

SARS-CoV-2-Nowcasting und -R-Schätzung

[Robert Koch-Institut](#) | RKI

Nordufer 20
13353 Berlin

[Matthias an der Heiden](#)

FG 34 | HIV/AIDS und andere sexuell oder durch Blut übertragbare Infektionen

Beitragende:

FG 32 | Surveillance und elektronisches Melde- und Informationssystem (DEMIS) | ÖGD
Kontaktstelle

[Justus Benzler](#) (Datenkuration)

MF 4 | Fach- und Forschungsdatenmanagement

[Hannes Wuensche](#) (Datenkuration)

Zitieren

an der Heiden, Matthias (2023): SARS-CoV-2-Nowcasting und -R-Schätzung, Berlin: Zenodo.
DOI:[10.5281/zenodo.7949551](https://doi.org/10.5281/zenodo.7949551).

Informationen zur Studie und zum Forschungskontext

Das SARS-CoV-2-Nowcasting erstellt eine Schätzung des Verlaufs der Anzahl von bereits erfolgten SARS-CoV-2-Erkrankungsfällen in Deutschland unter Berücksichtigung des Diagnose-, Melde- und Übermittlungsverzugs. Aufbauend auf dem Nowcasting kann eine Schätzung der zeitabhängigen Reproduktionszahl R (oder R -Wert) durchgeführt werden. Die Reproduktionszahl beschreibt, wie viele Menschen eine infizierte Person im Mittel ansteckt. Sie kann nicht alleine als Maß für die Notwendigkeit und Wirksamkeit von Maßnahmen herangezogen werden. Wichtig sind außerdem u.a. die absolute Zahl der täglichen Neuinfektionen sowie die Schwere der Erkrankungen. Die absolute Zahl der Neuinfektionen muss klein genug sein, um eine effektive Kontaktpersonennachverfolgung zu ermöglichen, und die Zahl der schweren Erkrankungen klein genug, um die Kapazitäten an Intensivbetten nicht zu überlasten.

Administrative und organisatorische Angaben

Der Datensatz "SARS-CoV-2-Nowcasting und -R-Schätzung" wird vom Robert Koch-Institut im Zusammenhang mit der SARS-CoV-2-Pandemie bereitgestellt. Autor des Datensatzes ist [Matthias an der Heiden](#), wissenschaftlicher Mitarbeiter des [Fachgebiet 34 | HIV/AIDS und](#)

[andere sexuell oder durch Blut übertragbare Infektionen](#) des RKI. Inhaltliche Fragen bezüglich des Nowcastings und der R-Schätzungen können an das RKI unter info@rki.de gestellt werden. Das Nowcasting und die R-Schätzung erfolgt seit März 2020 und wird täglich aktualisiert.

Die Datenkuration sowie das Qualitätsmanagement der (Meta-)Daten erfolgt durch [Justus Benzler, Fachgebiet 32 | FG 32 | Surveillance und elektronisches Melde- und Informationssystem \(DEMIS\) | ÖGD Kontaktstelle](#) und [Hannes Wuensche, Fachgebiet MF 4 | Fach- und Forschungsdatenmanagement](#) des RKI. Fragen zum Datenmanagement und zur Publikationsinfrastruktur können an das Open Data Team des Fachgebiets MF4 unter OpenData@rki.de gerichtet werden.

Inhalt und Aufbau des Datensatzes

Der Datensatz enthält die epidemiologischen Schätzungen über den Verlauf der SARS-CoV-2-Infektionen in Deutschland und der daraus resultierenden Reproduktionszahl. Weiterhin enthält er grundlegende Metadaten sowie die Datensatzdokumentation und die in ihr zitierten wissenschaftlichen Publikationen. Im Datensatz enthalten sind:

- Tabelle mit täglichen Fallzahlenschätzungen und R-Wert-Angaben
- Archiv mit der Sammlung bisheriger Fallzahlenschätzungen und R-Wert-Angaben
- Lizenz-Datei mit der Nutzungslizenz des Datensatzes in Deutsch und Englisch
- Datensatzdokumentation und auf den Datensatz bezogene Publikationen in deutscher Sprache
- Metadaten-Datei zum Import in Zenodo

Aufbereitung und Auswertung der Daten

Es besteht ein großes Interesse daran, das aktuelle Infektionsgeschehen und die zeitliche Entwicklung der Zahlen von SARS-CoV-2-Infektionen und Covid-19-Erkrankungsfällen in Deutschland zeitnah darzustellen und zu verstehen. Aufgrund unvermeidbarer Verzögerungen 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, dass nicht alle, die Symptome entwickeln, eine Arztpraxis aufsuchen, dass nicht alle, die zum Arzt gehen, getestet werden, und dass nicht alle, die positiv getestet werden, auch in einem Erhebungssystem erfasst werden. 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 diagnostizierenden Ärzten und Laboren an die zuständigen Gesundheitsämter gemeldet und von diesen über die zuständigen Landesbehörden an das Robert Koch-Institut übermittelt.

Methoden, Instrumente und Verlauf der Datengenerierung

Grundlage der Berechnungen des Nowcastings und der R-Schätzungen sind die aktuellen, durch die Gesundheitsämter an das Robert Koch-Institut gemeldeten, SARS-CoV-2-Infektionen in Deutschland. Die öffentlich zur Verfügung gestellten Daten der Infektionsmeldungen sind im [SARS-CoV-2 Daten-Dashboard des RKI](#), auf [GitHub](#) und in [Zenodo](#) abrufbar. Auf Basis der gemeldeten SARS-CoV-2-Infektionen erfolgen die Fallzahlschätzungen und die R-Wert-Berechnung in folgenden Schritten:

1. Multiple Imputation fehlender Information zum Erkrankungsbeginn von COVID-19-Fällen unter einer Missing-at-Random-Annahme
2. Korrektur der Anzahl von Neuerkrankungen für den Diagnose-, Melde- und Übermittlungsverzug mittels des Nowcasting-Verfahrens
3. Berechnung der zeitlich variierenden Reproduktionszahl unter der Annahme einer Generationszeit von 4 Tagen

Eine detaillierte Beschreibung der Methoden, Instrumente und des Verlaufs der Datengenerierung findet sich in folgenden Publikationen, die ebenfalls im Datensatz enthalten sind:

an der Heiden, M; Hamouda, O (2020): Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting. *Epid Bull*, S.10–16. DOI: [10.25646/6692.4](https://doi.org/10.25646/6692.4)

Robert Koch-Institut (2020): Erläuterung der Schätzung der zeitlich variierenden Reproduktionszahl R. RKI. DOI:[10.25646/8164](https://doi.org/10.25646/8164)

Daten und Datenaufbereitung

Zentrales Datum des Datensatzes ist die Tabelle mit dem Nowcasting der aktuellen Infektionen und der sich daraus ergebenden R-Werte. Diese sind im Hauptverzeichnis unter "Nowcast_R_aktuell.csv" abrufbar und werden täglich überschrieben. Im Archivordner sind das Nowcasting und die R-Schätzung unter den Dateinamen "Nowcast_R_YYYY-MM-TT.csv" enthalten. Im Dateinamen repräsentiert die Sequenz "YYYY-MM-TT" das Erstellungsdatum der Datei und damit gleichzeitig das Datum des enthaltenen Datenstands. "YYYY" steht dabei für das Jahr, "MM" für den Monat und "TT" für den Tag der Erstellung bzw. des enthaltenen Datenstands.

Nowcast_R_aktuell.csv
Archiv/Nowcast_R_YYYY-MM-TT.csv

Die Tabelle des Nowcastings und der R-Schätzung wird jeden Tag um die Zeile der Daten des vergangenen Tages erweitert. Die Fallzahlschätzungen und R-Wert-Angaben bilden einen tagesaktuellen Stand (00:00 Uhr) ab.

Die Schätzwerte zur Anzahl von Neuerkrankungen und der R-Schätzung zu früheren Tagen können von den Angaben in früheren Versionen der Tabelle abweichen, weil täglich der Gesamtverlauf, aufgrund der aktuell verfügbaren Daten, neu geschätzt wird.

Formatierung der Nowcasting-und-R-Schätzung-Tabelle

Das Nowcasting und die R-Schätzung sind im Datensatz als kommaseparierte .csv-Datei enthalten. Der verwendete Zeichensatz der .csv-Datei ist UTF-8. Trennzeichen der einzelnen Werte ist ein Komma ",". Datumsangaben sind im ISO-8601-Standard formatiert.

- Zeichensatz: UTF-8
- Datumsformat: ISO 8601
- .csv-Trennzeichen: Komma ","

Variablen des Nowcastings und der R-Schätzung

Drei zentrale Variablen sind im Nowcasting und in der R-Schätzung für jeden Tag JJJJ-MM-TT abgebildet:

- Punktschätzer der Anzahl an Neuerkrankungen (ohne Glättung)
 - ohne Glättung: ohne Bildung eines gleitenden Mittelwerts
- Punktschätzer der Anzahl an Neuerkrankungen (mit Glättung)
 - mit Glättung: unter Bildung eines gleitenden Mittelwerts über 4 Tage
 - jeder Wert mit den Werten der 3 vorhergehenden Tage gemittelt
- Punktschätzer des 7-Tage-R-Werts

Zu jeder dieser Variablen ist darüber hinaus das 95%-Prädiktionsintervall mit einer Ober- und einer Untergrenze angegeben. In der folgenden Tabelle sind die sich daraus ergebenden Variablen und deren Ausprägungen angegeben:

Merkmal	Ausprägung	Erläuterung
Datum	JJJJ-MM-TT	Datum der geschätzten Neuinfektionen und der sich ergebenden R-Werte. JJJJ entspricht der Jahreszahl, MM dem Monat und TT dem Tag.
PS_COVID_Faelle	Natürliche Zahl	Punktschätzer der Anzahl an Neuerkrankungen (ohne Glättung)
UG_PI_COVID_Faelle	Natürliche Zahl	Untere Grenze des 95%-Prädiktionsintervalls der Anzahl an Neuerkrankungen (ohne Glättung)
OG_PI_COVID_Faelle	Natürliche Zahl	Obere Grenze des 95%-Prädiktionsintervalls der Anzahl an Neuerkrankungen (ohne Glättung)
PS_COVID_Faelle_ma4	Natürliche Zahl	Punktschätzer der Anzahl an Neuerkrankungen (mit Glättung)
UG_PI_COVID_Faelle_ma4	Natürliche Zahl	Untere Grenze des 95%-Prädiktionsintervalls der Anzahl an Neuerkrankungen (mit Glättung)
OG_PI_COVID_Faelle_ma4	Natürliche Zahl	Obere Grenze des 95%-Prädiktionsintervalls der Anzahl an Neuerkrankungen (mit Glättung)
PS_7_Tage_R_Wert	Rationale Zahl	Punktschätzer des 7-Tage-R-Werts
UG_PI_7_Tage_R_Wert	Rationale Zahl	Untere Grenze des 95%-Prädiktionsintervalls des 7-Tage-R-Werts
OG_PI_7_Tage_R_Wert	Rationale Zahl	Obere Grenze des 95%-Prädiktionsintervalls des 7-Tage-R-Werts

Hinweise zur Nachnutzung der Daten

Offene Forschungsdaten des RKI werden auf GitHub.com sowie Zenodo.org bereitgestellt:

- <https://github.com/robert-koch-institut>
- <https://zenodo.org/communities/robertkochinstitut>

Metadaten

Die bereitgestellten Daten sind mit Metadaten beschrieben und wissenschaftlich zitierbar, u.a. durch die Vergabe einer DOI durch Zenodo.org. Die für den Import in Zenodo bereitgestellten Metadaten sind in folgender Datei hinterlegt:

Metadaten/zenodo.json

Die Dokumentation der einzelnen Metadatenvariablen ist unter <https://developers.zenodo.org/#representation> nachlesbar.

Lizenz

Der Datensatz "SARS-CoV-2-Nowcasting und -R-Schätzung" ist lizenziert unter der [Creative Commons Namensnennung 4.0 International Public License | CC-BY 4.0 International](#).

Die im Datensatz bereitgestellten Daten sind, unter Bedingung der Namensnennung des Autors als Quelle, frei verfügbar. Das bedeutet, jede_r hat das Recht die Daten zu verarbeiten und zu verändern, Derivate des Datensatzes zu erstellen und sie für kommerzielle und nicht-kommerzielle Zwecke zu nutzen. Weitere Informationen zur Lizenz finden sich in der LICENSE- bzw. LIZENZ-Datei des Datensatzes.