

## Leithema

Onkologie 2024 · 30:280–288  
<https://doi.org/10.1007/s00761-023-01426-6>  
Angenommen: 5. Oktober 2023  
Online publiziert: 25. Oktober 2023  
© The Author(s) 2023



# Forschung mit Krebsregistern – was geht, was (noch) nicht?

Alexander Katalinic<sup>1,2</sup> · Ron Pritzkuleit<sup>2</sup> · Klaus Kraywinkel<sup>3</sup>

<sup>1</sup> Institut für Sozialmedizin und Epidemiologie, Universität zu Lübeck, Lübeck, Deutschland

<sup>2</sup> Krebsregister Schleswig-Holstein, Lübeck, Deutschland

<sup>3</sup> Zentrum für Krebsregisterdaten, Robert Koch-Institut, Berlin, Deutschland

## In diesem Beitrag

### – Aufgaben der Register

### – Methoden

Daten der Krebsregister • Nutzungsanalyse von Krebsregisterdaten • Ermittlung von Forschungsmöglichkeiten mit Krebsregisterdaten

### – Ergebnisse

Wissenschaftliche Nutzung von Krebsregisterdaten • Möglichkeiten zur wissenschaftlichen Nutzung von Krebsregisterdaten • Rechtliche Grundlagen und Antragsverfahren

### – Diskussion

## Zusammenfassung

**Hintergrund:** Eine der Aufgaben der Landeskrebsregister ist die Bereitstellung von Daten für Forschung oder eigene Forschung. Daten werden im Krebsregister nicht primär für Forschung, sondern mit Meldepflicht und ohne Einwilligung der Betroffenen für die Erfüllung gesetzlicher Aufgaben erhoben. Sind Daten der Krebsregister für Forschung sinnvoll nutzbar, werden sie genutzt und welche Rahmenbedingungen zur Datennutzung sollten nachgebessert werden?

**Methoden:** Mittels PubMed®-Recherche wurde die Nutzung von Krebsregisterdaten in wissenschaftlichen Publikationen für Deutschland und für das US-SEER-Programm (Surveillance, Epidemiology, and End Results Program) ermittelt. Aus Landes- und Bundesgesetzen wurden Möglichkeiten zur Datennutzung ermittelt.

**Ergebnisse:** Für Deutschland wurden 2268 Artikel mit Krebsregisterdaten identifiziert, für 2022 etwa 250 Publikationen. Aus dem US-SEER-Programm gehen derzeit über 1500 Publikationen pro Jahr hervor. Aus den Krebsregistergesetzen wurden folgende wesentliche Nutzungsarten identifiziert: Bereitstellung und Nutzung von aggregierten und Einzelfalldaten, Kontaktierung von Krebskranken aus dem Registerbestand und Abgleich bestehender Kohorten mit dem Krebsregister. Während sich aggregierte und Einzelfalldaten i. d. R. mittels standardisierten Formulars beantragen lassen, erfordern die andern Nutzungsarten übliche wissenschaftliche Anträge mit Ethikvotum und ggf. weiteren Genehmigungen. Ausgewählte Praxisbeispiele zeigen die verschiedenen Nutzungsarten.

**Schlussfolgerung:** Die steigende Zahl an Publikationen belegt eine rege Forschungsnutzung. Dennoch ist in Deutschland das Potenzial nicht ausgereizt. Bundesländerübergreifende Projekte mit Verlinkung zu anderen Daten sind sehr aufwendig oder wegen bürokratischer Hürden oft nicht realistisch durchführbar. Mit Projekten und Gesetzesinitiativen wird derzeit versucht, solche Datennutzungen zu vereinfachen.

## Schlüsselwörter

Onkologie · Krebsregister · Onkologischer Basisdatensatz · Datenanalyse · Datensysteme

Sind Daten der Krebsregister für wissenschaftliche Projekte in der Onkologie sinnvoll nutzbar und werden Krebsregisterdaten in Deutschland für Forschung genutzt? Welche prinzipiellen Nutzungsmöglichkeiten gibt es? Und welche Rahmenbedingungen sollten nachgebessert werden, um Krebsregisterdaten wissenschaftlich besser nutzen zu können? Diese Fragen werden im Folgenden beantwortet und diskutiert.

## Aufgaben der Register

Landeskrebsregister haben originäre, gesetzliche Aufgaben im Bereich der öffentlichen Gesundheit, wie die Ermittlung der Krankheitslast an Krebs, Trends, Untersuchung verdächtiger Häufungen, Bewertung von Primärprävention und Früherkennung und in der Qualitätssicherung der onkologischen Versorgung von Patientinnen und Patienten [1–3]. Diese Aufgaben erfüllen die Register in einigen



QR-Code scannen & Beitrag online lesen

Bereichen schon seit vielen Jahren, z.T. seit Jahrzehnten, zuverlässig und in hoher Qualität. Damit ist Potenzial der Registerdaten aber noch nicht ausgeschöpft [4]. Die Forschung mit den Daten durch die Krebsregister selbst und die Bereitstellung der Daten und methodischer Expertise für Forschungsprojekte Dritter sind weitere wesentliche Aufgaben der Register [5]. Die Bereitstellung der mit großem Aufwand und öffentlichen Mitteln erhobenen Daten für eine wissenschaftliche Nutzung ist nicht nur in Bundes- und Landesgesetzen gesetzlich verankert, sondern auch ethisch geboten. Viele Forschungsfragen lassen sich entweder ausschließlich mit Krebsregisterdaten beantworten, oder Krebsregisterdaten bieten erhebliche Vorteile.

### » Da es sich um eine Vollerhebung handelt, entfällt das Problem der Stichprobenrepräsentativität

Da es sich um eine Vollerhebung handelt, entfällt das Problem der Stichprobenrepräsentativität. Die Daten sind bevölkerungsbezogen, repräsentativ und über lange Zeiträume vorhanden. Auch seltene Krebserkrankungen, seltene Patientengruppen oder seltene Ereignisse (z.B. Therapien) lassen sich sinnvoll beforschen. Der – neben der Lebensqualität – wichtigste Endpunkt in der Onkologie, das Versterben, wird systematisch und zeitnah erhoben.

Es muss aber berücksichtigt werden, dass Krebsregister keine forschungsgetriebenen Register sind, sondern für die Erfüllung definierter gesetzlicher Aufgaben eingerichtet wurden und dazu einen zwar klar definierten und konsentierten, aber relativ starren Datensatz nutzen. Das hat Implikationen für die wissenschaftliche Nutzung. Die Tiefe der Daten ist häufig nicht so hoch oder nicht so exakt, wie es in einer eigens geplanten Studie mit spezifischen Instrumenten möglich wäre. Eine weitere Limitation der Krebsregisterdaten ist, dass die Daten der an Krebs erkrankten Personen i.d.R. mit Meldepflicht, meist ohne (komplettes) Widerspruchsrecht und damit unter Einschränkung des Grundrechts auf informationelle Selbstbestimmung erhoben werden, womit die Daten wesentlich strenger datenschutzrechtlichen Restriktionen unterliegen. Dies ist

ein fundamentaler Unterschied zur klinischen Forschung am Menschen, die eine informierte Entscheidung zur Teilnahme und Datennutzung erfordert. Dieser Punkt kann die Nutzung von Registerdaten erschweren oder sogar einschränken.

Sind Daten der Krebsregister unter diesen Prämissen überhaupt für wissenschaftliche Projekte in der Onkologie sinnvoll nutzbar und wurden bzw. werden Krebsregisterdaten in Deutschland für Forschung genutzt? Welche prinzipiellen Nutzungsmöglichkeiten gibt es? Und welche Rahmenbedingungen sollten nachgebessert werden, um Krebsregisterdaten wissenschaftlich besser nutzen zu können? Diese Fragen werden im Folgenden beantwortet und diskutiert.

## Methoden

### Daten der Krebsregister

Um Daten der Krebsregister sinnvoll und effektiv nutzen zu können, müssen die Eigenschaften der Daten, deren Qualität sowie die Erhebungswege und -instrumente berücksichtigt werden. Ansonsten kann es bei Auswertungen zu schwerwiegenden Fehlinterpretationen kommen. Die folgenden Punkte sollten bei der Planung und Durchführung von Projekten mit Krebsregisterdaten und deren Interpretation berücksichtigt werden: Landeskrebsregister haben heute die Aufgabe, alle Krebserkrankten, die ihrem Einzugsgebiet (i. d. R. das Bundesland) wohnhaft sind bzw. deren Erkrankungen im Einzugsgebiet diagnostiziert oder behandelt wurden, zu erfassen. Hierdurch ergeben sich 2 Sichtweisen auf den Datenbestand der Register, einmal wohnort- und einmal behandlungsortbezogen, mit überschneidenden, aber unterschiedlichen Fällen und Fallzahlen. Entscheidend für beide Sichtweisen ist ein funktionierender registerübergreifender Datenaustausch, um Vollzähligkeit und Vollständigkeit der Daten zu gewährleisten. Außerdem ist zu beachten, dass bei einem Patienten mehrere Tumoren gemeldet werden können und dass jeder Tumor wiederum aus mehreren Meldungen (meist 5–10 Meldungen zu Diagnose, Therapie, Rezidiv, Tod usw.) bestehen kann. Die Summe aller Meldungen zu einem Tumor bildet den „Best-of-Datensatz“ in seinem

Hier steht eine Anzeige.



<b>Tab. 1</b> Vereinfachte Darstellung des onkologischen Basisdatensatzes und seiner Module (oBDS 3.0)	
<b>Basisdatensatz</b>	Datenfelder (etwa 130)
Personendaten	Krankenversicherungsnummer, Namen, Adresse, Geburtsdatum, Geschlecht
Angaben zur Meldestelle	Einrichtung, Name, Adresse
Tumordiagnose	ICD-10, ICD-O-Lokalisation, Diagnosedatum, Diagnosesicherung, Seitenlokalisation
Histologie	ICD-O-Morphologie, Grading, Anzahl Lymphknoten untersucht und befallen
Tumorklassifikation	TNM und weitere Klassifikationen
Genetische Variante	Art
Residualstatus	Nach Operation und Gesamtbewertung
Performance-Status	ECOG (bei Diagnosestellung)
Operation	Datum, Intention, OPS-Kodierung, Komplikationen
Strahlentherapie	Datum, Intention, Stellung zur Operation, Zielgebiet, Beginn, End, Dosis, Boost, Komplikationen (CTCAE)
Systemische Therapie	Datum, Intention, Stellung zur Operation, Art der Therapie (einschließlich aktiver Überwachung), Protokoll, Substanz, Beginn, Ende
Tumorkonferenz	Datum, Art
Follow-up	Datum, Gesamtbeurteilung von Tumor, Lymphknoten und Metastasen
Tod	Datum, Ursache
<i>Organspezifische Module</i>	Spezifische Inhalte
Brustkrebs	10 Variablen
Prostatakarzinom	10 Variablen
Darmkrebs	12 Variablen
Melanom der Haut	4 Variablen
Eine ausführliche Beschreibung der einzelnen Datenfelder findet sich unter: [6]	
CTCAE Common Terminology Criteria for Adverse Events; ICD-O International Classification of diseases – Oncology, OPS Operationen- und Prozedurenschlüssel	

Gesamtverlauf ab. Für Forschungsprojekte ist i. d. R. der „Best-of-Datensatz“ der Tumorebene relevant, aber für besondere Fragestellungen (z. B. Abgrenzung von First-, Second-, Third-Line-Therapien) kann auch eine Auswertung auf Meldungsebene notwendig werden.

### » Wichtig für die Forschungsnutzung sind Kenntnisse über die Datenqualität

Für die Planung eines Forschungsprojekts mit Krebsregisterdaten sollte man sich zunächst mit den Dateninhalten und Ausprägungen vertraut machen. Einen Überblick über den onkologischen Basisdatensatz (oBDS) und seine Module gibt **Tab. 1**. Pro Tumorerkrankung werden über 130 verschiedene Einzelangaben erfasst, die bei längeren Krankheitsverläufen auch mehrfach erhoben werden (Therapien, Folgeereignisse). Ausführliche

Informationen zum Datensatz finden sich auf der Web-Seite des oBDS [6].

Wichtig für die Forschungsnutzung sind Kenntnisse über die Datenqualität, insbesondere Vollzähligkeit (Erfassung möglichst aller Neuerkrankungsfälle in der Bezugsregion), Vollständigkeit (alle Angaben, die erhoben werden sollen, liegen dem Register auch vor) und Plausibilität (die Angaben zu einem Tumor sind in sich konsistent und in der zeitlich richtigen Reihenfolge). Für Studien zu klassisch epidemiologischen Fragestellungen, wie zu Inzidenz, Prävalenz, Trends oder relativem Überleben, ist eine hohe Vollzähligkeit erforderlich. Für die Beantwortung klinischer Fragestellungen ist die Vollständigkeit und Plausibilität der Daten meistens von großer Relevanz.

## Nutzungsanalyse von Krebsregisterdaten

Um die wissenschaftliche Nutzung von Krebsregisterdaten abzuschätzen, wurde eine orientierende Literatursuche mittels der National Library of Medicine (PubMed®) für deutsche und US-amerikanische Krebsregister durchgeführt. Als Suchstring wurde für deutsche Krebsregister („cancer registry“ and „Germany“ or „GEKID“ or „ZfKD“ or „German Childhood Cancer Registry“ or „Krebsregister“) und für die US-Register („SEER“ and „Cancer“) verwendet (Datum der Suche 28.07.2023). Zusätzlich wurde die Zahl der bewilligten Anträge für die wissenschaftliche Nutzung des Forschungsdatensatzes des Zentrums für Krebsregisterdaten (ZfKD), ein aus allen Krebsregistern bundesweit zusammengeführter Datensatz, ermittelt. Dazu wurde das öffentliche Register des ZfKD genutzt, das im Jahr 2021 eingerichtet wurde [7].

### Ermittlung von Forschungsmöglichkeiten mit Krebsregisterdaten

Mittels PubMed® und Google Scholar wurde eine explorative und themengeleitete Literatursuche mit den Suchworten „Krebsregister“ und „Forschung“ (Deutsch und Englisch) durchgeführt. Zusätzlich wurden Bundes- und Landeskrebsregistergesetze hinsichtlich der Nutzung der Registerdaten zur Forschung und zum Datenzugang analysiert.

## Ergebnisse

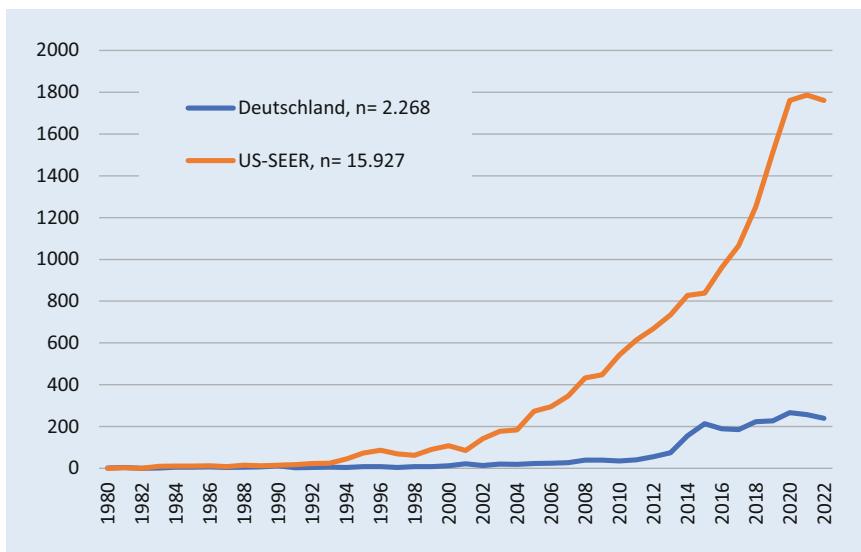
### Wissenschaftliche Nutzung von Krebsregisterdaten

Für Deutschland wurden 2268 Artikel mit Krebsregisterdaten in PubMed® identifiziert. Der Erste stammt aus dem Jahr 1965. Wildner und Eckert präsentieren eine Analyse von 1490 Fällen mit Lymphogranulomatose aus dem Krebskrankenregister der ehemaligen DDR [8]. Ab 2000 steigt die Zahl der Publikationen langsam an und liegt im Jahr 2014 erstmal über 100 pro Jahr.

Aktuell erscheinen etwa 250 Publikationen pro Jahr mit Bezug zu deutschen Krebsregisterdaten (**Abb. 1**). Für die USA wurde nur nach Publikationen mit

Hier steht eine Anzeige.





**Abb. 1 ▲** Anzahl an Publikationen pro Jahr, gelistet in der National Library of Medicine (PubMed®), mit Nennung von Krebsregistern in Deutschland bzw. der SEER-Datenbank (Surveillance, Epidemiology, and End Results Program) der US-amerikanischen Krebsregister

der zentralen Datenbank des SEER-Programms (Surveillance, Epidemiology, and End Results Program) [9], welches die Daten der SEER-Krebsregister zusammenfasst, gesucht. Erste Publikationen starten hier um das Jahr 1980. Schon im Jahr 2000 überschreitet die Anzahl der SEER-Publikationen die Marke von 100 und steigt dann rasant auf aktuell über 1700 Publikationen pro Jahr.

### » Aktuell erscheinen etwa 250 Publikationen pro Jahr mit Bezug zu deutschen Krebsregisterdaten

Für das Jahr 2022 fanden sich im öffentlichen Register des ZfKD 25 Forschungsanträge.

### Möglichkeiten zur wissenschaftlichen Nutzung von Krebsregisterdaten

In **Tab. 2** sind die prinzipiellen Möglichkeiten zur wissenschaftlichen Nutzung der Krebsregisterdaten zusammengefasst.

Im einfachsten Fall werden für ein Forschungsprojekt *aggregierte Daten* benötigt, also aufsummierte Fallzahlen nach definierten Eigenschaften, z.B. Fallzahlen nach Geschlecht, 5-Jahres-Altersgruppen und Diagnosejahr aggregiert. Solche Daten sind entweder direkt über die Krebsregisterdatenbanken abrufbar oder auf An-

frage von den Registern erhältlich. Eine formelle Antragstellung zur Nutzung der Daten i. d. R. nicht erforderlich, es sollte aber auf die richtige Zitierung der Daten geachtet werden.

Die meisten wissenschaftlichen Fragestellungen erfordern Datensätze mit einzelnen Fällen und ausgewählten Variablen aus dem oBDS. *Einzelfalldaten* (ohne Angabe von Namen oder Adressen der Krebskranken) können von den Krebsregistern und vom ZfKD auf Antrag bereitgestellt werden. Im Antrag sind die Forschungsfrage, Auswertungsmethodik, Ein- und Ausschlusskriterien für die Fallselektion (u.a. Diagnosen, Zeitraum) und die erforderlichen Variablen klar zu benennen. In **Tab. 3** wird ein Beispiel aus einem von der Krebshilfe geförderten Projekt zum Krebsüberleben gezeigt [10].

In einigen Fällen ist es erforderlich, in Forschungsprojekten an Krebs *erkrankte Personen* zu kontaktieren, um beispielsweise Fragen zu Krebsursachen, zur Lebensqualität oder Langzeitfolgen stellen zu können. Die meisten Krebsregister dürfen für solche Fragestellungen die Personendaten von ausgewählten Krebskranken heraussuchen und bereitstellen. Im Praxisbeispiel (**Tab. 3**) wurden Personen mit Darmkrebs aus dem Krebsregister identifiziert und von Forschergruppe zur Abgabe einer Bioprobe kontaktiert [11].

In Forschungsprojekten, in denen eine Gruppe von Personen (Kohorte) namentlich bekannt ist, kann es von Interesse sein, an Krebs erkrankte Personen im Krebsregister zu identifizieren und Informationen für diese Kohortenmitglieder aus dem Krebsregisterdatensatz zu ergänzen. Für diesen *Kohortenabgleich* übermittelt das Forschungsprojekt Namen und Adressen der Kohortenmitglieder an das Krebsregister, welches die übermittelten Angaben mit dem Registerbestand abgleicht. Ist eine Person der Kohorte im Register erfasst, kann das Register Informationen zu Krebskrankungen und ggf. zur Therapie an das Forschungsprojekt zurückmelden. In der Regel ist es für den Kohortenabgleich erforderlich, dass die Kohortenmitglieder in den Abgleich eingewilligt haben. Ein Beispiel wäre die Frage, ob und wann Kohortenmitglieder im zeitlichen Verlauf an Krebs erkrankt sind und wann und welche Erkrankung auftrat, ggf. wie diese behandelt wurde. Die **Tab. 3** zeigt hier ein Praxisbeispiel [12]. Die Verlinkung von externen Datenquellen mit den Krebsregisterdaten ist als eine Sonderform des Kohortenabgleichs zu sehen, wie das zweite Praxisbeispiel zeigt (**Tab. 3**, [13]).

### » Die Verlinkung externer Datenquellen mit Krebsregisterdaten ist eine Sonderform des Kohortenabgleichs

Neben den genannten Nutzungsmöglichkeiten gibt es einige Sonderfälle an der Grenze zwischen Forschung und Qualitätssicherung. Krebsregister können Meldestellen bei vorhandener gesetzlicher Grundlage das Follow-up (i. Allg. Versterben oder Rezidiv) der eigenen gemeldeten Patienten bereitstellen. Für die Evaluation des Mammographie-Screenings, z.B. zur Identifizierung von Intervallkarzinomen (**Tab. 3**, [14]) und für die Früherkennung von Darm- und Gebärmutterhalskrebs existieren landes- bzw. bundesgesetzliche Regelungen zur Verknüpfung von Screening- und Krebsregisterdaten.

### Rechtliche Grundlagen und Antragsverfahren

Die rechtlichen Grundlagen für die aufgeführten Nutzungsmöglichkeiten finden

**Tab. 2** Übersicht über grundsätzliche Forschungsmöglichkeiten mit Krebsregisterdaten

Nutzungsarten	Beschreibung	Üblicher Datenzugang
Aggregierte Daten	Fallzahlen in definierten Gruppen, z.B. nach Diagnose, Alter, Geschlecht, Region und Jahr	Internetangebote der KR und des ZfKD oder formlose Anfrage
Einzelfalldaten	Einzelfälle, nach definierten Ein- und Ausschlusskriterien mit definierten Variablen-Set	Standardisierter Kurzantrag (Formular) an KR oder ZfKD
Kontaktierung von Krebskranken	Kontaktaufnahme zu (lebenden) Patientinnen und Patienten zur Rekrutierung für Studien	Antrag an KR mit Studienprotokoll, Aufklärung und Einwilligungsformular, Votum Ethikkommission und ggf. Datenschutz
Kohortenabgleich	Abgleich einer Forschungskohorte (i.d.R. mit Einwilligung) zur Feststellung, ob Kohortenmitglieder an Krebs erkrankt oder an/mit Krebs verstorben sind	Antrag an KR mit Studienprotokoll, Aufklärung und Einwilligungsformular, Votum Ethikkommission und ggf. Datenschutz
<i>Sonderfälle</i>		
Bereitstellung von Daten für meldende Einrichtungen	Zurückspielen der kompletten Information zu gemeldeten Tumorfällen (inkl. Follow-up und Versterben)	Gesetzlich geregelt, auf Anfrage
Abgleich Mammographie-Screening (MS)	Abgleich der Teilnehmerinnen am Mammographie-Screening zur Identifikation von Intervallkarzinomen im MS-Programm	Gesetzlich auf Landesebene geregelt
Abgleich Darmkrebs- und Zervix-Screening	Abgleich mit Krankenkassendaten durch Vertrauensstelle nach § 299 SGB V	Gesetzlich auf Bundes- und Landesebene geregelt

KR Landeskrebsregister, SGB V Sozialgesetzbuch V, ZfKD Zentrum für Krebsregisterdaten am Robert Koch-Institut

**Tab. 3** Ausgewählte Beispiele für die wissenschaftliche Nutzung von Krebsregisterdaten nach den in **Tab. 2** genannten Nutzungsarten

Nutzungsarten	Projekt und ausgewählte Publikation	Beschreibung
Einzelfalldaten	Überleben nach Krebs in Deutschland <i>Survival from common and rare cancers in Germany in the early 21st century</i> [10]	Bevölkerungsbezogene Einzelfalldaten aus allen Landeskrebsregistern wurden zusammengefasst (1,1 Mio. Krebserkrankungen von 1961–2006) und hinsichtlich des Krebsüberlebens analysiert. Insgesamt sind etwa 50 Publikationen in diesem Themenbereich entstanden
Kontaktierung von Krebskranken	POPGEN <i>Genetic investigation of DNA-repair pathway genes PMS2, MLH1, MSH2, MSH6, MUTYH, OGG1 and MTH1 in sporadic colon cancer</i> [11]	Durch ein Landeskrebsregister wurden Fälle mit Darmkrebs identifiziert und der Forschergruppe Namen und Adressen der Betroffenen übermittelt. Diese konnten angesprochen und so eine Tumorprobe identifiziert und untersucht werden
Kohorten-abgleich/ Verlinkung von Daten mit Krebsregisterdaten	Qualitätsgesicherte Mammdiagnostik <i>Impact of the Quality assured Mamma Diagnostic (QuaMaDi) programme on survival of breast cancer patients</i> [12]	Eine Kohorte von 155.658 Frauen, die von 2001–2007 an einer Qualitätsgesicherten Mammographie teilnahmen, wurde an das Landeskrebsregister übermittelt, abgeglichen und Daten zu Treffern an die Forschergruppe weitergegeben
	Krebsinzidenz nach Diabetes <i>Cohort study of cancer incidence in patients with type 2 diabetes: record linkage of encrypted data from an external cohort with data from the epidemiological Cancer Registry of North Rhine-Westphalia</i> [13]	67.447 Versicherte einer Krankenkasse mit Diabetes wurden mit den Daten eines Landeskrebsregisters abgeglichen
Abgleich Mammographie-Screening (MS)	Intervallkarzinome im Mammographie-Screening <i>Time trend in programme sensitivity within the German mammography screening programme in North Rhine-Westphalia and Lower Saxony</i> [14]	Fast 4 Mio. Untersuchungen im Mammographie-Screening (2006–2011) wurden mit 2 Krebsregistern abgeglichen (im Sinne eines Kohortenabgleichs), um Intervallkarzinome zu identifizieren

sich in den jeweiligen Landesgesetzen, für die auf nationaler Ebene zusammengeführten Daten in Bundesgesetzen. Ob sich eine der genannten Nutzungsmöglichkeit tatsächlich umsetzen lässt, ist im konkreten Fall mit den zuständigen Krebsregistern oder dem ZfKD zu klären. Aggregierte Daten und anonymisierte Einzelfalldaten können bei den Landeskrebsregistern oder beim ZfKD beantragt werden. For-

schungsvorhaben, die Angaben zur Person voraussetzen, ist eine Datenbeschaffung über das ZfKD nicht möglich, sondern sie muss über die Landeskrebsregister erfolgen. Die Kontaktierung von Krebskranken ist in vielen Bundesländern möglich, wobei sich die Zugangswege unterscheiden. In einigen Ländern müssen zur Kontaktierung die meldenden Ärztinnen und Ärzte oder das Krebsregister eine Ein-

willigung bei den Betroffenen einholen, in anderen Ländern können die Adressen von registrierten Krebserkrankten direkt an die antragsstellende Forschungsstelle übermittelt werden. Der Abgleich von Personen aus einer Forschungskohorte mit dem Bestand eines Krebsregisters ist, bei vorliegender Einwilligung der abzugleichenden Person, in den meisten Krebsregistern möglich.

## Leitthema

Zur Beantragung von Daten aus den Landeskrebsregistern wurde von den Krebsregistern ein einheitliches Antragsformular entwickelt [15], ein vergleichbares Formular gibt es für die Beantragung vom Daten beim ZfKD [16]. In die Genehmigung der Anträge sind in vielen Ländern und beim ZfKD beratende Gremien (z.B. wissenschaftlicher Beirat) eingebunden. Bei Anträgen mit Personendaten, wie die Kontaktierung von Krebskranken oder dem Kohortenabgleich, sind i.d.R. ein ausführliches wissenschaftliches Studienprotokoll, das Votum einer Ethikkommission und ggf. weitere Genehmigungen (z.B. Datenschutzbehörde, Ministerium) erforderlich.

## Diskussion

Zunächst wurde gezeigt, dass Krebsregisterdaten in Deutschland bereits intensiv für die onkologische Forschung genutzt werden. Die Nutzungs frequenz (etwa 250 Publikationen pro Jahr) hat sich in den letzten Jahren deutlich gesteigert. Die Gründe für die stärkere Nutzung von deutschen Krebsregisterdaten sind vielfältig. Die Flächendeckung mit bevölkerungsbezogenen Krebsregister ist immer weiter angestiegen und deckt seit 2009 alle Bundesländer ab. Mit dem Bundeskrebsregisterdatengesetz [17] konnten ab 2009 alle epidemiologischen Daten der Krebsregister am Zentrum für Krebsregisterdaten zentralisiert werden, was die Beantragung von Datensätzen für ganz Deutschland deutlich erleichtert hat und gut mit dem Anstieg von Krebsregister-Publikationen in Deutschland übereinstimmt. Durch die Einführung einer bundesweiten zusätzlichen klinischen Krebsregistrierung ab dem Jahr 2013 [3] wurde die Attraktivität der Krebsregisterdaten für Forschung weiter gesteigert. Ab 2023 ist ein großer Teil der neuen klinischen Daten auch über das Zentrum für Krebsregisterdaten verfügbar.

Insofern ist abzusehen, dass die Nutzung von Krebsregisterdaten sich weiter positiv entwickeln wird. Trotzdem ist aber festzuhalten, dass die Nutzung der deutschen Krebsregisterdaten noch nicht den Grad erreicht hat wie bei den ähnlichen Daten aus dem US-SEER-Programm [9]. Das hat sicherlich mit der längeren Verfügbar-

keit kompletter klinischer Daten und der Bekanntheit des SEER-Programms zu tun.

Weiter ist festzustellen, dass in den verschiedenen Krebsregistergesetzen der Länder Grundlagen für die wissenschaftliche Nutzung von Krebsregisterdaten gelegt wurden. Dazu gehören auch die Möglichkeit, aus dem Registerbestand bestimmte Patientengruppen zu identifizieren (z.B. bestimmte Diagnosen, bestimmtes Alter, mit bestimmten Therapien oder mit einem bestimmten Folgeereignis) und diese direkt zu kontaktieren oder bereits bestehende Forschungskohorten mit dem Krebsregister abzugleichen. Praxisbeispiele zeigen, dass diese Forschungsmöglichkeiten auch genutzt werden. Die deutsche Gesundheitsstudie (NAKO, ehemals Nationale Kohorte, [18]) mit über 200.000 eingeschlossenen Personen nutzt beispielsweise den Kohortenabgleich zur Identifikation von neuen Krebskrankungen im Verlauf.

### » Es ist abzusehen, dass die Nutzung von Krebsregisterdaten sich weiter positiv entwickeln wird

Auch wenn die mit den Krebsregistern geschaffene Datenbasis und die Nutzungsmöglichkeiten bei kaum einem anderen Krankheitsbild vorhanden sind, gibt es weiteres Potenzial für eine effektivere und effizientere Forschungsnutzung. Dazu sind aber einige persistierende Problempunkte zu lösen:

*Länderübergreifende Projekte*, für welche der ZfKD Datensatz nicht ausreichend ist (das ZfKD erhält nur einen Teildatensatz aus den Landeskrebsregistern), sind äußerst aufwendig. Selbst eine einfache Datenstudie würde 15 separate und unterschiedliche Anträge (15, weil Berlin und Brandenburg ein gemeinsames Register betreiben) an die jeweiligen Landeskrebsregister bedeuten. Im Idealfall erhält man dann 15 verschiedene Bewilligungen (mit unterschiedlicher Bearbeitungsdauer und ggf. Auflagen) und 15 Einzeldatensätze. Bei komplexeren Studien, wie Patientenkontakte oder Kohortenabgleichen, sind neben den Anträgen weitere landespezifische Prüfungen (ggf. mit 15 Datenschutzbehörden, 15 Ministerien und 15 Ethikkommissionen) erforderlich. Werden die Projekte noch komplexer, z.B. bei

einer Verlinkung von Krebsregisterdaten mit anderen Daten (z.B. von Krankenkassen), ist der organisatorische Aufwand in diesem Setting einem annehmbaren Zeitrahmen kaum leistbar. Für solche Projekte wäre eine zentrale Struktur für die anlassbezogenen Datenzusammenführung wünschenswert, die zum einen ein vereinfachtes Antragswesen mit einem Antrag für alle Register beinhaltet, zum anderen die Einzeldatensätze der Register standardisiert zusammenführt und ggf. die Verlinkung mit anderen Datenbeständen organisiert. Im Gesetz zur Zusammenführung von Krebsregisterdaten (KRDaZuG) vom 18. August 2021 ist die Entwicklung einer solchen zentralisierten Struktur bis Ende 2024 bereits angedacht [19]. Inzwischen befassen sich verschiedene Projekte mit der Entwicklung entsprechender Strukturen [20–22].

### » Krebsregisterdaten werden ohne Einwilligung zur Nutzung für Forschung erhoben

Ungelöst, zumindest aber unklar ist die Abwägung von Datenschutz auf der einen Seite und der wissenschaftlichen Nutzbarkeit der Registerdaten auf der anderen Seite. Wie dargelegt, werden Krebsregisterdaten ohne Einwilligung zur Nutzung für Forschung erhoben. Das KRDaZuG sieht daher vor, dass nur anonymisierte Daten an die Forschenden herausgegeben werden. Dabei besteht das Problem, dass die komplexen Daten der Krebsregister auch nach Löschung eindeutiger Identifikatoren, wie Name, Adresse oder Krankenversicherungsnummer nicht als komplett anonym gelten. Durch personenbeziehbare Angaben im Datensatz (Alter, Geschlecht, Wohnort, Diagnose usw.) und Vorwissen besteht die theoretische Möglichkeit, einzelne Personen zu reidentifizieren. Ähnlich verhält es sich mit Studien, bei denen Daten der Krebsregister mit anderen Datenquellen (Krankenkassen, Biobanken usw.) auf Personenebene ohne vorliegende Einwilligung verknüpft werden sollen. Ohne personenidentifizierende Angaben werden solche Projekte kaum möglich sein, und die Problematik des Reidentifikationsrisikos dürfte hier noch präsenter sein. Dieser Zwiespalt zwischen der Notwendigkeit, diverse Datenquellen mit den Krebsregis-

terdaten zu verbinden, und den Datenschutzbelangen kann nicht allein von den Krebsregistern gelöst werden. Hier ist die Politik gefordert, einen geeigneten rechtlichen Rahmen vorzugeben.

Einige Maßnahmen mit dem Ziel, das Risiko einer Reidentifikation auf ein vertretbares Maß zu senken und gleichzeitig die Analysemöglichkeiten möglichst wenig einzuschränken, sind für Gesundheitsdaten, die ohne Einwilligung erhoben wurden, entweder bereits im Einsatz oder in der Erprobung. Dazu gehören die Beschränkung des Nutzerkreises (hier ist sicher zwischen öffentlich geförderten renommierten Forschungsinstituten, industriellen Einrichtungen oder Versicherungen zu unterscheiden), die Prüfung eines plausiblen wissenschaftlichen Interesses an den Daten, die strikte Prüfung des für eine Auswertung notwendigen Datenumfangs und der notwendigen Detailtiefe, v. a. von Schlüsselvariablen wie Alter und Wohnort, und Sanktionsmaßnahmen bei Bekanntwerden von Missbrauch bzw. fahrlässigem Umgang mit den Daten. Auch Verfahren wie das Verfremden („Verrauschen“) von Daten, die Bereitstellung von berechneten Variablen anstelle von Originaldaten (z.B. Alter statt Geburts- und Diagnosedatum) sowie die Erzeugung synthetischer Daten werden diskutiert. Bei dezentral vorliegenden Daten sind mittlerweile Methoden für „verteiltes Rechnen“ („federated learning“) entwickelt worden, die eine Auswertung ohne physische Zusammenführung der Daten erlauben.

Sollte für ein Forschungsprojekt keine angemessene Anonymisierung möglich sein, sieht schon das KRDaZuG die Bereitstellung von pseudonymisierten Daten in einer sicheren Verarbeitungsumgebung vor. Die kontrollierte Analyseumgebung geht v. a. auf die Bedenken von Datenschützenden bezüglich einer Erzeugung vielfacher Kopien der Originaldaten (bzw. Teildatensätzen) zu Forschungszwecken ein, die das Risiko durch Angriffe von außen oder das irrtümliche und fahrlässige Verbreiten von Daten erhöhen. Solche „sicheren Datenhäfen“ („data safe haven“) sind auch in den Entwürfen für einen Europäischen Gesundheitsdatenraum vorgesehen und werden derzeit im britischen Gesundheitssystem bereits etabliert [23]. In Deutschland fehlen derzeit noch gesetz-

liche Vorgaben und die notwendige Infrastruktur für solche Umgebungen. Beim Forschungsdatenzentrum Gesundheit am BfArM [24] wird seit einiger Zeit eine entsprechende Infrastruktur aufgebaut, die jedoch zunächst nur für die dort vorgehaltenen bundesweiten Abrechnungsdaten nach Datentransparenzverordnung nutzbar sein wird.

### » Das KRDaZuG verlangt für pseudonymisierte Daten eine kontrollierte Verarbeitungsumgebung

Für die Nutzung der Krebsregisterdaten ergibt sich ein Dilemma. Das KRDaZuG verlangt für pseudonymisierte Daten eine kontrollierte Verarbeitungsumgebung, die es aber nicht gibt und von der nicht absehbar ist, wann sie geschaffen sein wird.

Der Weg, Routinedaten wie die Krebsregisterdaten nur mit immer komplexeren Verfahren und Einschränkungen nutzen zu können, könnte sich leicht als Irrweg herausstellen. Wenn die Hürden aus Antragstellung und Datenzusammenführung so hoch werden, dass sie nicht zu finanzieren sind und Vorlaufzeiten vor der eigentlichen Forschung entstehen, die ein übliches Forschungsprojekt um Jahre übersteigt, ist abzusehen, dass die Daten nicht genutzt werden und aufwendig finanzierte Datenfriedhöfe entstehen. Ob das neue Gesundheitsdatennutzungsgesetz [25] die Forschungsnutzung der Krebsregister, wie beabsichtigt, verbessern wird, ist offen.

Auch gilt es die Verhältnismäßigkeit der beschriebenen Maßnahmen zu hinterfragen. In über 50 Jahren weltweiter Forschung mit Krebsregisterdaten ist kein Fall bekannt, in dem Daten von Forschenden missbräuchlich verwendet wurden und Personen zu Schaden kamen. Bedenkt man die Fortschritte durch die intensive Forschungsnutzung der Krebsregisterdaten, die unbestreitbar auch Leben gerettet haben, wären mehr pragmatische Lösungen angezeigt, die auch auf Vertrauen und Verantwortung von Wissenschaft und Forschung setzen und dem öffentlichen Interesse an einer möglichst ungehinderten wissenschaftlichen Nutzung der aus öffentlichen Mitteln erhobenen Daten Rechnung tragen.

### Fazit für die Praxis

- Die Daten der Landeskrebsregister werden bereits intensiv für onkologische Forschung genutzt.
- Für unterschiedliche Fragestellungen existieren geeignete Zugangsmöglichkeiten und Antragsverfahren.
- Die Register können aggregierte Daten oder Einzelfalldaten bereitstellen, Kontakt zu Krebskranken herstellen und den Abgleich von bestehenden Forschungskohorten unterstützen.
- Es empfiehlt sich bereits bei der Planung eines Forschungsprojekts und bei der Beantragung von Daten, frühzeitig Kontakt zu den Krebsregistern aufzunehmen und die dort vorhandene Expertise in Anspruch zu nehmen.
- Zusammengeführte Daten der Landeskrebsregister können auch beim Zentrum für Krebsregisterdaten am Robert Koch-Institut beantragt werden.

### Korrespondenzadresse

**Prof. Dr. Alexander Katalinic**  
Krebsregister Schleswig-Holstein  
Lübeck, Deutschland  
alexander.katalinic@uksh.de

**Funding.** Open Access funding enabled and organized by Projekt DEAL.

### Einhaltung ethischer Richtlinien

**Interessenkonflikt.** A. Katalinic, R. Pitzkuleit und K. Kraywinkel geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien. Die Autoren forschen seit vielen Jahren mit Krebsregisterdaten.

**Open Access.** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

## Abstract

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

1. Katalinic A (2004) Population-based cancer registration in Germany. Essentials and perspectives. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 47(5):422–428
2. Klinkhammer-Schalke M, Gerken M, Barlag H, Tillack A (2015) Quality management in oncology supported by clinical cancer registries. *Z Evid Fortbild Qual Gesundhwes* 109(6):452–458
3. Katalinic A, Halber M, Meyer M, Pflüger M, Eberle A, Nennecke A et al (2023) Population-based clinical cancer registration in Germany. *Cancers* 15(15):3934
4. Arndt V, Holleczeck B, Kajüter H, Luttmann S, Nennecke A, Zeissig SR et al (2020) Data from population-based cancer registration for secondary data analysis: methodological challenges and perspectives. *Gesundheitswesen* 82(1):S62–71
5. Stegmaier C, Hentschel S, Hofstädter F, Katalinic A, Tillack A, Klinkhammer-Schalke M (2019) The Cancer Registry Manual (Das Manual der Krebsregistrierung). W. Zuckschwerdt Verlag, München
6. Plattform65c Onkologischer Basidatensatz. <https://basidatensatz.de/basidatensatz>. Zugriffen: 17. Sept. 2023
7. Publikationsserver des Robert Koch-Instituts, Öffentliches Verzeichnis über bewilligte Anträge auf Nutzung von Krebsregisterdaten. <https://edoc.rki.de/handle/176904/8317>. Zugriffen: 17. Sept. 2023
8. Wildner GP, Eckert H (1965) Zur Morbiditätsstatistik der Lymphogranulomatose. Eine statistische Untersuchung von 1490 Fällen aus dem Krebskrankenregister der Deutschen Demokratischen Republik. *Arch Geschwulstforsch* 26(4):216–222
9. National Cancer Institute (2023) Surveillance, epidemiology, and end results program (SEER). <https://seer.cancer.gov/about/>. Zugriffen: 5. Juni 2023
10. Hiripi E, Gondos A, Emrich K, Holleczeck B, Katalinic A, Luttmann S et al (2012) Survival from common and rare cancers in Germany in the early 21st century. *Ann Oncol* 23(2):472–479
11. Schafmayer C, Buch S, Egberts JH, Franke A, Brosch M, El Sharawy A et al (2007) Genetic investigation of DNA-repair pathway genes PMS2, MLH1, MSH2, MSH6, MUTYH, OGG1 and MTH1 in sporadic colon cancer. *Int J Cancer* 121(3):555–558
12. Obi N, Waldmann A, Schäfer F, Schreier I, Katalinic A (2011) Impact of the quality assured mamma diagnostic (quamaDi) programme on survival of breast cancer patients. *Cancer Epidemiol* 35(3):286–292
13. Kajüter H, Geier AS, Wellmann I, Krieg V, Fricke R, Heidinger O et al (2014) Cohort study of cancer incidence in patients with type 2 diabetes: record linkage of encrypted data from an external cohort with data from the epidemiological Cancer Registry of North Rhine-Westphalia. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 57(1):52–59
14. Bokhof B, Khil L, Urbschat I, Gnas L, Hecht G, Heidinger O et al (2018) Time trend in programme sensitivity within the German mammography screening programme in North Rhine-Westphalia and Lower Saxony. *Bundesgesundheits-*

## Research with cancer registries—what is and is not (yet) possible?

**Background:** One of the tasks of state cancer registries is to provide data for research or one's own research. Data is not collected in the cancer registry primarily for research, but with mandatory reporting and without the consent of the persons concerned for the fulfilment of legal tasks. Can data from cancer registries be meaningfully used for research, are they used and which framework conditions for data use should be improved?

**Methods:** Using PubMed®, the use of cancer registry data in scientific publications was determined for Germany and for the US Surveillance, Epidemiology, and End Results (SEER) program. Possibilities for data use were identified from state and federal laws.

**Results:** For Germany, 2268 articles with cancer registry data were identified with about 250 publications in 2022. The US SEER program currently produces over 1500 publications per year. The following main types of use were identified from the cancer registry laws: provision and use of aggregated and individual case data, contacting cancer patients from the registry population, and matching existing cohorts to the cancer registry. While aggregated and individual case data can usually be applied for using standardized forms, other types of use require standard scientific applications with an ethics vote and, if necessary, further approval. Selected practical examples show the different types of use.

**Conclusion:** The increasing number of publications is evidence of lively research use. Nevertheless, the potential in Germany has not been exhausted. Projects across federal states with links to other data are very time-consuming or often not realistically feasible due to bureaucratic hurdles. Attempts are currently being made with projects and legislative initiatives to simplify use of such data.

### Keywords

Oncology · Cancer registry · Basic oncology dataset · Data analysis · Data systems

- blatt Gesundheitsforschung Gesundheitsschutz 61(12):1517–1527
15. Bayerisches Krebsregister, Antrag auf Nutzung von Krebsregisterdaten für Bayern. [https://www.lgl.bayern.de/downloads/gesundheit/krebsregister/doc/Datennutzung\\_Antragsformular.pdf](https://www.lgl.bayern.de/downloads/gesundheit/krebsregister/doc/Datennutzung_Antragsformular.pdf). Zugriffen: 22. Sept. 2023
16. Zentrum für Krebsregisterdaten, Antrag auf Datennutzung. [https://www.krebsdaten.de/Krebs/DE/Content/Forschungsdaten/Info\\_antrag/Antragsformulare/antragsformulare\\_anicht.html?nn=2640312](https://www.krebsdaten.de/Krebs/DE/Content/Forschungsdaten/Info_antrag/Antragsformulare/antragsformulare_anicht.html?nn=2640312). Zugriffen: 22. Sept. 2023
17. BGBI (2009) Bundeskrebsregisterdatengesetz (BKRG) vom 10.08.2009, S 2707 (Nr. 53)
18. Peters A, Peters A, Greiser KH, Göttlicher S, Ahrens W, Albrecht M et al (2022) Framework and baseline examination of the German national cohort (NAKO). *Eur J Epidemiol* 37(10):1107–1124
19. Cancer registry data consolidation act. <https://www.buizer.de/gezet/14935/index.htm>. Zugriffen: 5. Juni 2023
20. Bundesministerium für Gesundheit, Konzept zur Schaffung einer Plattform zur bundesweiten Auswertung von klinischen Krebsregisterdaten für Forschungsfragen und registerbasierte Studien (Plato2). <https://www.bundesgesundheitsministerium.de/ministerium/ressortforschung/handlungsfelder/digitalisierung/plato2.html>. Zugriffen: 22. Sept. 2023
21. Bundesministerium für Gesundheit, KI-unterstützte, versorgungsnahen Nutzung von Krebsregisterdaten (AI-CARE). <https://www.bundesgesundheitsministerium.de/ministerium/ressortforschung/handlungsfelder>
- forschungsschwerpunkte/krebsregisterdaten/ai-care.html. Zugriffen: 22. Sept. 2023
22. Bundesministerium für Gesundheit Krebs, Forschungsdatenzentrum – KI-gestützte Evidenzgenerierung aus versorgungsnahen Daten klinischer Krebsregister, GKV-Routinedaten, Klinikdaten und deren Linkage (onkoFDZ). <https://www.bundesgesundheitsministerium.de/ministerium/ressortforschung/handlungsfelder/forschungsschwerpunkte/krebsregisterdaten/onkofdz.html>
23. National Health Service England, Protecting and safely using data in the new NHS England. <https://www.england.nhs.uk/about/protecting-and-safely-using-data-in-the-new-nhs-england/>. Zugriffen: 22. Sept. 2023
24. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), FDZ Gesundheit. [https://www.bfarm.de/DE/Das-BfArM/Aufgaben/Forschungsdatenzentrum/\\_node.html](https://www.bfarm.de/DE/Das-BfArM/Aufgaben/Forschungsdatenzentrum/_node.html). Zugriffen: 22. Sept. 2023
25. Bundesministerium für Gesundheit Gesundheitsdatennutzungsgesetz (GDNG). <https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/detail/gesundheitsdatennutzungsgesetz.html>. Zugriffen: 22. Sept. 2023

**Hinweis des Verlags.** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.