

Ergebnisdatensatz BURDEN 2020 – Krankheitslast in Deutschland und seinen Regionen

Robert Koch-Institut | RKI

Nordufer 20
13353 Berlin

Beitragende

Janko Leddin¹ (Datenmanagement) und Michael Porst¹ (Morbiditätsanalysen, Statistik)

¹ Robert Koch-Institut | Abteilung 2

Zitieren

Robert Koch-Institut. (2025). Ergebnisdatensatz BURDEN 2020 – Krankheitslast in Deutschland und seinen Regionen [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.14750006>

Einleitung

Die vorliegende Datenveröffentlichung stellt die Ergebnisdaten des Projekts „BURDEN 2020 – Die Krankheitslast in Deutschland und seine Regionen“ (BURDEN 2020) zur Verfügung und beschreibt deren Inhalt und Struktur. Es wird zudem in das Projekt und seinem Thema Krankheitslast eingeführt. Die Beschreibung der Methoden und Berechnungsverfahren zur Bestimmung der Krankheitslastindikatoren sowie die Ergebnisdarstellung sind im Detail den bereits veröffentlichten BURDEN 2020 Projektpublikationen zu entnehmen. Auf sie wird an entsprechenden Stellen als integraler Bestandteil dieser Datenveröffentlichung verwiesen (siehe [Datengrundlagen und Methoden](#)). Die Ergebnisdaten wurden zur Veranschaulichung des Konzepts auf der Webseite www.daly.rki.de visualisiert.

Informationen zum Datensatz und Entstehungskontext

Der Ergebnisdatensatz entstand dem Pilotprojekt „BURDEN 2020 - Die Krankheitslast in Deutschland und seinen Regionen“ (BURDEN 2020). Ziel des Projektes war es, die Methode der Krankheitslastrechnung, welche im Rahmen der Global Burden of Disease (GBD)-Studie [\[1-3\]](#) verwendet wird, für Deutschland zu adaptieren und auf Basis nationaler Datenquellen anzuwenden. Es wurden mit Krankenkassenroutinedaten, offiziellen Statistiken und Primärerhebungen bundesweit und regional belastbare Daten für Deutschland herangezogen. Im Ergebnis wurde durch die Nutzung der bestehenden Datenvielfalt ein konsistentes Informationssystem zu den drei Kernindikatoren years of life lost (YLL), years lived with disability (YLD) und disability-adjusted life years (DALY) geschaffen. Es ermöglicht, Erkrankungen und Verletzungen nach ihrer Public Health-Relevanz zu priorisieren, Potenziale für eine bedarfsgerechte Prävention und Versorgung zu identifizieren und die Planung und Bewertung gesundheitspolitischer Maßnahmen zu unterstützen.

Weiter Informationen zum Projekthintergrund findet sich in Rommel et al. 2018:

Rommel, A., von der Lippe, E., Plaß, D. et al. BURDEN 2020—Burden of disease in Germany at the national and regional level. Bundesgesundheitsbl 61, 1159–1166 (2018). DOI: [10.1007/s00103-018-2793-0](https://doi.org/10.1007/s00103-018-2793-0)

Administrative und organisatorische Angaben

Das Projekt BURDEN 2020 wurde unter der Federführung der [Abteilung für Epidemiologie und Gesundheitsmonitoring \(Abteilung 2\)](#) des Robert Koch-Instituts (RKI) im Verbund mit dem Umweltbundesamt (UBA) und dem Wissenschaftlichen Institut der AOK (WIdO) durchgeführt (BURDEN 2020 study group).

Zu den Projektmitarbeitern der BURDEN 2020 study group zählten:

Abteilung 2 | Epidemiologie und Gesundheitsmonitoring

Dr. Aline Anton (Projektkoordination), Tim Fleckenstein (Website und Visualisierung), Janko Leddin (Datenmanagement), Dr. Elena von der Lippe (Methodische Leitung), Michael Porst (Morbiditätsanalysen, Statistik), Dr. Alexander Rommel (Projektleitung), Dr. Annelene Wengler (Mortalitätsanalysen, Statistik), Dr. Thomas Ziese (Projektleitung)

Umweltbundesamt | UBA

Heike Gruhl (umweltbezogene Risikoanalysen), Dr. Dietrich Plaß (Leitung UBA)

Wissenschaftliches Institut der AOK | WIdO

Dr. Jan Breitkreuz (Sekundärdatenanalysen), Dr. Katrin Schüssel (Sekundärdatenanalysen), Gabriela Brückner (Sekundärdatenanalysen), Helmut Schröder (Projektleitung WIdO)

Inhaltliche Fragen bezüglich der Berechnung der Indikatoren (Aufbereitung der Datenquellen, verwendete Methoden, etc.) sowie zur Qualitätssicherung der Ergebnisse oder Feedback zur Visualisierung der Ergebnisdaten auf der [Website](#) können direkt an das BURDEN 2020-Team am RKI unter daly@rki.de gerichtet werden.

Die Veröffentlichung der Daten, die Datenkuration 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.

Förderung

Das Projekt BURDEN 2020 wurde durch den Innovationsfonds beim Gemeinsamen Bundesausschuss (G-BA, Förderkennzeichen: 01VSF17007) gefördert. Es wurde durch einen wissenschaftlichen Beirat begleitet, der sich wie folgt zusammensetzte: Prof. Dr. Hajo Zeeb (Vorsitzender), Prof. Dr. Gabriele Doblhammer (stellvertretende Vorsitzende), Prof. Dr. Reinhard Busse, Dr. Conrad Franke, PhD Henk Hilderink, Prof. Dr. Tobia Lakes, Dr. Odile Mekel, Prof. Dr. Ralf Münnich, Prof. Dr. Annette Peters, Dr. Enno Swart und Prof. Dr. Susanne Wurm. Die Projektlaufzeit war von April 2018 bis März 2020.

Datengrundlagen und Methoden

Zur Bestimmung der Indikatoren wurden mehrere Primär- und Sekundärdatenquellen genutzt. Hinter den Kernindikatoren YLL, YLD und DALY stehen in BURDEN 2020 komplexe Modellierungen: So wurden nichtinformativ Todesursachen umverteilt, um allen Verstorbenen im Rahmen der Krankheitslastberechnung gültige Todesursachen zuzuweisen. Es wurden regionale Prävalenzen geschätzt und alters- sowie morbiditätsadjustiert für die Gesamtbevölkerung hochgerechnet. In die Berechnungen gehen spezifische Schweregradverteilungen und Erkrankungsdauern ein, um die Krankheitslast durch

Morbidität in der Bevölkerung quantifizieren zu können. Die Ergebnisse wurden für Multimorbidität adjustiert und ein Konzept entwickelt, um die statistische Unsicherheit aus verschiedenen Quellen abzubilden. Detaillierte Informationen dazu wurden aus dem Projekt heraus bereits an anderer Stelle publiziert und im Folgenden themenspezifisch darauf verwiesen.

Krankheitslast durch Mortalität - YLL

Das Akronym YLL steht für "years of life lost" und repräsentiert die Mortalitätskomponente der Krankheitslast, also die Krankheitslast durch Versterben. Je höher der errechnete Zahlenwert, desto höher ist die Krankheitslast durch diese Komponente zu bewerten.

Datengrundlagen

Grundlage zur Berechnung des Krankheitslastindikators YLL waren Daten des Statistischen Bundesamtes (Destatis), darunter die Todesursachenstatistik 2017 und Daten aus Sterbetafeln zur durchschnittlichen Lebenserwartung:

Todesursachenstatistik 2017. DOI: [10.21242/23211.2017.00.00.1.1.0](https://doi.org/10.21242/23211.2017.00.00.1.1.0)

Sterbetafel. Ergebnisse aus der laufenden Berechnung von Periodensterbetafeln für Deutschland und die Bundesländer. Wiesbaden. 2016/18. 2019. Durchschnittliche Lebenserwartung über [GENESIS-Online](#)

Methodik und Berechnung

Gemäß Sterbetafeln wird jedem Alter eine statistische Restlebenserwartung zugeordnet. Die Summen der Restlebenserwartungswerte aller Verstorbenen eines Berichtsjahres bildet den Krankheitslastwert durch Mortalität (YLL). Über die Todesursachenstatistik lässt sich dieser Wert nach Ursache, Region, Alter und Geschlecht ausdifferenzieren. Die Todesursachenstatistik wurde um nichtinformativ Todesursachen mittels Umverteilungsverfahren bereinigt.

Ergebnis- und Methodenpublikation:

Wengler et al. 2021: „Years of Life Lost to Death - A comprehensive analysis of mortality in Germany conducted as part of the BURDEN 2020 project“ DOI: [10.3238/arztebl.m2021.0148](https://doi.org/10.3238/arztebl.m2021.0148)

Zuordnung der ICD-10-Codes zu den mortalitätsspezifischen Krankheitslastursachen:

Johnson et al. Public health utility of cause of death data: applying empirical algorithms to improve data quality. BMC Med Inform Decis Mak 21, 175 (2021). DOI: [10.1186/s12911-021-01501-1](https://doi.org/10.1186/s12911-021-01501-1).
s. Additional file 1, Appendix, Figure 3

Diagnose und Umverteilung nichtinformativer Todesursachen:

Wengler, A., Rommel, A., Plaß, D. et al. ICD-Codierung von Todesursachen: Herausforderungen bei der Berechnung der Krankheitslast in Deutschland. *Bundesgesundheitsbl* 62, 1485–1492 (2019). DOI: [10.1007/s00103-019-03054-1](https://doi.org/10.1007/s00103-019-03054-1)

Wengler, A., Gruhl, H., Plaß, D. et al. Redistributing ill-defined causes of death – a case study from the BURDEN 2020-project in Germany. *Arch Public Health* 79, 33 (2021). DOI: [10.1186/s13690-021-00535-1](https://doi.org/10.1186/s13690-021-00535-1)

Weitere Literatur:

Mikkelsen, L., Iburg, K.M., Adair, T. et al. Assessing the quality of cause of death data in six high-income countries: Australia, Canada, Denmark, Germany, Japan and Switzerland. *Int J Public Health* 65, 17–28 (2020). DOI: [10.1007/s00038-019-01325-x](https://doi.org/10.1007/s00038-019-01325-x)

Krankheitslast durch Morbidität - YLD

Das Akronym YLD steht für "years lived with disability" und repräsentiert die Morbiditätskomponente der Krankheitslast, also die Krankheitslast durch Erkrankungen oder Verletzungen. Je höher der errechnete Zahlenwert, desto höher ist die Krankheitslast durch diese Komponente zu bewerten.

Datengrundlagen

Grundlage zur Berechnung des Krankheitslastindikators YLD waren verschiedene Primär- und Sekundärdatenquellen, darunter alters-, geschlechts- und morbiditätsadjustierte Krankenkassenroutinedaten der AOK-Versicherten. Diese wurden vom Wissenschaftlichen Institut der AOK im Rahmen des Projektes generiert und bereitgestellt:

Wissenschaftliches Institut der AOK (WIdO). (2024). Datenrepositorium Krankheitslage-Deutschland.de: WIdO-Ergebnisse für die Krankheitslastberechnung im Projekt BURDEN 2020 (1.1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.14500672>

Zudem wurden Befragungsdaten zu Schmerz- und Suchterkrankungen, die Routinedaten der amtlichen Straßenverkehrsunfallstatistik, sowie Vorarbeiten der GBD-Studie verwendet. Eine vollständige Liste der verwendeten Datenquellen kann folgender BURDEN 2020 Projektpublikation entnommen werden:

Porst et al. 2022: „Die Krankheitslast in Deutschland und seinen Regionen - Ergebnisse zu den disability-adjusted life years (DALY) aus der Studie BURDEN 2020“. DOI: [10.3238/arztebl.m2022.0314](https://doi.org/10.3238/arztebl.m2022.0314) s. Anhang: eTabelle 3 - Datengrundlage und Verfahren (Morbiditätskomponente)

Methodik und Berechnung

Lebensjahre in gesundheitlicher Einschränkung wurden durch (i) die Häufigkeit einer Erkrankung oder Verletzung, (ii) die Verteilung der Erkrankten nach Schweregraden (sog. severity distributions), (iii) bei periodisch auftretenden Erkrankungen zusätzlich durch die mittleren Erkrankungsdauern sowie (iv) durch schweregradspezifische Gewichte (sog. disability weights) bestimmt. Es wurden regionale Prävalenzen geschätzt und alters- sowie morbiditätsadjustiert für die Gesamtbevölkerung hochgerechnet. In die Berechnungen gehen spezifische Schweregradverteilungen und Erkrankungsdauern ein, um die tatsächliche Krankheitslast in der Bevölkerung messen zu können.

Porst et al. 2022: „Die Krankheitslast in Deutschland und seinen Regionen - Ergebnisse zu den disability-adjusted life years (DALY) aus der Studie BURDEN 2020“. DOI: [10.3238/arztebl.m2022.0314](https://doi.org/10.3238/arztebl.m2022.0314)

Porst et al. 2023: Methodenbericht zur Quantifizierung der Krankheitslastindikatoren im Projekt BURDEN 2020 – Krankheitshäufigkeiten, Schweregrade, Dauern, Disability weights sowie Sensitivitätsanalysen. Robert Koch-Institut, Berlin. DOI: [10.25646/10022.2](https://doi.org/10.25646/10022.2) (überarbeitete Fassung)

Weitere Literatur:

Breitkreuz, J., Brückner, G., Burgard, J.P. et al. Schätzung kleinräumiger Krankheitshäufigkeiten für die deutsche Bevölkerung anhand von Routinedaten am Beispiel von Typ-2-Diabetes. *StA Wirtsch Sozialstat Arch* 13, 35–72 (2019). DOI: [10.1007/s11943-019-00241-z](https://doi.org/10.1007/s11943-019-00241-z)

Breitkreuz et al. 2021: „Krankheitslastbestimmung mit Prävalenzen und Schweregraden auf Routinedatenbasis“. *GGW* 21 (1): 24-34

Porst M, Wengler A, Leddin J, et al. 2020: „Migräne und Spannungskopfschmerz in Deutschland. Prävalenz und Erkrankungsschwere im Rahmen der Krankheitslast-Studie BURDEN 2020“ DOI: [10.25646/6988.2](https://doi.org/10.25646/6988.2)

von der Lippe E, Krause L, Porst M, et al. 2021: „Prävalenz von Rücken- und Nackenschmerzen in Deutschland. Ergebnisse der Krankheitslast-Studie BURDEN 2020“ DOI: [10.25646/7854](https://doi.org/10.25646/7854)

Schüssel et al. 2022: „Nutzung von Krankenkassenroutinedaten zur Bestimmung von Krankheitshäufigkeiten im Projekt BURDEN 2020“ DOI: [10.1055/a-1806-2115](https://doi.org/10.1055/a-1806-2115)

Summe der Krankheitslast - DALY

Das Akronym steht für "disability-adjusted life years" und repräsentiert die Summe der Krankheitslast aus den Einzelkomponenten YLD und YLL. Je höher der Zahlenwert, desto höher ist die Krankheitslast zu bewerten.

Datengrundlagen

Ergebnisdatensätze zu den Einzelkomponenten YLL und YLD.

Methodik und Berechnung

Summe aus YLL und YLD nach Alter, Geschlecht, Region und Krankheitslastursache.

Limitationen des Datensatzes

Während der Ergebnisdatensatz zur Mortalitätskomponente (YLL) aufgrund seiner Genese für alle Krankheitslastursachen vollständig vorliegt (Wengler et al. 2021), wurde im Rahmen des Pilotprojekts für den Bereich Morbidität (YLD) - und damit für die gesamte Krankheitslast (DALY) - zunächst mit einer Vorauswahl an Krankheitslastursachen auf Ebene 3 gearbeitet ([eTabelle 2](#), [Porst et al. 2022](#)). Eine Berechnung der YLD und DALY auf höheren Ebenen einerseits, und die vollständige Abbildung des Klassifikationssystems der Ursachen andererseits, ist dadurch allerdings bislang noch nicht möglich. In der Bereitstellung der Daten zu den YLD und DALY wird deshalb auf die Ursachenebenen 0-2 verzichtet. Weitere Limitationen ergeben sich für einzelne Krankheitslastursachen im Morbiditätsbereich sowohl hinsichtlich der räumlichen als auch der altersgruppenspezifischen Auflösung der Daten ([eTabelle 3](#), [Porst et al. 2022](#)).

Aufbau und Inhalt des Datensatzes

Der Datensatz enthält Daten zur Krankheitslast in Deutschland und seinen Regionen des Projekts BURDEN2020. Im Datensatz enthalten sind:

- Ergebnisdaten des Projektes BURDEN 2020
- Schlüsseltabellen verwendeter Hierarchiesysteme
- Lizenzdatei mit der Nutzungslizenz des Datensatzes in Deutsch
- Datensatzdokumentation in deutscher Sprache
- Metadaten zur automatisierten Weiterverarbeitung

BURDEN 2020 Ergebnisdaten

Die Ergebnistabellentabelle beinhaltet alle berechneten Datenpunkte des Projekts BURDEN 2020 in strukturierter Form. Ein Datenpunkt ist definiert als das Kreuzprodukt aller möglichen Ausprägungen der Variablen: Berichtsjahr, Indikator, Ursache, Region, Geschlecht, Altersgruppe und Einheit. Jede Zeile der Tabelle stellt den Wert im Kontext des Datenpunktes dar. Die Ergebnistabelle enthält keine Leerzeilen (Datenpunkte, die nicht berechnet werden konnten, sind nicht enthalten). Die BURDEN 2020 Ergebnisdaten beziehen sich ausschließlich auf das Berichtsjahr 2017. Weitere Limitationen sind dem entsprechenden Abschnitt dieser Dokumentation zu entnehmen.

BURDEN2020_Ergebnisdaten.csv

Insgesamt enthält der Datensatz 2.519.913 berechnete Datenpunkte.

Die Ausprägungen der beiden Variablen Ursache und Region folgen hierarchischen Gliederungssystematiken, die mit den mitgelieferten Schlüsselstabellen korrespondieren. Jede Schlüsselstabelle bildet ihr jeweiliges zugrundeliegende Hierarchiesystem vollständig ab. Ihr jeweiliger Aufbau wird im entsprechenden Abschnitt dieser Dokumentation nochmal näher erläutert. Ergebnistabelle und Schlüsselstabellen können bei Bedarf über die entsprechenden ID Variablen zusammengeführt werden.

Der Datensatz erhebt strukturell keinen Anspruch typischer Normalformen der Datenhaltung.

Variablen und Variablenausprägung

Die Datei [BURDEN2020_Ergebnisdaten.csv](#) enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen. Ein maschinenlesbares Datenschema ist im [Data Package Standard](#) in [tableschema_BURDEN2020_Ergebnisdaten.json](#) hinterlegt:

tableschema_BURDEN2020_Ergebnisdaten.json

Variable	Typ	Ausprägungen	Beschreibung
Berichtsjahr	date	Werte: 2017 Format: YYYY	Gibt das Jahr an, für das der [Wert] berechnet wurde.
Indikator	string	Werte: YLL, YLD, DALY	Gibt die Krankheitslastkomponente an, für die der [Wert] berechnet wurde (siehe Datengrundlagen und Methoden)
Ursache_id	integer	Werte: 1000, 2000, 2001, 2002, 2007, 2014, 2015, ...	Eindeutige Schlüsselnummer der [Ursache].
Ursache	string	Werte: Alle Ursachen, Übertragbare, maternale, neonatale und ernährungsbedingte Erkrankungen, HIV/AIDS und sexuell übertragbare Erkrankungen, HIV/AIDS, Sexuell übertragbare Krankheiten ohne HIV, Atemwegsinfektionen, Tuberkulose, ...	Bezeichnung der Krankheitslastursache, für die der [Wert] berechnet wurde (siehe Keytable_Ursache.csv).
Ursachenebene	integer	Werte: 0, 1, 2, 3, 4	Nummer als Bezeichner der hierarchischen Ebene, in die die [Ursache] eingeordnet ist. Bsp.: 0 = Höchste Ebene.
Region_id	integer	Werte: 0, 1, 101, 102, 103, 104, 105, ...	Eindeutige Schlüsselnummer der [Region].
Region	string	Werte: Deutschland, Schleswig-Holstein, Schleswig-Holstein Mitte, Schleswig-Holstein Nord, Schleswig-Holstein Ost, Schleswig-Holstein Süd, Schleswig-Holstein Süd-West, ...	Name der Region, für die der [Wert] berechnet wurde (siehe Keytable_Region.csv).

Regionsebene	string	Werte: Land , Bundesland , Raumordnungsregion	Bezeichnung der hierarchischen Ebene, in die die [Region] eingeordnet ist. Bsp.: Land = höchste Ebene.
Geschlecht	string	Werte: männlich , weiblich , Gesamt	Geschlecht (biologisch), für das der [Wert] berechnet wurde.
Altersgruppe	string	Werte: 1 - 4 , 10 - 14 , 100 + , 15 - 19 , 20 - 24 , 25 - 29 , 30 - 34 , ...	<p>Bezeichnung der Altersgruppe, für die der [Wert] berechnet wurde.</p> <p>Standardmäßig werden die Ergebnisse in 5-Jahres-Altersgruppen bis 95+, Gesamt und Altersstandardisiert angegeben. Daneben stehen für YLL zusätzliche Altersgruppen zur Verfügung: > 1 , 1 - 4 , 95 - 99 und 100 +</p> <p>Altersstandardisiert wurde nach der europäischen Standardbevölkerung 2013.</p>
Einheit	string	Werte: Absolut , Prozent , Rate	<p>Bezeichnung der Einheit, in der der [Wert] ausgegeben wird.</p> <p>Absolut : [Wert] als absoluter Zahlenwert</p> <p>Prozent : Ableitung des Absolutwertes als Anteil in Prozent (%) bezogen auf die Krankheitslast aller Ursachen (Ursache_id=1000) je Region, Altersgruppe und Geschlecht.</p> <p>Rate : Ableitung des Absolutwertes als Rate pro 100.000 Einwohner.</p>
Wert	number	Werte: ≥0	Berechneter Zahlenwert des zeilenbezogenen Datenpunktes.
UI_unten	number	Werte: ≥0	95% Unsicherheitsintervall: untere Grenze
UI_oben	number	Werte: ≥0	95% Unsicherheitsintervall: obere Grenze

Schlüsseltabellen

Schlüsseltabelle Krankheitslastursache

Die Systematik zur hierarchischen Einteilung der Krankheitslastursachen (Ursachen) ist an Vorarbeiten der GBD-Studie angelehnt [4]. Die Ursachen ([Ursache]) sind dabei in 5 Ebenen ([Ursachenebene]) untergliedert, wobei 0 die höchste (Aggregations-)Ebene mit einer einzigen Ausprägung, nämlich „Alle Ursachen“, darstellt. Auf Ebene 1 sind die Ursachen in 3 Hauptgruppen untergliedert und werden ab Ebene 2 immer feiner aufgespalten:

Ursachenebene	Ausprägungen
0	Alle Ursachen
1	Übertragbare Erkrankungen Verletzungen Nicht-übertragbare Erkrankungen
2	22 weitere distinkte Ursachen
3	136 weitere distinkte Ursachen

Die umfangreichste Ebene ist Ebene 3, die detailreichste ist Ebene 4. Nicht jede Ursache auf Ebene 3 lässt sich auf Ebene 4 weiter ausdifferenzieren. Jede Ursache ist genau einer Ursache auf der nächsthöheren Ebene zugeordnet. Die Zuordnung ist über das Feld [Ursache_id_Stamm] gegeben.

Die Felder [Ursache] und [Ursachenebene] sind in der Ergebnistabelle bereits integriert. Eine separate Zuordnungen zwischen [Ursache_id], [Ursache] und [Ursachenebene] befindet sich in der Schlüsseltabelle:

[Kontextmaterialien/Keytable_Ursache.csv](#)

Variablen und Variablenausprägung

Die Schlüsseltabelle Krankheitslastursache enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Name	Typ	Ausprägung	Beschreibung
Ursache_id	integer	>0	Eindeutige Schlüsselnummer der [Ursache].
Ursache	string	Keytable_Ursache.csv	Bezeichnung der die [Ursache_id] zugeordneten Krankheitslastursache.
Ursachenebene	integer	0, 1, 2, 3, 4	Nummer als Bezeichnung der hierarchischen Ebene (Level), in der die [Ursache] eingeordnet ist. Dabei 0 = höchste Ebene.
Ursache_id_Stamm	integer	>0	[Ursache_id] der hierarchisch übergeordneten [Ursache] auf der nächst-höheren [Ursachenebene].

Schlüsseltabelle Region

Die räumliche Dimension ([Region]) ist in 3 Ebenen ([Regionsebene]) untergliedert, wobei Land die höchste Ebene darstellt.

Die einzelnen Ebenen in der Übersicht:

Regionsebene	Ausprägungen
Land	Deutschland
Bundesland	16 Bundesländer
Raumordnungsregion	96 Raumordnungsregionen

Die kleinräumigste Ebene ist Raumordnungsregion. Jede Region ist genau einer Region auf der nächsthöheren Ebene zugeordnet. Die Zuordnung ist über das Feld [Region_id_Stamm] gegeben.

Die Felder [Region] und [Regionsebene] sind in der Ergebnistabelle bereits integriert. Eine separate Zuordnungen zwischen [Region_id], [Region] und [Regionsebene] befindet sich in der Schlüsseltabelle:

[Kontextmaterialien/Keytable_Region.csv](#)

Variablen und Variablenausprägung

Die Schlüsseltabelle Region enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Name	Typ	Ausprägung	Beschreibung
Region_id	integer	≥ 0	Eindeutige Schlüsselnummer der [Region].
Region	string	Keytable_Region.csv	Name der Region.
Regionsebene	string	Land Bundesland Raumordnungsregion	Bezeichnung der hierarchischen Ebene, in der die [Region] eingeordnet ist. Bsp.: Land = höchste Ebene.
Region_id_Stamm	integer	≥ 0	[Region_id] der hierarchisch übergeordneten [Region] auf der nächsthöheren [Regionsebene].

Formatierung der Daten

Die 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
- CSV-Trennzeichen: Komma „,”

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 Metadatendatei, 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](#)

In der zenodo.json ist neben der Publikationsdatum ("publication_date") auch der Datenstand in folgendem Format enthalten (Beispiel):

```
"dates": [  
  {  
    "start": "2023-09-11T15:00:21+02:00",  
    "end": "2023-09-11T15:00:21+02:00",  
    "type": "Created",  
    "description": "Date when the Dataset was created"  
  }  
],
```

Hinweise zur Nachnutzung der Daten

Offene Forschungsdaten des RKI werden auf [Zenodo.org](https://zenodo.org), [GitHub.com](https://github.com), [OpenCoDE](https://opencode.de) und [Edoc.rki.de](https://edoc.rki.de) bereitgestellt:

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

Lizenz

Der Datensatz "Ergebnisdatsatz BURDEN 2020 – Krankheitslast in Deutschland und seinen Regionen" ist lizenziert unter der [Creative Commons Namensnennung 4.0 International Public License | CC-BY 4.0 International](https://creativecommons.org/licenses/by/4.0/).

Die im Datensatz bereitgestellten Daten sind, unter Bedingung der Namensnennung des Robert Koch-Instituts 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.

Referenzen

[1] Murray CJL, Lopez AD: The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge: Harvard School of Public Health on behalf of the World Health Organization and the World Bank 1996.

[2] Roth GA, Abate D, Abate KH, et al.: Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018; 392: 1736–88.

[3] James SL, Abate D, Abate KH, et al.: Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018; 392: 1789–858.

[4] Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Cause, REI, and Location Hierarchies. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2020.