

# SARS-CoV-2 Infektionen in Deutschland

ÄNDERE!!! aaaa

[Robert Koch-Institut \(https://grid.ac/institutes/grid.13652.33\)](https://grid.ac/institutes/grid.13652.33) | RKI  
Nordufer 20  
13353 Berlin

FG 32 | Surveillance  
[Michaela Diercke \(https://orcid.org/0000-0002-4678-1813\)](https://orcid.org/0000-0002-4678-1813) (Leitung)

FG 31 | Infektionsepidemiologische Fach-IT und Anwendungsentwicklung  
[Hermann Claus \(https://orcid.org/0000-0002-0120-1846\)](https://orcid.org/0000-0002-0120-1846) (Leitung)

MF 4 | Forschungsdatenmanagement  
[Linus Grabenhenrich \(https://orcid.org/0000-0002-9300-6625\)](https://orcid.org/0000-0002-9300-6625) (Leitung)  
[Hannes Wuensche \(https://orcid.org/0000-0002-8837-0326\)](https://orcid.org/0000-0002-8837-0326) (Datenkuration)

---

## DOI to replace

10.5072/zenodo.1138213

Robert Koch-Institut (2021): SARS-CoV-2 Infektionen in Deutschland, Berlin: Zenodo. DOI: [10.5281/zenodo.4681153](https://doi.org/10.5281/zenodo.4681153)  
(<https://doi.org/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 \(https://creativecommons.org/licenses/by/4.0/deed.de\)](https://creativecommons.org/licenses/by/4.0/deed.de)

## 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, der nach den Vorgaben des [Infektionsschutzgesetz - IfSG \(https://www.gesetze-im-internet.de/ifsg/index.html\)](https://www.gesetze-im-internet.de/ifsg/index.html) 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 \(https://www.rki.de/DE/Content/Institut/OrgEinheiten/Abt3/FG32/FG32\\_node.html\)](https://www.rki.de/DE/Content/Institut/OrgEinheiten/Abt3/FG32/FG32_node.html) des RKI.

Die Verarbeitung und Aufbereitung der im Meldesystem vorliegenden Rohdaten erfolgt durch das [Fachgebiet 31 | Infektionsepidemiologische Fach-IT und Anwendungsentwicklung \(https://www.rki.de/DE/Content/Institut/OrgEinheiten/Abt3/FG31/FG31\\_node.html\)](https://www.rki.de/DE/Content/Institut/OrgEinheiten/Abt3/FG31/FG31_node.html) des RKI.

Die Veröffentlichung der Daten, die Datenkuration sowie das Qualitätsmanagement der (Meta-)Daten erfolgt durch das Fachgebiet [MF 4 | Forschungsdatenmanagement \(https://www.rki.de/DE/Content/Institut/OrgEinheiten/MF/MF4/mf4\\_node.html\)](https://www.rki.de/DE/Content/Institut/OrgEinheiten/MF/MF4/mf4_node.html). Die Datensatzkuration und Betreuung Publikationsinfrastruktur erfolgt durch Hannes Wuensche. Fragen zum Datenmanagement können an das Fachgebiet MF4 gerichtet werden ([MF4@rki.de \(mailto:MF4@rki.de\)](mailto: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 Gemeindegliederungsschlüssel (AGS) des Quartals 2 2020, abgerufen im [Portal des Statistischen Bundesamtes \(https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/GVAuszugQ/AuszugGV2QAktuell.html\)](https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/GVAuszugQ/AuszugGV2QAktuell.html). 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 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 selbstständigen Gemeinden mit ausgewählten Merkmalen des Statistischen Bundesamtes Gebietsstand: 30.06.2020 \(2. Quartal\)](https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/) (<https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/>). 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](https://www.statistik-berlin-brandenburg.de/regionalstatistiken/r-gesamt_neu.asp?Ptyp=410&Sageb=33000&creg=BBB&anzwer=9) ([https://www.statistik-berlin-brandenburg.de/regionalstatistiken/r-gesamt\\_neu.asp?Ptyp=410&Sageb=33000&creg=BBB&anzwer=9](https://www.statistik-berlin-brandenburg.de/regionalstatistiken/r-gesamt_neu.asp?Ptyp=410&Sageb=33000&creg=BBB&anzwer=9)). 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](https://www.statistik-berlin-brandenburg.de/regionalstatistiken/r-gesamt_neu.asp?Ptyp=410&Sageb=33000&creg=BBB&anzwer=9) ([https://www.statistik-berlin-brandenburg.de/regionalstatistiken/r-gesamt\\_neu.asp?Ptyp=410&Sageb=33000&creg=BBB&anzwer=9](https://www.statistik-berlin-brandenburg.de/regionalstatistiken/r-gesamt_neu.asp?Ptyp=410&Sageb=33000&creg=BBB&anzwer=9)). Zum Bevölkerungsstand der Berliner Bezirke sind keine Daten verfügbar. In den Geodaten sind daher die Daten des [Einwohnerregisters](https://www.statistik-berlin-brandenburg.de/grundlagen/Einwohnerregister.asp?Ptyp=50&Sageb=12041&creg=BBB&anzwer=10) (<https://www.statistik-berlin-brandenburg.de/grundlagen/Einwohnerregister.asp?Ptyp=50&Sageb=12041&creg=BBB&anzwer=10>) 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  $\text{km}^2$  (Flaeche)
- 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)
	Name (Zusatz)	Name des Landkreises, gegeben falls mit erläuterndem Zusatz z. B. "Landeshauptstadt", "Hansestadt", etc.
Flaeche	Zahl	Fläche des Landkreises in $\text{km}^2$ 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 \(https://creativecommons.org/licenses/by/4.0/deed.de\)](https://creativecommons.org/licenses/by/4.0/deed.de)

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) (<https://doi.org/10.5281/zenodo.4681153>).