen in Deutschland darzustellen und zu verstehen. Naturgemäß kann niemand die tatsächliche Anzahl der heute oder in der vergangenen Woche erfolgten Infektionen genau wissen oder bestimmen. Erst wenn die betroffenen Personen positiv getestet wurden, kann deren Anzahl in einem Erhebungssystem erfasst und analysiert werden.

Ganz allgemein gilt jedoch, dass nicht alle infizierten Personen Symptome entwickeln, nicht alle die Symptome entwickeln suchen eine Arztpraxis auf, nicht alle die zum Arzt gehen werden getestet und nicht alle die positiv getestet werden, werden auch in einem Erhebungssystem erfasst. Außerdem vergeht zwischen all diesen einzelnen Schritten eine gewisse Zeit, so dass kein Erhebungssystem, und sei es noch so gut, ohne zusätzliche Annahmen und Berechnungen eine Aussage über das aktuelle Infektionsgeschehen machen kann.

In Deutschland werden gemäß der Meldepflicht nach Infektionsschutzgesetz (IfSG) Infektionen mit SARS-CoV-2 von den Ärzten und Laboren an die zuständigen Gesundheitsämter gemeldet und von diesen über die zuständigen Landesbehörden an das Robert Koch-Institut (RKI) übermittelt.

Zum aktuellen Datenstand (13.4.2020, 00:00 Uhr) lagen 123.016 SARS-CoV-2-Fälle vor. Darunter waren 59.306 (48,2 %) Männer und 63.287 (51,5 %) Frauen. Bei 420 (0,3 %) weiteren Fällen war das Geschlecht entweder divers, nicht erhoben oder nicht bekannt. Der Altersmedian lag bei 50 Jahren (Interquartilsabstand [IQR]: 33–62 Jahre), zu 201 Fällen lagen keine Altersangaben vor. Trägt man die Fälle nach dem Datum des Eingangs am RKI auf, so ergeben sich die Kurven in Abbildung 1. Insgesamt wurden von den Gesundheitsämtern bis zu 6.000 Fälle pro Tag an das RKI übermittelt, in den letzten Tagen weniger – möglicherweise aufgrund der Osterfeiertage. Eine getrennte Darstellung dieser Entwicklung nach Geschlecht und Altersgruppen zeigt vor allem eine deutlich ansteigende Anzahl von neuen Erkrankungsfällen in der Altersgruppe (80+).

Anzahl neuer SARS-CoV-Fälle



**Abb. 1 |** Entwicklung der Anzahl von neuen SARS-CoV-2 Fällen in Deutschland nach Datum der Übermittlung an das RKI. Die dunkleren vertikalen Linien kennzeichnen den Start der in Tab. 1 (S. 15) genannten Maßnahmen am 9. März, 16. März und 23. März 2020.

Die dem RKI von den Gesundheitsämtern täglich übermittelten Fälle spiegeln, wie oben ausgeführt, nicht direkt das aktuelle Infektionsgeschehen wider. Der zeitliche Verlauf des aktuellen Infektionsgeschehens wäre am besten durch die Anzahl der täglich erfolgten Infektionen darstellbar. Der genaue Infektionszeitpunkt ist aber in den allermeisten Fällen nicht bekannt oder ermittelbar.

### Imputation von fehlenden Werten zum Erkrankungsbeginn

Um die aktuelle Entwicklung der SARS-CoV-2-Epidemie darzustellen ist daher das Erkrankungsdatum (Datum des Symptombeginns) der am besten geeignete und aus den Meldedaten zur Verfügung stehende Parameter. Der Erkrankungsbeginn wurde bei 76.914 (62,5%) Fällen von den Gesundheitsämtern angegeben. In manchen Fällen einer bestätigten SARS-CoV-2-Infektion entwickelt sich ein asymptomatischer Verlauf, so dass es nie zu einem Erkrankungsbeginn kommt. Trotzdem wird auch diesen Fällen in Rahmen unserer Analyse ein künstlicher Erkrankungsbeginn zugeordnet, sie werden behandelt, als handele es sich um eine fehlende Angabe des Erkrankungsbeginns. In 522 Fällen war der zeitliche Abstand zwischen dem Datum der Übermittlung an das RKI und dem Erkrankungsbeginn negativ oder lag über 30 Tage\*. Diese Fälle wurden bei der nachfolgenden Analyse und der Imputation des Erkrankungsbeginns nicht miteinbezogen.

Als Verfahren für das Ersetzen dieser fehlenden Angaben wurde eine sogenannte *multiple Imputation* durchgeführt,<sup>1</sup> bei der die fehlenden Datenwerte aufgrund der statistischen Zusammenhänge der bekannten Daten geschätzt werden. Als wichtigste Angabe zur Ermittlung des fehlenden Erkrankungsbeginns wurde dabei das Datum des Eingangs der Fallmeldung am RKI verwendet, die fehlenden Werte wurden getrennt nach Geschlecht und Altersgruppe geschätzt.

Die Verteilung der Dauer zwischen Erkrankungsbeginn und dem Datum der Übermittlung der Meldung an das RKI zeigt, dass 50% der Fälle nach 7 Tagen übermittelt wurden. Für die meisten Fälle

liegt diese Dauer zwischen 5 und 10 Tagen. Eine Analyse der zeitlichen Dynamik dieser Verteilung nach dem Tag des Eingangs am RKI ergab: Der Mittelwert stieg zwischen dem 12. und 21. März von 5,3 Tagen auf 6,6 Tage. Zwischen dem 22. und 28. März lag er bei etwa 8 Tagen, zwischen dem 29. und 31. März bei etwa 9 Tagen. Seit dem 1. April sinkt die Dauer von Erkrankungsbeginn bis zur Übermittlung der Meldung an das RKI eher wieder und lag zuletzt bei etwa 7,6 Tagen. Diese Verschiebungen werden bei der Imputation der fehlenden Werte des Erkrankungsbeginns berücksichtigt.

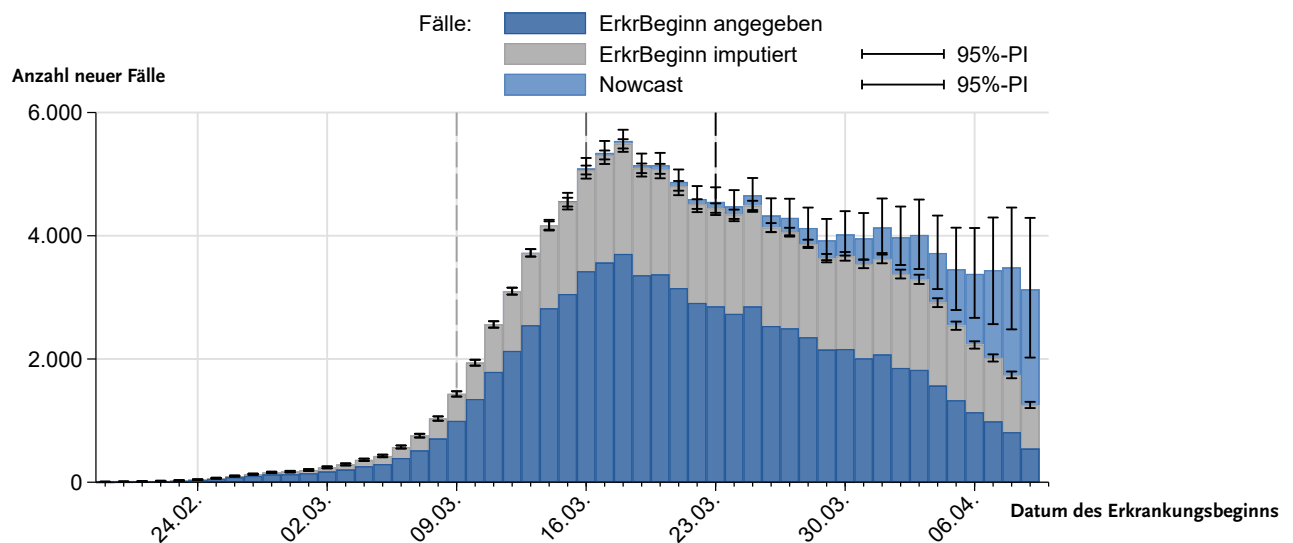
Zur Durchführung der *multiple Imputation* wurden (getrennt nach Geschlecht und Altersgruppe) jeweils 200 Realisationen aus der empirischen Verteilung der Dauer zwischen Erkrankungsbeginn und Übermittlungsdatum den Fällen ohne Erkrankungsbeginn zugeordnet. Die Differenz zwischen dem Übermittlungsdatum und diesem Abstand ergibt dann die verschiedenen Realisierungen des simulierten Erkrankungsbeginns. Damit erhalten wir eine Schätzung des Erkrankungsbeginns der bereits übermittelten Fälle (s. Abb. 2 „Erkrankungsbeginn imputiert“, S. 12).

### Erläuterung des Nowcastings

Das Nowcasting<sup>2</sup> erstellt eine Schätzung des Verlaufs der Anzahl von bereits erfolgten SARS-CoV-2-Erkrankungsfällen in Deutschland unter Berücksichtigung des Diagnose-, Melde- und Übermittlungsverzugs. Dazu ermitteln wir den Anteil an Fällen, der nach einer bestimmten Anzahl von Tagen,  $x$ , nach Erkrankungsbeginn gemeldet wurde. Dieser Anteil wird verwendet um die Anzahl der übermittelten Meldungen mit Erkrankungsbeginn  $x$  Tage vor dem Stand der Analyse zu korrigieren. Dabei muss beachtet werden, dass für erst kürzlich erkrankte Fälle nur diejenigen mit kurzen Diagnose-, Melde- und Übermittlungsabständen bereits erfasst wurden.<sup>3</sup>

Um zeitliche Veränderungen der Korrekturverteilung zu berücksichtigen, setzen wir die Informationen der aktuell gemeldeten Fälle in einen Kontext von Fällen innerhalb eines Fensters von 7 Tagen des Erkrankungsbeginns. Die obige Anteilsverteilung wird hierbei wie in Lawless 1994<sup>3</sup> beschrieben als Produkt bestimmter bedingter Wahrscheinlichkeiten geschätzt. Ein Beispiel ist die Wahrscheinlichkeit,

\*Anmerkung: Eine Dauer von 0 Tagen oder sogar kleine negative Dauern lassen sich durch Fälle erklären, die im Rahmen einer Kontaktpersonen-Nachverfolgung eines bestätigten Falles getestet wurden und erst nach dem positiven Test Symptome entwickelt haben.



**Abb. 2** | Geschätzte Entwicklung der Anzahl von neuen SARS-CoV-2-Fällen in Deutschland (Nowcast) aufgrund teilweise imputiertem Datum des Erkrankungsbeginns und adjustiert für Diagnose- und Meldeverzug mit 95%-Prädiktionsintervallen (95%-PI). Die gestrichelten vertikalen Linien kennzeichnen den Start bestimmter Maßnahmen am 9. März, 16. März und 23. März, s. Tab. 1 (S. 15). Dargestellt bis zum 9. April; 3 Tage vor dem Stand der Analyse.

dass ein Fall, der mit einer Verzögerung,  $D$ , von höchstens  $d$  Tagen übermittelt wird, genau nach  $d$  Tagen übermittelt wird,  $g(d) = P(D = d \mid D \leq d)$ . Wir benennen das Datum der Analyse mit  $T$ . Um  $g$  für einen Verzug von  $d$  Tagen zu schätzen, verwenden wir alle Fälle mit Erkrankungsbeginn im Bereich  $T-d$  und  $T-d-7$ . Denn Fälle mit späterem Erkrankungsbeginn sind nach den vereinbarten  $d$  Tagen noch nicht vollständig gemeldet und Fälle mit früherem Erkrankungsbeginn gehören vermutlich schon zu einer anderen Anteilsverteilung. Das 7-Tages-Fenster wurde gewählt, weil es zu stabilen Ergebnissen führt und Unterschiede zwischen den Wochentagen ausgleicht. Darüber hinaus war die Anzahl der übermittelten Fälle nach Altersgruppe und Geschlecht im Zeitraum von 7 Tagen ausreichend hoch.

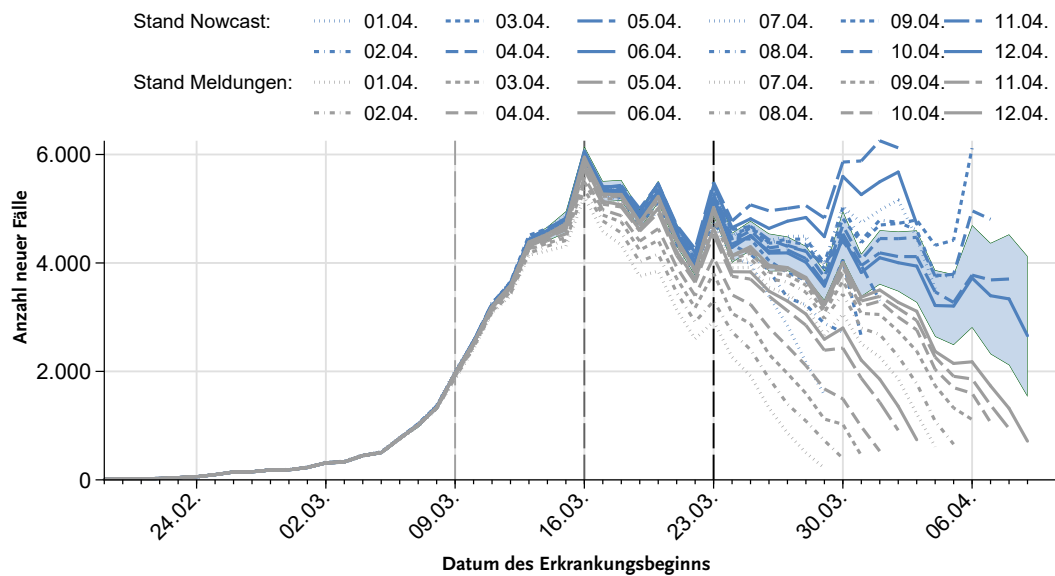
### Ergebnisse des Nowcastings

Der Verlauf der Kurve der an das RKI übermittelten Fälle (dunkelblau und grau) liegt vor allem in den letzten 10 Tagen deutlich unter dem geschätzten Verlauf der bereits symptomatischen Fälle (hellblau) in Deutschland, die einen Erkrankungsbeginn vor wenigen Tagen hatten und erst zu einem kleinen Teil diagnostiziert, gemeldet und übermittelt wurden (s. Abb. 2). Der für den Melde- und Übermittlungsverzug korrigierte Verlauf der zu erwartenden neuen Fälle steigt zunächst stetig an und erreicht

ein Maximum am 18. März mit etwa 5.500 neuen Erkrankungsfällen. Danach fällt die Anzahl neuer Fälle pro Tag auf etwa 4.000. Seit dem 4. April kommt es möglicherweise zu einem weiteren Rückgang, dies ist aber noch mit höherer Unsicherheit verbunden und könnte sich in den nächsten Tagen noch ändern (s. Abb. 2).

Die 95%-Prädiktionsintervalle zeigen die Unsicherheit aufgrund der Adjustierung nach dem Diagnose- und Meldeverzug als auch aufgrund der teilweise fehlenden Angaben zum Erkrankungsbeginn. Das Nowcasting verhält sich instabil für Fälle mit einem Erkrankungsbeginn 3 Tage oder weniger vor dem Stand der Analyse, da innerhalb von 3 Tagen ein zu kleiner Teil dieser Fälle erfasst wird. Insgesamt tendiert das Nowcasting dazu relativ sensitiv auf Schwankungen der Anzahl neuer Fälle in der Nähe des Stands der Auswertung zu reagieren, stabilisiert sich dann aber nach wenigen Tagen, wenn sich der Anstieg oder Rückgang nicht bestätigt (s. Abb. 3, S. 13).

Eine Betrachtung der Entwicklung nach Geschlecht und Altersgruppen (0–19, 20–39, 40–59, 60–79 und 80+) zeigt, dass die prognostizierte Anzahl von Fällen pro 100.000 Einwohner in der Altersgruppe (80+) besonders stark ansteigt. Dies wird sich vermutlich auch in einem stärkeren Anstieg der Anzahl von hospitalisierten Fällen und intensivpflicht-



**Abb. 3 |** Vergleich der geschätzten Entwicklung der Anzahl von SARS-CoV-2-Fällen in Deutschland (Nowcast) zu verschiedenen Datenständen, die aktuelle Schätzung mit Prädiktionsintervall. Die gestrichelten vertikalen Linien kennzeichnen den Start bestimmter Maßnahmen am 9. März, 16. März und 23. März, s. Tab. 1 (S. 15).

tigen Fällen zeigen. In absoluten Zahlen dominieren die Erwachsenen im Alter von 20–79 Jahren, auf Grund des hohen Anteils an der Gesamtbevölkerung vor allem die 40- bis 59-Jährigen.

### Schätzung der Reproduktionszahl $R$

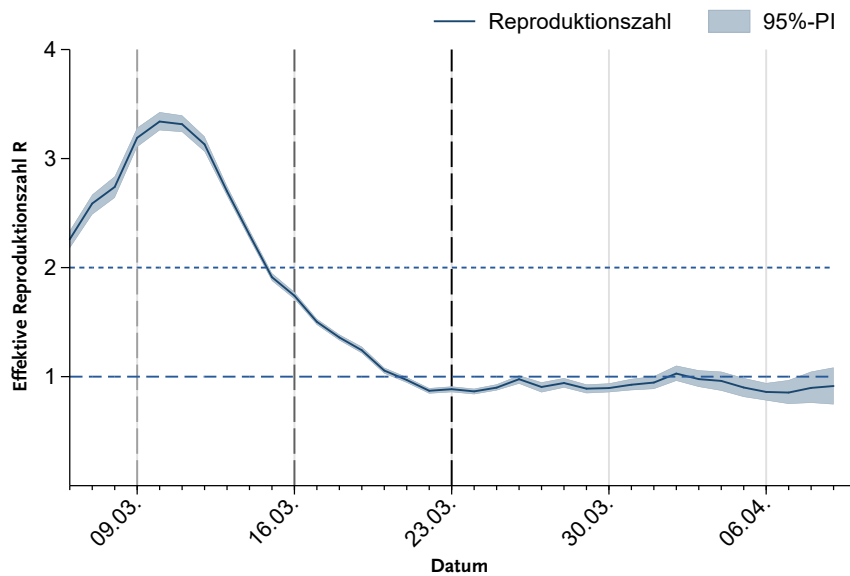
Aufbauend auf dem Nowcasting kann eine Schätzung der zeitabhängigen Reproduktionszahl  $R$  durchgeführt werden. Die Reproduktionszahl ist die Anzahl der Personen, die im Durchschnitt von einem Indexfall angesteckt werden. Basierend auf dem aktuellen Wissensstand gehen wir davon aus, dass zwischen der Ansteckung und dem Beginn der ersten Symptome im Mittel 5 Tage vergehen. Vermutlich sind infizierte Personen aber bereits etwa 2 Tage vor dem Symptombeginn infektiös und können also bereits 3 Tage nach der eigenen Exposition weitere Personen anstecken. Die Generationszeit beschreibt die mittlere Zeitspanne von der Infektion einer Person bis zur Infektion der von ihr angesteckten Folgefälle. Sie entspricht etwa dem seriellen Intervall, das die mittlere Dauer zwischen dem Erkrankungsbeginn eines Falles und dem Erkrankungsbeginn seiner Folgefälle angibt. Diese Zeitspanne schätzen wir auf etwa 4 Tage, weil die Infektiosität zu Beginn der Infektion besonders hoch ist und sich die infizierte Person vor dem Symptombeginn nicht darüber bewusst ist, dass sie bereits andere anstecken kann. Die Generationszeit ist dabei

keine stabile Eigenschaft des Erregers, sondern hängt ebenso wie die Reproduktionszahl von verschiedenen Faktoren ab und kann sich über die Zeit verändern. Zum Beispiel führen Maßnahmen zur Isolation von bestätigten Fällen und Quarantäne von Kontaktpersonen nicht nur zu einer Verringerung der Anzahl von Folgefällen, sondern auch zu einer Verkürzung der Generationszeit, weil die wenigen Ansteckungen direkt am Anfang der Infektion passieren.

Wenn jeder Fall im Durchschnitt 2 Folgefälle ansteckt ( $R = 2$ ), dann verdoppelt sich die Anzahl der neuen Infektionen jeweils nach einer Generationszeit. Dagegen halbiert sich die Anzahl neuer Infektionen bei einer Reproduktionszahl  $R = 0,5$ . Genau diese Dynamik kann man umgekehrt nun verwenden um die effektive Reproduktionszahl aus den Daten zu bestätigten SARS-CoV-2-Neuerkrankungen zu schätzen.

Bei einer konstanten Generationszeit von 4 Tagen, ergibt sich  $R$  als Quotient der Anzahl von Neuerkrankungen in zwei aufeinander folgenden Zeitabschnitten von jeweils 4 Tagen<sup>#</sup>. Hat sich die Anzahl der Neuerkrankungen im zweiten Zeitabschnitt erhöht, so liegt das  $R$  über 1. Ist die Anzahl der Neuerkrankungen

<sup>#</sup>Wenn die Generationszeit als Verteilung geschätzt werden kann, ergibt sich eine etwas kompliziertere Formel mit ähnlicher Grundstruktur.

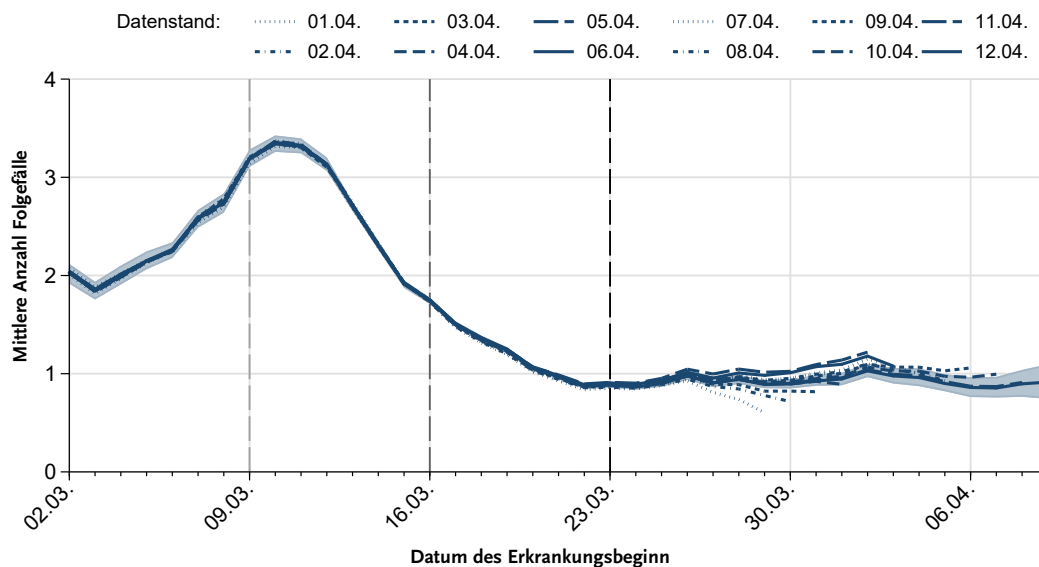


**Abb. 4** | Schätzung der effektiven Reproduktionszahl  $R$  für eine angenommene Generationszeit von 4 Tagen. Die gestrichelten vertikalen Linien kennzeichnen den Start der in Tab. 1 (S. 15) genannten Maßnahmen am 9. März, 16. März und 23. März 2020.

kungen in beiden Zeitabschnitten gleich groß, so liegt die Reproduktionszahl bei 1. Dies entspricht dann einem linearen Anstieg der Fallzahlen. Wenn dagegen nur jeder zweite Fall eine weitere Person ansteckt, also  $R = 0,5$  ist, dann halbiert sich die Anzahl der neuen Infektionen innerhalb der Generationszeit.

Die  $R$ -Schätzung ergibt für Anfang März Werte im Bereich von  $R = 3$ , die danach absinken, und sich etwa seit dem 22. März um  $R = 1$  stabilisieren (s. Abb. 4). Am 9. April lag der Wert von  $R$  bei 0,9 (95 %-PI: 0,8–1,1). Ein Grund dafür, dass der Rück-

gang der Neuerkrankungen trotz der gravierenden Maßnahmen nur relativ langsam passiert, ist, dass sich das Virus nach dem 18. März stärker auch unter älteren Menschen ausbreitet und wir zunehmend auch Ausbrüche in Pflegeheimen und Krankenhäusern beobachten. Ein weiterer Aspekt ist aber auch, dass in Deutschland die Testkapazitäten deutlich erhöht worden sind und durch stärkeres Testen ein insgesamt größerer Teil der Infektionen sichtbar wird. Dieser strukturelle Effekt und der dadurch bedingte Anstieg der Meldezahlen, kann dazu führen dass der aktuelle  $R$ -Wert das reale Geschehen etwas



**Abb. 5** | Vergleich der Schätzung der effektiven Reproduktionszahl  $R$  für eine angenommene Generationszeit von 4 Tagen zu unterschiedlichen Datenständen. Die gestrichelten vertikalen Linien kennzeichnen den Start der in Tab. 1 (S. 15) genannten Maßnahmen am 9. März, 16. März und 23. März 2020.

Start der Maßnahme	Maßnahme
9. März	Absage großer Veranstaltungen in verschiedenen Bundesländern (bei über 1.000 Teilnehmer)
16. März	Bund-Länder-Vereinbarung zu Leitlinien gegen die Ausbreitung des Coronavirus
23. März	Bundesweit umfangreiches Kontaktverbot

**Tab. 1** | Start von Maßnahmen zur Eindämmung der SARS-CoV-2-Epidemie in Deutschland, 2020.

überschätzt. Eine Adjustierung für die höheren Testraten ist nicht ohne weiteres möglich, da keine ausreichend differenzierten Testdaten vorliegen.

Eine Stabilitätsanalyse der R-Schätzung zeigt, dass sich der R-Wert insgesamt stabiler als das Nowcasting selber verhält (s. Abb. 5, S. 14). Dennoch finden sich auch hier einzelne kleine Ausreißer, die aber nach wenigen Tagen wieder verschwinden.

In den Abbildungen zum Nowcast und zur R-Schätzung wird zur Orientierung das Datum des Starts wichtiger Maßnahmen zur Eindämmung der SARS-CoV-2-Epidemie in Deutschland dargestellt. In die Schätzung des Nowcasts selber gehen diese Zeitpunkte aber nicht ein. Neben dem Test von Verdachtsfällen, der Isolation von bestätigten Fällen und der Quarantänisierung der engen Kontaktpersonen von bestätigten Fällen, sind dies allgemeine kontaktreduzierende Maßnahmen zur Verringerung der Ausbreitung des Virus (s. Tab. 1).

Eine R-Schätzung ist auch möglich aufgrund des Verlaufs der neuen Fälle etwa nach dem Meldedatum, und bei internationalen Daten dürfte das auch oft die einzige Möglichkeit sein. Grundsätzlich sollte dies zu ähnlichen Ergebnissen führen, aber zeitlich etwas verschoben und noch etwas anfälliger für Meldeartefakte sein.

## Literatur

- 1 Little RJ, Rubin DB: Statistical analysis with missing data: John Wiley & Sons; 2019: 3. Auflage 2020
- 2 Höhle M, an der Heiden M: Bayesian nowcasting during the STEC O104: H4 outbreak in Germany, 2011. *Biometrics* 2014;70(4):993–1002
- 3 Lawless J: Adjustments for reporting delays and the prediction of occurred but not reported events. *Canadian Journal of Statistics* 1994;22(1):15–31

## Autor

<sup>a)</sup> Dr. Matthias an der Heiden | <sup>b)</sup> Dr. Osamah Hamouda

<sup>a)</sup> Robert Koch-Institut | Abteilung 3 | FG 34 HIV/AIDS und andere sexuell oder durch Blut übertragbare Infektionen

<sup>b)</sup> Robert Koch-Institut | Abteilung 3

**Korrespondenz:** [anderHeidenM@rki.de](mailto:anderHeidenM@rki.de)

## Vorgeschlagene Zitierweise

an der Heiden M, Hamouda O: Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting.

Epid Bull 2020;17:10–15 | DOI 10.25646/6692.2

(Eine 1. Fassung ist am 9.4.2020 online vorab erschienen.)

(Dieser Artikel ist am 15.4.2020 online vorab erschienen.)

## Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

## Danksagung

Wir danken Prof. Dr. Michael Höhle von der Universität Stockholm für hilfreiche Diskussionen und Anregungen und die kritische Durchsicht des Manuskriptes.