

SARS-CoV-2 Infektionen in Deutschland

ÄNDERE!!! aaaa

[Robert Koch-Institut](#) | RKI

Nordufer 20

13353 Berlin

FG 32 | Surveillance

[Michaela Diercke](#) (Leitung)

FG 31 | Infektionsepidemiologische Fach-IT und Anwendungsentwicklung

[Herrmann Claus](#) (Leitung)

MF 4 | Forschungsdatenmanagement

[Linus Grabenhenrich](#) (Leitung)

[Hannes Wuensche](#) (Datenkuration)

DOI to replace

LatestDOI

10.5072/zenodo.1165354

Some older DOI

10.5072/zenodo.1165354

Some random DOI

10.5074/zenodo.9951513

Robert Koch-Institut (2021): SARS-CoV-2 Infektionen in Deutschland, Berlin: Zenodo.

DOI:[10.5281/zenodo.4681153](#).

Der Datensatz "SARS-CoV-2 Infektionen in Deutschland" ist lizenziert unter der [Creative Commons Namensnennung 4.0 International Public License](#) | [CC-BY 4.0 International](#)

Informationen zum Datensatz und Entstehungskontext

HELLO! Knut Super Branch coming soon!!!!
IM FRESH! PLZ



Administrative und organisatorische Angaben

Im Datensatz "SARS-CoV-2 Infektionen in Deutschland" werden die tagesaktuellen Fallzahlen, die nach den Vorgaben des [Infektionsschutzgesetz - IfSG](#) von den Gesundheitsämtern in Deutschland gemeldeten positiven SARS-CoV-2 Infektionen, Todes- und Genesungsfälle bereitgestellt.

Die zugrundeliegenden Daten werden an das Robert Koch-Institut (RKI) über das Meldesystem gemäß IfSG übermittelt. Zuständig für den Betrieb des Meldesystems ist das [Fachgebiet 32 | Surveillance](#) des RKI.

Die Verarbeitung und Aufbereitung der im Meldesystem vorliegenden Rohdaten erfolgt durch das [Fachgebiet 31 | Infektionsepidemiologische Fach-IT und Anwendungsentwicklung](#) des RKI.

Die Veröffentlichung der Daten, die Datenkuration sowie das Qualitätsmanagement der (Meta-)Daten erfolgt durch das Fachgebiet [MF 4 | Forschungsdatenmanagement](#). Die Datensatzkuration und Betreuung Publikationsinfrastruktur erfolgt durch Hannes Wuensche. Fragen zum Datenmanagement können an das Fachgebiet MF4 gerichtet werden (MF4@rki.de).

Inhalt und Aufbau des Datensatzes

Der Datensatz enthält epidemiologische Daten über den Verlauf der SARS-CoV-2 Infektionen in Deutschland. Im Datensatz enthalten sind:

- Fallzahlendaten mit tagesaktuellen Meldungen von SARS-CoV-2 Infektionen
- Geodaten mit Geo- und Zensusinformationen der erhobenen Landkreise
- Archiv mit der Sammlung aller bisherigen Fallzahlentabellen
- Lizenz Datei mit der Nutzungslizenz des Datensatzes
- Datensatzdokumentation in deutscher Sprache
- Metadaten Datei zum Import in Zenodo

Aufbereitung der Daten

Daten und Datenaufbereitung

Die Fallzahlendaten bilden einen tagesaktuellen Stand (00:00 Uhr) aller bisherig gemeldeten Infektionsfälle in Deutschland ab. Das bedeutet, dass alle, bis 00:00 Uhr des Tages JJJJ-MM-TT, von den Gesundheitsämtern, über die zuständigen Landesbehörden, an das Meldesystem des RKIs übermittelten SARS-CoV-2 Infektionen im Datenstand enthalten sind. Die Daten werden täglich vollständig neu erzeugt und dieser Datenstand ersetzt den Datenstand des Vortages.

Die Fallzahlendaten enthalten als einzige Geoinformation die Landkreis ID. Diese richtet sich nach dem Amtlichen Gemeindeschlüssel (AGS) des Quartal 2 2020, abgerufen im [Portal des Statistischen Bundesamtes](#). Für eine genauere Darstellung des Landkreises Berlin, werden die 12 Stadtbezirke als eigene "Landkreise" aufgegliedert. Hier wird von den Vorgaben des AGS abgewichen. Für die weitere Verknüpfung mit zugehörigen Geo- und Zensusinformationen werden weitere Geodaten bereitgestellt (siehe Abs. Geodaten).

Fallzahlendaten

JJJJ-MM-TT_Deutschland_SARSCoV2_Infektionen.csv

Zentrales Datum des Datensatzes sind die aktuellen Fallzahlendaten. Diese erhalten den Dateinamen "JJJJ-MM-TT_Deutschland_SARSCoV2_Infektionen". Im Dateinamen repräsentiert die Sequenz "JJJJ-MM-TT" das Erstellungsdatum der Datei und damit gleichzeitig das Datum des enthaltenen Datenstands. "JJJJ" steht dabei für das Jahr, "MM" für den Monat und "TT" für den Tag der Erstellung bzw. des enthaltenen Datenstands.

Formatierung der Fallzahlen Daten

Die Fallzahlen Daten 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 ISO8601 Standard formatiert.

- Zeichensatz: UTF-8
- Datumsformat: ISO8601
- .csv Trennzeichen: Komma ","

Merkmale der Fallzahlen Daten

In der .csv Fallzahlentabelle differenzieren die Spalten die verschiedenen Merkmale einer Fallgruppe. Pro Zeile ist eine eindeutige Fallgruppe abgebildet. Eine Fallgruppe umfasst keine Einzelfälle. Jedoch ist es möglich, dass in der Fallgruppe nur ein Fall enthalten ist. Eine Fallgruppe wird grundlegend durch folgende Eigenschaften charakterisiert (in den Klammern finden sich die Merkmale dieser Eigenschaften):

- Ort der Infektionen (IdLandkreis)
- Personengruppe (Geschlecht, Altersgruppe)
- Meldezeitpunkt der Infektion (Meldedatum)
- Erkrankungsbeginn (Refdatum, IstErkrankungsbeginn)
- Mächtigkeit der Gruppe (AnzahlFall, AnzahlTodesfall, AnzahlGenesen)
- Meldestatus (NeuerFall, NeuerTodesfall, NeuGenesen)

Eine Fallgruppe nimmt eine eindeutige Ausprägung hinsichtlich ihrer Anzahl von Fällen ("AnzahlFall"), "Altersgruppe", "Geschlecht", ihres Landkreises ("IdLandkreis"), "Meldedatum"s, Erkrankungsdatums ("Refdatum") und der Informationen ob das Erkrankungsdatum bekannt ist "IstErkrankungsbeginn", an.

Weiterhin wird die "AnzahlTodesfall" oder "AnzahlGenesen" jeder Fallgruppe angegeben, wobei nur eines der beiden Merkmale "AnzahlTodesfall" oder "AnzahlGenesen" angenommen werden kann. Das heißt, sofern es in einer Fallgruppe Todesfälle oder Genesene gibt, werden die Anzahl der Todesfälle oder die Anzahl der genesenen Fälle in einer neuen Gruppe angegeben. Treten z. B. beide Fälle in einer Fallgruppe auf, teilt sich die Fallgruppe in zwei weitere Gruppen auf, und zwar in eine Gruppe der Todesfälle und eine Gruppe der Genesenen.

Beispiel

Es wird eine neue Fallgruppe w registriert (IdLandkreis, Geschlecht, Altersgruppe, Meldedatum, Refdatum, IstErkrankungsbeginn sind konstant). Diese enthält eine Fallgruppe zu Beginn:

Fallgruppe w: 5 Infizierte, 0 Todesfälle und 0 genesene Fälle

Sterben 1 und genesen 2 der Fälle so spaltet sich die Fallgruppe w in 3 Gruppen:

Fallgruppe x: 2 Infizierte, 0 Todesfälle und 0 genesene Fälle

Fallgruppe y: 1 Infizierte, 1 Todesfälle und 0 genesene Fälle

Fallgruppe z: 2 Infizierte, 0 Todesfälle und 2 genesene Fälle

Die Merkmale des Meldestatus geben an, ob, bezogen auf den Vortag, in einer Fallgruppe Veränderungen bei den Infektionsfällen, Todesfällen und Genesenen entstanden sind. Das ermöglicht die Veränderungen zum Vortag nachzuvollziehen. Diese entstehen durch Neumeldungen von Infektionen (inklusive Nachmeldungen), Korrekturen (z. B. durch irrtümliche Meldungen, aber auch Korrekturen bzgl. Landkreis, Alter, Geschlecht oder Erkrankungsbeginn) und Veränderung des Gesundheitszustands (genesen, verstorben). Die Ausprägungen des Meldestatus spalten Fallgruppen temporär auf. Die Aufspaltung erfolgt temporär, da sie nur die Veränderungen vom Publikationstag zum Vortag abbilden. Neue Fälle bilden für den Tag der Neumeldung eine eigene Fallgruppe. Da ein Fall nur an einem Tag neu gemeldet, neu genesen oder neu verstorben oder korrigiert wird, folgt auf die temporäre Aufspaltung der Fallgruppe am Tag der Neumeldung des Meldestatus, eine Zusammenlegung der Gruppen am Folgetag. Eine genauere Erläuterung zu diesem Prozess wird im folgenden Abschnitt gegeben.

Merkmalsausprägungen

Die Fallzahlendaten enthalten die in der folgenden Tabelle abgebildeten Merkmale und deren Ausprägungen:

Merkmalsausprägungen	Merkmalsausprägungen	Merkmalsausprägungen
Merkmalsausprägungen	Merkmalsausprägungen	Merkmalsausprägungen
IdLandkreis	1001 bis 16077	Identifikationsnummer des Landkreises basierend auf dem Amtlichen Gemeindeschlüssel (AGS) zuzüglich der 12 Bezirke Berlins (11001 bis 11012); Gebietsstand: 30.06.2020 (2. Quartal)
Geschlecht	W, M, unbekannt	Geschlecht der Fallgruppe: weiblich (W), männlich (M) und (unbekannt)

Altersgruppe	A00-04, A05-14, A15-A34, A35-A59, A60-A79, A80+, unbekannt	Altersspanne der in der Gruppe enthaltenen Fälle, stratifiziert nach 0-4 Jahren, 5-14 Jahren, 15-34 Jahren, 35-59 Jahren, 60-79 Jahren, 80+ Jahren sowie unbekannt
Meldedatum	JJJJ-MM-TT	Datum, wann der Fall dem Gesundheitsamt bekannt geworden ist. JJJJ entspricht der Jahreszahl, MM dem Monat und TT dem Tag.
Refdatum	JJJJ-MM-TT	Datum des Erkrankungsbeginns. Wenn das nicht bekannt ist, das Meldedatum.
IstErkrankungsbeginn	0, 1	1: Refdatum ist der Erkrankungsbeginn 0: Refdatum ist das Meldedatum
AnzahlFall	Natürliche Zahl	Anzahl der Fälle in der entsprechenden Fallgruppe für NeuerFall = -1, ist die Anzahl als negativ zu interpretieren. Es handelt sich um eine Korrektur der Fallgruppe. Die AnzahlFall für NeuerFall = -1 gibt an, wie viele Infektionen zu viel gemeldet worden sind
AnzahlTodesfall	Natürliche Zahl	Anzahl der Todesfälle in der entsprechenden Fallgruppe für NeuerTodesfall = -1, ist die Anzahl als negativ zu interpretieren. Es handelt sich um eine Korrektur der Fallgruppe. Die AnzahlTodesfall für NeuerTodesfall = -1 gib an, wie viele Todesfälle zu viel gemeldet worden sind
AnzahlGenesen	Natürliche Zahl	Anzahl der genesenen Fälle in der entsprechenden Fallgruppe für NeuGenesen = -1, ist die Anzahl als negativ zu interpretieren. Es handelt sich um eine Korrektur der Fallgruppe. Die AnzahlGenesen für NeuGenesen = -1 gib an, wie viele genesene Fälle zu viel gemeldet worden sind

NeuerFall, NeuerTodesfall, NeuGenesen	0, 1, -1	<p>0 : Fälle der Gruppe sind in der Publikation für den aktuellen Tag und in der für den Vortag enthalten. Das bedeutet diese Fälle sind seit mehr als einem Tag bekannt.</p> <p>1 : Fälle der Gruppe sind erstmals in der aktuellen Publikation enthalten. Das heißt, es sind für den Publikationstag neu übermittelte oder entsprechend neu bewertete Fälle.</p> <p>-1: Fälle der Gruppe sind in der Publikation des Vortags enthalten, werden jedoch nach dem aktuellen Tag aus den Fallzahlendaten entfernt. Das heißt, es sind Fälle die ab dem aktuellen Tag wegfallen. Eine solche Fallgruppe kann beispielsweise durch fälschliche Meldungen entstehen, die so als Korrektur angezeigt werden.</p>
NeuerTodesfall, NeuGenesen	-9	<p>Fälle in der Gruppe sind weder in der Publikation für den aktuellen Tag, noch in der Publikation des Vortags, als genesen ("NeuGenesen") oder verstorben ("NeuerTodesfall") gemeldet. Das bedeutet, dass zu den Fällen in der Gruppe keine Information über den Gesundheitsverlauf der Infektion bekannt ist. Das ist zum Beispiel häufig der Fall, wenn eine Fallgruppe gerade erst als infiziert gemeldet worden ist.</p>

Die temporäre Aufspaltung der Fallgruppen durch die Merkmale des Meldestatus wird im folgenden Beispiel verdeutlicht. Temporäre Gruppen sind durch ein 'g' gekennzeichnet. Neumeldungen wird bei Betrachtung der Ausprägungen der Merkmale deutlich:

Beispiel

Es wird eine neue Fallgruppe am Tag TT registriert (IdLandkreis, Geschlecht, Altersgruppe, Meldedatum, Refdatum, IstErkrankungsbeginn sind konstant), so nimmt sie den Meldestatus NeuerFall = [1] an. Sind noch keine Genesenen oder Todesfälle bekannt, sind in der Fallgruppe gemeldet, sind NeuerTodesfall und NeuGenesen = [-9]:

Die Fälle der Fallgruppe w' sind im Datensatz von Tag TT neu enthalten (NeuerFall [1]), die Fälle der Gruppe sind keine Todes- oder Genesungsfälle (NeuerTodesfall [-9], NeuGenesen [-9]).

Fallgruppe w' :

Infizierte [4], Todesfälle [0] und genesene Fälle [0]
NeuerFall [1], NeuerTodesfall [-9], NeuGenesen [-9]

Am nächsten Tag, $TT+1$ sind die Fälle aus Fallgruppe w' nicht mehr neu. Ihr Meldestatus ändert sich daher von [1] auf [0]. Die temporäre Fallgruppe w' (NeuerFall [1]) wird zur stetigen Fallgruppe w (NeuerFall [0]):

Fallgruppe w :

Infizierte [4], Todesfälle [0] und genesene Fälle [0]
NeuerFall [0], NeuerTodesfall [-9], NeuGenesen [-9]

An Tag $TT+1$ wird ein zusätzlicher, neuer Fall in der Fallgruppe w registriert. Da es sich um einen neuen Fall handelt, bildet er wieder eine temporäre, eigene Gruppe w' :

Fallgruppe w' :

Infizierte [1], Todesfälle [0] und genesene Fälle [0]
NeuerFall [1], NeuerTodesfall [-9], NeuGenesen [-9]

Am nächsten Tag, $TT+2$ sind auch die Fälle der Fallgruppe $w'(TT+1)$ nicht mehr neu, ihr Meldestatus ändert sich wie am Tag zuvor bei Fallgruppe $w'(TT)$. Durch die Änderung des Meldestatus in $w'(TT+1)$, geht w' in w auf. Die Anzahl der Infizierten beider Fallgruppen wird addiert.

Fallgruppe w :

Infizierte [5], Todesfälle [0] und genesene Fälle [0]
NeuerFall [0], NeuerTodesfall [-9], NeuGenesen [-9]

Ähnlich wie mit neuen Infektionsmeldungen verhält es sich mit Meldungen von Todes- oder Gesundungsfällen. Diese bilden temporäre Fallgruppen y' und z' welche später in stetige Fallgruppen y und z übergehen:

Tag $TT+3$

Fallgruppe w :

Infizierte [4], Todesfälle [0] und genesene Fälle [0]
NeuerFall [0], NeuerTodesfall [-9], NeuGenesen [-9]

Fallgruppe y':

Infizierte [1], Todesfälle [1] und genesene Fälle [0]

NeuerFall [0], NeuerTodesfall [1], NeuGenesen [-9]

Tag TT+4

Fallgruppe w:

Infizierte [2], Todesfälle [0] und genesene Fälle [0]

NeuerFall [0], NeuerTodesfall [-9], NeuGenesen [-9]

Fallgruppe y:

Infizierte [1], Todesfälle [1] und genesene Fälle [0]

NeuerFall [0], NeuerTodesfall [0], NeuGenesen [-9]

Fallgruppe z':

Infizierte [2], Todesfälle [0] und genesene Fälle [2]

NeuerFall [0], NeuerTodesfall [-9], NeuGenesen [1]

Hinweis zu Genesenen

Anhand der dem RKI von den Gesundheitsämtern übermittelten Detailinformationen zu einem Erkrankungsfall wird für jeden Fall eine Dauer der Erkrankung geschätzt. Für Fälle, bei denen nur Symptome angegeben sind, die auf einen leichten Erkrankungsverlauf schließen lassen, wird eine Dauer der Erkrankung von 14 Tagen angenommen. Bei hospitalisierten Fällen oder Fällen mit Symptomen, die auf einen schweren Verlauf hindeuten (z. B. Pneumonie) wird eine Dauer der Erkrankung von 28 Tagen angenommen. Ausgehend vom Beginn der Erkrankung, bzw. wenn dieser nicht bekannt ist, vom Meldedatum ergibt sich ein geschätztes Datum der Genesung für jeden Fall. Da im Einzelfall auch deutlich längere Erkrankungsverläufe möglich sind, bzw. die hier genutzten Informationen nicht bei allen Fällen dem RKI übermittelt werden, sind die so berechneten Daten nur grobe Schätzungen für die Anzahl der Genesenen und sollten daher auch nur unter Berücksichtigung dieser Limitationen verwendet werden.

Geodaten

2020-06-30_Deutschland_Landkreise_GeoDemo.csv

Die bereitgestellten Geo- und Zensusdaten (im Folgenden nur als Geodaten bezeichnet) ermöglichen die Verbindung der Fallzahlen mit den geographischen und demografischen Angaben der Landkreise. Die Verbindung beider Daten ist über die Landkreis ID ("IdLandkreis") möglich. Diese ist in den Fallzahlen und Geodaten konsistent enthalten.

Primärquelle der Geodaten ist das [Gemeindeverzeichnis aller politisch selbständigen Gemeinden mit ausgewählten Merkmalen des Statistischen Bundesamtes Gebietsstand: 30.06.2020 \(2. Quartal\)](#). Die darin enthaltenen geographischen Angaben befinden sich auf dem Zuordnungsstand vom 31.12.2019. Die demographischen Angaben befinden sich auf dem Zuordnungsstand vom 31.12.2018.

Für die genauere Darstellung Berlins ist der Landkreis Berlin (Landkreis ID 11000) auf Bezirksebene aufgeschlüsselt. Die Zuordnung der Bezirke zur den entsprechenden Landkreis IDs erfolgt nach dem [Schema des Amt für Statistik Berlin-Brandenburg](#). Die letzten zwei Stellen der Landkreis ID bilden die Bezirksnummern der Berliner Bezirke ab. Im Bereich von 11001 bis 11012 entspricht z.B. Landkreis ID 11001 dem Bezirk Berlin Mitte, 11012 dem Bezirk Berlin Reinickendorf.

Die Zuordnung der Flächenanteile der Bezirke erfolgt nach [Gebietsstand 2019 des Amts für Statistik Berlin-Brandenburg](#). Zum Bevölkerungsstand der Berliner Bezirke sind keine Daten verfügbar. In den Geodaten sind daher die Daten des [Einwohnerregisters](#) mit Gebietsstand vom 31.12.2019 angegeben.

Formatierung der Geodaten

Die Geodaten sind im Datensatz als kommaseparierte .csv Datei enthalten. Der verwendete Zeichensatz der .csv Datei ist UTF-8. Trennzeichen der einzelnen Werte ist ein Komma ",".

- Zeichensatz: UTF-8
- .csv Trennzeichen: Komma ","

Merkmale der Geodaten

In der .csv Geodatentabelle differenzieren die Spalten die verschiedenen Merkmale eines Landkreises. Pro Zeile ist ein Landkreis eindeutig abgebildet. Ein Landkreis wird durch folgende Eigenschaften beschrieben (in den Klammern finden sich die Merkmale dieser Eigenschaften):

- Landkreis ID auf Basis des amtlichen Gemeindeschlüssels AGS, zuzüglich der 12 Bezirke Berlin (IdLandkreis)
- Landkreisname (Landkreis)
- Fläche des Landkreises in km² (Fläche)
- Einwohner_innenzahl (EW_gesamt, EW_weiblich, EW_maennlich)

Merkmalsausprägungen der Geodaten

Die Geodaten enthalten die in der folgenden Tabelle abgebildeten Merkmale und deren Ausprägungen:

Merkmal	Ausprägung	Erläuterung
IdLandkreis	1001 bis 16077	Identifikationsnummer des Landkreises basierend auf dem Amtlichen Gemeindeschlüssel (AGS); Gebietsstand: 30.06.2020 (2. Quartal)
Landkreis	Name (Zusatz)	Name des Landkreises, gegeben falls mit erläuterndem Zusatz z. B. "Landeshauptstadt", "Hansestadt", etc.
Flaeche	Zahl	Fläche des Landkreises in km ² auf 2 Nachkommastellen gerundet.
EW_gesamt	Zahl	Gesamtanzahl der Einwohner_innen des Landkreises. Datenstand vom 31.12.2019
EW_weiblich	Zahl	Gesamtanzahl der weiblichen Einwohner_innen des Landkreises. Datenstand vom 31.12.2019
EW_maennlich	Zahl	Gesamtanzahl der männlichen Einwohner_innen des Landkreises. Datenstand vom 31.12.2019

Hinweis zur Angabe der Einwohner_innenzahlen

Das "Gesetz zur Änderung der in das Geburtenregister einzutragenden Angaben" ist zum 22. Dezember 2018 in Kraft getreten. Mit dem Gesetz wird das Personenstandsgesetz (PStG) angepasst und ermöglicht Menschen die Möglichkeit, im Geburtenregister neben den Angaben "männlich", "weiblich" die weitere Bezeichnung "divers" zu wählen. In den Angaben der Einwohner_innenzahlen, mit Datenstand vom 31.12.2018, ist Kategorie "divers", des Geschlechtseintrags, noch nicht abgebildet.

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 beschreiben 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:

.zenodo.json

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

Lizenz

Der Datensatz "SARS-CoV-2 Infektionen in Deutschland" 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 Robert Koch-Instituts als Quelle, frei verfügbar. Das bedeutet, dass jede_r das Recht hat, 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/LIZENZ Datei des Datensatzes.

Die empfohlene Zitierweise ist:

Robert Koch-Institut (2021): SARS-CoV-2 Infektionen in Deutschland, Berlin: Zenodo.
DOI:[10.5281/zenodo.4681153](https://doi.org/10.5281/zenodo.4681153).