

Schätzung der aktuellen Entwicklung neuer Erkrankungsfälle von SARS-CoV-2 in Deutschland - Nowcasting

Matthias an der Heiden
anderHeidenM@rki.de

07.04.2020

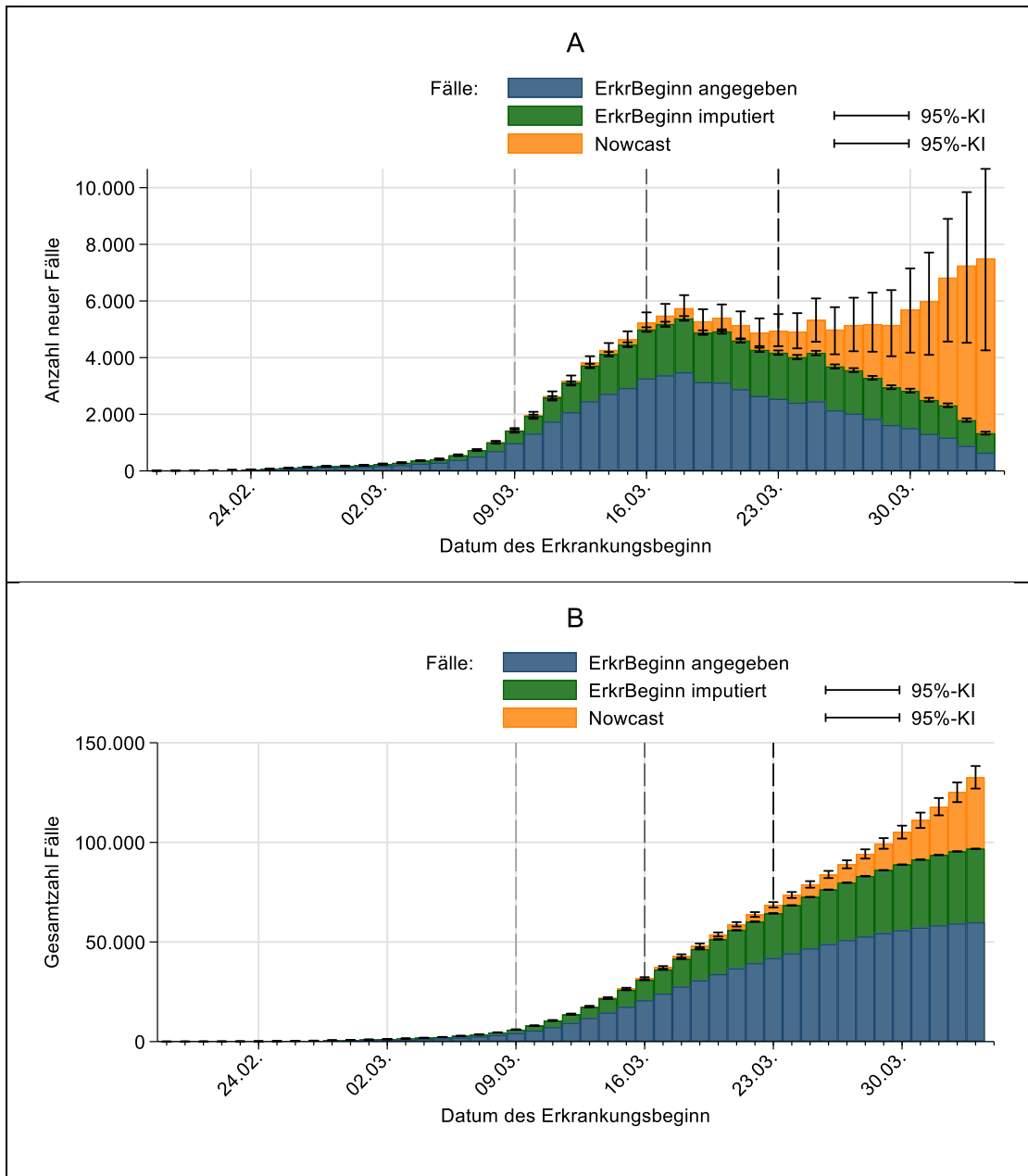


Abb. 1. Geschätzte Entwicklung der Anzahl von SARS-CoV-2 Fällen in Deutschland (Nowcast) aufgrund teilweise imputiertem Datum des Erkrankungsbeginns und adjustiert für Diagnose- und Meldeverzug. Die gestrichelten Linien kennzeichnen den Start bestimmter Maßnahmen am 9. März, 16. März und 23. März, siehe Tab. 1. Die obere Abbildung zeigt den Verlauf der neuen Fälle, die untere die kumulative Fallzahl. Dargestellt bis zum Tag des letzten Eingangsdatums.

Das Nowcasting erstellt eine Schätzung des Verlaufs der Anzahl von SARS-CoV-2 Erkrankungsfällen in Deutschland unter Berücksichtigung des Diagnose- und Meldeverzugs, siehe Abb. 1. Der Verlauf der Kurve der gemeldeten Fälle (grün) liegt deutlich unter dem geschätzten Verlauf der bereits erkrankten Fälle (rot) in Deutschland, die bei einem Erkrankungsbeginn vor wenigen Tagen, eben erst zu einem kleinen Teil diagnostiziert, gemeldet und übermittelt wurden. Im Median dauert es 7 Tage vom Erkrankungsbeginn einer Person bis zur Übermittlung der Meldung des bestätigten Falles an das Robert Koch-Institut, das heißt nach 7 Tagen sind 50% der Fälle gemeldet. Bei der Hälfte der übermittelten Fälle lag diese Dauer zwischen 5 und 10 Tagen.

Der korrigierte Verlauf der neuen Fälle steigt zunächst stetig an, stabilisiert sich aber dem 16. März auf einem Niveau von etwa 5.000 bis 6.000 neuen Fällen pro Tag. Seit dem 30. März steigt diese Anzahl möglicherweise auf etwa 7.500 neue Fälle pro Tag, siehe Abb. 1 oben.

Es wird sowohl die Unsicherheit der Adjustierung nach dem Diagnose- und Meldeverzug als auch die Unsicherheit aufgrund der teilweise fehlenden Angaben zum Erkrankungsbeginn angegeben. Aufgrund der hohen Fallzahlen fällt das Konfidenz-Intervall trotzdem relativ schmal aus und gibt nicht die gesamte Unsicherheit wider. Der Vergleich der Nowcasts über verschiedene Datenstände seit dem 20. März 2020 in Abb. 2 zeigt, dass die einzelnen Prognosen bis etwa zum 1. April untereinander relativ stimmig sind. Am 2. und 3. April zeigt das Nowcasting deutliche höhere Anzahlen, während es in der aktuellen Version zuletzt wieder abfällt. Generell ist die Prognose in den letzten 3 Tagen vor dem Datenstand instabil, da in diesem Zeitraum erst ein kleiner Teil der Fälle gemeldet wird.

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 intensivpflichtigen Fällen zeigen. In absoluten Anzahlen dominieren nach wie vor die Erwachsenen im Alter von 20 bis 79 Jahren, auf Grund des hohen Anteils an der Gesamtbevölkerung vor allem die 40-59-jährigen.

Aufbauend auf dem Nowcasting kann eine Schätzung der effektiven Reproduktionszahl durchgeführt werden. Die Reproduktionrate scheint deutlich gefallen zu sein, siehe Abb. 5. Geht man von einer relativ niedrigen Generationszeit von 4 Tagen aus, so ergab die R-Schätzung Anfang März Werte im Bereich von $R = 3$, die danach absinken, und sich etwa seit dem 20. März um $R = 1$ stabilisieren. Passend zur Entwicklung der Anzahl von Neuerkrankungen steigt auch die R-Schätzung nach dem 30. März bis auf etwa 1.6 an.

Schaut man den Verlauf der Reproduktionszahl getrennt nach Alter und Geschlecht an, zeigt sich, dass in der Altersgruppe (80+) die Werte meist im Bereich $R = 1,5$ bis 2 liegen, bei den Frauen noch etwas höher als bei den Männern. Auch bei den Kindern scheint es einen stärkeren Anstieg zu geben. In allen Gruppen werden zuletzt R-Werte über 1 beobachtet. Dies könnte auch teilweise an gründlicherer Testung liegen, die einen größeren Anteil der SARS-CoV-2 Infektionen sichtbar macht.

In die Abbildungen zum Nowcast und zur R-Schätzung wird zur Orientierung das Datum des Start 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 (Tab. 1).

Tab. 1. Start von Maßnahmen zur Eindämmung der SARS-CoV-2 Epidemie in Deutschland, 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 Ausbreitung des Coronavirus
23. März	Bundesweit umfangreiches Kontaktverbot

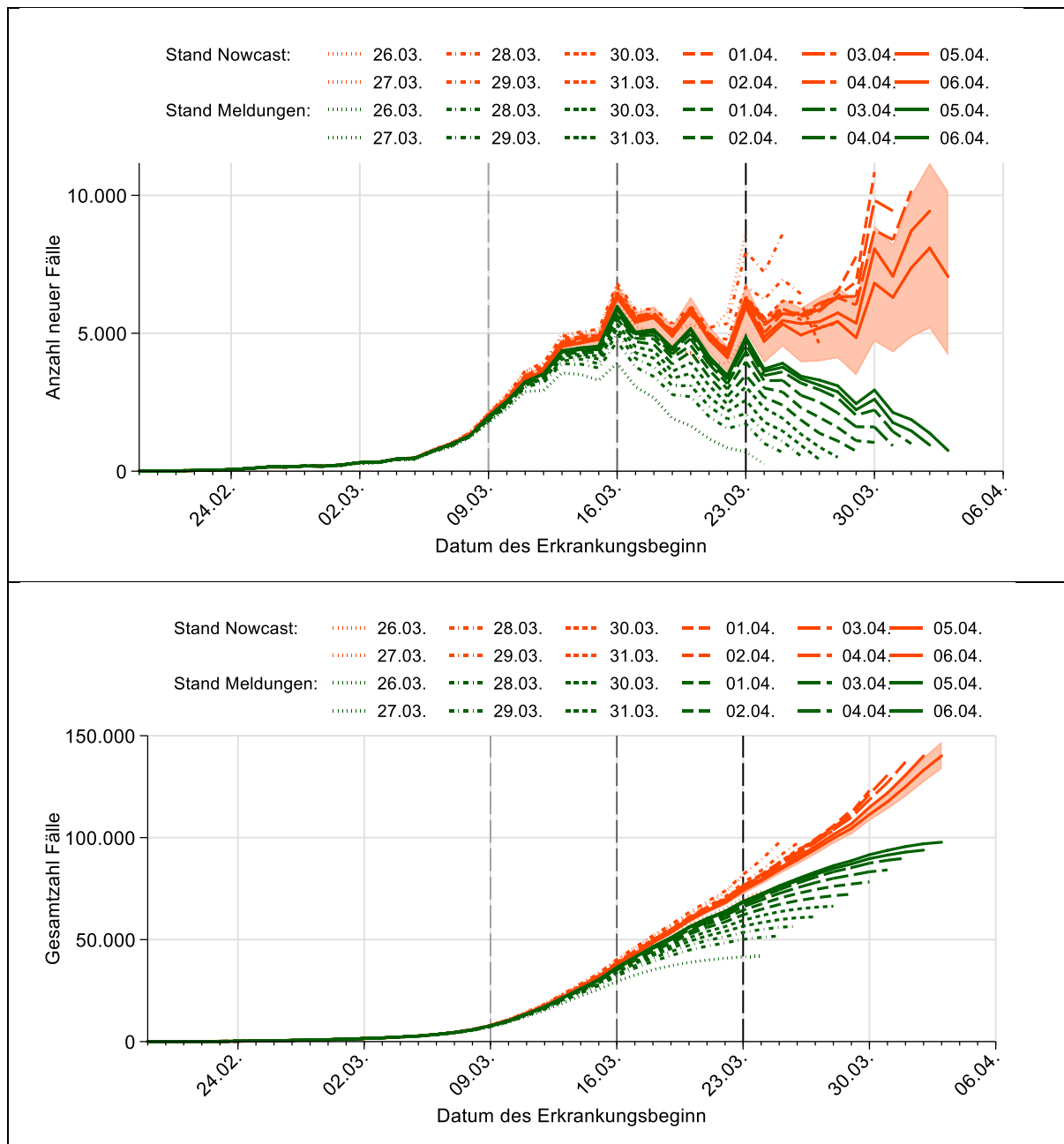


Abb. 2. Vergleich der geschätzten Entwicklung der Anzahl von SARS-CoV-2 Fällen in Deutschland (Nowcast) zu verschiedenen Datenständen, die aktuelle Schätzung mit Konfidenz-Intervall. Die gestrichelten Linien kennzeichnen den Start bestimmter Maßnahmen am 9. März, 16. März und 23. März, siehe Tab. 1.

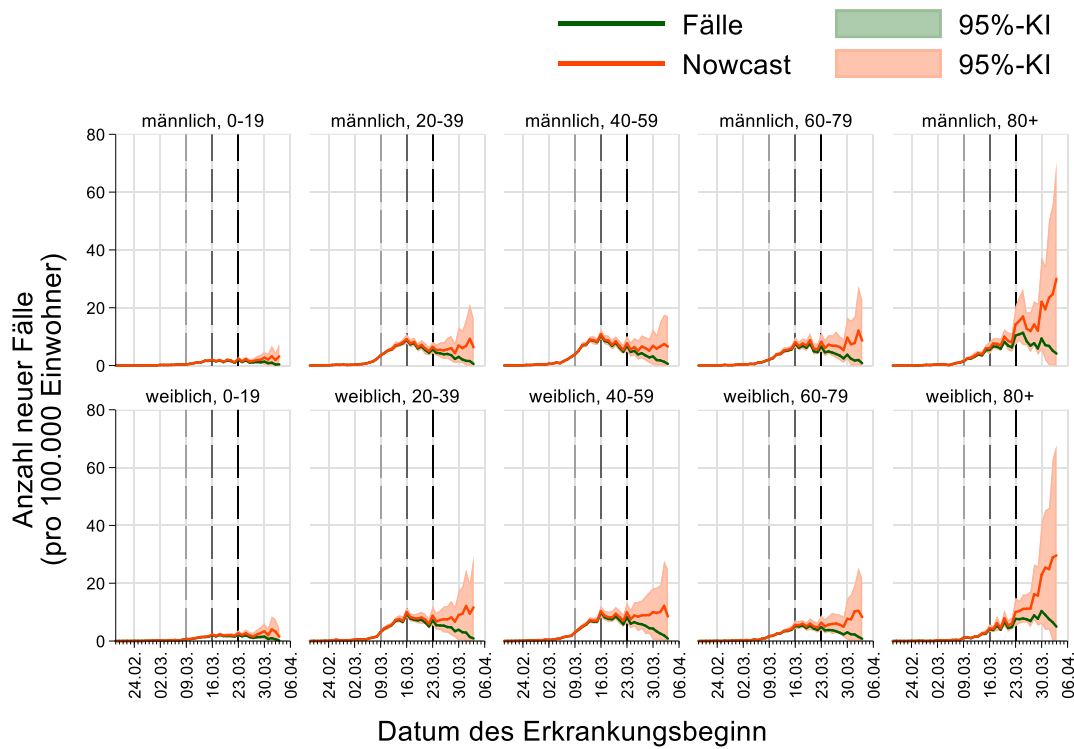


Abb. 3. Geschätzte Entwicklung der Anzahl von gemeldeten SARS-CoV-2 Fällen pro 100.000 in Deutschland nach teilweise imputierten Datum des Erkrankungsbeginns und für Meldeverzug adjustierte Kurve (Nowcast) nach Geschlecht und Altersgruppe. Die gestrichelten Linien kennzeichnen den Start bestimmter Maßnahmen am 9. März, 16. März und 23. März, siehe Tab. 1.

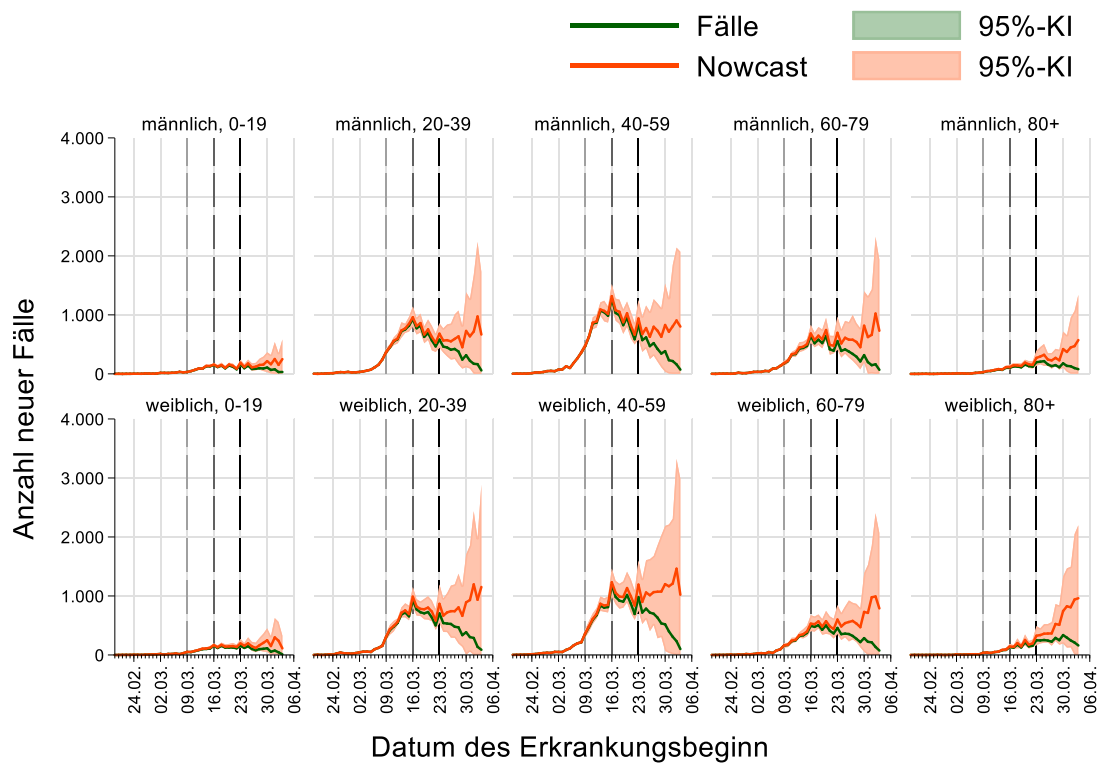


Abb. 4. Geschätzte Entwicklung der Anzahl von gemeldeten SARS-CoV-2 Fällen in Deutschland nach teilweise imputiertem Datum des Erkrankungsbeginns und für Meldeverzug adjustierte Kurve (Nowcast) nach Geschlecht und Altersgruppe. Die gestrichelten Linien kennzeichnen den Start von bevölkerungsbezogenen Maßnahmen am 9. März, 16. März und 23. März, siehe Tab. 1.

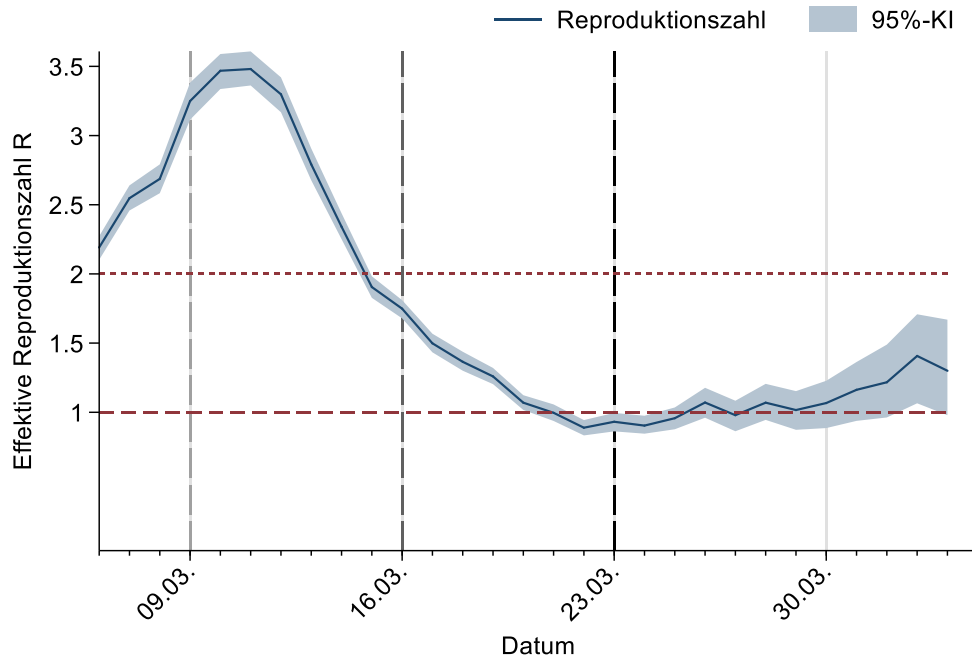


Abb. 5. Schätzung der effektiven Reproduktionszahl R für eine angenommene Generationszeit von 4 Tagen.

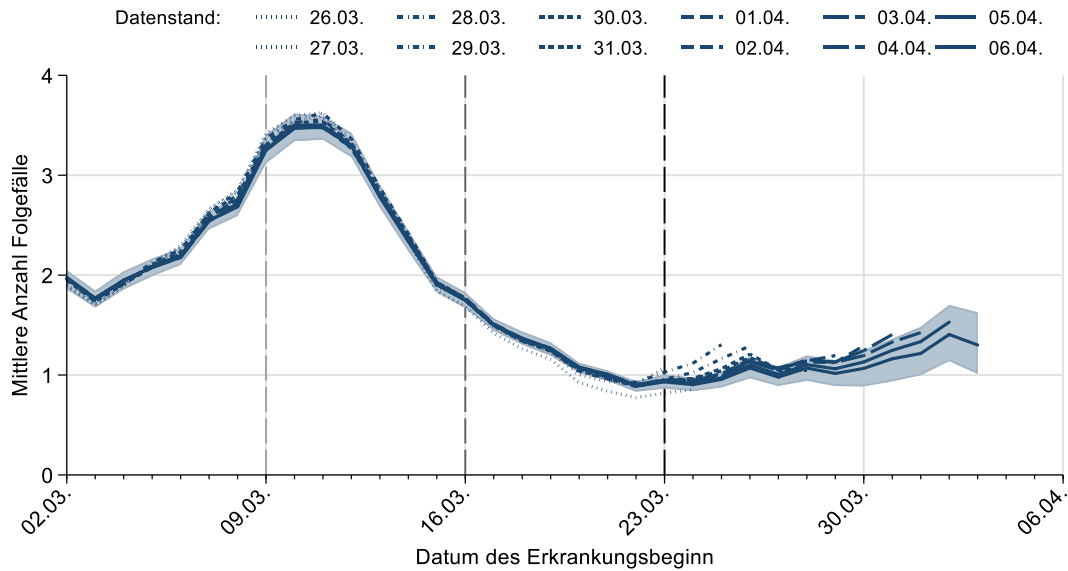


Abb. 5b. Vergleich der Schätzung der effektiven Reproduktionszahl R für eine angenommene Generationszeit von 4 Tagen zu unterschiedlichen Datenständen.

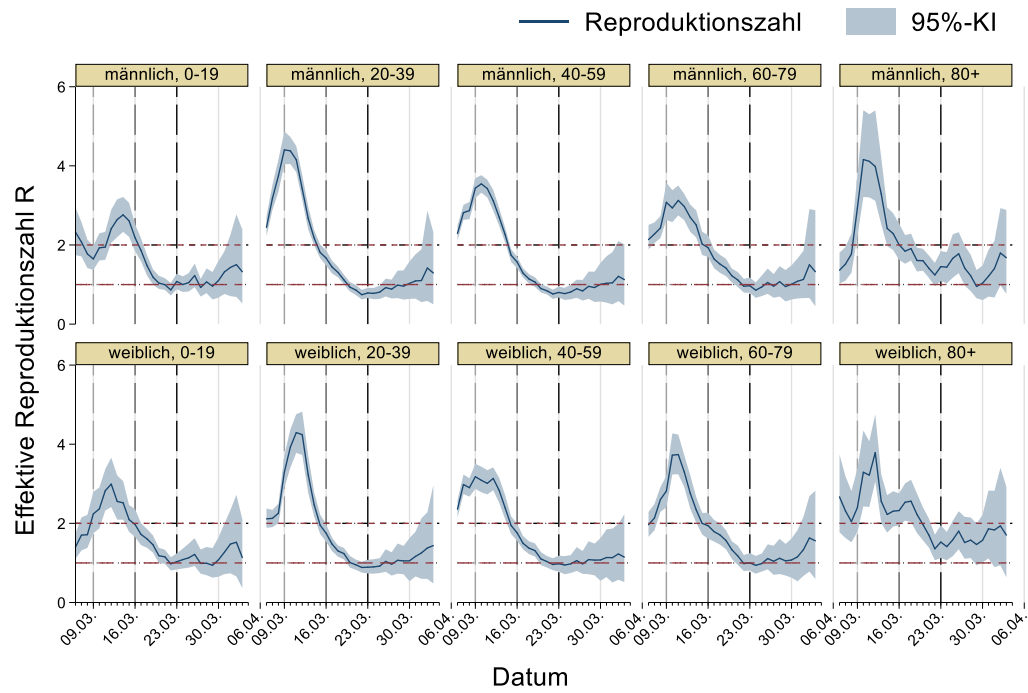


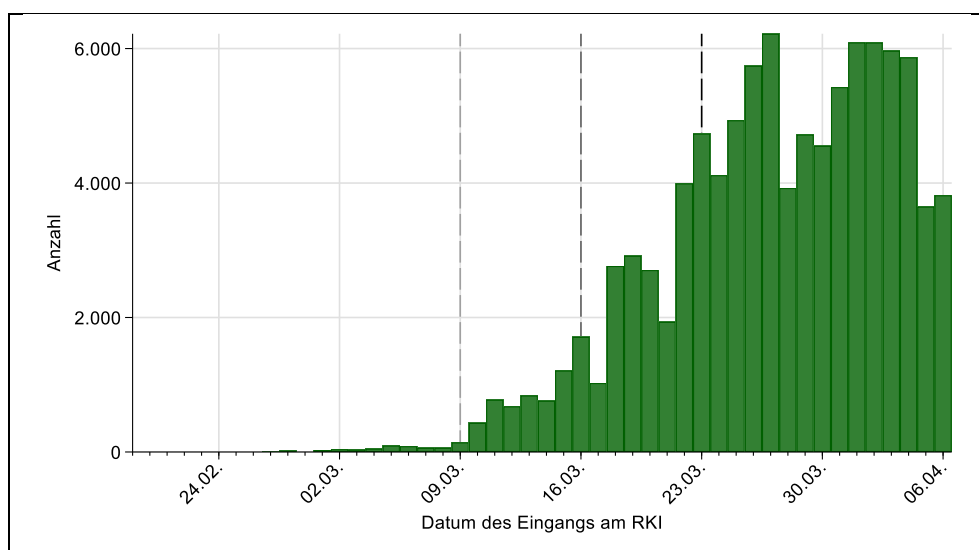
Abb. 6. Schätzung der effektiven Reproduktionszahl R für verschiedene angenommene Generationszeiten nach Geschlecht und Altersgruppe.

Erläuterung des Nowcastings

Über die Meldepflicht nach IfSG werden Infektionen mit SARS-CoV-2 in Deutschland von den Ärzten und Laboren an die Gesundheitsämtern gemeldet und von diesen über die zuständigen Landesbehörden an das Robert Koch-Institut übermittelt. In der folgenden Darstellung werden nur bestätigte Fälle aufgenommen, die vom Robert Koch-Institut validiert wurden.

Zum aktuellen Datenstand (07.04.2020, 10:00 Uhr) lagen 99.104 SARS-CoV-2 Fälle vor. Darunter waren 49.985 Männer und 49.742 Frauen. Bei 377 weiteren Fällen war das Geschlecht entweder divers, nicht erhoben oder nicht bekannt. Der Altersmedian lag bei 49 Jahren (Interquartilsabstand [IQR]: 33–61 Jahre), zu 203 Fällen lagen keine Altersangaben vor. Trägt man die Fälle nach dem Datum des Eingangs am RKI auf so ergeben sich die Kurven in Abb. 7. Insgesamt stabilisiert sich die übermittelte Anzahl von Fällen bei etwa 5.000 bis 6.000 pro Tag, ist in der Tendenz aber immer noch leicht ansteigend. Eine getrennte Darstellung dieser Entwicklung nach Geschlecht und Altersgruppen zeigt eine deutliche ansteigende Anzahl von neuen Erkrankungsfällen in der Altersgruppe (80+), aber auch auf niedrigerem Niveau bei den Kindern, siehe Abb. 8.

Diese Entwicklung kann auch direkt für eine Schätzung der zeitabhängigen Reproduktionszahl R verwendet werden, siehe Abb. 9 und 10. Hier zeigen sich bei Annahme einer Generationszeit von 4 Tagen im Vergleich zur Schätzung basierend auf dem Nowcasting deutlich höhere Werte von R besonders in der Woche zwischen dem 9. und 16. März, die auch langsamer abfallen und um den 30. März herum den Wert $R = 1$ erreichen. Danach zeigt sich wieder ein leichter Anstieg. Analysiert man diese Entwicklung getrennt nach Geschlecht und Altersgruppe zeigen sich zuletzt wieder höhere R -Werte in der Altersgruppe (80+).



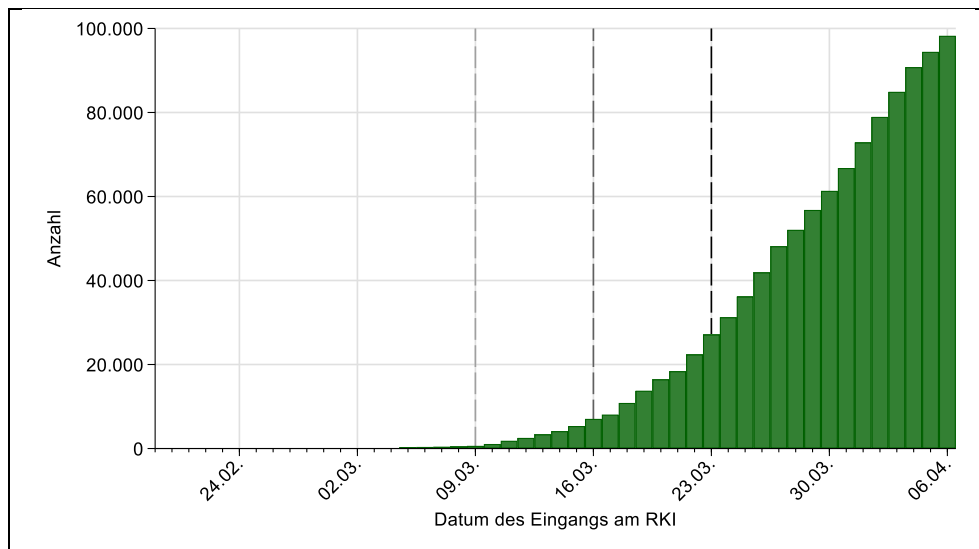


Abb. 7. Entwicklung der Anzahl von gemeldeten SARS-CoV-2 Fällen in Deutschland nach Datum der Übermittlung an das Robert Koch-Institut. Die gestrichelten Linien kennzeichnen den Start der in Tab. 1 genannten Maßnahmen am 9. März, 16. März und 23. März 2020. Die Abbildung oben zeigt die Entwicklung der neuen Fälle, unten wird die kumulative Anzahl von Fällen dargestellt.

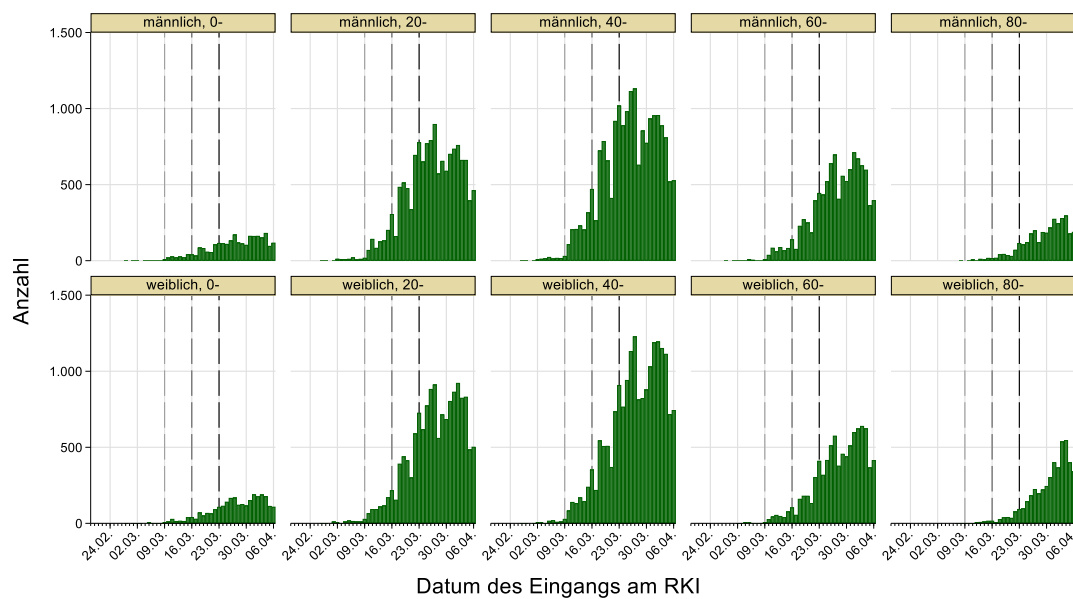


Abb. 8. Entwicklung der Anzahl von neu gemeldeten SARS-CoV-2 Fällen in Deutschland nach Datum der Übermittlung an das Robert Koch-Institut, getrennt nach Geschlecht und Altersgruppen. Die gestrichelten Linien kennzeichnen den Start der in Tab. 1 genannten Maßnahmen am 9. März, 16. März und 23. März 2020.

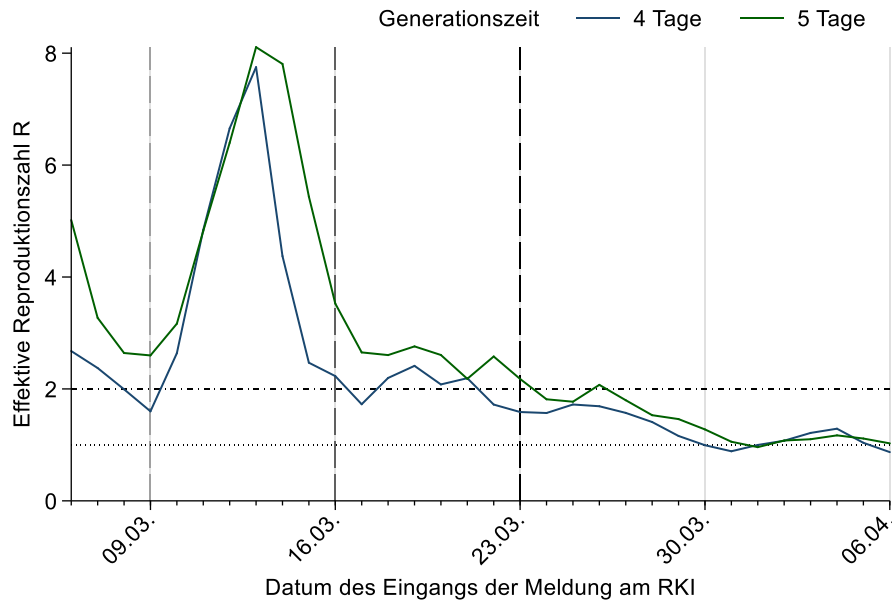


Abb. 9. Schätzung der effektiven Reproduktionszahl R für verschiedene angenommene Generationszeiten aufgrund der Zeitreihe der an das Robert Koch-Institut übermittelten Fälle.

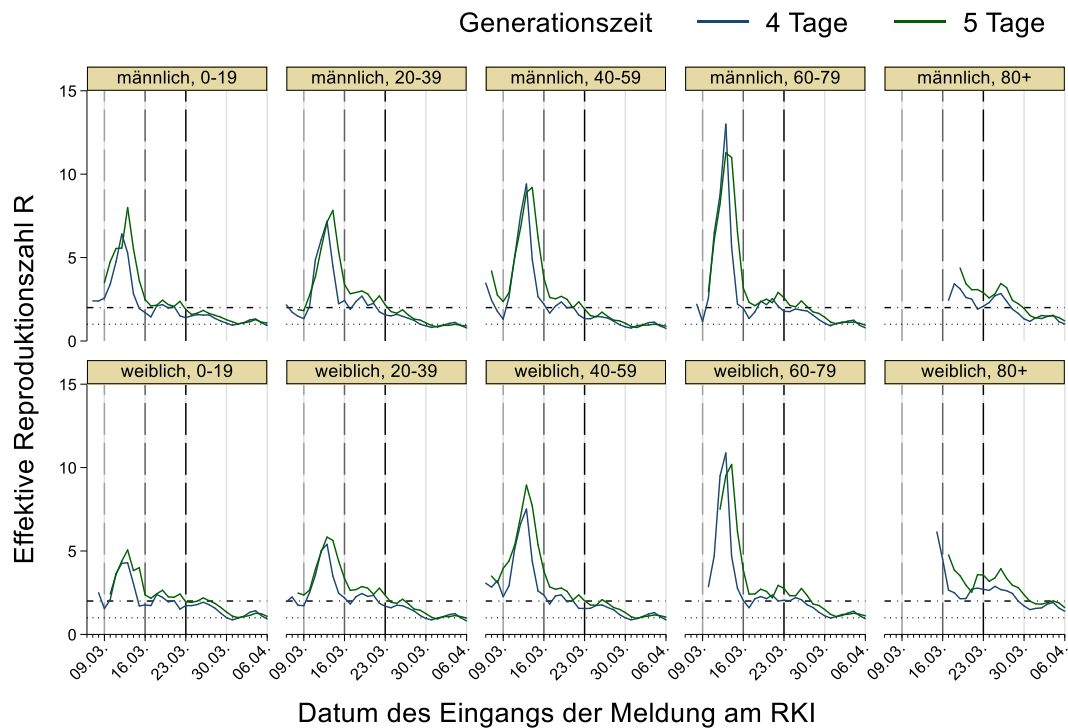


Abb. 10. Schätzung der effektiven Reproduktionszahl R für verschiedene angenommene Generationszeiten aufgrund der Zeitreihe der an das Robert Koch-Institut übermittelten Fälle nach Geschlecht und Altersgruppe.

Um diese Entwicklung besser zu verstehen erfolgt zunächst eine Darstellung dieser Entwicklung nach dem Erkrankungsbeginn. Der Erkrankungsbeginn war in 61.067 Fällen angegeben, in 38.037 Fällen war er nicht 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 470 Fälle war der zeitliche Abstand zwischen dem Datum der Übermittlung an das Robert Koch-Institut und dem Erkrankungsbeginn negativ oder lag bei über 30 Tagen¹. 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, bei der die fehlenden Datenwerte aufgrund der statistischen Zusammenhänge der bekannten Daten geschätzt werden. Als wichtigste Angabe zur Ermittlung des Erkrankungsbeginns wurde dabei das Meldedatum verwendet, die fehlenden Werte wurden getrennt nach Geschlecht und Altersgruppe geschätzt. Eine Analyse der zeitlichen Dynamik dieser Verteilung nach dem Tag des Eingangs am RKI ergibt eine kontinuierliche Verschiebung dieser Verteilung zu längeren Dauern, siehe Abb. 8. Der Mittelwert der Verzögerung stieg seit dem 12. März von 5,3 Tagen auf 6,6 Tage am 21. März. 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 8 Tagen. Diese Verschiebung wird bei der Imputation der fehlenden Werte des Erkrankungsbeginns berücksichtigt.

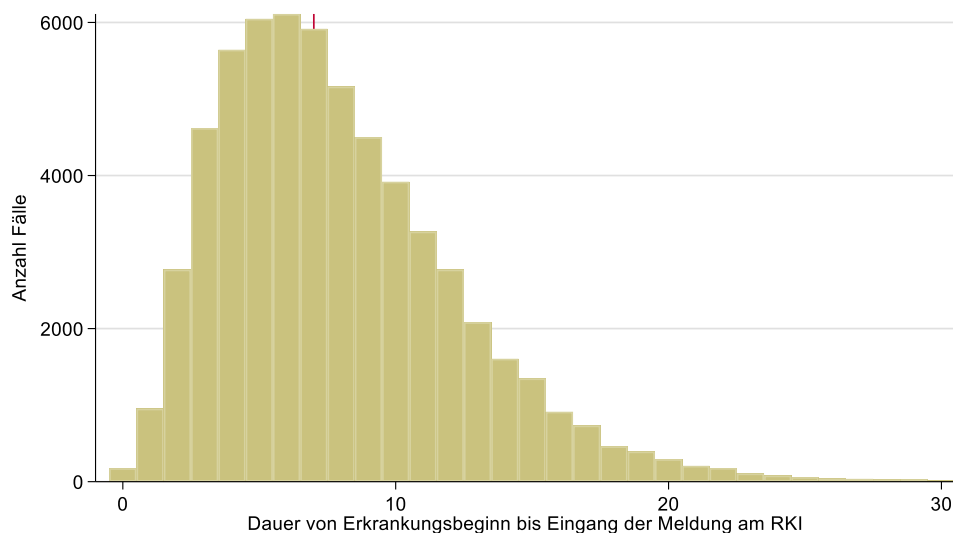


Abb. 11. Verteilung des zeitlichen Abstands zwischen Eingang der Meldung am RKI und Erkrankungsbeginn. Die rote Linie kennzeichnet den Median der Verteilung, der bei 7 Tagen liegt, etwa 50% der Werte liegen im Bereich zwischen 5 und 10 Tagen.

¹ Anmerkung: Eine Dauer von 0 Tagen oder sogar kleine negative Dauern lassen sich durch Fälle erklären, die im Rahmen einer Kontaktpersonen-Nachverfolgen eines bestätigten Falles getestet wurden und erst nach dem positiven Test Symptome entwickelt haben.

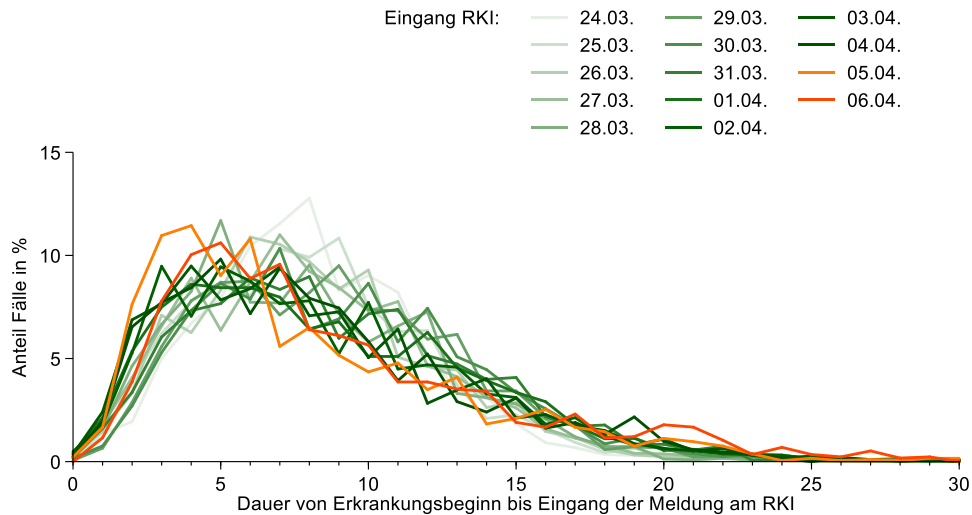


Abb. 12. Verteilung des zeitlichen Verzugs zwischen Erkrankungsbeginn und Übermittlung der Meldung an das Robert Koch-Institut getrennt nach dem Datum der Übermittlung an das Robert Koch-Institut.

Zur Durchführung der multiplen Imputation wurden (getrennt nach Geschlecht und Altersgruppe) jeweils 200 Realisation aus der empirischen Verteilung der Dauer zwischen Erkrankungsbeginn und Meldedatum den Fällen ohne Erkrankungsbeginn zugeordnet. Die Differenz zwischen dem Meldedatum und diesem Abstand ergibt dann die verschiedenen Realisierungen des simulierten Erkrankungsbeginns. Damit erhalten wir eine Entwicklung der gemeldeten Fälle nach Erkrankungsbeginn.

Zur Schätzung der Anzahl von bereits an SARS-CoV-2 erkrankten Personen ermitteln wir nun den Anteil an Fällen, der nach einer bestimmten Anzahl x von Tagen typischerweise gemeldet wurde, zur Korrektur der Anzahl gemeldeter Fälle, die vor x Tagen an das Robert Koch-Institut gemeldet wurden. Dabei muss allerdings beachtet werden, dass es durch die Systematik der Meldung automatisch zu einer verzerrten Schätzung dieses Anteils kommt, wenn alle vorhandenen Meldungen gleichmäßig in die Schätzung miteinbezogen würden. Das liegt daran, dass unter den rezenten Erkrankungen nur diejenigen mit einer schnellen Meldung zum Zeitpunkt der Analyse bereits übermittelt wurden und somit kurze zeitliche Abstände überrepräsentiert sind. Um dieser Verzerrung zu entgehen wird eine Korrektur-Methode von Lawless, 1994 (1) verwendet, die den Verzug auf Grundlage von bedingten Wahrscheinlichkeit schätzt. Dabei wird die vereinfachende Annahme getroffen, dass die Meldeverzugsverteilung über die Zeit konstant ist. Abbildung 9 evaluiert diese Annahme und zeigt, dass die geschätzte Verteilung des Anteils über die Zeit sich zu einer vollständigeren Übermittlung nach einer gegebenen Anzahl von Tagen hin entwickelt. Dies betrifft vor allem die Fälle mit Dauern von länger als 10 Tagen.

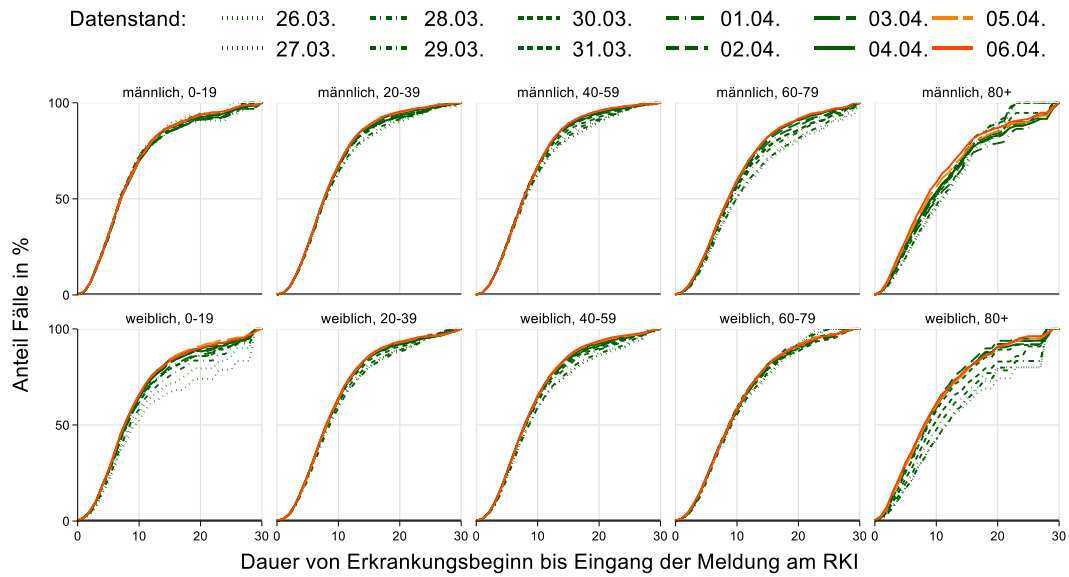


Abb. 13. Verteilung des Anteils von Meldungen, die bereits an das Robert Koch-Institut übermittelt wurden nach Datum des Nowcasts, getrennt nach Geschlecht und Altersgruppe.

Literatur

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