

# Open Data Sandbox

Vorname Name<sup>1</sup>

<sup>1</sup> Robert Koch-Institut

## Zitieren

Name, V. (2026). Open Data Sandbox [Data set]. Zenodo. <https://doi.org/10.5072/zenodo.318172>

## Zusammenfassung

## Inhaltsverzeichnis

- Informationen zum Datensatz und Entstehungskontext
  - Erhebung und Aufbereitung der Daten
  - Aufbau und Inhalt des Datensatzes
  - Hinweise zur Nachnutzung der Daten
  - Danksagung

## Informationen zum Datensatz und Entstehungskontext

Adipositas ist bereits im Kindesalter mit einem erhöhten Risiko für Folgeerkrankungen verbunden. Als beeinflussbare Gesundheitsrisiken spielen Übergewicht und Adipositas daher eine zentrale Rolle im Gesundheitsmonitoring.

Eine potenziell wertvolle Quelle von kleinräumigen Daten zu Übergewicht und Adipositas im Vorschulalter stellen Schuleingangsuntersuchungen (SEU) dar. Diese sind deutschlandweit verpflichtend und die Daten liegen auf Kreisebene vor.

Im Pilotprojekt AdiRaum wurde die Nutzbarkeit von SEU-Daten zu Übergewicht und Adipositas für das bundesweite Gesundheitsmonitoring geprüft und die Daten in einem Dashboard visualisiert.

Details finden sich in Kühnelt C, Starker A, Varnaccia G, Schienkiewitz A (2023) Schuleingangsuntersuchungen als kleinräumige Datenquelle für ein Monitoring der Kindergesundheit am Beispiel Adipositas. J Health Monit 8(2): 6–20. DOI [10.25646/11297](https://doi.org/10.25646/11297)

## Administrative und organisatorische Angaben

Das Pilotprojekt AdiRaum ("Erschließung kleinräumig aussagekräftiger Daten aus den Schuleingangsuntersuchungen für das bundesweite Gesundheitsmonitoring am Beispiel der Übergewichts- und Adipositasprävalenzen") wurde im Zeitraum 01.05.2022 bis 31.12.2022 durchgeführt und aus Sonderforschungsmitteln des Robert Koch-Instituts finanziert.

Das Nachfolgeprojekt [AdiRaum 2.0](#) wird derzeit für den Zeitraum 01.12.2023 bis 30.11.2025 vom [Bundesministerium für Gesundheit \(BMG\)](#) gefördert (Förderkennzeichen ZMII2-2523KIG913).

Die Projekte wurden in Zusammenarbeit mit den Landesgesundheitsämtern oder vergleichbaren Institutionen durchgeführt. Vertreter:innen der Gesundheitsberichterstattung der folgenden Länder waren aktiv an der Projektgestaltung und den Projektergebnissen beteiligt.

- Baden-Württemberg
- Bayern
- Berlin
- Brandenburg
- Bremen
- Hamburg
- Hessen
- Mecklenburg-Vorpommern
- Niedersachsen
- Nordrhein-Westfalen (Pilotprojekt)
- Rheinland-Pfalz
- Saarland
- Sachsen
- Sachsen-Anhalt
- Schleswig-Holstein
- Thüringen

An dieser Stelle werden die Daten der Länder veröffentlicht, die dem RKI die erforderlichen Nutzungsrechte eingeräumt haben.

Die Projektleitung und -koordination, das Datenmanagement und die Datenauswertung erfolgten durch das Fachgebiet [FG 27 | Gesundheitsverhalten](#).

Die Veröffentlichung der Daten, die Datenkuration sowie das Qualitätsmanagement der (Meta-)Daten erfolgten durch das Fachgebiet [MF 4 | Fach- und Forschungsdatenmanagement](#).

Fragen zum Datenmanagement und zur Publikationsinfrastruktur können an das Open Data Team des Fachgebiets MF4 unter [OpenData@rki.de](mailto:OpenData@rki.de) gerichtet werden.

## Erhebung und Aufbereitung der Daten

---

Der Datensatz enthält Daten aus Schuleingangsuntersuchungen und Daten zu gesundheitsrelevanten Risiko- und Schutzfaktoren aus öffentlichen Datenbanken, die auf Kreisebene verknüpft wurden.

### 1. Daten aus den Schuleingangsuntersuchungen (SEU)

Die Teilnahme an den SEU ist für Kinder vor der Einschulung verpflichtend. Aufgrund der föderalen Strukturen bilden die Landesgesetze, u.a. die Schulgesetze, die gesetzliche Grundlage. In der Regel werden die SEU vom Kinder- und Jugendgesundheitsdienst (KJGD) des öffentlichen Gesundheitsdienstes (ÖGD) durchgeführt. Ihr primäres Ziel liegt in der Einschätzung des Entwicklungs- und Gesundheitszustandes und schulrelevanter Stärken und Schwächen und der Feststellung von Förderbedarfen vor dem Schuleintritt. Unter anderem erfolgt eine standardisierte körperlich-medizinische Untersuchung, in der Körpergröße und -gewicht gemessen werden.

Mithilfe von Angaben zu Körpergröße und -gewicht, Alter und Geschlecht wird der BMI und daraus die Anzahl der Kinder mit Übergewicht (BMI-Perzentile > P90 nach Kromeyer-Hauschild) bzw. Adipositas (BMI-Perzentile > P97) bestimmt.

Im Rahmen der Projekte wurden dem RKI für die Kreise, kreisfreien Städte bzw. Bezirke ([Landkreis\\_Name](#), [Kreisschlüssel](#)) der teilnehmenden Länder ([Bundesland\\_Name](#)) für die verfügbaren Einschulungsjahrgänge

- die Anzahl der untersuchten Kinder ([Anzahl\\_Untersucht](#))
- die Anzahl der Kinder mit Übergewicht ([Anzahl\\_Uebergewicht](#))
- die Anzahl der Kinder mit Adipositas ([Anzahl\\_Adipositas](#))

differenziert nach Alter (4-7 Jahre, [Alter](#)) und Geschlecht ([Geschlecht](#)) übermittelt.

Daraus wurden je Kreis und Einschulungsjahrgang die Übergewichts- und Adipositasprävalenzen je Altersgruppe und Geschlecht bestimmt.

**Hinweis Geheimhaltungsregeln der Länder:** Für geringe Fallzahlen (größer 0) wurde unter Berücksichtigung der Geheimhaltungsregeln der Länder (z.B. ≤ 5) der Mittelpunkt imputiert.

### 2. Daten zu gesundheitsrelevanten Risiko- und Schutzfaktoren

Die Daten aus den SEU wurden mit Indikatoren aus verschiedenen Datenquellen ergänzt. Es handelt sich dabei um einen um Risiko- und Schutzfaktoren für Adipositas im Kindesalter, die im Rahmen einer systematischen Literaturrecherche im [AdiMon-Projekt](#) ("Bevölkerungsweites Monitoring adipositasrelevanter Einflussfaktoren im Kindes- und Jugendalter") identifiziert und in Abstimmung mit

den an AdiRaum teilnehmenden Ländern ausgewählt wurden. Zum anderen handelt es sich um siedlungsstrukturelle Merkmale, die einen Vergleich der Kreise ermöglichen sollen.

Es werden die folgenden Indikatoren dargestellt:

- Siedlungsstruktureller Kreistyp ( `Kreistyp` ): Dieser Indikator zur Siedlungsstruktur stammt vom Bundesinstitut für Bau-, Stadt- und Raumforschung
- Regionale sozioökonomische Deprivation des Landkreises ( `GISD` ): Dieser Indikator zur Soziodemografie stammt vom Robert Koch-Institut
- Situation von Kinderarmut im Landkreis ( `Bedarfsgemeinschaften` ): Dieser Indikator zur Soziodemografie stammt aus der Statistik der Bundesagentur für Arbeit
- Schulabgänger:innen ohne Schulabschluss ( `Ohne_Schulabschluss` ): Dieser Bildungs-Indikator stammt aus der Statistik der allgemeinbildenden Schulen
- Mitgliedschaften in Sportvereinen ( `Vereinsmitgliedschaft` ): Dieser Indikator zur Ausprägung der sportlichen Aktivität im Landkreis stammt von den Landessportbünden und dem Amt für Statistik Berlin Brandenburg
- Betreuungsumfang ( `Ganztagsbetreuung` ): Dieser Indikator stammt aus der Statistik der Kinder- und Jugendhilfe
- Verfügbarkeit von Bewegungsräumen im Landkreis ( `Erholungsflaeche` ): Dieser Indikator stammt von den Statistischen Ämtern der Länder und des Bundes

## Aufbau und Inhalt des Datensatzes

---

Der Datensatz enthält Daten zu Übergewicht und Adipositas aus den SEU der am Projekt beteiligten Bundesländer und die in der Datenverarbeitung unterstützenden Kontextmaterialien. Im Datensatz enthalten sind:

- Daten zu Adipositas und Übergewicht aus den Schuleingangsuntersuchungen
- Datenschema der Adipositasdaten
- Lizenz Datei mit der Nutzungs Lizenz des Datensatzes
- Datensatzdokumentation in deutscher Sprache
- Metadaten Datei zum Import in Zenodo

### Daten zu Adipositas aus den Schuleingangsuntersuchungen

Die Daten zu Übergewicht und Adipositas aus Schuleingangsuntersuchungen der Landkreise in den beteiligten Ländern sowie relevante kleinräumige Indikatoren sind im Hauptverzeichnis unter "Landkreise\_Adipositas\_Schuleingangsuntersuchungen.tsv" abrufbar.

| [Landkreise\\_Adipositas\\_Schuleingangsuntersuchungen.tsv](#)

### Variablen und Variablenausprägungen

Die Datei [Landkreise\\_Adipositas\\_Schuleingangsuntersuchungen.tsv](#) enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen. Ein maschinenlesbares Datenschema ist im [Data Package Standard](#) in [tableschema\\_Landkreise\\_Adipositas\\_Schuleingangsuntersuchungen.json](#) hinterlegt:  
> [tableschema\\_Landkreise\\_Adipositas\\_Schuleingangsuntersuchungen.json](#)

Variable	Typ	Ausprägungen	Beschreibung
Bundesland_Name	string	Werte: Baden-Württemberg , Bayern, Berlin, Brandenburg, Bremen, Saarland, Sachsen-Anhalt, ...	Bundesländer Deutschlands
Landkreis_Name	string	Werte: Aichach-Friedberg , Alb-Donau-Kreis , Altmarkkreis Salzwedel , Altötting, Amberg-Sulzbach , Anhalt-Bitterfeld , Ansbach, ...	Name des Landkreises bzw. der kreisfreien Stadt (inkl. Berliner Bezirke)
Landkreis_ID	string	Werte: 1001, 1002, 1003, 1004, 10041, 10042, 10043, ...  Fehlende Werte: leeres Feld	Identifikationsnummer des Landkreises basierend auf dem Amtlichen Gemeindeschlüssel zuzüglich der 12 Bezirke Berlins (11000001 bis 11000012)
Einschulungsjahr	integer	Werte: 2006 - 2025	Einschulungsjahrgang. Die verfügbaren Einschulungsjahrgänge, für welche die Länder Daten zur Verfügung stellten
Geschlecht	string	Werte: Junge, Mädchen  Fehlende Werte: Jahr nicht verfügbar , Daten des Kreises nicht übermittelt	Sex (biologisches Geschlecht). Ausprägung „divers“ wurde nicht in allen Ländern und somit nicht einheitlich bzw. vergleichbar erfasst.

Variable	Typ	Ausprägungen	Beschreibung
Anzahl_Untersucht	integer	Werte: $\geq 0$ Fehlende Werte: Altersgruppe nicht übermittelt , Daten des Kreises nicht übermittelt , Geheimhaltungs regel , Jahr nicht verfügbar	Anzahl untersuchter Kinder (mit vollständigen Daten zu Alter, Geschlecht, Körpergröße und Körpergewicht)
Anzahl_Uebergewicht	number	Werte: $\geq 0$ Fehlende Werte: Altersgruppe nicht übermittelt , Daten des Kreises nicht übermittelt , Geheimhaltungs regel , Jahr nicht verfügbar , keine untersuchten Kinder	Anzahl Kinder mit Übergewicht (>90. Perzentil nach Kromeyer Hauschild)
Anzahl_Adipositas	number	Werte: $\geq 0$ Fehlende Werte: Altersgruppe nicht übermittelt , Daten des Kreises nicht übermittelt , Geheimhaltungs regel , Jahr nicht verfügbar , keine untersuchten Kinder	Anzahl Kinder mit Adipositas (>97. Perzentil nach Kromeyer Hauschild)

Variable	Typ	Ausprägungen	Beschreibung
Alter	string	Werte: 4 Jahre, 5 Jahre, 6 Jahre, 7 Jahre Fehlende Werte: Jahr nicht verfügbar , Daten des Kreises nicht übermittelt	Alter der untersuchten Kinder
GISD	string	Werte: niedrig, mittel, hoch Fehlende Werte: geänderte Kreisstruktur , Jahr nicht verfügbar , Kontextfaktor nicht verfügbar	German Index of Socioeconomic deprivation. Kategorisierung basiert auf den drei Kategorien des GISD <a href="https://www.doi.org/10.25646/10640">https://www.doi.org/10.25646/10640</a> . Daten dazu unter <a href="https://www.doi.org/10.5281/zenodo.6840303">https://www.doi.org/10.5281/zenodo.6840303</a>
Kreistyp	string	Werte: dünn besiedelter ländlicher Kreis , kreisfreie Großstadt , ländlicher Kreis mit Verdichtungsan sätzen , städtischer Kreis Fehlende Werte: Jahr nicht verfügbar , geänderte Kreisstruktur	Siedlungsstruktureller Kreistyp (BBSR). Kategorisierung nach <a href="https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/kreise/siedlungsstrukturelle-kreistypen/kreistypen.html">https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/kreise/siedlungsstrukturelle-kreistypen/kreistypen.html</a>
Ganztagsbetreuung	number	Werte: 0 - 100 Fehlende Werte: leeres Feld	Anteil der in Ganztagsbetreuung (> 7h/Tag) betreuten Kinder an den Kindern gleichen Alters in % (3-5 Jahre).

Variable	Typ	Ausprägungen	Beschreibung
Ganztagsbetreuung_Kategorie	string	<p>Werte:</p> <p>&lt; 20% ,</p> <p>20 bis &lt; 40% ,</p> <p>40 bis &lt; 60% ,</p> <p>60 bis &lt; 80% ,</p> <p>≥ 80%</p> <p>Fehlende Werte:</p> <p>geänderte Kreisstruktur</p> <p>,</p> <p>Jahr nicht verfügbar</p> <p>,</p> <p>Kontextfaktor nicht verfügbar</p>	<p>Anteil der in Ganztagsbetreuung (&gt; 7h/Tag) betreuten Kinder an den Kindern gleichen Alters in % (3-5 Jahre). Diese 5 Kategorien wurden bereits im Ländermonitor Frühkindliche Bildungssysteme verwendet</p> <p><a href="https://www.laendermonitor.de/de/fokus-regionale-daten/kinder-und-eltern/betreuungsumfang-in-kitas-2">https://www.laendermonitor.de/de/fokus-regionale-daten/kinder-und-eltern/betreuungsumfang-in-kitas-2</a></p>
Ohne_Schulabschluss	number	<p>Werte: 0 - 100</p> <p>Fehlende Werte:</p> <p>leeres Feld</p>	<p>Anteil Schulabgänger:innen ohne Hauptschulabschluss an Schulabgänger:innen allgemeinbildender Schulen</p>
Ohne_Schulabschluss_Kategorie	string	<p>Werte:</p> <p>≤ 6% , &gt; 6%</p> <p>Fehlende Werte:</p> <p>geänderte Kreisstruktur</p> <p>,</p> <p>Jahr nicht verfügbar</p> <p>,</p> <p>Kontextfaktor nicht verfügbar</p>	<p>Anteil Schulabgänger:innen ohne Hauptschulabschluss an Schulabgänger:innen allgemeinbildender Schulen.</p> <p>Einteilung auf Grundlage des deutschlandweiten Medians in viel (Wert liegt über dem Median) oder wenig (Wert liegt unter dem Median)</p>
Vereinsmitgliedschaft	number	<p>Werte: ≥0</p> <p>Fehlende Werte:</p> <p>leeres Feld</p>	<p>Anzahl Mitgliedschaften in Sportvereinen je 100 Kinder</p>
Vereinsmitgliedschaft_Kategorie	string	<p>Werte:</p> <p>≤ 25 , &gt; 25</p> <p>Fehlende Werte:</p> <p>geänderte Kreisstruktur</p> <p>,</p> <p>Jahr nicht verfügbar</p> <p>,</p> <p>Kontextfaktor nicht verfügbar</p>	<p>Anzahl Mitgliedschaften in Sportvereinen je 100 Kinder.</p> <p>Einteilung auf Grundlage des deutschlandweiten Medians in viel (Wert liegt über dem Median) oder wenig (Wert liegt unter dem Median)</p>

Variable	Typ	Ausprägungen	Beschreibung
Bedarfsgemeinschaften	number	Werte: 0 - 100 Fehlende Werte: leeres Feld	Anteil Kinder in SGB-II-Bedarfsgemeinschaften an der Altersgruppe in %
Bedarfsgemeinschaften_Kategorie	string	Werte: <code>&lt; 5%</code> , <code>10 bis &lt; 15%</code> , <code>15 bis &lt; 20%</code> , <code>5 bis &lt; 10%</code> , <code>≥ 20%</code> Fehlende Werte: <code>geänderte Kreisstruktur</code> , <code>Jahr nicht verfügbar</code>	Anteil Kinder in SGB-II-Bedarfsgemeinschaften an der Altersgruppe in %. Diese 5 Kategorien wurde bis 2021 im Deutschlandatlas verwendet <a href="https://www.deutschlandatlas.bund.de/DE/Karten/Wie-wir-lernen/181/_node.html">https://www.deutschlandatlas.bund.de/DE/Karten/Wie-wir-lernen/181/_node.html</a>
Erholungsflaeche	number	Werte: ≥0 Fehlende Werte: leeres Feld	Sport-, Freizeit- und Erholungsfläche (ha) je 1.000 Kinder
Erholungsflaeche_Kategorie	string	Werte: <code>&gt; 82</code> , <code>≤ 82</code> Fehlende Werte: <code>geänderte Kreisstruktur</code> , <code>Jahr nicht verfügbar</code> , <code>Kontextfaktor nicht verfügbar</code>	Sport-, Freizeit- und Erholungsfläche (ha) je 1.000 Kinder. Einteilung auf Grundlage des deutschlandweiten Medians in <code>viel</code> (Wert liegt über dem Median) oder <code>wenig</code> (Wert liegt unter dem Median)

## Formatierung

Die Daten sind im Datensatz als .xlsx-Datei und Tab-separierte .tsv-Datei enthalten. Der verwendete Zeichensatz der .tsv-Datei ist UTF-8.

- Zeichensatz: UTF-8
- .tsv-Trennzeichen: Tab "

## 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](https://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 dem 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 published data was created"
  }
],
```

Zusätzlich beschreiben wir tabellarische Daten mithilfe des [Data Package Standards](#). Ein Data Package ist eine strukturierte Sammlung von Daten und zugehörigen Metadaten, die den Austausch und die Wiederverwendung von Daten erleichtert. Es besteht aus einer `datapackage.json`-Datei, die zentrale Informationen wie die enthaltenen Ressourcen, ihre Formate und Schema-Definitionen beschreibt.

Der Data Package Standard wird von der [Open Knowledge Foundation](#) bereitgestellt und ist ein offenes Format, das eine einfache, maschinenlesbare Beschreibung von Datensätzen ermöglicht.

Die Liste der in diesem Repository enthaltenen Daten ist in folgender Datei hinterlegt:

#### datapackage.json

Für tabellarische Daten definieren wir zusätzlich ein [Table Schema](#), das die Struktur der Tabellen beschreibt, einschließlich Spaltennamen, Datentypen und Validierungsregeln. Diese Schema-Dateien finden sich unter:

#### Metadaten/schemas/

## 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 "Open Data Sandbox" 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, 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.

## Danksagung

---

Für die Erhebung und Bereitstellung der Daten und der Untertützung des Projektes danken wir den Gesundheitsämtern bzw. den Kinder- und Jugendgesundheitsdiensten der Kreise und kreisfreien Städte in Baden-Württemberg, Bayern, Berlin, Brandenburg, Bremen, Hamburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, im Saarland, Sachsen, Sachsen-Anhalt, und Schleswig-Holstein und Thüringen.