



StopptCOVID-Studie - Daten, Analyse und Ergebnisse

[Andreas Hicketier](#)¹ und [Matthias an der Heiden](#)²

¹ [Robert Koch-Institut](#) | Fachgebiet 32 | Surveillance und elektronisches Melde- und Informationssystem (DEMIS) | ÖGD-Kontaktstelle

² [Robert Koch-Institut](#) | Fachgebiet 34 | HIV/AIDS und andere sexuell oder durch Blut übertragbare Infektionen

Zitieren

Hicketier A und an der Heiden M (2024): StopptCOVID-Studie - Daten, Analyse und Ergebnisse. Berlin: Zenodo. [DOI: 10.5281/zenodo.10888033](#)

Informationen zum Projekt und Entstehungskontext

Die getroffenen Maßnahmen zur Kontrolle von Severe Acute Respiratory Syndrome Coronavirus Type 2 (SARS-CoV-2) haben während der Coronavirus Disease 2019-(COVID-19-) Pandemie zu starken Einschränkungen des öffentlichen Lebens in Deutschland geführt. Das übergeordnete Ziel des Projekts "StopptCOVID" bestand darin, die Evidenzgrundlage für die Beurteilung der Effektivität verschiedener antipandemischer, nicht-pharmazeutischer Maßnahmen (NPI) zu verbessern. Dabei war die Frage, inwiefern verordnete Maßnahmen einen Anstieg der COVID-19-Inzidenz bremsen konnten. An dieser Stelle veröffentlichen wir Daten und Code für die Analyse der NPI in Deutschland.

Das StopptCOVID-Gesamtprojekt bestand aus zwei sich ergänzenden Teilprojekten, die mit unterschiedlichen Methoden durchgeführt wurden: Die CoViRiS-Fallkontrollstudie (Corona-Virus Risiko- und Schutzfaktoren im Alltag in Deutschland) und die StopptCOVID Analyse.

Zusätzlich wurde durch die Universität Bielefeld der „Covid-19 Pandemic Policy Monitor“ (COV-PPM) erstellt, der die NPI auf europäischer Ebene dokumentiert. In einer weiteren Analyse wurde von der Universität Bielefeld der Zusammenhang zwischen dem Deprivationsscore sowie dem Anteil der ausländischen Bevölkerung und der COVID-19-Inzidenz auf Ebene der Kreise in Deutschland untersucht.

Im vorliegenden Datensatz werden die Daten, Analysen und Ergebnisse der StopptCOVID-Studie zur Wirksamkeit NPI in Deutschland bereitgestellt. Der Datensatz bezieht sich direkt auf den vom RKI publizierten Abschlussbericht: "[Wirksamkeit und Wirkung von anti-epidemischen Maßnahmen auf die COVID-19-Pandemie in Deutschland \(StopptCOVID-Studie\)](#)".

an der Heiden M, Hicketier A und Bremer V (2024): Wirksamkeit und Wirkung von anti-epidemischen Maßnahmen auf die COVID-19-Pandemie in Deutschland (StopptCOVID-Studie). Berlin: RKI. [DOI: 10.25646/12007.2](#)

Administrative und Organisatorische Angaben

Die Erhebung, Analyse und Bereitstellung der Daten erfolgt durch [Fachgebiet 32 | Surveillance | ÖGD-Kontaktstelle](#) und [Fachgebiet 34 | HIV/AIDS und andere sexuell oder durch Blut übertragbare Infektionen](#) des RKI. Inhaltliche Fragen bezüglich der StopptCOVID-Studie können an das RKI unter info@rki.de gestellt werden. Die Veröffentlichung der Daten, die Datenkuratation sowie das Qualitätsmanagement der (Meta-)Daten erfolgt durch das Fachgebiet [MF 4 | Fach- und Forschungsdatenmanagement](#). Fragen zum Datenmanagement können an das Open Data Team des Fachgebiets MF4 gerichtet werden OpenData@rki.de.

Daten und Datenauswertung

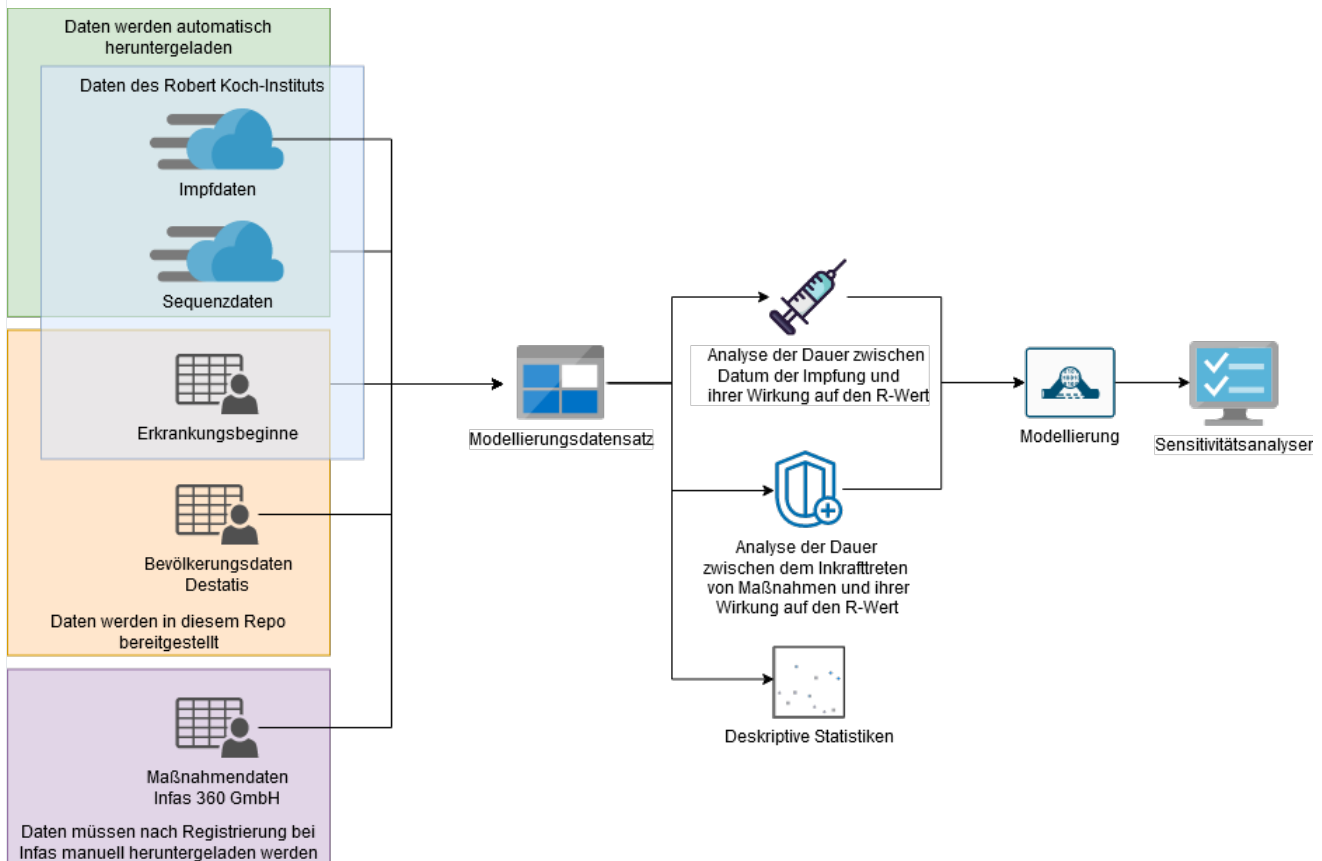
Der bereitgestellte Datensatz war Grundlage unserer Analyse und kann zur Reproduktion unserer Ergebnisse verwendet werden. Um das zu ermöglichen, werden Daten des Robert Koch-Instituts zu den COVID-19 Meldedaten nach Erkrankungsbeginn bereit gestellt, sowie R-Skripte, die Daten zu NPI auf Ebene der Landkreise, zur Bevölkerung, zu Impfungen sowie zu Genom-Sequenzdaten für die Analyse aufbereiten und integrieren. Eine detaillierte Beschreibung der Datenquellen ist im Abschnitt [Datenquellen und Datenaufbereitung](#) zu finden.

Der Einfluss der NPI auf den effektiven Reproduktionswert (R-Wert) der COVID-19-Epidemie in Deutschland wurde anhand eines Regressionsmodells untersucht. Als Zielvariable der Analyse wurde der R-Wert gewählt, da die NPI zur Reduktion von (infektiösen) Kontakten führen und somit die Anzahl von Folgeinfektionen pro Fall verringern sollten. Die Auswertung wurde adjustiert für den Einfluss der COVID-19 Impfquote, der vorherrschenden Erregervarianten sowie von Saisonalität und Schulferien.

Da die NPI in verschiedenen Bereichen teilweise zu sehr ähnlichen Zeiten verschärft und auch wieder gelockert wurden, ihr Aktivierungsprofil also stark korreliert war, konnten diese Bereich nur kombiniert ausgewertet werden.

Für die Auswertung wurde insbesondere die Dauer zwischen dem Inkrafttreten von Verordnungen zur Verschärfung oder Lockerung von NPI und deren Wirkung auf den R-Wert analysiert, sowie die Dauer zwischen dem Datum der 1. und 2. Impfung und deren Wirkung auf der R-Wert. Schließlich wurden die Stabilität des Hauptmodells analysiert und es wurden verschiedene Sensitivitätsanalysen durchgeführt.

Eine Übersicht und Beschreibung der erstellten R-Skripte befindet sich im Abschnitt [Datenauswertung](#).



Datenquellen und Datenaufbereitung

COVID-19 Impfdaten, RKI

Die Impfdaten werden vom Robert Koch-Institut im Datensatz "COVID-19-Impfungen in Deutschland" bereitgestellt und umfassen tägliche Informationen auf Ebene der Bundesländer für die Altersgruppen 0-17 Jahre, 18-59 Jahre und 60+ Jahre. Der automatische Download der Daten mit Datenstand vom 2022-05-18 ist im Skript [import_impfdaten.R](#) implementiert. Weitere spezifische Details zum verwendeten Datensatz sind der Datenpublikation zu entnehmen und können unter folgendem Link eingesehen werden:

Robert Koch-Institut, Fachgebiet 33 (2022). COVID-19-Impfungen in Deutschland (2022-05-18) [Data set]. Zenodo. [DOI: 10.5281/zenodo.6559081](https://doi.org/10.5281/zenodo.6559081)

Sequenzdaten, RKI

Die Anzahl der besorgniserregenden SARS-CoV-2-Virusvarianten (Variants of Concern; VOC) in Deutschland wird vom Robert Koch-Institut im Datensatz "SARS-CoV-2 Sequenzdaten" bereitgestellt. Der automatische Download der Daten mit Datenstand vom 2024-03-13 ist im Skript [import_seq_datan.R](#) implementiert. Weitere spezifische Details zum verwendeten Datensatz sind der Datenpublikation zu entnehmen und können unter folgendem Link eingesehen werden:

Robert Koch-Institut. (2024). SARS-CoV-2 Sequenzdaten aus Deutschland (2024-03-13) [Data set]. Zenodo. [DOI: 10.5281/zenodo.10813808](https://doi.org/10.5281/zenodo.10813808)

COVID-19 Fälle nach Erkrankungsbeginn, RKI

Gemäß dem Infektionsschutzgesetz (IfSG), werden Meldedaten zu COVID-19-Fällen an das Robert Koch-Institut übermittelt. Die Meldungen enthalten unter anderem Informationen zum Krankheitsbeginn der gemeldeten Fälle. Für Fälle, die keinen Krankheitsbeginn enthalten, wird ein Krankheitsbeginn imputiert (siehe an der Heiden und Hamouda 2020). Die Altersgruppen sind 0-19 Jahre, 20-59 Jahre und 60-120 Jahre, was eine Näherung an die Altersgruppen in den Impfdaten darstellt. Die berechneten Daten sind im Datensatz in der Datei [IfSG_COVID-19_Erkrankungsbeginn_Erwartungswert.csv](#) bereitgestellt.

an der Heiden M und Hamouda O (2020): Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting. Berlin: Epidemiologisches Bulletin. [DOI: 10.25646/6692.4](#)

Bevölkerung, DESTATIS

Die in der Analyse verwendeten Bevölkerungsdaten werden über die von DESTATIS betriebene Plattform GENESIS-Online, mit Datenstand 2020-12-31, bezogen. Datenquelle ist die Tabelle 12411-0017: Bevölkerung: Kreise, Stichtag, Altersgruppen. Die Daten sind im Datensatz als [GENESIS-Online_Bevölkerung_Kreise_Altersgruppen.csv](#) enthalten.

<https://www-genesis.destatis.de/genesis/online?operation=table&code=12411-0017>

NPI-Daten, Infas 360 GmbH

Daten zu nicht-pharmazeutischen Interventionen (NPI) auf Landkreisebene wurden, im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi), von der Infas 360 GmbH dokumentiert. Die Daten umfassen tagesgenaue Informationen zu NPI auf Landkreisebene, siehe Tabellen unter [Maßnahmen Unterkategorien Kreise](#). Die Daten sind nicht frei verfügbar. Eine vorherige, kostenlose Registrierung ist erforderlich, um auf die Daten zuzugreifen.

https://www.healthcare-datenplattform.de/dataset/massnahmen_unterkategorien_kreise

Die folgende Tabelle enthält eine Übersicht über die R-Skripte, die die Daten laden und für die Analyse aufbereiten. Das Skript [cr_measure_data.R](#) codiert die NPI-Daten von Infas, wie in Tabelle 1 des Abschlussberichts ausgeführt. Informationen zur Ausführung der Skripte finden sich im Abschnitt [Skripte](#).

Datei	Aufgabe	Beschreibung
import_impfdaten.R	Datenimport und Datenaufbereitung	Lädt die COVID-19 Impfdaten
import_seq_datan.R	Datenimport und Datenaufbereitung	Lädt die Sequenzdaten
process_measures.R	Datenimport und Datenaufbereitung	Lädt die Maßnahmendaten von Infas, aus dem Ordner Daten/infas des Datensatzes
cr_measure_data.R	Datenimport und Datenaufbereitung	Bereitet die Maßnahmendaten auf
holidays.R	Datenimport und Datenaufbereitung	Erzeugt Variablen zu Schulferien und Feiertagen
cr_modeling_data.R	Datenimport und Datenaufbereitung	Aggregiert Maßnahmendaten auf Bundeslandebene und kombiniert sie mit den übrigen Daten
cr_modeling_ag_data.R	Datenimport und Datenaufbereitung	Bereitet Daten für die Altersgruppen auf

Datenauswertung

Eine detaillierte Beschreibung der Datenauswertung der StopptCOVID-Studie zur Wirksamkeit der NPI in Deutschland befindet sich im Abschnitt Methoden des Abschlussberichts "Wirksamkeit und Wirkung von anti-epidemischen Maßnahmen auf die COVID-19-Pandemie in Deutschland (StopptCOVID-Studie)".

an der Heiden M, Hicketier A und Bremer V (2024): Wirksamkeit und Wirkung von anti-epidemischen Maßnahmen auf die COVID-19-Pandemie in Deutschland (StopptCOVID-Studie). Berlin: RKI. [DOI: 10.25646/12007.2](#)

Um die Modellierung und Sensitivitätsanalysen detailliert nachvollziehbar zu machen, werden im Datensatz neben den Informationen zu den Datenquellen auch die R-Skripte der Analysen bereitgestellt. Die Ergebnisse und Analysen des Abschlussberichts, lassen sich mit den bereitgestellten R-Skripten reproduzieren. Aus Perspektive der Datenauswertung lassen sich die Skripte grob in zwei Aufgaben-Bereiche einteilen:

1. Deskriptive Statistiken

- o Hier werden diverse Plots generiert, die die Datengrundlagen beschreiben, z.B. Korrelationsmatrizen, R-Werte, der Maßnahmenscore etc.

2. Analyse

- o Hier werden die eigentlichen Modelle angepasst und deren Ergebnisse in einer Reihe von Grafiken und Textdateien ausgegeben.

In der folgenden Tabelle ist eine Übersicht über die Skripte der deskriptiven Statistiken und Analyse gegeben. Um die Entstehung der im Abschlussbericht enthaltenen Abbildungen nachvollziehen zu können, sind diese den erzeugenden Skripten zugeordnet:

Datei	Aufgabe	Beschreibung	Erzeugt folgende Abbildungen im Abschlussbericht
corelation_matrix_main.R	Deskriptive Statistiken	Berechnet die Korrelationsmatrix der NPI Aktivitätsprofile	Abb. 6-8
describe_data.R	Deskriptive Statistiken	Beschreibt die Daten	Abb. 4-5, 9-15
optim_lag_vacc.R	Analyse	Analysiert den Verzug zwischen dem Datum der 1. und 2. Impfung und ihrer Wirkung auf den R-Wert	Abb. 21
optim_lag_npi.R	Analyse	Analysiert den Verzug zwischen dem Inkrafttreten von Verordnungen zu NPI und deren Wirkung auf den R-Wert	Abb. 16-20
Main_model.R	Analyse	Fittet das Hauptmodell an die Daten	Abb. 22-23
Models_rangeLag.R	Analyse	Erstellt eine Stabilitätsanalyse für das Hauptmodell und vergrößerte Modelle	Abb. 25-27
Cum_main_model.R	Analyse	Berechnet kumulative Effekte über alle NPI im Hauptmodell	Abb. 24
Sens_Main_model.R	Analyse	Berechnet Sensitivitätsanalysen zu den verschiedenen Wellen, zur Gewichtung, etc.	Abb. 28-31

Abweichungen vom Abschlussbericht

Bei Reproduktion der Ergebnisse unserer Analysen mit dem hier bereitgestellten Datensatz, kommt es zu minimalen Abweichung im Vergleich zu den Ergebnissen des Abschlussberichts. Diese Abweichungen basieren auf Änderungen in zwei der unterliegenden Datenquellen:

- Die ursprünglichen Analysen nutzten einen Datensatz der Infas 360 GmbH vom 3. März 2022. Dieser Datensatz ist nicht mehr verfügbar und die [Lizenz](#) schließt eine Weitergabe unsererseits aus. Der derzeit unter <https://www.healthcare-datenplattform.de/> erhältliche Datensatz weicht an 14 Tagen, in einer Variable, in einem Kreis von dem ursprünglich von uns verwendeten Datensatz ab.
- Der Abschlussbericht basierte auf Virus-Varianten-Daten, die im Rahmen der [Virusvariantenberichte](#) veröffentlicht worden waren. Diese Daten sind nicht mehr öffentlich verfügbar. Die jetzt in den Skripten verwendeten Daten zu [besorgniserregenden SARS-CoV-2-Virusvarianten](#) weichen minimal von den ursprünglichen Virus-Varianten-Daten.

Diese Änderungen sind insbesondere in folgender Abbildung zu erkennen:

- Ranking der Effekte der Maßnahme: Die erstellte Grafik `ranking_main_model_AG` weicht minimal von ihrem Pendant, Abb. 13 im Abschlussbericht, ab.

Darüber hinaus sind die R^2 -Werte des Hauptmodells und mehrerer Modelle der Sensitivitätsanalyse deutlich höher, und damit besser, als im Abschlussbericht dargestellt. Das R^2 ist ein Gütemaß für die Anpassung eines Modells an die Daten. Die Ursache ist bislang unklar. Es erscheint unwahrscheinlich, dass dies auf den o.g. Änderungen der Datensätze basiert.

Inhalt und Aufbau der bereitgestellten Daten

Die im Abschnitt [Daten und Datenauswertung](#) beschriebenen Daten und die verarbeitenden R-Skripte werden als offene Daten zur Verfügung gestellt. Der folgende Abschnitt beschreibt die Struktur des Datensatzes im Detail:

- Daten zu
 - COVID-19 Fälle nach Erkrankungsbeginn
 - Bevölkerung nach Kreisen, Stichtag, Altersgruppen
- R-Skripte für Import, Aufbereitung und Datenauswertung
- Metadaten

Daten zu COVID-19-Erkrankungsbeginn und Bevölkerung

Die Daten zu [COVID-19- Impfungen und besorgniserregenden SARS-CoV-2-Virusvarianten](#) sind als Open Data verfügbar und können direkt über die Skripte [import_impfdaten.R](#) und [import_seq_daten.R](#) für die Analyse eingelesen werden.

Für die Daten der COVID-19 Fälle nach Erkrankungsbeginn und Bevölkerung nach Kreisen, Stichtag, Altersgruppen ist dies nicht direkt möglich. Die entsprechenden Daten sind daher gesondert im Datensatz enthalten:

[Daten/IfSG_COVID-19_Erkrankungsbeginn_Erwartungswert.csv](#) [Daten/GENESIS-Online_Bevoelkerung_Kreise_Altersgruppen.csv](#)

Variablen und Variablenausprägungen

Die Variablen und Variablenausprägungen der [IfSG_COVID-19_Erkrankungsbeginn_Erwartungswert.csv](#) sind in folgender Tabelle abgebildet:

Variable	Typ	Ausprägungen	Beschreibung
Bundesland	Text	Baden-Wuerttemberg , Bayern , Berlin , Brandenburg , Bremen , Hamburg , Hessen , Mecklenburg-Vorpommern , Niedersachsen , Nordrhein-Westfalen , Rheinland-Pfalz , Saarland , Sachsen , Sachsen-Anhalt , Schleswig-Holstein , Thuringen	Bundesland des gemeldeten COVID-19 Falls
Datum	Datum	yyyy-mm-dd	Erkrankungsbeginn des gemeldeten Falls oder des imputierter Erkrankungsbeginns (falls dieser nicht berichtet wurde) im ISO8601 Format
Altersgruppe	Text	00-17 , 18-59 , 60+	Altersgruppe der gemeldeten Fälle
EW_Fallzahl	Fließkommazahl	≥0.00	Erwartungswert der Fallzahl unter Berücksichtigung der Imputation des Erkrankungsbeginns (siehe an der Heiden und Hamouda 2020)

Die Variablen und Variablenausprägungen der [GENESIS-Online_Bevoelkerung_Kreise_Altersgruppen.csv](#) sind in folgender Tabelle abgebildet:

Variable	Typ	Ausprägungen	Beschreibung
Bundesland	Text	Baden-Wuerttemberg , Bayern , Berlin , Brandenburg , Bremen , Hamburg , Hessen , Mecklenburg-Vorpommern , Niedersachsen , Nordrhein-Westfalen , Rheinland-Pfalz , Saarland , Sachsen , Sachsen-Anhalt , Schleswig-Holstein , Thuringen	Bundesland
Kreis	Text	LK Ahrweiler , ..., SK Zweibrücken	Name des Landkreises
Kreisschlüssel	Text	07131 , 09771 , ...	Amtlicher Kreisschlüssel
Altersgruppe	Text	00-19 , 20-59 , 60+	Altersgruppe
Bevölkerung	natürliche Zahl	≥0	Anzahl der Einwohner nach Landkreis und Altersgruppe

Formatierung der Daten

Die Daten sind im Datensatz als semikolon-separierte .csv Datei enthalten. Der verwendete Zeichensatz der CSV-Dateien ist UTF-8. Trennzeichen der einzelnen Werte ist ein Semikolon ";". Datumsangaben sind im ISO8601 Standard formatiert.

- Zeichensatz: UTF-8
- CSV-Trennzeichen: Semikolon ";"
- Kennzeichnung fehlender Werte: „NA

Skripte

Die R-Skripte für die gesamte Reproduktion der Analysen sind im Datensatz im Ordner [Skripte](#) bereitgestellt.

[Skripte/](#)

Die einzelnen Skripte können über [StopptCOVID_main.R](#) gesammelt ausgeführt werden. Diese befindet sich im Hauptverzeichnis des Datensatzes.

[StopptCOVID_main.R](#)

Die NPI-Daten der Infas 360 GmbH sind nicht im Repository enthalten und müssen vor dem Ausführen von [StopptCOVID_main.R](#) in den Unterordner [Daten/infas](#) heruntergeladen werden. Alle anderen Daten sind im Datensatz enthalten oder werden automatisch heruntergeladen (siehe [Datenquellen und Datenaufbereitung](#)).

Das Ausführen aller Skripte dauert ca. eine halbe Stunde, hauptsächlich wegen der Analyse der Verzögerungen in [optim_lag_vacc.R](#) und [optim_lag_npi.R](#). Bei Bedarf können diese Skripte durch Auskommentierung ausgelassen werden.

Die bereitgestellten R-Skripte lassen sich grob in drei Aufgaben-Bereiche einteilen:

1. Datenimport und Datenaufbereitung
 - Hier werden die oben genannten Datensätze eingelesen und für die Analyse aufbereitet, z.B. durch Berechnung der R-Werte, Aufbereitung und Kodierung der NPI etc.
2. Deskriptive Statistiken
 - Hier werden diverse Plots generiert, die die Datengrundlagen beschreiben, z.B. Korrelationsmatrizen, R-Werte, der Maßnahmenscore etc.
3. Analyse
 - Hier werden die eigentlichen Modelle angepasst und deren Ergebnisse in einer Reihe von Grafiken und Textdateien ausgegeben.

Metadaten

Zur Erhöhung der Auffindbarkeit sind die bereitgestellten Daten mit Metadaten beschrieben. Über GitHub Actions werden Metadaten an die entsprechenden Plattformen verteilt. Für jede Plattform existiert eine spezifische Metadatenfile, diese sind im Metadatenordner hinterlegt:

[Metadaten/](#)

Versionierung und DOI-Vergabe erfolgt über [Zenodo.org](#). Die für den Import in Zenodo bereitgestellten Metadaten sind in der [zenodo.json](#) hinterlegt. Die Dokumentation der einzelnen Metadatenvariablen ist unter <https://developers.zenodo.org/representation> nachlesbar.

[Metadaten/zenodo.json](#)

Hinweise zur Nachnutzung der Daten

Offene Forschungsdaten des RKI werden auf [GitHub.com](#), [Zenodo.org](#) und [Edoc.rki.de](#) bereitgestellt:

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

Lizenz

Der Datensatz "StopptCOVID-Studie - Daten, Analyse und Ergebnisse" 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 und des Umweltbundesamtes als Quelle, frei verfügbar. Das bedeutet, jede Person 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.