

Zusammenführung von umfangreichen Daten und Vollständigkeit der Informationen zu Diagnose, Behandlung und zum Erkrankungsverlauf aus mehreren populationsbezogenen Krebsregistern in Deutschland – erste Erfahrungen am Beispiel von Lungenkrebs

Zusammenfassung

Hintergrund: In Deutschland existiert eine flächendeckende Krebsregistrierung. Populationsbezogene Krebsregister sind auf Bundeslandebene organisiert. Bislang ist eine wissenschaftsgetriebene, anlassbezogene Zusammenführung von ausführlichen (klinischen) Daten zu Diagnose, Therapie und Verlauf aus mehreren Landeskrebsregistern nur selten erfolgt. Wir prüften die Machbarkeit und den Aufwand der Datenzusammenführung sowie die Eignung der Daten zur Deskription der onkologischen Versorgung.

Methoden: Daten zu Lungenkrebs (ICD-10 C34) aus den Diagnosejahren 2016–2019 wurden bei vier Landeskrebsregistern angefragt, aufbereitet und zusammengeführt. Sofern mehrere fallzuordnungsfähige Meldungen zu einer Therapie vorlagen, wurden diese im Rahmen des Pilotprojektes zur „besten Information“ aggregiert. Die Daten wurden deskriptiv und explorativ ausgewertet.

Ergebnisse: Die Datenaufbereitung und Zusammenführung waren zeitaufwändig, aber technisch gut machbar. Der Auswertungsdatensatz umfasst Informationen zu mehr als 61.000 Erkrankungsfällen mit mindestens einem Tag Nachbeobachtungszeit. Zu 74% liegen Informationen zu mindestens einer Therapieart (Operation, Strahlentherapie, systemische Therapie) vor. Die bereitgestellten Angaben zu den Details der Therapien sind überwiegend vollständig oder weisen nur wenig fehlende Werte auf.

Diskussion: Trotz z.T. föderaler Strukturen der Krebsregistrierung in den Ländern waren Unterschiede bezüglich der Datenverteilung und der Datenqualität im gepoolten Datensatz gering. Der Anteil mit vollständig fehlenden Angaben zur Therapie ist mit 26% als unseres Erachtens hoch zu bewerten, aber vergleichbar mit anderen Registerstudien. Auf Basis der vorhandenen Informationen können die Grundzüge einzelner onkologischer Therapien beschrieben und für die onkologische Versorgungsforschung genutzt werden.

Schlüsselwörter: populationsbezogene Krebsregister, Datenzusammenführung, Datenqualität, Therapie, Nachbeobachtung

Annika Waldmann¹

Louisa Labohm¹

Hannah Baltus¹

Christine Eisfeld²

Lina Jansen³

Imma Löhden⁴

Alice Nennecke⁴

Florian Oesterling²

Ron Pritzkuleit⁵

Alexander Katalinic¹

1 Institut für Sozialmedizin und Epidemiologie, Universität zu Lübeck, Lübeck, Deutschland

2 Landeskrebsregister NRW gGmbH, Bochum, Deutschland

3 Epidemiologisches Krebsregister Baden-Württemberg, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Deutschland

4 Hamburgisches Krebsregister, Freie und Hansestadt Hamburg, Behörde für Wissenschaft, Forschung, Gleichstellung und Bezirke, Hamburg, Deutschland

5 Krebsregister Schleswig-Holstein, Registerstelle, Lübeck, Deutschland

Einleitung

Mit dem Gesetz zur Weiterentwicklung der Krebsfrüherkennung und zur Qualitätssicherung durch klinische Krebsregister wurde 2013 in Deutschland die Basis zur flächendeckenden klinischen Krebsregistrierung geschaffen, die heute in § 65c des Sozialgesetzbuches V abgebildet ist [1], [2]. Landesspezifische Krebsregistergesetze [3], [4], [5], [6] und § 8 des Bundeskrebregisterdatengesetzes (BKRG) [7] regeln u.a. Datenverarbeitung, Meldewege und -anlässe. Der onkologische Basisdatensatz (oBDS) und ergänzende Module [8], [9], [10], [11] regeln Art und Umfang der zu registrierenden Daten.

Das Gesetz zur Zusammenführung von Krebsregisterdaten aus dem Jahr 2021 [12] bildet die Grundlage für die bundesweite Zusammenführung eines Teils der klinischen Daten am Zentrum für Krebsregisterdaten (ZfKD) in den ‚bundesweiten klinischen Krebsregisterdatensatz‘ [13]. Als erste Stufe wurden im Jahr 2023 erstmals Daten zu Therapie und Verlauf aus den Landeskrebsregistern zusammengeführt [13]. Auf Antrag können die bundesweiten Daten als Einzelfalldaten bereitgestellt werden. Diese Daten umfassen nicht alle Variablen des oBDS, so fehlen z.B. Angaben aus den organspezifischen Modulen sowie zu Biomarkern und zu Leistungserbringern. Für tiefergehende Analysen müssen deshalb weiterhin Landeskrebsregisterdaten genutzt werden. In einer zweiten Stufe sollen anlassbezogene, patienten- und leistungserbringerbezogene, registerübergreifende Zusammenführungen von Krebsregisterdaten realisiert werden [12]. Für diese Stufe steht ein konkretes Konzept noch aus.

Eine Zusammenführung von ausführlichen klinischen Daten zur Therapie und zum Krankheitsverlauf aus mehreren deutschen Landeskrebsregistern erfolgte – mit Ausnahme der Auswertungen, die regelmäßig anlässlich des Deutschen Krebsskongresses vorgenommen werden und dort unter dem Begriff „Bundesweite Onkologische Qualitätskonferenz“ vorgestellt werden [14] – zunächst selten. Inzwischen beginnen einzelne Forschungsprojekte nationale klinische Daten zusammenzustellen [15], [16]. In einem Pilotprojekt wollten wir Machbarkeit und Aufwand einer Zusammenführung sowie die Dateneignung zur Deskription der onkologischen Versorgung prüfen. Es wurden Tumor-, Therapie- und Verlaufsdaten zu Personen mit Lungenkrebs aus vier Landeskrebsregistern angefordert, aufbereitet und zur Beantwortung folgender Fragestellungen genutzt:

- Wie groß ist der Aufwand der Datenzusammenführung?
- Sind die im Register bekannten Behandlungen vollständig dokumentiert, sodass sie für eine Deskription der onkologischen Versorgung geeignet sind?

Methoden

Datenanforderung

Für das Pilotprojekt wurden die Landeskrebsregister von Baden-Württemberg, Hamburg, Nordrhein-Westfalen und Schleswig-Holstein kontaktiert, da sie entsprechende Daten mindestens seit dem Jahr 2016 erfassen. Bei der Auswahl dieser Register spielten außerdem folgende Überlegungen eine Rolle: es sollten (1) maximal zwei unterschiedliche Krebsregister-Softwaretypen verwendet werden, (2) Krebsregister mit großer wie auch mit kleiner zugrundeliegender Bevölkerung und (3) Stadtstaaten und Flächenländer zur Datenbasis beitragen.

Zum Jahreswechsel 2020/2021 wurde das Vorgehen der Datenanforderung abgestimmt und zusätzlich zu den registerspezifischen Formularen ein allgemeines Studienprotokoll und eine Variablenliste an die Register gesendet. Die Liste enthielt die angefragten Variablen aus dem oBDS mit Wunschformat (numerisch, String).

Beantragt wurden Informationen zu Erkrankungsfällen mit Lungenkrebs (ICD-10 C34) aus den Diagnosejahren 2016–2019. Die Erkrankten mussten mindestens 18 Jahre sein und im Register mussten Mindestangaben zur Diagnose vorliegen. Außerdem mussten die Erkrankten bei Diagnose im jeweiligen Bundesland leben – dies entsprach einer Selektion auf so genannte wohnortbezogene Fälle, die eine systematische Erhebung des Vitalstatus über einen Abgleich mit Meldeamtinformationen bzw. Mortalitätsdaten erlauben [17].

Datenbereitstellung und Aufbereitung

In den Registern liegen die Daten in relationalen Datenbanken in zwei unterschiedlichen Formen vor. Zum einen enthalten die Datenbanken „Rohdaten“ in Form von fallzuordnungsfähigen Meldungen, wie sie von Meldenden an das Register übermittelt werden. Diese Meldungen erfolgen zu bestimmten Meldeanlässen (z.B. *Diagnose, Operation, Start einer Strahlentherapie* etc.) und enthalten üblicherweise nur Daten, die diesem Anlass zuzuordnen sind (z.B. *enthalten OP-Meldungen keine Angaben über eine begleitende systemische Therapie*). Zu einem Meldeanlass können mehrere Meldungen von unterschiedlichen Meldenden vorliegen (z.B. *eine Diagnosemeldung aus einer Praxis und einer spezialisierten Klinik*). Zum anderen enthalten die Krebsregister-Datenbanken sogenannte „Best-of-Informationen“. Dies sind Zusammenfassungen von mehreren Meldungen (zur Best-of-Bildung siehe nächster Abschnitt), die letztendlich den Verlauf eines Erkrankungsfalls abbilden. Ein Best-of-Datensatz hat verschiedene 1:n-Verbindungen – 1 Patient : n Tumoren, 1 Tumor : n Therapien (Abbildung 1).

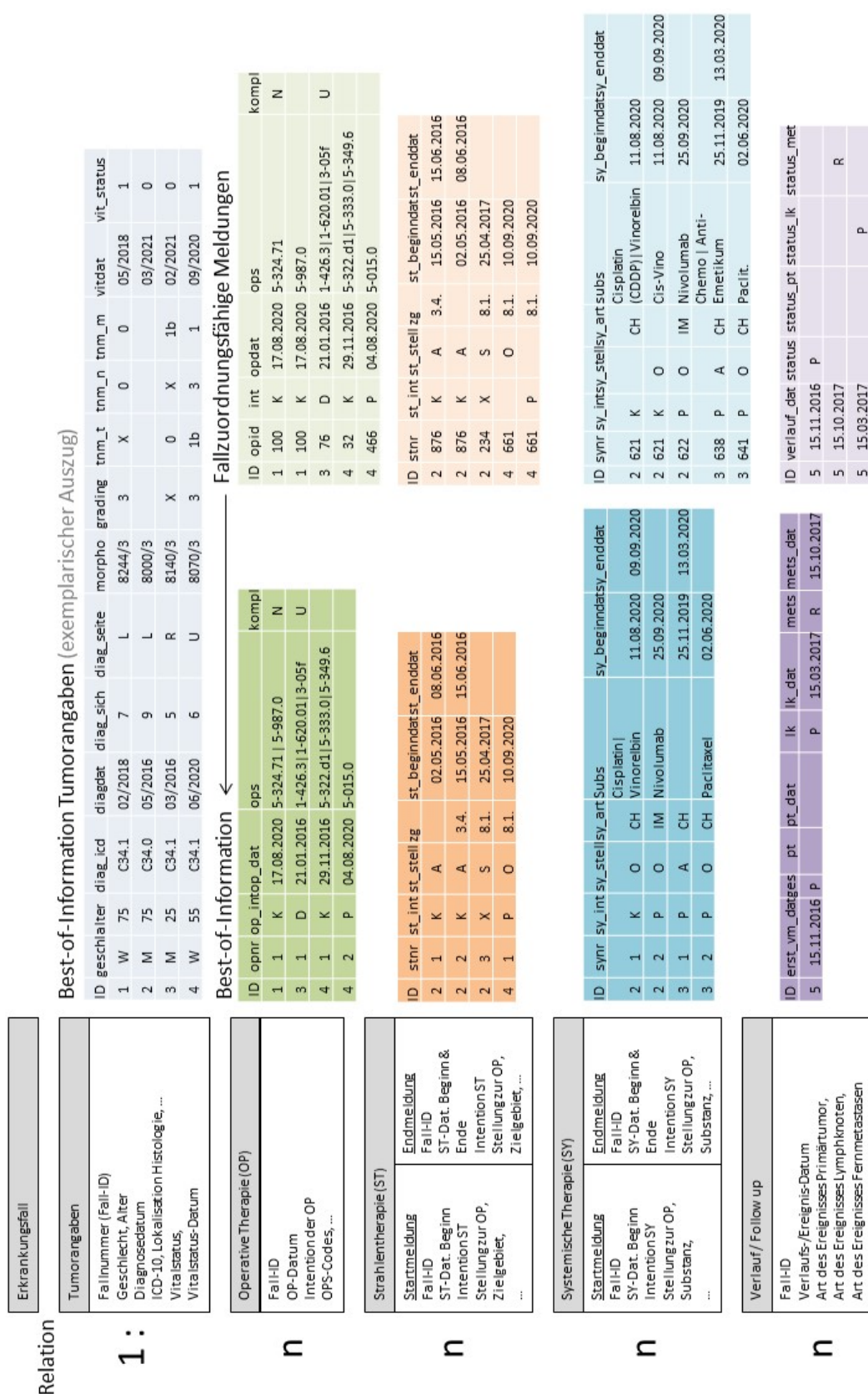


Abbildung 1: Teil-Datensätze der flächendeckenden klinischen/behandlungsortbezogenen Krebsregistrierung – Relation der Datensätze zueinander sowie exemplarischer Inhalt und Skizzierung des Vorgehens bei der Bildung der „besten Information“ auf Basis von fallzuordnungsfähigen Meldungen

Im Teil-Datensatz „Best-of Information Tumorangaben“ gibt es für jeden Erkrankungsfall genau eine Zeile im Datensatz. In den Best-of-Datensätzen zur Therapie gibt es für jeden Therapieanlass (z.B. Operation des Primärtumors, Revisionsoperation, eine Teilbestrahlung im Rahmen einer Strahlentherapie) genau eine Zeile, allerdings kann ein Erkrankungsfall bei Vorhandensein mehrerer Therapieanlässe mehrfach im Datensatz vorkommen (1:n-Relation).

Bei Datenbeantragung lagen in den Registern noch nicht für alle Episoden eines Krankheitsverlaufs Best-of-Informationen vor, so dass z.T. fallzuordnungsfähige Meldungen übermittelt wurden und die Aufgabe der Best-of-Bildung im Pilotprojekt lag. Angaben zur Diagnose, also der Beschreibung des Primärtumors (Tumordaten), wurden aus allen Registern als Best-of-Informationen geliefert, bei den Therapien und Verläufen war das nicht immer der Fall (Tabelle 1). Die zur Verfügung gestellten Teil-Datensätze wurden im „long-Layout“ (Daten in gestapelter Struktur; d.h. wurde ein Erkrankungsfall zweimal operiert, dann gab es für diesen Erkrankungsfall zwei Zeilen im Datensatz) oder im „wide-Layout“ (ungestapelte Struktur; d.h. wurde ein Erkrankungsfall zweimal operiert, dann wurden die Spalten mit den Informationen zur Operation verdoppelt und jeder Erkrankungsfall war mit einer Zeile im Datensatz enthalten) bereitgestellt. Im Zuge der Aufbereitung wurden alle Teil-Datensätze in ein gestapeltes Layout überführt.

Angaben zu Substanzen oder Protokollen der systemischen Therapie werden den Registern über Freitextfelder gemeldet, wobei laut Datensatzbeschreibung des oBDS [11] Substanzen und Handelsnamen der Produkte erlaubt sind. Für die Aufbereitung dieser Angaben wurde eine von den Krebsregistern abgestimmte Substanz-Referenzliste verwendet, die neben der Standardisierung von Substanzen auch die Zuordnung zu den Therapiearten (Chemo-, Immuntherapie etc.) erlaubt. Die Angaben zur Art des operativen Eingriffs gelangen über Textfelder, in die entsprechend der Datensatzbeschreibung [11] so genannte OPS-Codes (Operationen- und Prozedurenschlüssel [18]) eingetragen werden sollen, in die Register. Für die Entscheidung, ob ein OPS-Code auf eine tumorresezierende Operation hinweist, wurde ebenfalls eine von den Krebsregistern abgestimmte OPS-Referenzliste verwendet.

Bildung von Best-of-Informationen

Angaben zur gleichen Information von mehreren Meldenden müssen zu so genannten Best-of-Informationen zusammengefasst werden (Abbildung 1). Für die Diagnoseinformationen (Tabelle Tumordaten) gibt es seit vielen Jahren Empfehlungen zur (epidemiologischen) Best-of-Bildung – basierend auf internationalen Richtlinien und Übereinkünften der Landeskrebsregister – und die Umsetzung der Empfehlung ist in den Registern etabliert [17]. Da solche Richtlinien für Therapieangaben zu Beginn der flächendeckenden klinischen Krebsregistrierung in Deutschland nicht existierten, hat eine Arbeitsgruppe der Krebsregister Empfehlungen zur Best-of-Bildung in den Registern erarbeitet (Datum der Verabschiedung für die Empfehlungen zur OP: 07.05.2019; Strahlentherapie: 05.11.2019; systemische Therapie: 11.01.2022). Sofern registerseits keine Best-of-Therapieinformationen bereitgestellt wurden, wurden sie auf Basis der Einzelmeldungen und der AG-Empfehlungen (Stand Frühjahr 2021) in einem zeit- und programmieraufwändigen Verfahren generiert. Hierbei wurde z.B. bei Vorliegen von

Start- und Endmeldung einer Therapie mit diskordanten Informationen zu einem Merkmal das vorgeschlagene „Ranking“ von Ausprägungen umgesetzt.

Datenverarbeitung und statistische Auswertungen

Die bereitgestellten Datensätze wurden mit dem Open Source Statistikprogramm R (Version 4.1.3) [19] aufbereitet, zusammengeführt und ausgewertet. Die Aufbereitungsschritte (u.a. Normierung von Variablenamen, -formaten und ggf. der Kodierungen, ggf. Best-of-Bildung) wurden für jedes Register separat vorgenommen. Sowohl die Roh-Datensätze als auch die durch die Bearbeitung und Zusammenführung entstandenen Datensätze wurden auf Datenqualität und Plausibilität geprüft. Die Variablen in den gepoolten Datensätzen wurden zusätzlich auf mögliche Registereffekte, d.h. auf Unterschiede in der Häufigkeitsverteilung der Merkmalsausprägungen zwischen den Registern, geprüft.

Zur Beurteilung der Datenhomogenität werden für die prozentualen Häufigkeiten und Mediane das Minimum und das Maximum der Werte aus den vier Registern angegeben.

Für die Analysen wurden Fälle, die dem Register nur auf Basis eines Leichenschauzeins (DCO-Fall) bekannt waren, Fälle mit 0 Tagen Überleben, Fälle mit fehlenden Informationen zum Vitalstatus und Fälle mit Informationen zu Therapien, die zeitlich mehr als 31 Tage vor der Lungendixdiagnose lagen und die als Hinweis auf weitere Krebserkrankungen gedeutet werden können, ausgeschlossen.

Eine statistische Hypothesentestung war nicht intendiert.

Ethik

Das Vorhaben wurde bei der Ethikkommission der Universität zu Lübeck angezeigt und zustimmend zur Kenntnis genommen (Az 20-483). Im Deutschen Register Klinischer Studien ist es unter der Nummer DRKS00025080 registriert [20].

Ergebnisse

Datenbereitstellung und Aufbereitung

Tabelle 1 verdeutlicht die Heterogenität der Datenbereitstellung. Zwei Register übermittelten die Daten zeitnah, während ein Register sieben Monate benötigte. Die Anzahl der bereitgestellten Teil-Datensätze und deren Layout unterschied sich u.a. aufgrund vager Vorgaben im Antrag auf Datenübermittlung.

Informationen zu operativen Eingriffen wurden von zwei Registern und Informationen zur Strahlentherapie wurden von einem Register als Best-of-Information zur Verfügung gestellt. Die Informationen zu systemischen bzw. abwartenden Therapien (im Folgenden „systemische Therapie“)

Tabelle 1: Beschreibung der bereitgestellten Daten und von Besonderheiten bei der Datenbereitstellung

	Besonderheiten bei der Datenbereitstellung	Zeitspannen bzw. Spannweiten der Fallzahlen und relativen Anteile von Fällen der jeweiligen vier Register
Call for Data verschickt		Dezember 2020 – Januar 2021
Erhalt der Daten 1. Datenlieferung		März 2021 – August 2021*
Wenn mehrere Datenlieferungen erfolgten, Zeitraum der finalen Lieferung	Von drei Registern wurden im Nachgang korrigierte bzw. aktualisierte Daten versendet.	August 2021 – August 2022
Anzahl übermittelter Datendateien		4–7
Datei mit Tumorangaben	Bei zwei Registern waren in den Dateien mit Tumorangaben zusätzlich die Informationen zu Verlaufereignissen und zu Metastasen (sowohl zum Primärtumor als auch im Verlauf) enthalten.	Spannweite der Anzahl der Erkrankungsfälle** je Register: 4.200–39.823
Datei mit Angaben aus der Meldungsart „Verlauf“ (Progression, Rezidiv, Metastase)	s. o.	Spannweite des Anteils mit mindestens einem Verlaufereignis an den jeweiligen Fällen eines Registers: 16,2%–33,2%
Datei mit Angaben aus der Meldungsart „Metastasen“	Ein Register lieferte die Angaben zu Metastasen zum Primärtumor und im Verlauf in einer Datei. Ein Register lieferte die Angaben zu Metastasen zum Primärtumor und im Verlauf in zwei separaten Dateien. Zwei Register übermittelten diese Angaben in der Datei mit den Tumorangaben.	Spannweite des Anteils mit Information zu mindestens einer Metastase zum Diagnosezeitpunkt an den jeweiligen Fällen eines Registers: 33,3%–51,5% Spannweite des Anteils mit Information zu mind. einer Metastase im Krankheitsverlauf an den jeweiligen Fällen eines Registers: 0%–11,0%
Datei mit Angaben zur Operation (OP)	Zwei Register lieferten Informationen in Form von fallzuordnungsfähigen Meldungen. Zwei Register lieferten Best-of-Informationen.	Nach Best-of-Bildung: Spannweite des Anteils mit Information zur OP an den jeweiligen Fällen eines Registers: 19,3%–30,1%
Datei mit Angaben zur Strahlentherapie (Teilbestrahlungen; ST)	Drei Register lieferten Informationen in Form von fallzuordnungsfähigen Meldungen. Ein Register lieferte Best-of-Informationen.	Nach Best-of-Bildung: Spannweite des Anteils mit Information zur ST an den jeweiligen Fällen eines Registers: 23,7%–38,1%
Datei mit Angaben zur Systemischen Therapie (SY)	Vier Register lieferten Informationen in Form von fallzuordnungsfähigen Meldungen.	Nach Best-of-Bildung: Spannweite des Anteils mit Information zur SY an den jeweiligen Fällen eines Registers: 38,9%–51,0%
Letzte systematische Erfassung des Vitalstatus durch Abgleich der Daten mit Meldeamts- bzw. Mortalitätsdaten aus dem Gesundheitsamt		Dezember 2019 – April 2021
Anteil an Fällen, die dem Register nur über Leichenschauschein bekannt sind (DCO-Fälle), an allen Erkrankungsfällen		Spannweite des Anteils an DCO-Fällen an den jeweiligen Fällen eines Registers: 4,0%–12,1%
Anteil an Fällen mit 0 Tagen Nachbeobachtungszeit		Spannweite des Anteils mit 0 Tagen Nachbeobachtung an den jeweiligen Fällen eines Registers: 5,3%–17,8%
Anteil an Fällen mit fehlendem Vitalstatus		Spannweite des Anteils mit fehlendem Vitalstatus an den jeweiligen Fällen eines Registers: 0%–0,8%

* In einem Register mussten sich zum Zeitpunkt der Datenbeantragung sowohl der Beirat als auch der wissenschaftliche Fachausschuss über den Antrag auf Überlassung von Individualdaten beraten und eine Empfehlung abgeben. Beide Gremien tagen mindestens einmal pro Jahr gemäß gesetzlicher Grundlage, faktisch tagen sie halbjährlich. Mitte 2023 wurde das Prozedere geändert und aktuell muss nur noch der wissenschaftliche Fachausschuss eine Empfehlung abgeben.

** Fälle = Tumoren

wurden von allen Registern als fallzuordnungsfähige Meldungen übermittelt.

Alle hier einbezogenen Register erheben den Vitalstatus systematisch und in ähnlicher Weise durch einen regelmäßigen Abgleich der Krebsregisterdaten mit Meldeamtsinformationen bzw. Mortalitätsdaten. Der jeweils letzte Abgleich für den hier genutzten Datensatz erfolgte in einem Register im Dezember 2019, in einem weiteren im Dezember 2020 und in den zwei anderen Registern im Jahr 2021.

Datengrundlage: Gepoolter Datensatz mit „Best-of-Informationen“

Nach Aufbereitung, Best-of-Bildung und Zusammenführung lag eine Datenbank mit sechs Teil-Datensätzen für insgesamt 70.821 Erkrankungsfälle vor (Tabelle 1, Abbildung 2): Je ein Datensatz mit Best-of-Informationen zur Diagnose (und zum Vitalstatus), Operationen, Strahlentherapien, systemischen Therapien (*die entweder nur aus einer Therapieart wie bspw. Chemotherapie oder einer Kombination aus Therapiearten bestehen können*) bzw. mit Verlaufsinformationen und ein weiterer mit Informationen zu Metastasen zum Zeitpunkt des Primärtumors und im Verlauf.

Die Auswertungen beruhen auf 61.806 Fällen nach Anwendung der Ausschlusskriterien.

Vollständigkeit der dokumentierten Therapien: Erkrankungsfälle mit und ohne bekannte Therapie

Bei 45.465 der 61.806 Fälle (73,8%) lagen Informationen zu mindestens einem Therapieereignis vor. Diese Erkrankungsfälle waren im Vergleich zu Fällen ohne Informationen zur Therapie im Median 4 Jahre jünger, hatten seltener ein fehlendes Tumorstadium, seltener großzellige und etwas häufiger kleinzellige Lungenkarzinome und verstarben seltener im Beobachtungszeitraum (Tabelle 2). Bei bekannter Therapie lagen am häufigsten Informationen zur alleinigen systemischen Therapie ($n=12.560$; 27,6%), zu einer Kombination von Strahlen- und systemischer Therapie ($n=10.081$; 22,2%) und zur alleinigen Operation ($n=8.641$; 19,0%) vor (Abbildung 3).

Vollständigkeit der Angaben zu den dokumentierten Therapien

Insgesamt lagen Informationen zu 19.917 operativen Eingriffen vor. Mindestens ein **operativer Eingriff** war für 27,7% der Erkrankungsfälle in der Auswertungspopulation bekannt (Abbildung 2, Tabelle 3). Dieser Anteil variierte zwischen den Registern im Bereich von 20,2% bis 31,1%. Rund 79,2% aller Eingriffe erfolgten innerhalb von 6 Monaten nach Diagnosestellung (92,2% aller Ersteingriffe). Nach Bildung der Best-of-Informationen fehlten selten konkrete Angaben zur Intention der Operation (fehlende Angabe oder Ausprägung „unbekannt“: 12,2%)

und zu Operationsprozeduren (1,0%), während Angaben zu Komplikationen der Operation zu mehr als 50% fehlend oder unbekannt waren. Für die Register waren die Verteilungen der Merkmalsausprägungen überwiegend recht ähnlich (Tabelle 3 – Abschnitt A). In rund 95% der Fälle mit Informationen zur Operation waren gültige Angaben zu Datum, Intention und den erfolgten Prozeduren enthalten. Dieser Anteil variierte zwischen den Registern im Bereich von 85,0% bis 100%.

Für 33,0% der Erkrankungsfälle in der Auswertungspopulation lag mindestens eine Information zur Teilbestrahlung im Rahmen einer **Strahlentherapie** vor (Abbildung 2, Tabelle 3). Rund 50,6% aller Teilbestrahlungen wurden innerhalb von 6 Monaten nach Diagnosestellung begonnen (80,1% aller Erst-Teilbestrahlungen). Für mehr als 90% lagen Angaben zur Gesamtdosis der Teilbestrahlung vor. Mit Ausnahme der Angaben zur Seite des Zielgebietes (unbekannt: 22,8%, fehlend: 14,7%) lag der Anteil fehlender Werte bei maximal 15%. Auch für die Angaben aus den Meldungen zu den Teilbestrahlungen gilt, dass die Häufigkeitsverteilungen im Registervergleich überwiegend recht ähnlich waren (Tabelle 3 – Abschnitt B). Insgesamt enthielten rund 72% dieser Best-of-Informationen gültige Angaben zu Start und Ende der Therapie, Stellung zur Operation, Applikationsart, Intention und Zielgebiet (Streuung zwischen den Registern: 65,0%–92,7%).

Mindestens eine Best-of-Information zur **systemischen Therapie** konnte für 48,4% der Erkrankungsfälle gebildet werden (Abbildung 2, Tabelle 3). Rund 47,4% aller systemischen Therapien wurden innerhalb von 6 Monaten nach Diagnosestellung begonnen (92,6% aller Erst-Therapien). Information zum Enddatum der Therapie und zum Beendigungsgrund (unbekannt oder fehlend) fehlten zu je ca. 40%. In rund 93% der Best-of-Informationen konnte mindestens eine Substanz der systemischen Therapie über die Substanz-Referenzliste den Therapiearten Chemo-, Immuntherapie oder Zielgerichtete Therapie zugewiesen werden. Im Registervergleich zeigte sich, dass die Angaben zur systemischen Therapie überwiegend recht ähnlich waren (Tabelle 3– Abschnitt C). Größere Unterschiede zeigten sich für die Informationen zu den Protokollen der systemischen Therapie – sowohl der Befüllungsgrad als auch der Inhalt des Feldes variierte deutlich. Angaben zu den Protokollen waren nur selten zu finden, stattdessen wurden häufig die Angaben aus dem Feld Substanzen wiederholt oder ergänzt z.B. um weitere Substanzen oder Dosierungen bzw. Applikationsarten. Insgesamt enthielten rund 88% der Best-of-Informationen zur systemischen Therapie gültige Angaben zu Start, Stellung zur Operation, Intention und verabreichten Substanzen (Streuung zwischen den Registern: 83,9%–95,2%).

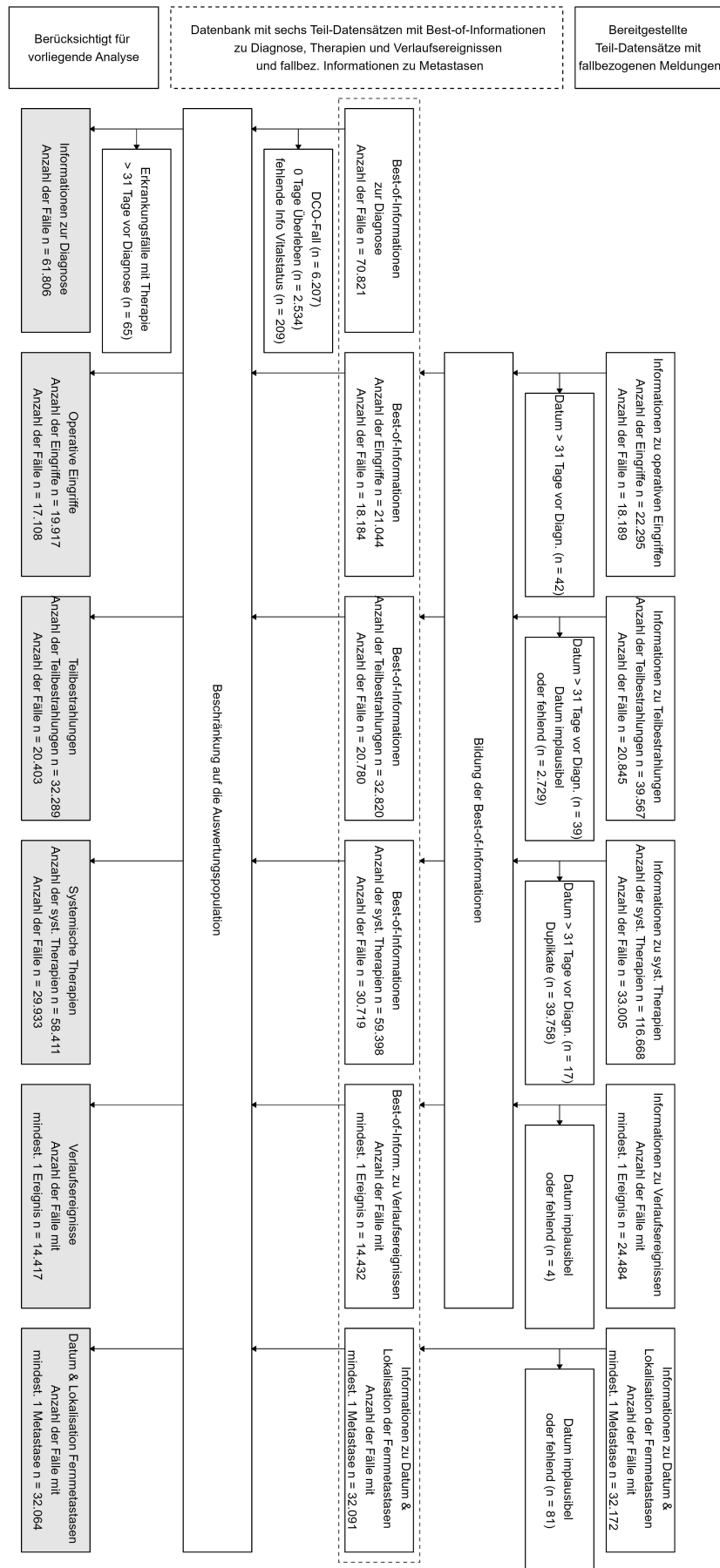


Tabelle 2: Beschreibung der Lungenkrebsfälle (ICD-10 C34, Diagnosejahre 2016–2019) insgesamt und nach Vorhandensein von mind. 1 Therapieinformation
(absolute und relative Häufigkeit bezogen auf die jeweilige Zeile, sofern nicht anders angegeben)

	Erkrankungs- fälle insgesamt (Spalten- prozente)	Erkrankungs- fälle mit mind. 1 bekannten Therapie- ereignis	Spannweite der Werte zwischen den vier Registern	Erkrankungs- fälle ohne bekanntes Therapie- ereignis	Spannweite der Werte zwischen den vier Registern
Anzahl Erkrankungsfälle, die für die Analysen berücksichtigt wurden	61.806 (100%)	45.465 (73,6%)	69,8%–79,8%	16.341 (26,4%)	20,2%–30,2%
Diagnosejahr					
2016	14.209 (23,0%)	10.238 (72,1%)	66,5%–81,9%	3.971 (27,9%)	18,1%–33,5%
2017	16.126 (26,1%)	11.632 (72,1%)	67,1%–80,0%	4.494 (27,9%)	20,0%–32,9%
2018	16.241 (26,3%)	11.938 (73,5%)	70,5%–77,5%	4.303 (26,5%)	22,5%–29,5%
2019	15.230 (24,6%)	11.657 (76,5%)	72,6%–81,1%	3.573 (23,5%)	18,9%–27,4%
Geschlecht					
Männer	37.172 (60,1%)	27.252 (73,3%)	69,4%–79,6%	9.920 (26,7%)	20,4%–30,6%
Frauen	24.634 (39,9%)	18.213 (73,9%)	70,5%–79,9%	6.421 (26,1%)	20,1%–29,5%
Medianes Alter bei Diagnose in Jahren (25er–75er Perzentile)	68 (61–75)	67 (60–74)	Mediane: 67–68	71 (63–78)	Mediane: 70–74
Stadium					
Early Stage (UICC Stadium IA bis IIB)	11.387 (18,4%)	9.105 (80,0%)	75,8%–86,4%	2.282 (20,0%)	13,6%–24,2%
Locally Advanced (UICC Stadium IIIA bis IIIC)	11.554 (18,6%)	9.144 (79,1%)	74,8%–84,9%	2.410 (20,9%)	15,1%–25,2%
Metastatic Disease (UICC Stadium IV)	28.659 (46,5%)	20.972 (73,2%)	69,5%–79,7%	7.687 (26,8%)	20,3%–30,5%
Fehlend	6.296 (10,2%)	3.452 (54,8%)	49,5%–71,1%	2.844 (45,2%)	28,9%–50,5%
Nicht TNM-fähig	3.910 (6,3%)	2.792 (71,4%)	48,7%–83,2%	1.118 (28,6%)	16,8%–51,3%
Grading*					
G1	1.629 (2,8%)	1.204 (73,9%)	70,5%–78,7%	425 (26,1%)	21,3%–29,5%
G2	14.093 (24,1%)	10.806 (76,7%)	71,8%–84,2%	3.287 (23,3%)	15,8%–28,2%
G3	22.058 (37,8%)	16.769 (76,0%)	72,5%–83,4%	5.289 (24,0%)	16,6%–27,5%
G4	3.349 (5,7%)	2.631 (78,6%)	72,3%–84,1%	718 (21,4%)	15,9%–27,7%
GX	2.075 (3,6%)	1.595 (76,9%)	30,0%–78,8%	480 (23,1%)	21,2%–70,0%
U	15.041 (25,8%)	9.945 (66,1%)	63,3%–73,6%	5.096 (33,9%)	26,4%–36,7%
T	130 (0,2%)	93 (71,5%)	28,6%–82,4%	37 (28,5%)	17,6%–71,4%
Fehlend	3.431 (5,6%)	2.422 (70,6%)	65,4%–85,1%	1.009 (29,4%)	14,9%–34,6%

(Fortsetzung)

Tabelle 2: Beschreibung der Lungenkrebsfälle (ICD-10 C34, Diagnosejahre 2016–2019) insgesamt und nach Vorhandensein von mind. 1 Therapieinformation
(absolute und relative Häufigkeit bezogen auf die jeweilige Zeile, sofern nicht anders angegeben)

	Erkrankungs- fälle insgesamt (Spalten- prozente)	Erkrankungs- fälle mit mind. 1 bekannten Therapie- ereignis	Spannweite der Werte zwischen den vier Registern	Erkrankungs- fälle ohne bekanntes Therapie- ereignis	Spannweite der Werte zwischen den vier Registern
Histologiegruppe					
Karzinome					
– Adenokarzinom	26.675 (43,2%)	19.839 (74,4%)	70,1%–81,7%	6.836 (25,6%)	18,3%–29,9%
– Plattenepithelkarzinom (SCC)	13.604 (22,0%)	9.975 (73,3%)	70,1%–79,4%	3.629 (26,7%)	20,6%–29,9%
– Kleinzelliges Lungen- karzinom (SCLC)	9.103 (14,7%)	7.139 (78,4%)	74,5%–8,05%	1.964 (21,6%)	15,0%–25,5%
– Großzelliges Lungen- karzinom (LCLC)	2.515 (4,1%)	1.379 (54,8%)	49,6%–6,07%	1.136 (45,2%)	33,0%–50,4%
– Andere spez. Karzinome	5.940 (9,6%)	4.303 (72,4%)	67,3%–80,5%	1.637 (26,6%)	19,5%–32,7%
Andere spez. mal. Neoplasien	51 (0,1%)	28 (54,9%)	25,0%–100%	23 (45,1%)	37,9%–75,0%
Andere unspez. mal. Neoplasien	3.854 (6,2%)	2.766 (71,8%)	48,8%–84,0%	1.088 (28,2%)	16,0%–51,2%
Sarkome	62 (0,1%)	35 (56,5%)	37,5%–100%	27 (43,5%)	38,1%–62,5%
Fehlender Morphologie- Code	2 (<0,1%)	1 (50,0%)	50,0%–50,0%	1 (50,0%)	50,0%–50,0%
Mediane Anzahl an Mona- ten unter Beobachtung (25er–75er Perzentile)	9,0 (3,0–19)	10 (4,0–21)	Mediane: 9–15	3,9 (1,0–12)	Mediane: 3–4
Progress, Rezidiv, Fernmetastase im Beobachtungszeitraum	15.082 (24,4%)	13.132 (28,9%)	22,0%–37,0%	1.950 (11,9%)	6,3%–14,3%
Verstorben im Beobachtungszeitraum	41.613 (67,3%)	28.872 (69,4%)	66,3%–75,0%	12.741 (30,6%)	25,0%–33,7%
Medianes Alter bei Tod in Jahren (25er–75er Perzentile)	70 (63–77)	69 (62–76)	Mediane: 68–70	73 (65–79)	Mediane: 72–75

* 82 Tumoren mit Angabe Grading = L wurden der Gruppe G2, 30 mit Angabe M wurden G3 und 178 mit Angabe H wurden G4 zugeordnet.

Diskussion

In zunehmendem Maße werden von den Krebsregistern in Europa Behandlungsdaten erhoben und berichtet, allerdings mit derzeit noch großer Heterogenität bezüglich der erfassten Daten und der Datenqualität [21]. In Deutschland liegen in den behandlungsortbezogenen (klinischen) Landeskrebsregistern umfangreiche Daten zu Diagnose, Therapie und zum Erkrankungsverlauf von Krebserkrankungen, die im Einzugsgebiet des jeweiligen Krebsregisters aufgetreten und therapiert worden sind, vor. Eine Datenzusammenführung von (ausführlichen klinischen) Registerdaten auf Basis des oBDS war zum Zeitpunkt der Studie noch nicht geschehen. Aus diesem Grund prüften wir in einem Pilotprojekt Machbarkeit und Aufwand der Datenzusammenführung sowie die Dateneignung mit einem Fokus auf Vollständigkeit der Angaben zur Deskription der onkologischen Versorgung am Beispiel von Lungenkrebs.

Datenzusammenführung und Bildung der besten Information

Zum Zeitpunkt der Projektplanung und -durchführung gab es noch keinen ‚bundesweiten klinischen Krebsregisterdatensatz‘ mit Informationen zur Therapie und zu Verlaufsereignissen [13], der beim ZfKD angefordert werden konnte, so dass die Daten bei den Krebsregistern beantragt werden mussten.

Außerdem gab es weder ein einheitliches Antragsformular noch ein gemeinsames Antragsportal, was die Beantragung erschwerte. Mittlerweile ist ein einheitliches Antragsformular von der Plattform § 65c entwickelt worden (Beschlussnummer 2024/70/06), wodurch sich der Aufwand für die Datenbeantragung verringert hat. Die Entwicklung und dauerhafte Implementierung eines gemeinsamen Antragsportals für die Daten aus unterschiedlichen Krebsregistern, wie sie derzeit im Projekt AI-Care ange-

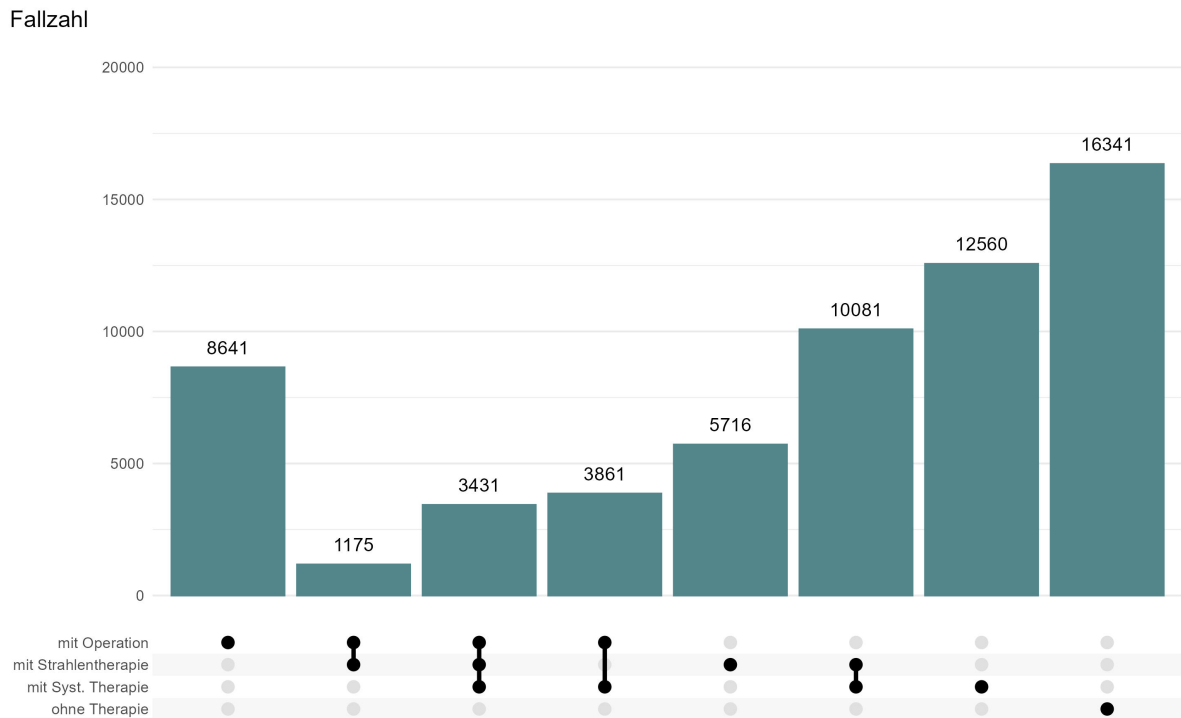


Abbildung 3: Anzahl an Erkrankungsfällen ohne bzw. mit Best-of-Informationen zur Therapie
 Datengrundlage sind die 61.806 Erkrankungsfälle unter Anwendung der Ausschlusskriterien. Für einen Erkrankungsfall können entweder keine, eine oder eine Kombination mehrerer Therapiearten im Krebsregister bekannt sein.

strebt wird [22], ist wünschenswert und könnte den Aufwand der Datenbeantragung zukünftig weiter reduzieren. Der Prozess der Datenaufbereitung, Harmonisierung und Zusammenführung war mit einer Dauer von neun Monaten von der Datenbeantragung bis zur ersten Version einer qualitätsgesicherten Datenbank mit sechs (Teil-) Datensätzen sehr zeitaufwändig. Unter anderem trugen zu vage formulierte Vorgaben bei der Datenbeantragung zu Struktur und Inhalt von Teil-Datensätzen sowie die noch nicht überall umgesetzte Bildung der „besten Therapie-Information“ und die abweichenden Zeitfenster „Beantragung bis Übermittlung der Daten“ zum hohen Aufwand bei. Der Grund für die starke Abweichung der Zeitfenster liegt u.a. darin, dass zum Zeitpunkt der Datenbeantragung in einem Bundesland im Falle eines Antrags auf Überlassung von Individualdaten zunächst ein Beirat und ein wissenschaftlicher Fachausschuss der Datenübermittlung über den Antrag beraten und eine Empfehlung abgeben mussten. Beide Gremien tagen gemäß gesetzlicher Grundlage mindestens einmal pro Jahr [5] – aktuell tagen sie regelhaft halbjährlich. Mitte 2023 wurde dieses Prozedere geändert und aktuell muss in diesem Bundesland nur noch der wissenschaftliche Fachausschuss über den Antrag beraten.

In Bezug auf die Struktur und das Format der Datensätze waren im Antrag auf Datenbereitstellung nur implizit Vorgaben gemacht worden, u.a. um es den Krebsregistern zu erlauben, auf gegebenenfalls vorhandene Datenextraktionsroutinen zurückgreifen zu können. Um den Arbeitsaufwand bei der Datenvereinheitlichung und -zusammenführung klein zu halten, empfiehlt sich eine intensive

Absprache mit den Registern zu Anzahl, Format und konkretem Inhalt der angefragten Datensätze sowie zu den Ein- und Ausschlusskriterien.

Start und Ende der Strahlen- und der systemischen Therapie sind in der behandlungsortbezogenen Krebsregistrierung in vielen Bundesländern getrennte Meldeanlässe. Deswegen muss bei der Bildung der besten Information geprüft werden, ob erwartungsgemäß zwei Meldungen zu einem Therapieanlass vorliegen und ob Angaben logisch zusammenpassen bzw. in welcher Meldung überhaupt eine oder die „bessere“ Information steht. Bei Vorliegen von Start- und Endmeldung, aber abweichenden Angaben zu einem Merkmal kann auf Basis definierter Algorithmen (z.B. *anhand der von den Krebsregistern erarbeiteten Empfehlungen zur Best-of-Bildung*) entschieden werden, welche Angabe jeweils in die Best-of-Information übernommen wird. Die Therapiedaten wurden überwiegend (im Falle der systemischen Therapie von allen Registern) in Form von fallzuordnungsfähigen Meldungen bereitgestellt. Die Komplexität der Programmierung der Best-of-Bildung der Therapieereignisse trug erheblich zum Aufwand der Datenaufbereitung bei. Sobald die Register für alle Therapiearten Best-of-Informationen bereitstellen können, wird sich der Aufwand deutlich reduzieren. Da seit dem Jahr 2022 fallbasierte (und nicht meldungsbasierte) Informationen durch die Krebsregister an das ZfKD geliefert werden [13], ist davon auszugehen, dass die Best-of-Bildung in den Krebsregistern aktuell weitestgehend flächendeckend und für alle Meldungsarten umgesetzt ist. Ob diese auch die komplexe Abgren-

Tabelle 3: Anzahl an Fällen mit Best-of-Informationen und Vollständigkeit der Angaben zur Therapie
(absolute und relative Häufigkeit, sofern nicht anders angegeben)

	Häufigkeit	Spannweite der Werte zwischen den vier Registern
Abschnitt A: operative Eingriffe		
Anzahl an Erkrankungsfällen mit mind. einem bekannten operativen Eingriff	17.108 (27,7%)	20,2%–31,1%
Mediane Anzahl an Operationen pro Erkrankungsfall Bezogen auf alle 61.806 Fälle (25er–75er Perzentile) Bezogen auf alle Fälle mit Best-of-Informationen zur entsprechenden Therapie (25er–75er Perzentile)	0 (0–1) 1 (1–2)	Mediane: 0–0 Mediane: 1–1
Anteil der operativen Eingriffe innerhalb von 6 Monaten nach Diagnose Erste OP innerhalb von 6 Monaten Erste OP erst nach 6 Monaten Folge-OP (ohne Berücksichtigung des Datums)	15.774 (79,2%) 1.334 (6,7%) 2.809 (14,1%)	76,6%–89,4% 3,9%–8,3% 5,5%–15,3%
Intention der Therapie Bekannt (oBDS Ausprägungen: K, P, D, R, S) Unbekannt (oBDS Ausprägung: X) Fehlend	17.482 (87,8%) 1.599 (8,0%) 836 (4,2%)	79,3%–93,4% 6,6%–20,4% 0%–12,1%
Tumorresezierender OPS-Code* Ja, relevanter Code Nein, sonstige OPS-Codes angegeben Fehlende OPS-Codes in Therapiemeldung	11.570 (58,1%) 8.142 (40,9%) 205 (1,0%)	49,9%–70,3% 29,7%–50,1% 0%–3,1%
Komplikationen nach OP Ja Nein Unbekannt Keine Angabe / fehlend	1.112 (5,6%) 6.677 (33,5%) 1.930 (9,7%) 10.198 (51,2%)	3,6%–8,4% 22,6%–72,9% 0,7%–22,5% 0%–58,1%
Anteil der Meldungen mit gültigen Informationen zum Datum, zur Intention und zu den OPS-Codes Inkl. Ausprägung „Unbekannt“ bei Intention Ohne Ausprägung „Unbekannt“ bei Intention	18.887 (94,8%) 17.290 (86,8%)	85,0%–100% 77,5%–93,3%
Abschnitt B: Teilbestrahlungen / Strahlentherapie		
Anzahl an Erkrankungsfällen mit mind. einer bekannten Teilbestrahlung (Strahlentherapie)	20.403 (33,0%)	28,4%–39,4%
Mediane Anzahl an Teilbestrahlungen pro Erkrankungsfall Bezogen auf alle 61.608 Fälle (25er–75er Perzentile) Bezogen auf alle Fälle mit Best-of-Informationen zur entsprechenden Therapie (25er–75er Perzentile)	1 (1–2) 2 (1–3)	Mediane: 0–1 Mediane: 1–2
Anteil der Teilbestrahlungen mit Start innerhalb von 6 Monaten nach Diagnose Erste Bestrahlung innerhalb von 6 Monaten Erste Bestrahlung erst nach 6 Monaten Folge-Bestrahlung (ohne Berücksichtigung des Startdatums)	16.339 (50,6%) 4.064 (12,6%) 11.886 (36,8%)	46,4%–65,7% 10,9%–12,9% 23,3%–41,0%
Enddatum der Therapie fehlt	1.672 (5,2%)	3,0%–6,7%
Beendigungsgrund der Therapie Bekannt (oBDS Ausprägungen: A, E, V, P, S) Unbekannt (oBDS Ausprägung: U) Fehlend	29.283 (90,7%) 348 (1,1%) 2.658 (8,2%)	85,1%–94,3% 0,1%–1,9% 3,8%–14,6%
Stellung zur operativen Therapie Bekannt (oBDS Ausprägungen: O, A, N, I) Sonstiges (oBDS Ausprägung: S) Fehlend	29.160 (90,3%) 1.840 (5,7%) 1.289 (4,0%)	84,2%–96,4% 2,9%–7,2% 0,1%–8,6%
Applikationsart Bekannt (oBDS Ausprägungen: P, K, I, M) Sonstiges (oBDS Ausprägung: S) Fehlend	27.625 (85,6%) 4.115 (12,7%) 549 (1,7%)	82,3%–99,9% 0%–14,5% 0%–3,2%
Intention der Therapie Bekannt (oBDS Ausprägungen: K, P, S) Unbekannt (oBDS Ausprägung: X) Fehlend	30.484 (94,4%) 1.714 (5,3%) 91 (0,3%)	89,2%–99,4% 0,5%–10,8% 0%–0,6%
Zielgebiet Vorhanden Fehlend	27.529 (85,3%) 4.760 (14,7%)	73,3%–99,7% 0,3%–26,7%

(Fortsetzung)

Tabelle 3: Anzahl an Fällen mit Best-of-Informationen und Vollständigkeit der Angaben zur Therapie
(absolute und relative Häufigkeit, sofern nicht anders angegeben)

	Häufigkeit	Spannweite der Werte zwischen den vier Registern
Abschnitt B: Teilbestrahlungen / Strahlentherapie (Fortsetzung)		
Seite Bekannt (oBDS Ausprägungen: L, R, B, M) Unbekannt (oBDS Ausprägung: U) Fehlend	21.194 (65,6%) 7.365 (22,8%) 3.730 (11,6%)	61,3%–84,3% 11,7%–36,3% 2,1%–25,8%
Gesamtdosis Vorhanden Fehlend	29.492 (91,3%) 2.797 (8,7%)	90,0%–95,1% 4,9%–10,0%
Anteil der Meldungen mit gültigen Informationen zum Therapieende (Datum, Grund), zur Stellung, Applikationsart, Intention und zum Zielgebiet Inkl. der Ausprägung „Unbekannt“ bei Intention / Beendigungsgrund und „Sonstiges“ bei Stellung und Applikationsart Ohne Ausprägung „Unbekannt“ bei Intention / Beendigungsgrund	23.327 (72,2%) 23.179 (71,8%)	65,0%–92,7% 64,4%–92,6%
Abschnitt C: Systemische Therapie		
Anzahl an Erkrankungsfällen mit mind. 1 einer bekannten systemischen Therapie	29.933 (48,4%)	46,1%–52,9%
Mediane Anzahl an systemischen Therapien pro Erkrankungsfall bezogen auf alle 61.806 Fälle (25er–75er Perzentile) bei Vorliegen von Best-of Informationen zur entsprechenden Therapie (25er–75er Perzentile)	1 (0–3) 2 (1–4)	Mediane: 1–2 Mediane: 2–3
Anteil der systemischen Therapien mit Start innerhalb von 6 Monaten nach Diagnose Erste Therapie innerhalb von 6 Monaten Erste Therapie erst nach 6 Monaten Folge-Therapie (ohne Berücksichtigung des Startdatums)	27.706 (47,4%) 2.227 (3,8%) 28.478 (48,8%)	45,6%–59,8% 3,1%–4,3% 37,1%–51,0%
Enddatum der Therapie fehlt	23.372 (40,0%)	21,5%–54,4%
Beendigungsgrund der Therapie Bekannt (oBDS Ausprägungen: A, E, V, R, P, S) Unbekannt (oBDS Ausprägung: X) Fehlend	35.504 (60,8%) 22.626 (38,7%) 281 (0,5%)	53,1%–70,8% 29,1%–46,2% 0,1%–0,7%
Stellung zur operativen Therapie Bekannt (oBDS Ausprägungen: O, A, N, I) Sonstiges (oBDS Ausprägung: S) Fehlend	48.232 (82,6%) 6.663 (11,4%) 3.516 (6,0%)	73,0%–93,7% 4,0%–26,4% 0,5%–13,4%
Intention der Therapie Bekannt (oBDS Ausprägungen: K, P, S) Unbekannt (oBDS Ausprägung: X) Fehlend	53.841 (92,2%) 4.570 (7,8%) 0 (0%)	87,4%–97,0% 3,0%–12,6% 0%–0%
Angaben im Feld Protokoll der systemischen Therapie sind vorhanden	20.318 (34,8%)	0%–79,6%
Angaben zu den Substanzen der systemischen Therapie sind vorhanden	54.424 (93,2%)	91,2%–95,7%
Angegebene Substanz entspricht Therapieart Chemotherapie, Immuntherapie, Zielgerichtete Substanz Ja Nein (andere Substanz) Fehlende Angabe zur Substanz	53.393 (91,4%) 1.031 (1,8%) 3.987 (6,8%)	90,7%–94,9% 0,5%–4,1% 4,3%–8,8%
Anteil der Meldungen mit gültiger Information zum Therapieende (Datum, Grund), zur Stellung, Intention und zur Substanz der systemischen Therapie inkl. Ausprägung „Unbekannt“ bei Intention / Beendigungsgrund und „Sonstiges“ bei Stellung ohne Ausprägung „Unbekannt“ bei Intention / Beendigungsgrund	31.709 (54,3%) 27.065 (46,3%)	41,8%–70,8% 35,4%–60,7%
Anteil der Meldungen mit gültiger Information zur Stellung, Intention und zur Substanz der systemischen Therapie** inkl. Ausprägung „Unbekannt“ bei Intention und „Sonstiges“ bei Stellung ohne Ausprägung „Unbekannt“ bei Intention	51.305 (87,8%) 48.076 (82,3%)	83,9%–95,2% 76,3%–88,2%

* Segmentresektion und Bisegmentresektion der Lunge (OPS 5-323), Einfache (5-324) bzw. Erweiterte Lobektomie und Bilobektomie der Lunge (5-325), Einfache (5-327) bzw. Erweiterte (Pleuro-)Pneum(on)ektomie (5-328)

** Da einige systemische Therapien über einen sehr langen Zeitraum gegeben werden können und dieser sich ggf. über den Zeitpunkt der Datenanforderung hinaus erstrecken kann, wird der Anteil der Meldungen mit „vollständigen“ Informationen zusätzlich berechnet, ohne dass die Angaben zum Therapieende (Datum und Grund) erwartet werden.

zung von Mehrfachtherapien, wie First- und Second-Line-Therapien umfassen wird, bleibt abzuwarten.

Die Analysen auf Ebene der Best-of-Informationen erlauben neben Aussagen zur Vollständigkeit der Informationen auch – sofern die Therapien vollzählig gemeldet werden – Aussagen über typische Behandlungsmuster und Überleben nach bestimmten Therapien. Dabei gilt es zu bedenken, dass sich die Krebsregisterdaten mit ihrer vorgegebenen Struktur und definierten Merkmalsausprägungen – anders als individuelle klinische Fallakten – der klinischen Komplexität eines Falles nur annähern können.

Vollzähligkeit der dokumentierten Therapien

Der Anteil an Fällen mit Informationen zu mindestens einer Therapieart (74%) ist unseres Erachtens aktuell noch zu niedrig, liegt aber in vergleichbarer Höhe wie bspw. in den Daten aus dem SEER-Register (78,5% bei nichtkleinzelligem Lungenkrebs [NSCLC] bzw. 75,1% bei NSCLC in der Altersgruppe 80 Jahre und älter) [23], [24]. Ob bei den Fällen ohne bekannte Therapie lediglich die Information (Meldung) zum Zeitpunkt der Datenübermittlung im Register fehlt oder ob die Therapie nicht durchgeführt wurde, bleibt derzeit offen. Für Personen mit NSCLC im Stadium III kann auf Basis einer Analyse aus Zentraleuropa davon ausgegangen werden, dass rund 5% mit „Best Supportive Care“ versorgt werden [25] – dies sind Therapien, die möglicherweise nicht konsequent an die Krebsregister gemeldet werden. Die Vollzähligkeit von Therapiemeldungen soll im Rahmen einer weiteren Analyse ausführlicher untersucht werden.

Vollständigkeit der Angaben zu den dokumentierten Therapien

Sofern dem Register eine Therapie bekannt ist, kann insbesondere für die operativen Eingriffe von einer hohen Vollständigkeit (rund 95%) der wesentlichen Informationen ausgegangen werden. In der Gesamtschau liegt nur bei wenigen Variablen ein hoher Anteil fehlender Werte vor. Zu nennen sind hier die Angaben zu Komplikationen nach operativen Eingriffen und zu den Nebenwirkungen einer Strahlen- oder systemischen Therapie. Ein „Reporting Bias“ (*Verzerrung durch selektive Dokumentation*) kann diverse Ursachen haben und an verschiedenen Stellen des Dokumentationsprozesses entstehen: Das Auftreten unerwünschter Ereignisse ist kein eigener Meldeanlass nach landesrechtlichen Bestimmungen der involvierten Krebsregister [3], [4], [5], [6]). Zudem sieht der oBDS ohnehin nur die Meldung von unerwünschten Ereignissen, die innerhalb von 90 Tagen nach Therapie auftreten, vor; während diese in der klinischen Realität häufig erst später auftreten. Weiterhin sind die Angaben in den involvierten Registern keine „Mindestangabe“ bzw. kein Pflichtfeld bei der Meldung [26], [27], [28], daher kann ein selektives Nichtdokumentieren von „keine

Komplikation“ zum Zeitpunkt der Meldung zu Datenverzerrungen führen. Eine kürzlich veröffentlichte Studie zu NSCLC und Immuntherapie deutet auf weitere Gründe für den hohen Anteil fehlender Werte hin: Neben der Angst vor dem Abbruch der Therapie sind hauptsächlich die individuelle Einschätzung der Schweregrade und die Machbarkeit des Umgangs mit ihnen dafür verantwortlich, dass Komplikationen oder Nebenwirkungen von Erkrankten nicht gegenüber dem behandelnden Team geäußert und damit auch nicht dokumentiert werden [29]. Hinweise auf ein systematisches Fehlen von Angaben finden sich in den Daten in der US-amerikanischen National Cancer Database. Hier geht ein hoher Anteil fehlender Angaben mit einem schlechteren Überleben einher [30].

Fehlen die Informationen zum Enddatum oder zum Beendigungsgrund in den Endmeldungen einer Therapie oder in den Best-of-Informationen (bei Vorliegen von Start- und/oder Endmeldung) oder fehlen diese Endmeldungen selber, kann dies als Qualitätsdefizit gewertet werden. Beim Antrag auf Datenbereitstellung sollte daher der Meldeanlass für die Therapien (Start oder Ende) angefragt werden – mit dieser Angabe lässt sich unterscheiden, ob Qualitätsdefizite für die Endmeldung vorliegen oder Endmeldungen für die Therapien fehlen. Für das Pilotprojekt wurde aus Datensparsamkeitsgründen auf diese Angabe verzichtet, so dass das Defizit nur indirekt überprüft werden kann. Meldungen zu systemischen Therapien, die in den Jahren 2020 und 2021 initiiert wurden, wiesen erwartungsgemäß deutlich höhere Anteile fehlender Werte zum Beendigungsgrund (49% bzw. 76%) auf als solche für Therapien mit Start in 2016 bis 2019 (35%–40%).

Die Krebsregistergesetze in Schleswig-Holstein und Nordrhein-Westfalen sehen eine Meldung innerhalb von 6 und in Hamburg von 8 Wochen nach Bekanntwerden der Diagnose bzw. Therapie vor, in Baden-Württemberg hat die Meldung spätestens im Folgequartal zu erfolgen [3], [4], [5], [6]. Deswegen und wegen der medianen Dauer der ersten bekannten systemischen Therapie von 64 Tagen bzw. einem Abschluss der Therapie bei 75% aller Fälle innerhalb von 105 Tagen, wäre zu erwarten gewesen, dass das Ende der Therapie und der Beendigungsgrund bei Datenlieferung im Jahr 2021 bekannt sein sollten. Jedoch kann es aufgrund von fehlenden oder noch ausstehenden Meldungen seitens der Behandelnden bzw. aufgrund von Bearbeitungsrückstau im Register zu fehlenden Informationen insbesondere bei aktuelleren Therapien kommen.

Leider melden die an der onkologischen Versorgung beteiligten Kliniken und Praxen Informationen zur Therapie noch nicht in dem Maße und in der Qualität, die für eine valide Deskription wünschenswert wären. Es fehlen aktuell beispielsweise bei 15% der Best-of-Informationen zu den Teilbestrahlungen die Angaben zum Zielgebiet der Bestrahlung und bei 9% die Angaben zur Gesamtstrahldosis. Dies mag einerseits an mangelnden Ressourcen in den Kliniken und Praxen liegen, andererseits ist aber denkbar, dass zum Zeitpunkt der Meldung noch nicht alle Informationen zur Verfügung gestanden haben. In

den gepoolten Datensätzen zeigte sich, dass die Register unterschiedlich streng bei der Daten-Entgegennahme sind und in variierendem Ausmaß fehlende oder zunächst als „unbekannt“ kodierte Informationen nachrecherchieren. So gibt es je nach Register unterschiedliche Variablen, in denen die Merkmalsausprägung „unbekannt“ im Vergleich zu den anderen Registern auffällig häufig vorkommen.

Während die SEER-Daten nur eine limitierte Anzahl an Merkmalen zu den Primärtherapien, die innerhalb der ersten 24 Monate in Kliniken begonnen wurden, beinhalten [31] und damit nicht das vollständige Bild der onkologischen Versorgung widerspiegeln [32], [33], umfassen die Daten aus den deutschen Krebsregistern nicht nur die Primär- sondern auch die weiteren Therapien sowohl aus dem stationären als auch dem ambulanten Sektor. Hierin liegt eine Besonderheit, die eine umfassende Deskription der onkologischen Versorgung auf Basis der Krebsregisterdaten erlaubt – unter der Annahme, dass alle Therapien (*vollzählige Meldung*) mit vollständigen und validen Angaben zu den per oBDS definierten Merkmalen (*vollständige Meldungen*) in die Landeskrebsregister gelangen.

Datenumfang des ‚bundesweiten klinischen Krebsregisterdatensatz‘ im Vergleich zu selbst beantragten Daten bei den Krebsregistern

Der aktuell zu beantragende ‚bundesweite klinische Krebsregisterdatensatz‘ umfasst nicht alle Variablen des oBDS, so fehlen z.B. Angaben zu Leistungserbringern und aus den organspezifischen Modulen sowie zu Biomarkern. Weiterhin fehlen einzelne Angaben zu den Therapien wie unerwünschte Ereignisse und Komplikationen oder auch die Dosisangaben zur Strahlentherapie und der Residualstatus nach Abschluss der Primärtherapie [13]. Für tiefergehende Analysen müssen daher auch weiterhin selbst beantragte Daten der Landeskrebsregister genutzt werden.

Ein weiterer Nachteil des ‚bundesweiten klinischen Krebsregisterdatensatzes‘ ist, dass das früheste verfügbare Diagnose- und Behandlungsjahr in diesem Datensatz das Jahr 2020 ist. Sofern die Therapie und das Überleben vor und nach Zulassung neuartiger Therapeutika untersucht werden soll (wie bspw. von Personen mit nicht-kleinzelligem Lungenkrebs vor und nach Zulassung der Behandlung mit Osimertinib bei vorliegender EGFR-Mutation [34] oder einer Kombination aus Nivolumab, Ipilimumab und zwei Zyklen platinbasierter Chemotherapie beim metastasierten NSCLC [35]), muss auch weiterhin auf die Daten der Landeskrebsregister zurückgegriffen werden, in denen Behandlungs- und Verlaufsdaten auch für den Zeitraum vor dem Jahr 2020 vorliegen.

Fazit

In unserem Pilotprojekt zeigte sich, dass die Zusammenführung von klinischen Daten aus den Landeskrebsregistern zwar noch sehr (zeit-)aufwändig, aber technisch gut machbar war. Länderspezifische Unterschiede und unklare Angaben bei der Datenanforderung erschwerten und verlangsamten die Zusammenführung der Daten u.a. aufgrund unterschiedlicher Antragsprozeduren und Formate der bereitgestellten Daten insbesondere aber aufgrund der noch nicht vollständig umgesetzten Best-of-Bildung zu den Therapien in den Registern.

Die Datengenerierung unterliegt in den hier involvierten Bundesländern z.T. abweichenden (unterschiedliche Landesgesetzgebungen, Krebsregister-Software, Meldestrukturen und Register Routinen), z.T. gemeinsamen Rahmenbedingungen (Meldepflicht, oBDS). Bei der Interpretation der Daten sollten diese berücksichtigt werden. Die Frage der Vollständigkeit bzw. das Fehlen von Angaben in Datenfeldern (Variablen) muss sorgfältig analysiert werden, insbesondere weil unklar ist, ob Daten zufällig oder systematisch in Abhängigkeit vom Krankheitsstadium, Allgemeinzustand oder Alter oder in Abhängigkeit des Versorgungsbereiches (ambulantes/stationäres Setting) fehlen.

In den gepoolten Datensätzen waren die Unterschiede zwischen den Bundesländern bezüglich Datenverteilung und Datenqualität gering. Mit vertretbarem Aufwand konnte eine bevölkerungsbezogene Datenbank generiert werden, die mit mehr als 61.000 Fällen hinreichend groß ist, um auch für seltene histologische Subgruppen oder Altersgruppen mit wenigen Erkrankungsfällen – unter der Voraussetzung valider und vollzähliger Informationen – aussagekräftige Auswertungen vornehmen zu können. Selbst bei Beschränkung auf Erkrankungsfälle mit Informationen zu mindestens einer Therapieart bzw. auf Erkrankungsfälle mit vollständigen (wesentlichen) Informationen der Therapie ergeben sich immer noch sehr hohe Fallzahlen. Diese Fallzahlen sind vergleichbar mit aktuell laufenden onkologischen Langzeit-Registerstudien [36]. Es bleibt zu hoffen, dass wegen der gesetzlich vorgeschriebenen Lieferung an das ZfKD Best-of-Informationen für die Therapien zeitnah in den Registern vorliegen werden. Dies wird dazu führen, dass die Zusammenführung von Daten zur Therapie und zum Verlauf aus den Landeskrebsregistern deutlich einfacher werden wird. Es bleibt abzuwarten, ob durch die Stufe 2 des Bundesgesetzes [12] die anlassbezogene Datenzusammenführung und -bereitstellung mittels einer Zentralen Antrags- und Registerstelle schneller vonstattengehen wird. Aber auch dann wird eine gute Kommunikation mit den Registern für die Datenauswertung und -interpretation bedeutsam sein. Der Anteil mit gänzlich fehlenden Informationen zur Therapie ist unseres Erachtens mit ca. 26% noch recht hoch, aber vergleichbar zu anderen Registerstudien [23], [24]. Sofern Informationen zu den Therapien vorliegen, sind diese Angaben überwiegend hinreichend vollständig. Daher können die Grundzüge einzelner Therapien gut beschrieben und für die onkologische Versorgungsfor-

schung genutzt werden. Mögliche Verzerrungen durch fehlende Angaben müssen aber berücksichtigt werden, insbesondere dann, wenn Schlussfolgerungen für die onkologische Versorgung aus Populationssicht gezogen werden sollen.

Anmerkungen

Danksagung

Die Autor*innen danken Volker Arndt aus dem Krebsregister Baden-Württemberg für seine kritischen und hilfreichen Hinweise zu früheren Versionen dieses Manuskripts.

Interessenkonflikte

Die Autoren erklären, dass sie keine Interessenkonflikte in Zusammenhang mit diesem Artikel haben. Sie weisen darauf hin, dass sie (mit Ausnahme von AW, LL, HB) in Landeskrebsregistern angestellt sind oder diese leiten.

Ethik

Für diesen Beitrag wurden keine Studien an Menschen durchgeführt. Die Richtlinien für das Auswerten von anonymisierten Daten wurden eingehalten. Das Forschungsvorhaben wurde bei der Ethikkommission der Universität zu Lübeck angezeigt.

Literatur

1. Bundesregierung. Gesetz zur Weiterentwicklung der Krebsfrüherkennung und zur Qualitätssicherung durch klinische Krebsregister (Krebsfrüherkennungs- und -registergesetz – KFRG) vom 3. April 2013
2. Katalinic A, Halber M, Meyer M, Pflüger M, Eberle A, Nennecke A, Kim-Wanner SZ, Hartz T, Weitmann K, Stang A, Justenhoven C, Hollecsek B, Piontek D, Wittenberg I, Heßmer A, Kraywinkel K, Spix C, Pritzkuleit R. Population-Based Clinical Cancer Registration in Germany. *Cancers (Basel)*. 2023 Aug;15(15):3934. DOI: 10.3390/cancers15153934
3. Landesregierung Baden-Württemberg. Gesetz über die Krebsregistrierung in Baden-Württemberg (Landeskrebsregistergesetz – LKrebsRG) vom 7. März 2006.
4. Senat Freie und Hansestadt Hamburg. Hamburgisches Krebsregistergesetz (HmbKrebsRG) vom 27. Juni 1984.
5. Landesregierung Nordrhein-Westfalen. Gesetz über die klinische und epidemiologische Krebsregistrierung im Land Nordrhein-Westfalen (Landeskrebsregistergesetz – LKRG NRW) vom 2. Februar 2016.
6. Landesregierung Schleswig-Holstein. Gesetz über das Krebsregister des Landes Schleswig-Holstein (Krebsregistergesetz – KRG SH) vom 4. November 2015.
7. Bundesregierung. Bundeskrebregisterdatengesetz (BKRG) vom 10. August 2009 (BGBl. I S. 2702,2707) – zuletzt geändert durch Artikel 2 des Gesetzes vom 18. August 2021 (BGBl. I S. 3890).
8. Bundesministerium für Gesundheit. Bekanntmachung: Module zur Dokumentation des Brust- und Darmkrebses in Ergänzung des aktualisierten einheitlichen onkologischen Datensatzes der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) vom 28. Oktober 2015. BAnz AT. 26.11.2015;(B1):1-10.
9. Bundesministerium für Gesundheit. Bekanntmachung: Modul zur Dokumentation des Prostatakrebses in Ergänzung des aktualisierten einheitlichen onkologischen Datensatzes der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) vom 9. August 2017. BAnz AT. 29.08.2017;(B6):1-6.
10. Bundesministerium für Gesundheit. Bekanntmachung: Modul zur Dokumentation des Malignen Melanoms in Ergänzung des einheitlichen onkologischen Basisdatensatzes der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) vom 25. Mai 2020. BAnz AT. 26.06.2020;(B4):1-4.
11. Bundesministerium für Gesundheit. Bekanntmachung: Aktualisierter einheitlicher onkologischer Basisdatensatz der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) vom 10. Mai 2021. BAnz AT. 12.07.2021;(B4):1-31.
12. Bundesregierung. Gesetz zur Zusammenführung von Krebsregisterdaten. 2021.
13. Meisegeier S, Imhoff M, Berg K, Kraywinkel K. Bundesweiter klinischer Krebsregisterdatensatz – Datenschema und Klassifikationen (oBDS_v3.0.0.8a_RKI). Zenodo; 2023. DOI: 10.5281/zenodo.10022040
14. Arbeitsgemeinschaft Deutscher Tumorzentren e.V. Die ADT als Vertrauensstelle für Bundesweite Qualitätskonferenzen in der Onkologie. [last accessed 2024 Dec 13]. Available from: https://adt-netzwerk.de/Vertrauensstelle/Bundesweite_Qualitaetskonferenzen/
15. Roessler M, Schmitt J, Bobeth C, Gerken M, Kleihues-van Tol K, Reissfelder C, Rau BM, Distler M, Piso P, Günster C, Klinkhammer-Schalke M, Schoffer O, Bierbaum V. Is treatment in certified cancer centers related to better survival in patients with pancreatic cancer? Evidence from a large German cohort study. *BMC Cancer*. 2022 Jun;22(1):621. DOI: 10.1186/s12885-022-09731-w
16. Rudolph C, Germer S, Nennecke A, Kusche H, Labohm L, Rath N, Rausch K, Hollecsek B, Handels H, Katalinic A. Künstliche Intelligenz (KI) in der Krebsregistrierung: Methoden, Herausforderungen und erste Ergebnisse des AI-Care-Datensatzes. In: Gesundheit – gemeinsam. Kooperationstagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP), Deutschen Gesellschaft für Epidemiologie (DGEpi), Deutschen Gesellschaft für Medizinische Soziologie (DGMS) und der Deutschen Gesellschaft für Public Health (DGPH). Dresden, 08.-13.09.2024. Düsseldorf: German Medical Science GMS Publishing House; 2024. DocAbstr. 867. DOI: 10.3205/24gmds742
17. Stegmaier C, Hentschel S, Hofstädter F, Katalinic A, Tillack A, Klinkhammer-Schalke M, editors. Das Manual der Krebsregistrierung. München: Zuckschwerdt; 2019.
18. Operationen- und Prozedurenschlüssel (OPS). [last accessed 2024 Jan 24]. Available from: https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/_node.html
19. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.

20. Therapie des Lungenkarzinoms – eine Versorgungsforschungsstudie auf Basis von Krebsregisterdaten nach § 65c SGB V. DRKS00025080. In: Deutsches Register Klinischer Studien (DRKS). [last accessed 2024 Jan 24]. Available from: <https://drks.de/search/de/trial/DRKS00025080>
21. Giusti F, Martos C, Trama A, Bettio M, Sanvisens A, Audisio R, Arndt V, Francisci S, Dochez C, Ribes J, Fernández LP, Gavin A, Gatta G, Marcos-Gragera R, Lievens Y, Allemani C, De Angelis R, Visser O, Van Eycken L; ENCR Working Group on Treatment Data Harmonisation. Cancer treatment data available in European cancer registries: Where are we and where are we going? *Front Oncol.* 2023;13:1109978. DOI: 10.3389/fonc.2023.1109978
22. Institut für Krebsepidemiologie e.V. (IKE) der Universität zu Lübeck. KI-unterstützte, versorgungsnahe Nutzung von Krebsregisterdaten – AI-CARE. [last accessed 2024 Dec 6]. Available from: <https://ai-care-cancer.de/>
23. Bei Y, Chen X, Raturi VP, Liu K, Ye S, Xu Q, Lu M. Treatment patterns and outcomes change in early-stage non-small cell lung cancer in octogenarians and older: a SEER database analysis. *Aging Clin Exp Res.* 2021 Jan;33(1):147-56. DOI: 10.1007/s40520-020-01517-z
24. Ganti AK, Klein AB, Cotala I, Seal B, Chou E. Update of Incidence, Prevalence, Survival, and Initial Treatment in Patients With Non-Small Cell Lung Cancer in the US. *JAMA Oncol.* 2021 Dec;7(12):1824-32. DOI: 10.1001/jamaoncol.2021.4932
25. Zemanova M, Pirker R, Petruzella L, Zbozínková Z, Jovanovic D, Rajer M, Bogos K, Purkalne G, Ceriman V, Chaudhary S, Richter I, Kufa J, Jakubikova L, Zemaitis M, Cernovska M, Koubkova L, Vilasova Z, Dieckmann K, Farkas A, Spasic J, Fröhlich K, Tiefenbacher A, Hollosi V, Kultán J, Kolarová I, Votruba J. Care of patients with non-small-cell lung cancer stage III – the Central European real-world experience. *Radiol Oncol.* 2020 May;54(2):209-20. DOI: 10.2478/raon-2020-0026
26. Krebsregister Baden-Württemberg. Krebsregistrierung in Baden-Württemberg: Datenkatalog mit Merkmalsausprägungen nach ADT/GEKID Basisdatensatz 2.0.0 für Tumorzentren, Onkologische Schwerpunkte, Krankenhäuser und niedergelassene Ärzte. 2017.
27. Krebsregister Schleswig-Holstein. Meldepflicht – Mindestangaben (zwingend notwendige Angaben zur Erfüllung der Vollständigkeit einer Meldung und zur Auszahlung der Meldevergütung). 2018.
28. Landeskrebsregister NRW. Melder-Broschüre Krebsregistrierung in Nordrhein-Westfalen: Meldepflicht – Meldepflichtige Erkrankungen – Vergütung. Bochum; 2020.
29. Martin ML, Chung H, Rydén A. Willingness to report treatment-related symptoms of immunotherapy among patients with non-small cell lung cancer. *Qual Life Res.* 2022 Apr;31(4):1147-55. DOI: 10.1007/s11136-021-02966-3
30. Yang DX, Khara R, Miccio JA, Jairam V, Chang E, Yu JB, Park HS, Krumholz HM, Aneja S. Prevalence of Missing Data in the National Cancer Database and Association With Overall Survival. *JAMA Netw Open.* 2021 Mar;4(3):e211793. DOI: 10.1001/jamanetworkopen.2021.1793
31. SEER Acknowledgment of Treatment Data Limitations. For the 1975-2019 Data (November 2021 Submission). [last accessed 2023 Dec 14]. Available from: <https://seer.cancer.gov/data-software/documentation/seerstat/nov2021/treatment-limitations-nov2021.html>
32. Du XL, Key CR, Dickie L, Darling R, Delclos GL, Waller K, Zhang D. Information on chemotherapy and hormone therapy from tumor registry had moderate agreement with chart reviews. *J Clin Epidemiol.* 2006 Jan;59(1):53-60. DOI: 10.1016/j.jclinepi.2005.06.002
33. Noone AM, Lund JL, Mariotto A, Cronin K, McNeel T, Deapen D, Warren JL. Comparison of SEER Treatment Data With Medicare Claims. *Med Care.* 2016 Sep;54(9):e55-64. DOI: 10.1097/MLR.0000000000000073
34. EGFR-mutiertes (EGFRmut) NSCLC: Zulassung für Osimertinib firstline. *Im Focus Onkologie.* 2018;21(7):80. DOI: 10.1007/s15015-018-4126-4
35. Zulassung für Opdivo(R) / Yervoy(R): Erstlinie für NSCLC. *Dtsch Ärztebl.* 2020 Nov;117(Suppl. Pneumologie & Allergologie 2):40.
36. Nicht-interventionelle Studie (Anwendungsbeobachtung) NIS-Nr.: 356. CRISP – Clinical Research platform Into molecular testing, treatment and outcome of (non-)Small cell lung carcinoma Patients. [last accessed 2023 Dec 14]. Available from: <https://www.pei.de/SharedDocs/awb/nis-0301-0400/0356.html>

Korrespondenzadresse:

Prof. Dr. Annika Waldmann
Institut für Sozialmedizin und Epidemiologie, Universität zu Lübeck, Ratzeburger Allee 160 (Haus 409), 23562 Lübeck, Deutschland, Tel.: +49 451 50051240
Annika.Waldmann@uksh.de

Bitte zitieren als

Waldmann A, Labohm L, Baltus H, Eisfeld C, Jansen L, Löhden I, Nennecke A, Oesterling F, Pritzkeleit R, Katalinic A. Zusammenführung von umfangreichen Daten und Vollständigkeit der Informationen zu Diagnose, Behandlung und zum Erkrankungsverlauf aus mehreren populationsbezogenen Krebsregistern in Deutschland – erste Erfahrungen am Beispiel von Lungenkrebs. *GMS Med Inform Biom Epidemiol.* 2025;21:Doc02. DOI: 10.3205/mibe000274, URN: urn:nbn:de:0183-mibe0002746

Artikel online frei zugänglich unter

<https://doi.org/10.3205/mibe000274>

Veröffentlicht: 30.05.2025

Copyright

©2025 Waldmann et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.

Merging of comprehensive data and completeness of information on diagnosis, treatment and progression from several population-based cancer registries in Germany – initial experiences using the example of lung cancer

Abstract

Background: Germany has a nationwide cancer registration system. Population-based cancer registries are organised at federal state level. To date, research-driven merging of detailed (clinical) data on diagnosis, therapy and progression from several state cancer registries has only rarely been conducted. We examined the feasibility and effort of merging the data as well as the suitability of the data for describing oncological care.

Methods: Data on lung cancer (ICD-10 C34) from the diagnosis years 2016–2019 were requested from four population-based cancer registries, processed and merged. If several case-assignable reports on one therapy event were available, these were aggregated to provide the “best information” as part of the pilot study. The data is analysed in a descriptive and explorative manner.

Results: Data preparation and collation were time-consuming, but technically feasible. The analysis dataset includes over 61,000 cases of disease with follow-up. Information on at least one type of therapy (surgery, radiotherapy, systemic therapy) is available for 74%. The provided information on the details of the therapies is mostly complete or has only a few missing values.

Discussion: Despite some differences in the general conditions of cancer registration in the federal states at the time of this study, differences in data distribution and data quality in the pooled data set were small. The proportion with completely missing information on therapy (26%) could be regarded as high in our opinion, but comparable to other registry studies. Based on the available information, the main features of individual cancer therapies can be described and the information could be used for oncological health services research.

Keywords: population-based cancer registries, data merging, data quality, therapy, follow-up

Annika Waldmann¹

Louisa Labohm¹

Hannah Baltus¹

Christine Eisfeld²

Lina Jansen³

Imma Löhden⁴

Alice Nennecke⁴

Florian Oesterling²

Ron Pritzkuleit⁵

Alexander Katalinic¹

1 Institut für Sozialmedizin und Epidemiologie, Universität zu Lübeck, Lübeck, Germany

2 Landeskrebsregister NRW gGmbH, Bochum, Germany

3 Epidemiologisches Krebsregister Baden-Württemberg, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

4 Hamburgisches Krebsregister, Freie und Hansestadt Hamburg, Behörde für Wissenschaft, Forschung, Gleichstellung und Bezirke, Hamburg, Germany

5 Krebsregister Schleswig-Holstein, Registerstelle, Lübeck, Germany

Introduction

In 2013, the Act on the Further Development of Early Cancer Detection and Quality Assurance through Clinical Cancer Registries laid the basis for nationwide clinical cancer registration in Germany, which is now set out in Section 65c of the German Social Security Code V (SGB V) [1], [2]. State-specific cancer registry laws [3], [4], [5], [6] and Section 8 of the Federal Cancer Registry Data Act (BKRG) [7] regulate data processing, reporting channels and occasions, among other things. The oncological basic data set (oBDS) and supplementary modules [8], [9], [10], [11] regulate the type and scope of the data to be registered. The law on the consolidation of cancer registry data from 2021 [12] forms the basis for the nationwide consolidation of part of the clinical data at the Center for Cancer Registry Data (ZfKD) into the 'nationwide clinical cancer registry dataset' [13]. As a first step, data on therapy and progression from the state cancer registries were merged for the first time in 2023 [13]. On request, the nationwide data can be provided as individual case data. This data does not include all variables of the oBDS, e.g. information from the organ-specific modules as well as on biomarkers and health-care providers is missing. State cancer registry data must therefore continue to be used for more in-depth analyses. In a second stage, event-related, patient- and health-care provider-related, cross-registry compilations of cancer registry data are to be realised [12]. A detailed concept for this stage is still pending.

Merging of detailed clinical data on therapy and disease progression from several German state cancer registries was initially rare – with the exception of the evaluations that are regularly carried out on the occasion of the German Cancer Congress and presented there under the term “Nationwide Oncological Quality Conference” [14]. In the meantime, individual research projects are beginning to merge national clinical data [15], [16].

In a pilot project, we wanted to test the feasibility and cost of merging data as well as its suitability for describing oncological care. Tumour, therapy and progression data on individuals with lung cancer were requested from four federate state cancer registries, processed and used to answer the following questions:

- How much effort is required to merge the data?
- Are the treatments known in the register fully documented so that they are suitable for describing oncological care?

Methods

Request of data

The federate state cancer registries of Baden-Württemberg, Hamburg, North Rhine-Westphalia and Schleswig-Holstein were contacted for the pilot project, as they have been collecting corresponding data since at least 2016.

The following considerations also played a role in the selection of these registries: (1) a maximum of two different types of cancer registry software should be used, (2) cancer registries with both large and small underlying populations and (3) city states and territorial states should contribute to the database.

At the turn of the year 2020/2021, the data request procedure was coordinated and, in addition to the registry-specific forms, a general study protocol and a list of variables were sent to the registries. The list contained the requested variables from the oBDS in the desired format (numeric, string).

Information was requested on cases of lung cancer (ICD-10 C34) from the diagnosis years 2016–2019. The patients had to be at least 18 years of age and the registry had to have minimum information on the diagnosis. In addition, the patients had to live in the respective federal state at the time of diagnosis – selection on so-called residence-related cases, which allows a systematic survey of vital status via a comparison of data with registration office information or mortality data [17].

Data provision and processing

In the registries, the data is available in relational databases in two different forms. On the one hand, the databases contain “raw data” in the form of case-assignable reports, as transmitted to the registry by reporters. These reports are made for specific reporting events (e.g. *diagnosis, surgery, start of radiotherapy*, etc.) and usually only contain data that can be assigned to this event (e.g. *surgery reports do not contain any information on concomitant systemic therapy*). Several reports from different reporters/physicians may be available for one reporting occasion (e.g. *a diagnosis report from a practice and a specialised clinic*). On the other hand, the cancer registry databases contain so-called best-of information. These are summaries of several reports (see next section for best-of formation), which ultimately depict the course of a case of disease. A best-of dataset has various 1:n connections – 1 patient : n tumours, 1 tumour : n therapies (Figure 1).

At the time of data application, best-of information was not yet available in the registries for all episodes of a disease course, so that in some cases case-assignable reports were transmitted and the task of best-of formation lay with the pilot project. Information on the diagnosis, i.e. the description of the primary tumour (tumour data), was provided from all registries as best-of information; this was not always the case for the treatments and courses of disease (Table 1). The partial datasets provided were made available in “long layout” (data in a stacked structure; i.e. *if a disease case was operated twice, there were two rows in the dataset for this disease case*) or in “wide layout” (unstacked structure; *if a disease case was operated twice, the columns with the information on the surgeries were doubled and each disease case was included in the dataset with one row*). In

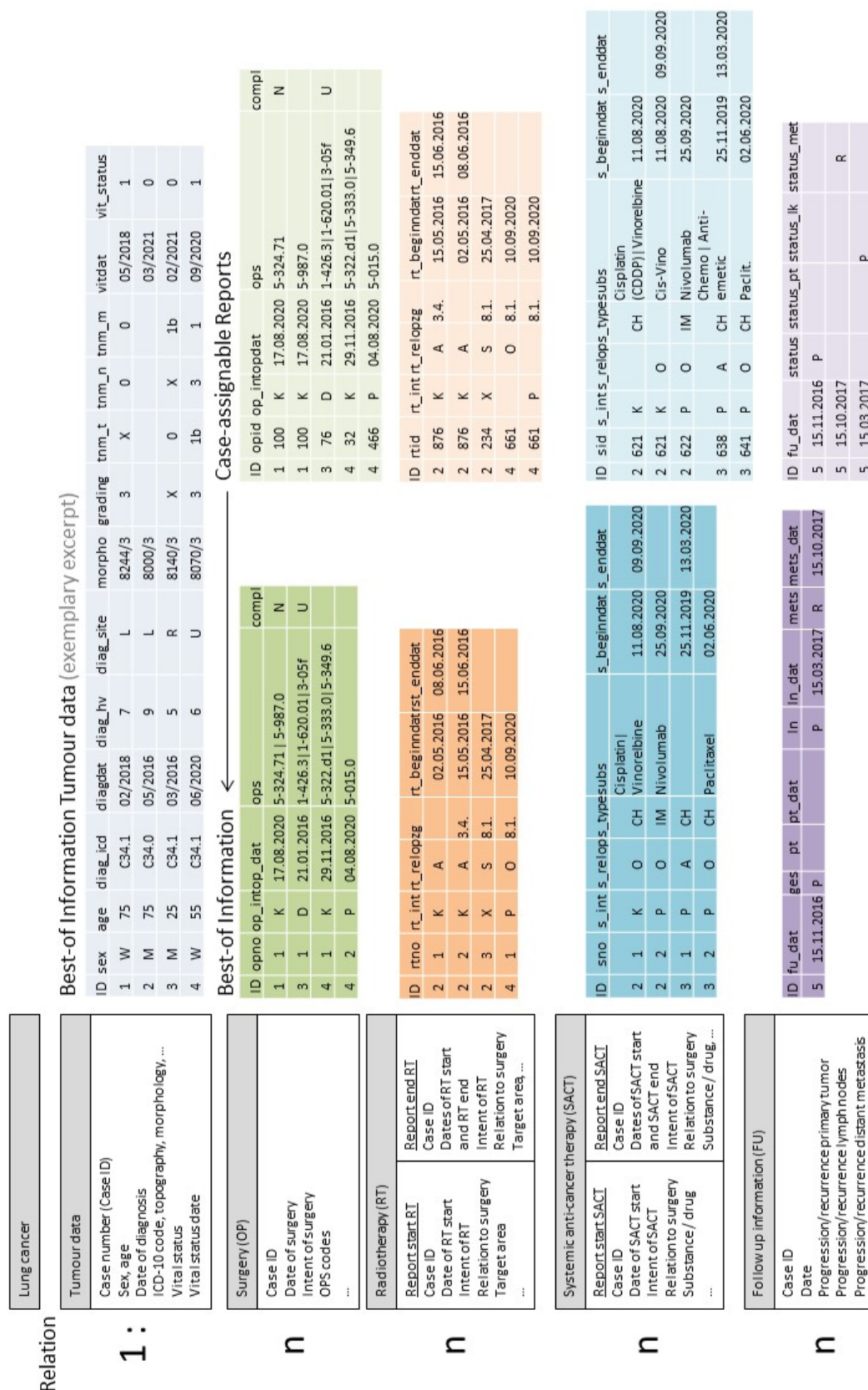


Figure 1: Partial data sets of the nationwide clinical cancer registration – relation of the data sets to each other as well as exemplary content and outline of the procedure for the formation of the best-of information on the basis of case-assignable reports

In the sub-dataset 'Best-of information tumour data', there is exactly one row in the dataset for each case. In the best-of datasets on therapy, there is exactly one row for each therapy occasion (e.g. surgery on the primary tumour, revision surgery; partial irradiation as part of radiotherapy), although a lung cancer case can occur several times in the dataset if there are several therapy occasions (1:n relation).

Table 1: Description of the data provided and special features of the data provision

	Special features of data provision	Time spans or ranges of the case numbers and relative proportions of cases in the four registries
Call for data send		December 2020 – January 2021
Receipt of data 1 st data delivery		March 2021 – August 2021*
If several data deliveries were made, period of the final delivery	Corrected or updated data was subsequently sent from three registers.	August 2021 – August 2022
Number of data files transmitted		4–7
File with tumour information	In two registries, the files with tumour data also contained information on progression events and metastases (both for the primary tumour and during progression).	Range of the number of cases** per register: 4,200–39,823
File with information from the notification type 'Progression' (progression, recurrence, metastasis)	See above	Range of the proportions with at least one progression event in the respective cases of a registry: 16.2%–33.2%
File with information from the notification type 'Metastases'	One registry provided information on metastases to the primary tumour and in the course of the disease in one file. One registry provided the information on metastases to the primary tumour and in the course in two separate files. Two registries provided this information in the file with the tumour information.	Range of the proportions of cases in a registry with information on at least one metastasis at the time of diagnosis: 33.3%–51.5% Range of the proportions with information on at least one metastasis in the course of the disease in the respective cases of a registry: 0%–11.0%
File with details of the surgery (Surg)	Two registers provided information in the form of case-assignable reports. Two registers provided best-of information.	After best-of formation: Range of the proportions of cases in a registry with information on surgery in the respective cases of a register: 19.3%–30.1%
File with information on radiotherapy (partial irradiation; RT)	Three registers provided information in the form of case-assignable reports. One register provided best-of information.	After best-of formation: Range of the proportions with information on RT in the respective cases of a register: 23.7%–38.1%
File with information on systemic therapy (SACT)	Four registers provided information in the form of case-assignable reports.	After best-of formation: Range of the proportion with information on SACT in the respective cases of a register: 38.9%–51.0%
Last systematic recording of vital status by comparing the data with registration office and mortality data from the public health department		December 2019 – April 2021
Proportion of cases known to the registry only via death certificates (DCO cases) in relation to all cases		Range of the proportions of DCO cases in the respective cases of a registry: 4.0%–12.1%
Proportion of cases with 0 days follow-up time		Range of the proportions with 0 days follow-up in the respective cases of a registry: 5.3%–17.8%
Proportion of cases with missing vital status		Range of the proportions with missing vital status in the respective cases of a registry: 0%–0.8%

* In one registry, both the advisory board and the scientific expert committee had to discuss the application for the transfer of individual data at the time of the data application and make a recommendation. Both committees meet at least once a year in accordance with the legal basis; in practice, they meet every six months. The procedure was changed in mid-2023 and currently only the scientific committee has to make a recommendation.

** Cases = tumours

the course of processing, all partial datasets were transferred to a stacked layout.

Information on substances or protocols of systemic therapy are reported to the registries via free text fields, whereby substances and trade names of the products are permitted according to the data set description of the oBDS [11]. A substance reference list agreed upon by the cancer registries was used for the preparation of this information, which, in addition to the standardisation of substances, also allows assignment to the types of therapy (chemotherapy, immunotherapy, etc.). The information on the type of surgical intervention is entered via text fields in which so-called OPS codes (surgery and procedure codes [18]) are to be entered in accordance with the data record description [11]. An OPS reference list agreed by the cancer registries was also used to decide whether an OPS code indicates a surgery with tumour resection.

Creation of best-of information

Data on the same information from several reporters/physicians must be combined into so-called best-of information (Figure 1). For many years, there have been recommendations for the (epidemiological) best-of formation of diagnostic information (tumour data table) – based on international guidelines and agreements of the state cancer registries – and the implementation of the recommendations has been established in the registries [17]. Since such guidelines for therapy information did not exist at the beginning of nationwide clinical cancer registration in Germany, a working group of the cancer registries has developed recommendations for best-of formation in the registries (date of adoption for the recommendations for surgery: May 7, 2019; radiotherapy: November 5, 2019; systemic therapy: January 11, 2022). If no best-of-therapy information was provided by the register, it was generated on the basis of the individual reports and the working groups' recommendations (as of spring 2021) in a time-consuming and programming-intensive process. Here, for example, the proposed "ranking" of characteristics was implemented if there were start and end reports of a therapy with discordant information on a characteristic.

Data processing and statistical evaluations

The data sets provided were processed, merged and evaluated using the open source statistical program R (version 4.1.3) [19]. The processing steps (including standardisation of variable names, formats and, if necessary, coding, best-of formation if necessary) were carried out separately for each register. Both the raw datasets and the datasets resulting from processing and merging were checked for data quality and plausibility. The variables in the pooled data sets were also checked for possible registry effects, i.e. differences in the frequency

distribution of the characteristic values between the registries.

To assess data homogeneity, the minimum and maximum values from the four registries are given for the percentage frequencies and medians.

For the analyses, cases that were only known to the registry on the basis of a death certificate (DCO case), cases with 0 days survival, cases with missing information on vital status and cases with information on therapies that occurred more than 31 days before the lung cancer diagnosis and that could be interpreted as an indication of other cancers were excluded.

Statistical hypothesis testing was not intended.

Ethics

The project was reported to the Ethics Committee of the University of Lübeck and approved (Ref. 20-483). It is registered in the German Register of Clinical Studies under the number DRKS00025080 [20].

Results

Data provision and processing

Table 1 illustrates the heterogeneity of data provision. Two registries transmitted the data promptly, while one registry took seven months. The number of partial datasets provided and their layout differed, partly due to vague specifications in the application for data transmission. Information on surgical interventions was provided by two registries and information on radiotherapy was provided by one registry as best-of information. The information on systemic or awaiting therapies (hereinafter referred to as "systemic therapy") was provided by all registries as case-assignable reports.

All registries included here collect vital status systematically and in a similar way by regularly comparing cancer registry data with information from the registration office or mortality data. The last comparison for the data set used here took place in one registry in December 2019, in another in December 2020 and in the two other registries in 2021.

Data basis: Pooled data set with best-of information

After processing, best-of formation and merging, a database with six partial data sets containing a total of 70,821 cases with lung cancer was available (Table 1, Figure 2): One dataset each with best-of information on diagnosis (and vital status), surgery, radiotherapy, systemic therapies (which can either consist of only one type of therapy such as chemotherapy or a combination of therapy types) or with follow-up information and another with information on metastases at the time of the primary tumour and during follow-up.

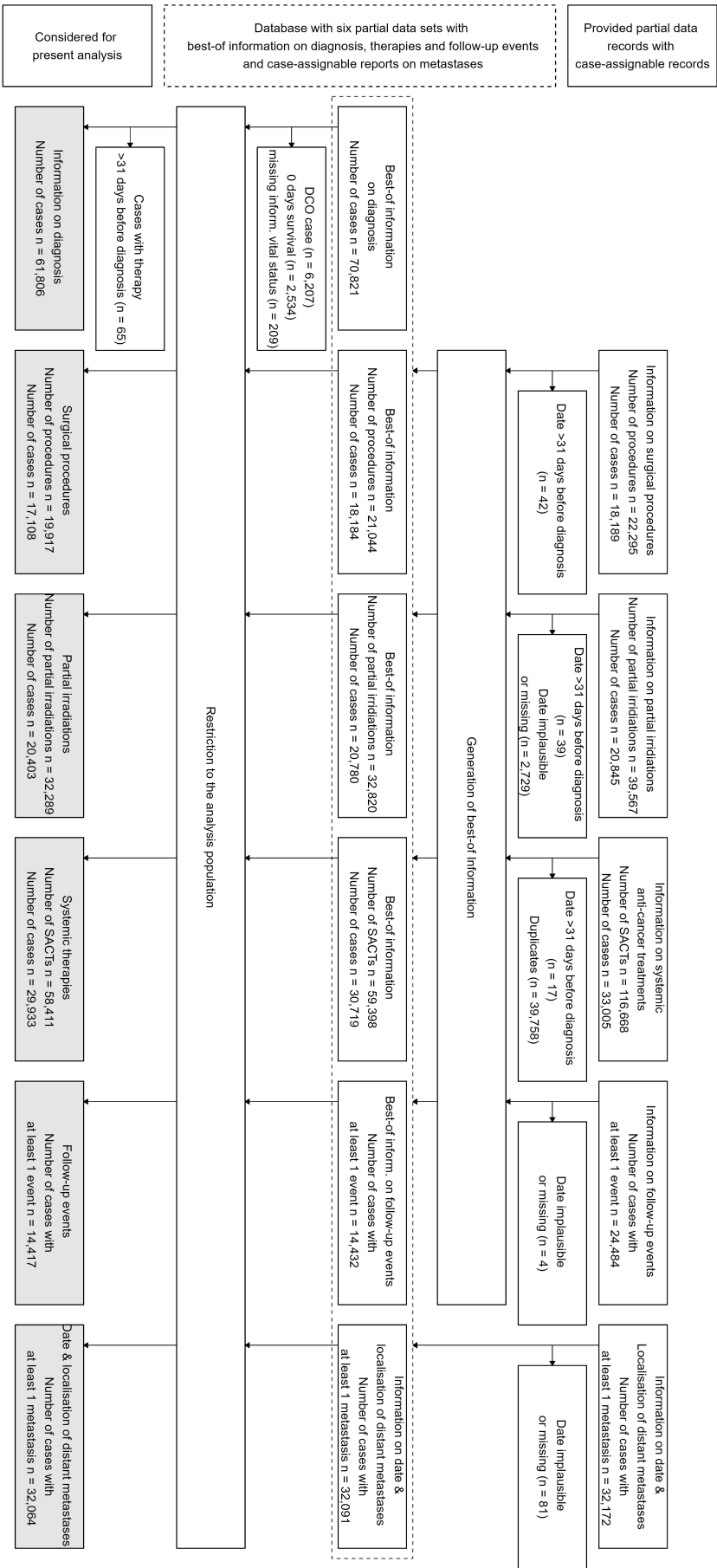


Figure 2: Flow chart

The analyses are based on 61,806 cases after application of the exclusion criteria (= evaluation population).

Completeness of documented therapies: Lung cancer cases with and without known therapy

In 45,465 of the 61,806 cases (73.8%), information on at least one treatment event was available. Compared to cases without information on treatment, these cases were on average 4 years younger, had a missing tumour stage less frequently, had large cell lung cancer less frequently and small cell lung cancer slightly more frequently, and died less frequently during the follow up period (Table 2). Where treatment was known, information was most frequently available on systemic therapy alone (n=12,560; 27.6%), on a combination of radiotherapy and systemic therapy (n=10,081; 22.2%) and on surgery alone (n=8,641; 19.0%) (Figure 3).

Completeness of the information on the documented therapies

In total, information on 19,917 surgical interventions was available. At least one **surgical procedure** was known for 27.7% of the cases in the evaluation population (Figure 2, Table 3). This proportion varied between the registries in the range of 20.2% to 31.1%. Around 79.2% of all interventions were performed within 6 months of diagnosis (92.2% of all initial interventions). After forming the best-of information, specific information on the intention of the surgery was rarely missing (missing information or "unknown": 12.2%) and on surgical procedures (1.0%), while information on complications of the surgery was missing or unknown in more than 50% of cases. The frequency distributions across the four registries were mostly quite similar (Table 3 – section A). Around 95% of the cases with surgery information contained valid information on the date, intention and procedures performed. This proportion varied between the registries in the range of 85.0% to 100%.

For 33.0% of the lung cancer cases in the evaluation population, at least one piece of information on partial **radiotherapy** was available (Figure 2, Table 3). Around 50.6% of all (partial) irradiation were started within 6 months of diagnosis (80.1% of all initial partial irradiations). Information on the total dose of partial irradiation was available for more than 90%. With the exception of the information on the side of the target area (unknown: 22.8%, missing: 14.7%), the proportion of missing values was a maximum of 15%. The frequency distributions across the registries were also largely quite similar for the information from the reports on partial irradiation (Table 3 – section B). Overall, around 72% of this best-of information contained valid details on the start and end of treatment, position regarding surgery, type of application, intention and target area (spread between the registries: 65.0%–92.7%).

At least one best-of information on **systemic anti-cancer therapy** could be formed for 48.4% of cases (Figure 2, Table 3). Around 47.4% of all systemic therapies were started within 6 months of diagnosis (92.6% of all initial therapies). Information on the end date of therapy and the reason for discontinuation (unknown or missing) was missing in around 40% of cases. In around 93% of the best-of information, at least one substance of the systemic therapy could be assigned to the one of the types chemotherapy, immunotherapy or targeted therapy via the substance reference list. The information on systemic therapy across the registries was predominantly quite similar (Table 3 – section C). There were major differences in the information on the protocols of systemic therapy – both, the degree of completion and the content of the field, varied significantly. Information on the protocols was rarely found; instead, the information from the substances field was often repeated or supplemented, e.g. with additional substances or dosages or application types. Overall, around 88% of the best-of information on systemic therapy contained valid information on start, position in relation to the surgery, intention and substances administered (variation between the registries: 83.9%–95.2%).

Discussion

Cancer registries in Europe are increasingly collecting and reporting treatment data, although there is currently still great heterogeneity with regard to the data collected and data quality [21]. In Germany, comprehensive data on the diagnosis, treatment and course of disease of cancers that have occurred and been treated in the catchment area of the respective federate-state cancer registry are available in the treatment-related (clinical) federate-state cancer registries. At the time of the study, (detailed clinical) registry data had not yet been merged on the basis of the oBDS. For this reason, we conducted a pilot project to examine the feasibility and effort involved in merging the data as well as the suitability of the data with a focus on the completeness of the information on the description of oncological care using the example of lung cancer.

Data consolidation and creation of the best-of information

At the time of project planning and implementation, there was no 'nationwide clinical cancer registry dataset' with information on therapy and progression events [13] that could be requested from the ZfKD, so the data had to be requested from the cancer registries.

In addition, there was neither a standardised application form nor a common application portal, which made the application process more difficult. In the meantime, a standardised application form has been developed by the § 65c platform (resolution number 2024/70/06), which has reduced the effort involved in applying for data.

Table 2: Description of lung cancer cases (ICD-10 C34, diagnosis years 2016–2019) in total and according to the presence of at least 1 treatment information
(absolute and relative frequency in relation to the respective row, unless otherwise stated)

	Total cases (column percentages)	Cases with at least 1 known treatment event	Range of values between the four registries	Cases without a known treatment event	Range of values between the four registries
Number of cases included in the analyses	61,806 (100%)	45,465 (73.6%)	69.8%–79.8%	16,341 (26.4%)	20.2%–30.2%
Year of diagnosis					
2016	14,209 (23.0%)	10,238 (72.1%)	66.5%–81.9%	3,971 (27.9%)	18.1%–33.5%
2017	16,126 (26.1%)	11,632 (72.1%)	67.1%–80.0%	4,494 (27.9%)	20.0%–32.9%
2018	16,241 (26.3%)	11,938 (73.5%)	70.5%–77.5%	4,303 (26.5%)	22.5%–29.5%
2019	15,230 (24.6%)	11,657 (76.5%)	72.6%–81.1%	3,573 (23.5%)	18.9%–27.4%
Sex					
Male	37,172 (60.1%)	27,252 (73.3%)	69.4%–79.6%	9,920 (26.7%)	20.4%–30.6%
Female	24,634 (39.9%)	18,213 (73.9%)	70.5%–79.9%	6,421 (26.1%)	20.1%–29.5%
Median age at diagnosis in years (25 th –75 th percentile)	68 (61–75)	67 (60–74)	Median age: 67–68	71 (63–78)	Median age: 70–74
Stage					
Early stage (UICC stage IA to IIB)	11,387 (18.4%)	9,105 (80.0%)	75.8%–86.4%	2,282 (20.0%)	13.6%–24.2%
Locally advanced (UICC stage IIIA to IIIC)	11,554 (18.6%)	9,144 (79.1%)	74.8%–84.9%	2,410 (20.9%)	15.1%–25.2%
Metastatic disease (UICC stage IV)	28,659 (46.5%)	20,972 (73.2%)	69.5%–79.7%	7,687 (26.8%)	20.3%–30.5%
Missing	6,296 (10.2%)	3,452 (54.8%)	49.5%–71.1%	2,844 (45.2%)	28.9%–50.5%
TNM not applicable	3,910 (6.3%)	2,792 (71.4%)	48.7%–83.2%	1,118 (28.6%)	16.8%–51.3%
Grading*					
G1	1,629 (2.8%)	1,204 (73.9%)	70.5%–78.7%	425 (26.1%)	21.3%–29.5%
G2	14,093 (24.1%)	10,806 (76.7%)	71.8%–84.2%	3,287 (23.3%)	15.8%–28.2%
G3	22,058 (37.8%)	16,769 (76.0%)	72.5%–83.4%	5,289 (24.0%)	16.6%–27.5%
G4	3,349 (5.7%)	2,631 (78.6%)	72.3%–84.1%	718 (21.4%)	15.9%–27.7%
GX	2,075 (3.6%)	1,595 (76.9%)	30.0%–78.8%	480 (23.1%)	21.2%–70.0%
U	15,041 (25.8%)	9,945 (66.1%)	63.3%–73.6%	5,096 (33.9%)	26.4%–36.7%
T	130 (0.2%)	93 (71.5%)	28.6%–82.4%	37 (28.5%)	17.6%–71.4%
Missing	3,431 (5.6%)	2,422 (70.6%)	65.4%–85.1%	1,009 (29.4%)	14.9%–34.6%
Histology group					
Carcinomas					
– Adenocarcinoma	26,675 (43.2%)	19,839 (74.4%)	70.1%–81.7%	6,836 (25.6%)	18.3%–29.9%
– Squamous cell carcinoma (SCC)	13,604 (22.0%)	9,975 (73.3%)	70.1%–79.4%	3,629 (26.7%)	20.6%–29.9%
– Small cell lung carcinoma (SCLC)	9,103 (14.7%)	7,139 (78.4%)	74.5%–8.05%	1,964 (21.6%)	15.0%–25.5%
– Large cell lung carcinoma (LCLC)	2,515 (4.1%)	1,379 (54.8%)	49.6%–6.07%	1,136 (45.2%)	33.0%–50.4%
– Other spec. carcinomas	5,940 (9.6%)	4,303 (72.4%)	67.3%–80.5%	1,637 (26.7%)	19.5%–32.7%
Other spec. mal. neoplasms	51 (0.1%)	28 (54.9%)	25.0%–100%	23 (45.1%)	37.9%–75.0%
Other non-spec. mal. neoplasms	3,854 (6.2%)	2,766 (71.8%)	48.8%–84.0%	1,088 (28.2%)	16.0%–51.2%
Sarcomas	62 (0.1%)	35 (56.5%)	37.5%–100%	27 (43.5%)	38.1%–62.5%
Missing morphology code	2 (<0.1%)	1 (50.0%)	50.0%–50.0%	1 (50.0%)	50.0%–50.0%
Median number of months of follow-up (25 th –75 th percentile)	9.0 (3.0–19)	10 (4.0–21)	Median no. months: 9–15	3.9 (1.0–12)	Median no. months: 3–4
Progression, recurrence, distant metastasis during the observation period	15,082 (24.4%)	13,132 (28.9%)	22.0%–37.0%	1,950 (11.9%)	6.3%–14.3%
Deceased during the observation period	41,613 (67.3%)	28,872 (69.4%)	66.3%–75.0%	12,741 (30.6%)	25.0%–33.7%
Median age at time of death in years (25 th –75 th percentile)	70 (63–77)	69 (62–76)	Median age: 68–70	73 (65–79)	Median age: 72–75

* 82 tumours with grading = L were assigned to group G2, 30 with M were assigned to G3 and 178 with H were assigned to G4.

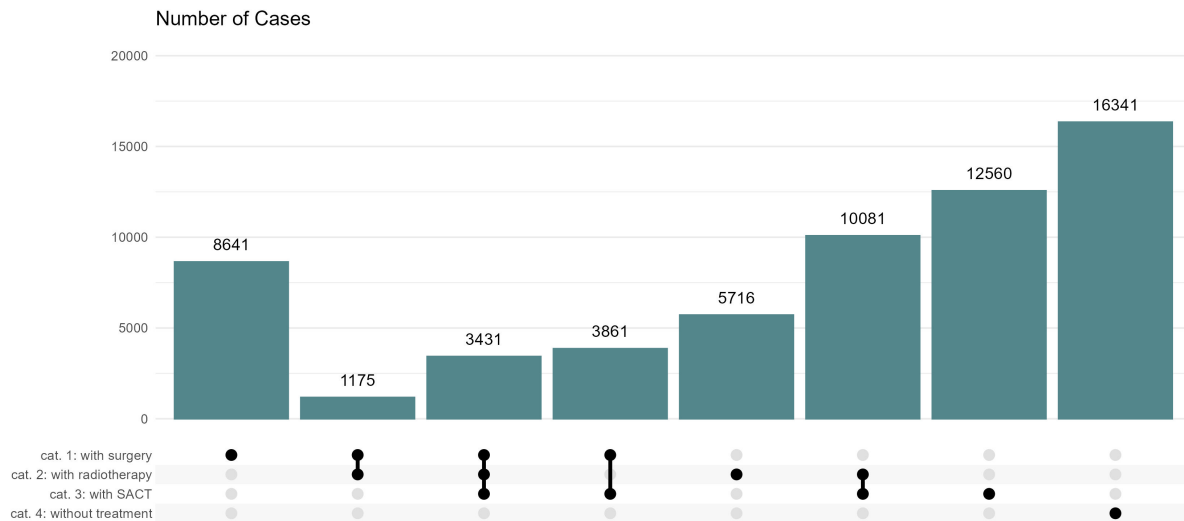


Figure 3: Number of cases without or with best-of information on treatment

The data is based on the 61,806 lung cancer cases using the exclusion criteria. For a case, either none, one or a combination of several types of therapy may be known in the cancer registry.

The development and permanent implementation of a common application portal for data from different cancer registries, as is currently being pursued in the AI-Care project [22], is desirable and could further reduce the effort involved in applying for data in the future.

The process of data preparation, harmonisation and consolidation was very time-consuming, taking nine months from the data application to the first version of a quality-assured database with six (partial) datasets. Among other things, too vaguely formulated specifications in the data application regarding the structure and content of partial datasets as well as the formation of the best-of information for treatments, which has not yet been implemented everywhere, and the deviating time windows “application to provision of the data” contributed to the high effort. The reason for the large deviation in the time frames is, among other things, that at the time of the data application in a federal state, in the case of an application for the transfer of individual patient level data, an advisory board and a scientific expert committee had to first discuss the application and make a recommendation. Both committees meet at least once a year [5] in accordance with the legal basis – they currently meet regularly every six months. This procedure was changed in mid-2023 and currently only the scientific committee has to discuss the application in this federal state.

With regard to the structure and format of the datasets, only implicit specifications were made in the application for data provision, among other things to allow the cancer registries to fall back on any existing data extraction routines. In order to keep the amount of work involved in data standardisation and merging to a minimum, intensive consultation with the registries regarding the number, format and specific content of the requested datasets as well as the inclusion and exclusion criteria is recommended.

In many federal states, the start and end of radiotherapy and systemic therapy are separate reporting events in cancer registration. Therefore, when determining the best-of information, it must be checked whether, as expected, two reports are available for one treatment occasion and whether the information logically matches or which report contains one or the “better” information. If there are start and end reports but differing information on a characteristic, a decision can be made on the basis of defined algorithms (e.g. *using the recommendations for best-of formation developed by the cancer registries*) as to which information should be included in the best-of information. The therapy data was predominantly provided (in the case of systemic therapy by all registries) in the form of case-assignable reports. The complexity of programming the best-of formation of therapy events contributed significantly to the effort involved in data preparation. As soon as the registries can provide best-of information for all therapy types, the effort will be significantly reduced. Since case-based (and not notification-based) information has been provided to the ZfKD by the cancer registries since 2022 [13], it can be assumed that best-of formation in the cancer registries is currently largely implemented in all cancer registries and for all notification types. It remains to be seen whether this will also include the complex delineation of multiple therapies, such as first and second-line treatments.

In addition to statements on the completeness of the information, the analyses at the level of best-of information also allow statements to be made on typical treatment patterns and survival after certain therapies, provided that the therapies are reported in full. It should be borne in mind that, unlike individual clinical case files, cancer registry data with its predefined structure and defined characteristics can only approximate the clinical complexity of a case.

Table 3: Number of cases with best-of information and completeness of treatment information
(absolute and relative frequency, unless otherwise stated)

	Frequency	Range of values between the four registries
Section A: Surgical procedures		
Number of cases with at least 1 known surgical intervention	17,108 (27.7%)	20.2%–31.1%
Median number of surgeries per case Based on all 61,806 cases (25 th –75 th percentile) For cases with best-of information on the corresponding therapy (25 th –75 th percentile)	0 (0–1) 1 (1–2)	Medians: 0–0 Medians: 1–1
Percentage of surgical interventions within 6 months of diagnosis First surgery within 6 months First surgery after 6 months Follow-up surgery (without taking the date into account)	15,774 (79.2%) 1,334 (6.7%) 2,809 (14.1%)	76.6%–89.4% 3.9%–8.3% 5.5%–15.3%
Intent of surgery Known (oBDS expressions: K, P, D, R, S) Unknown (oBDS expression: X) Missing	17,482 (87.8%) 1,599 (8.0%) 836 (4.2%)	79.3%–93.4% 6.6%–20.4% 0%–12.1%
OPS code indicates tumour resection* Yes, relevant code No, other OPS codes specified Missing OPS codes in treatment report	11,570 (58.1%) 8,142 (40.9%) 205 (1.0%)	49.9%–70.3% 29.7%–50.1% 0%–3.1%
Complications after surgery Yes No Unknown Not specified / missing	1,112 (5.6%) 6,677 (33.5%) 1,930 (9.7%) 10,198 (51.2%)	3.6%–8.4% 22.6%–72.9% 0.7%–22.5% 0%–58.1%
Percentage of reports with valid information on the date, intention and OPS codes Incl. 'Unknown' for intent Without 'Unknown' for intent	18,887 (94.8%) 17,290 (86.8%)	85.0%–100% 77.5%–93.3%
Section B: Partial irradiation / radiotherapy		
Number of cases with at least 1 known partial irradiation (radiotherapy)	20,403 (33.0%)	28.4%–39.4%
Median number of partial irradiation sessions per case Based on all 61,608 cases (25 th –75 th percentile) Based on all cases with best-of information on the corresponding therapy (25 th –75 th percentile)	1 (1–2) 2 (1–3)	Medians: 0–1 Medians: 1–2
Proportion of partial irradiation treatments started within 6 months of diagnosis First radiotherapy within 6 months First radiotherapy only after 6 months Follow-up radiotherapy (without taking the start date into account)	16,339 (50.6%) 4,064 (12.6%) 11,886 (36.8%)	46.4%–65.7% 10.9%–12.9% 23.3%–41.0%
End date of therapy is missing	1,672 (5.2%)	3.0%–6.7%
Reason for end of therapy Known (oBDS expressions: A, E, V, P, S) Unknown (oBDS expression: U) Missing	29,283 (90.7%) 348 (1.1%) 2,658 (8.2%)	85.1%–94.3% 0.1%–1.9% 3.8%–14.6%
Relation to surgical procedure Known (oBDS expressions: O, A, N, I) Other (oBDS expression: S) Missing	29,160 (90.3%) 1,840 (5.7%) 1,289 (4.0%)	84.2%–96.4% 2.9%–7.2% 0.1%–8.6%
Type of application Known (oBDS expressions: P, K, I, M) Other (oBDS expression: S) Missing	27,625 (85.6%) 4,115 (12.7%) 549 (1.7%)	82.3%–99.9% 0%–14.5% 0%–3.2%
Intent of therapy Known (oBDS expressions: K, P, S) Unknown (oBDS expression: X) Missing	30,484 (94.4%) 1,714 (5.3%) 91 (0.3%)	89.2%–99.4% 0.5%–10.8% 0%–0.6%
Target area Available Missing	27,529 (85.3%) 4,760 (14.7%)	73.3%–99.7% 0.3%–26.7%

(Continued)

Table 3: Number of cases with best-of information and completeness of treatment information
(absolute and relative frequency, unless otherwise stated)

	Frequency	Range of values between the four registries
Section B: Partial irradiation / radiotherapy (continued)		
Site Known (oBDS expressions: L, R, B, M) Unknown (oBDS expression: U) Missing	21,194 (65.6%) 7,365 (22.8%) 3,730 (11.6%)	61.3%–84.3% 11.7%–36.3% 2.1%–25.8%
Total dose Available Missing	29,492 (91.3%) 2,797 (8.7%)	90.0%–95.1% 4.9%–10.0%
Percentage of reports with valid information on the end of treatment (date, reason), position, application type, intention and target area Incl. 'Unknown' for intent/reason for termination and 'Other' for position and type of application Without 'Unknown' for intent / reason for termination	23,327 (72.2%) 23,179 (71.8%)	65.0%–92.7% 64.4%–92.6%
Section C: Systemic anti-cancer therapy (SACT)		
Number of cases with at least 1 known SACT	29,933 (48.4%)	46.1%–52.9%
Median number of SACTs per case Based on all 61,806 cases (25 th –75 th percentile) If best-of information on the corresponding therapy is available (25 th –75 th percentile)	1 (0–3) 2 (1–4)	Medians: 1–2 Medians: 2–3
Proportion of SACTs started within 6 months of diagnosis First therapy within 6 months First therapy after 6 months Follow-up therapy (without consideration of the start date)	27,706 (47.4%) 2,227 (3.8%) 28,478 (48.8%)	45.6%–59.8% 3.1%–4.3% 37.1%–51.0%
End date of therapy is missing	23,372 (40.0%)	21.5%–54.4%
Reason for end of therapy Known (oBDS expressions: A, E, V, R, P, S) Unknown (oBDS expression: X) Missing	35,504 (60.8%) 22,626 (38.7%) 281 (0.5%)	53.1%–70.8% 29.1%–46.2% 0.1%–0.7%
Relation to surgical procedure Known (oBDS expressions: O, A, N, I) Other (oBDS expression: S) Missing	48,232 (82.6%) 6,663 (11.4%) 3,516 (6.0%)	73.0%–93.7% 4.0%–26.4% 0.5%–13.4%
Intent of therapy Known (oBDS expressions: K, P, S) Unknown (oBDS expression: X) Missing	53,841 (92.2%) 4,570 (7.8%) 0 (0%)	87.4%–97.0% 3.0%–12.6% 0%–0%
Information in the SACT protocol field is available	20,318 (34.8%)	0%–79.6%
Information on the substances used in SACT is available	54,424 (93.2%)	91.2%–95.7%
Specified substance corresponds to therapy type Chemotherapy, Immunotherapy, Targeted substance Yes No (other substance) Missing information on substance	53,393 (91.4%) 1,031 (1.8%) 3,987 (6.8%)	90.7%–94.9% 0.5%–4.1% 4.3%–8.8%
Percentage of reports with valid information on the end of therapy (date, reason), position, intention and substance of SACT incl. 'Unknown' for intention/reason for termination and 'Other' for position without 'Unknown' for intent / reason for termination	31,709 (54.3%) 27,065 (46.3%)	41.8%–70.8% 35.4%–60.7%
Proportion of reports with valid information on the position, intention and substance of SACT** incl. 'Unknown' for intention and 'Other' for position without 'Unknown' for intent	51,305 (87.8%) 48,076 (82.3%)	83.9%–95.2% 76.3%–88.2%

* Segmental resection and bisegmental resection of the lung (OPS 5-323), simple (5-324) or extended lobectomy and bilobectomy of the lung (5-325), simple (5-327) or extended (pleuro-)pneum(ectomy) (5-328)

** Since some systemic therapies can be given over a very long period of time and this may extend beyond the time of the data request, the proportion of reports with 'complete' information is also calculated without expecting information on the end of therapy (date and reason).

Completeness of the documented therapies

In our opinion, the proportion of cases with information on at least one type of therapy (74%) is currently still too low, but is comparable to the data from the SEER registry (78.5% for non-small cell lung cancer [NSCLC] and 75.1% for NSCLC in the age group 80 years and older) [23], [24]. Whether the information (the report) is simply missing in the registry at the time of data provision in cases without known therapy or whether the therapy was not carried out remains open at present. Based on an analysis from Central Europe, it can be assumed that around 5% of people with stage III NSCLC are receiving 'best supportive care' [25] – these are therapies that may not be consistently reported to the cancer registries. The completeness of therapy reports is to be investigated in more detail as part of a further analysis.

Completeness of the information on the documented therapies

If a therapy is known to the registry, a high degree of completeness (around 95%) of the key information can be assumed, particularly for surgical interventions. Overall, there is only a high proportion of missing values for a few variables. These include information on complications following surgery and on the side effects of radiotherapy or systemic therapy. A 'reporting bias' (*distortion due to selective documentation*) can have various causes and arise at different points in the documentation process: The occurrence of adverse events is not a separate reporting cause according to the national legal provisions of the cancer registries involved [3], [4], [5], [6]. In addition, the oBDS only provides for the reporting of adverse events that occur within 90 days of treatment, whereas in clinical reality these often occur later. Furthermore, the information in the registries involved is not a 'minimum' or mandatory field for reporting [26], [27], [28], so selective non-documentation of 'no complication' at the time of reporting can lead to information and misclassification bias. A recently published study on NSCLC and immunotherapy points to further reasons for the high proportion of missing values: In addition to the fear of discontinuing therapy, the individual assessment of severity and the feasibility of dealing with them are mainly responsible for the fact that complications or side effects of patients are not expressed to the health-care providers and thus not documented [29]. Indications of a systematic lack of information can be found in the data in the US National Cancer Database. Here, a high proportion of missing information is associated with poorer survival [30].

If the information on the end date or the reason for termination is missing in the final reports of a therapy or in the best-of information (if start and/or end reports are available) or if these final reports themselves are missing, this can be considered a quality deficit. When applying

for data provision, the reason for reporting the treatments (start or end) should therefore be requested – this information can be used to distinguish whether there are quality deficits for the final report or whether final reports for the treatments are missing. For the pilot project, this information was omitted for reasons of data economy, so that the deficit can only be checked indirectly. As expected, reports on systemic therapies initiated in 2020 and 2021 showed significantly higher proportions of missing values for the reason for termination (49% and 76%, respectively) than those for therapies initiated in 2016 to 2019 (35–40%).

The cancer registry laws in Schleswig-Holstein and North Rhine-Westphalia stipulate reporting within 6 and in Hamburg within 8 weeks after the diagnosis or therapy becomes known; in Baden-Württemberg, reporting must take place in the following quarter at the latest [3], [4], [5], [6]. For this reason and due to the median duration of the first known systemic therapy of 64 days or a completion of therapy in 75% of all cases within 105 days, it would have been expected that the end of therapy and the reason for termination should be known when the data was provided in 2021. However, due to missing or outstanding reports from the treating physicians or due to processing backlogs in the registry, information may be missing, especially for more recent therapies.

Unfortunately, the clinics and practices involved in oncological care do not yet report treatment information to the extent and in the quality that would be desirable for a valid description. For example, 15% of the best-of information on partial irradiation currently lacks information on the target area of radiotherapy and 9% lack information on the total radiation dose. This may be due to a lack of resources in the clinics and practices, but it is also conceivable that not all information was available at the time of reporting. The pooled data sets showed that the registries differed in the strictness with which they accepted data. Depending on the registry, there are different variables in which the characteristic 'unknown' occurs conspicuously frequently compared to the other registries. While the SEER data only contain a limited number of characteristics on primary therapies that were started in hospitals within the first 24 months [31] and thus do not reflect the complete picture of oncological care [32], [33], the data from the German cancer registries include not only primary therapies but also further therapies from both the inpatient and outpatient sector. This is a special feature that allows a comprehensive description of oncological care on the basis of the cancer registry data – assuming that all therapies (*completeness of reporting*) with complete and valid information on the characteristics defined by the oBDS (*completeness of information*) are included in the state cancer registries.

Data scope of the ‘nationwide clinical cancer registry dataset’ compared to data requested from the cancer registries themselves

The ‘nationwide clinical cancer registry dataset’ currently to be applied for does not include all variables of the oBDS, e.g. information on service providers and from the organ-specific modules as well as on biomarkers is missing. Furthermore, individual details on therapies such as adverse events and complications or dose information on radiotherapy and the residual status after completion of primary therapy are missing [13]. For more in-depth analyses, therefore, self-reported data from the state cancer registries must continue to be used.

Another disadvantage of the ‘nationwide clinical cancer registry dataset’ is that the earliest available diagnosis and treatment year in this dataset is 2020. If therapy and survival are to be analysed before and after the approval of novel therapeutic agents (such as, for example if the treatment and survival of people with non-small cell lung cancer before and after approval of treatment with osimertinib in the presence of an EGFR mutation [34] or a combination of nivolumab, ipilimumab and two cycles of platinum-based chemotherapy in metastatic NSCLC [35] are to be analysed), the data from the state cancer registries, in which treatment and progression data are also available for the period before 2020, must continue to be used.

Conclusion

Our pilot project showed that although the merging of clinical data from the state cancer registries was still very (time) consuming, it was technically feasible. Country-specific differences and unclear information in the data request made it more difficult and slower to merge the data, partly due to different application procedures and formats of the data provided, but in particular due to the fact that the best-of information for the therapies in the registries has not yet been fully implemented.

Data generation in the federal states involved here is subject to partly differing (different state legislation, cancer registry software, reporting structures and registry routines) and partly common framework conditions (reporting obligation, oBDS). These should be taken into account when interpreting the data. The question of completeness or the absence of information in data fields (variables) must be carefully analysed, in particular because it is unclear whether data are missing randomly or systematically depending on the stage of disease, general condition or age or depending on the area of care (out-patient/inpatient setting).

In the pooled data sets, the differences between the federate-states registries in terms of data distribution and data quality were small. With reasonable effort, it was possible to generate a population-based database

that, with more than 61,000 cases, is sufficiently large to be able to carry out meaningful analyses even for rare histological subgroups or age groups with few cases of disease – provided that valid and complete information is available. Even when restricting the analysis to cases of disease with information on at least one type of therapy or to cases of disease with complete (essential) information on the therapy, the number of cases is still very high. These case numbers are comparable with currently ongoing long-term oncological registry studies [36].

It is to be hoped that, due to the legally required delivery to the ZfKD, best-of information for the therapies will be available in the registries in a timely manner. This will make it much easier to merge data on treatment and progression from the state cancer registries. It remains to be seen whether stage 2 of the federal law [12] will speed up the process of merging and providing event-related data by means of a central application and registry centre. But even then, good communication with the registries will be important for data evaluation and interpretation.

In our opinion, the proportion with completely missing information on treatment is still quite high at approx. 26%, but comparable to other registry studies [23], [24]. If information on the therapies is available, this information is mostly sufficiently complete. Therefore, the main features of individual therapies can be described well and used for oncological health services research. However, possible biases due to missing information must be taken into account, especially if conclusions are to be drawn for oncological care from a population perspective.

Notes

Acknowledgement

The authors would like to thank Volker Arndt from the Baden-Württemberg Cancer Registry for his critical and helpful comments on earlier versions of this manuscript.

Competing interests

The authors declare that they have no competing interests. They point out that they (with the exception of AW, LL, HB) are employed in or manage state cancer registries.

Ethics

No human studies were conducted for this article. The guidelines for the evaluation of anonymised data were adhered to. The research project was reported to the ethics committee of the University of Lübeck.

References

1. Bundesregierung. Gesetz zur Weiterentwicklung der Krebsfrüherkennung und zur Qualitätssicherung durch klinische Krebsregister (Krebsfrüherkennungs- und -registergesetz – KFRG) vom 3. April 2013
2. Katalinic A, Halber M, Meyer M, Pflüger M, Eberle A, Nennecke A, Kim-Wanner SZ, Hartz T, Weitmann K, Stang A, Justenhoven C, Hollecsek B, Piontek D, Wittenberg I, Heßmer A, Kraywinkel K, Spix C, Pritzkeleit R. Population-Based Clinical Cancer Registration in Germany. *Cancers (Basel)*. 2023 Aug;15(15):3934. DOI: 10.3390/cancers15153934
3. Landesregierung Baden-Württemberg. Gesetz über die Krebsregistrierung in Baden-Württemberg (Landeskrebsregistergesetz – LKrebsRG) vom 7. März 2006.
4. Senat Freie und Hansestadt Hamburg. Hamburgisches Krebsregistergesetz (HmbKrebsRG) vom 27. Juni 1984.
5. Landesregierung Nordrhein-Westfalen. Gesetz über die klinische und epidemiologische Krebsregistrierung im Land Nordrhein-Westfalen (Landeskrebsregistergesetz – LKRG NRW) vom 2. Februar 2016.
6. Landesregierung Schleswig-Holstein. Gesetz über das Krebsregister des Landes Schleswig-Holstein (Krebsregistergesetz – KRG SH) vom 4. November 2015.
7. Bundesregierung. Bundeskrebregisterdatengesetz (BKRG) vom 10. August 2009 (BGBl. I S. 2702,2707) – zuletzt geändert durch Artikel 2 des Gesetzes vom 18. August 2021 (BGBl. I S. 3890).
8. Bundesministerium für Gesundheit. Bekanntmachung: Module zur Dokumentation des Brust- und Darmkrebses in Ergänzung des aktualisierten einheitlichen onkologischen Datensatzes der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) vom 28. Oktober 2015. *BAnz AT*. 26.11.2015;(B1):1-10.
9. Bundesministerium für Gesundheit. Bekanntmachung: Modul zur Dokumentation des Prostatakrebses in Ergänzung des aktualisierten einheitlichen onkologischen Datensatzes der Arbeitsgemeinschaft Deutscher Tumorzentren e. V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e. V. (GEKID) vom 9. August 2017. *BAnz AT*. 29.08.2017;(B6):1-6.
10. Bundesministerium für Gesundheit. Bekanntmachung: Modul zur Dokumentation des Malignen Melanoms in Ergänzung des einheitlichen onkologischen Basisdatensatzes der Arbeitsgemeinschaft Deutscher Tumorzentren e. V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e. V. (GEKID) vom 25. Mai 2020. *BAnz AT*. 26.06.2020;(B4):1-4.
11. Bundesministerium für Gesundheit. Bekanntmachung: Aktualisierter einheitlicher onkologischer Basisdatensatz der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) vom 10. Mai 2021. *BAnz AT*. 12.07.2021;(B4):1-31.
12. Bundesregierung. Gesetz zur Zusammenführung von Krebsregisterdaten. 2021.
13. Meisegeier S, Imhoff M, Berg K, Kraywinkel K. Bundesweiter klinischer Krebsregisterdatensatz – Datenschema und Klassifikationen (oBDS_v3.0.0.8a_RKI). *Zenodo*; 2023. DOI: 10.5281/zenodo.10022040
14. Arbeitsgemeinschaft Deutscher Tumorzentren e.V. Die ADT als Vertrauensstelle für Bundesweite Qualitätskonferenzen in der Onkologie. [last accessed 2024 Dec 13]. Available from: https://adt-netzwerk.de/Vertrauensstelle/Bundesweite_Qualitaetskonferenzen/
15. Roessler M, Schmitt J, Bobeth C, Gerken M, Kleihues-van Tol K, Reissfelder C, Rau BM, Distler M, Piso P, Günster C, Klinkhammer-Schalke M, Schoffer O, Bierbaum V. Is treatment in certified cancer centers related to better survival in patients with pancreatic cancer? Evidence from a large German cohort study. *BMC Cancer*. 2022 Jun;22(1):621. DOI: 10.1186/s12885-022-09731-w
16. Rudolph C, Germer S, Nennecke A, Kusche H, Labohm L, Rath N, Rausch K, Hollecsek B, Handels H, Katalinic A. Künstliche Intelligenz (KI) in der Krebsregistrierung: Methoden, Herausforderungen und erste Ergebnisse des AI-Care-Datensatzes. In: *Gesundheit – gemeinsam. Kooperationstagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP), Deutschen Gesellschaft für Epidemiologie (DGEPi), Deutschen Gesellschaft für Medizinische Soziologie (DGMS) und der Deutschen Gesellschaft für Public Health (DGPH)*. Dresden, 08.-13.09.2024. Düsseldorf: German Medical Science GMS Publishing House; 2024. DocAbstr. 867. DOI: 10.3205/24gmids742
17. Stegmaier C, Hentschel S, Hofstädter F, Katalinic A, Tillack A, Klinkhammer-Schalke M, editors. *Das Manual der Krebsregistrierung*. München: Zuckschwerdt; 2019.
18. Operationen- und Prozedurenschlüssel (OPS). [last accessed 2024 Jan 24]. Available from: https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/_node.html
19. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.
20. Therapie des Lungenkarzinoms – eine Versorgungsforschungsstudie auf Basis von Krebsregisterdaten nach § 65c SGB V. DRKS00025080. In: *Deutsches Register Klinischer Studien (DRKS)*. [last accessed 2024 Jan 24]. Available from: <https://drks.de/search/de/trial/DRKS00025080>
21. Giusti F, Martos C, Trama A, Bettio M, Sanvisens A, Audisio R, Arndt V, Francisci S, Dochez C, Ribes J, Fernández LP, Gavin A, Gatta G, Marcos-Gragera R, Lievens Y, Allemani C, De Angelis R, Visser O, Van Eycken L; ENCR Working Group on Treatment Data Harmonisation. Cancer treatment data available in European cancer registries: Where are we and where are we going? *Front Oncol*. 2023;13:1109978. DOI: 10.3389/fonc.2023.1109978
22. Institut für Krebs Epidemiologie e.V. (IKE) der Universität zu Lübeck. KI-unterstützte, versorgungsnahe Nutzung von Krebsregisterdaten – AI-CARE. [last accessed 2024 Dec 6]. Available from: <https://ai-care-cancer.de/>
23. Bei Y, Chen X, Raturi VP, Liu K, Ye S, Xu Q, Lu M. Treatment patterns and outcomes change in early-stage non-small cell lung cancer in octogenarians and older: a SEER database analysis. *Aging Clin Exp Res*. 2021 Jan;33(1):147-56. DOI: 10.1007/s40520-020-01517-z
24. Ganti AK, Klein AB, Cotala I, Seal B, Chou E. Update of Incidence, Prevalence, Survival, and Initial Treatment in Patients With Non-Small Cell Lung Cancer in the US. *JAMA Oncol*. 2021 Dec;7(12):1824-32. DOI: 10.1001/jamaoncol.2021.4932
25. Zemanova M, Pirker R, Petruzelka L, Zbožínková Z, Jovanovic D, Rajer M, Bogos K, Purkalne G, Ceriman V, Chaudhary S, Richter I, Kuřa J, Jakubikova L, Zemaitis M, Cernovska M, Koubkova L, Vilasova Z, Dieckmann K, Farkas A, Spasic J, Fröhlich K, Tiefenbacher A, Hollosi V, Kultan J, Kolarová I, Votruba J. Care of patients with non-small-cell lung cancer stage III – the Central European real-world experience. *Radiol Oncol*. 2020 May;54(2):209-20. DOI: 10.2478/raon-2020-0026
26. Krebsregister Baden-Württemberg. Krebsregistrierung in Baden-Württemberg: Datenkatalog mit Merkmalsausprägungen nach ADT/GEKID Basisdatensatz 2.0.0 für Tumorzentren, Onkologische Schwerpunkte, Krankenhäuser und niedergelassene Ärzte. 2017.

27. Krebsregister Schleswig-Holstein. Meldepflicht – Mindestangaben (zwingend notwendige Angaben zur Erfüllung der Vollständigkeit einer Meldung und zur Auszahlung der Meldevergütung). 2018.
28. Landeskrebsregister NRW. Melder-Broschüre Krebsregistrierung in Nordrhein-Westfalen: Meldepflicht – Meldepflichtige Erkrankungen – Vergütung. Bochum; 2020.
29. Martin ML, Chung H, Rydén A. Willingness to report treatment-related symptoms of immunotherapy among patients with non-small cell lung cancer. *Qual Life Res.* 2022 Apr;31(4):1147-55. DOI: 10.1007/s11136-021-02966-3
30. Yang DX, Khera R, Miccio JA, Jairam V, Chang E, Yu JB, Park HS, Krumholz HM, Aneja S. Prevalence of Missing Data in the National Cancer Database and Association With Overall Survival. *JAMA Netw Open.* 2021 Mar;4(3):e211793. DOI: 10.1001/jamanetworkopen.2021.1793
31. SEER Acknowledgment of Treatment Data Limitations. For the 1975-2019 Data (November 2021 Submission). [last accessed 2023 Dec 14]. Available from: <https://seer.cancer.gov/data-software/documentation/seerstat/nov2021/treatment-limitations-nov2021.html>
32. Du XL, Key CR, Dickie L, Darling R, Delclos GL, Waller K, Zhang D. Information on chemotherapy and hormone therapy from tumor registry had moderate agreement with chart reviews. *J Clin Epidemiol.* 2006 Jan;59(1):53-60. DOI: 10.1016/j.jclinepi.2005.06.002
33. Noone AM, Lund JL, Mariotto A, Cronin K, McNeel T, Deapen D, Warren JL. Comparison of SEER Treatment Data With Medicare Claims. *Med Care.* 2016 Sep;54(9):e55-64. DOI: 10.1097/MLR.0000000000000073
34. EGFR-mutiertes (EGFRmut) NSCLC: Zulassung für Osimertinib firstline. *Im Focus Onkologie.* 2018;21(7):80. DOI: 10.1007/s15015-018-4126-4
35. Zulassung für Opdivo(R) / Yervoy(R): Erstlinie für NSCLC. *Dtsch Arztebl.* 2020 Nov;117(Suppl. Pneumologie & Allergologie 2):40.
36. Nicht-interventionelle Studie (Anwendungsbeobachtung) NIS-Nr.: 356. CRISP – Clinical Research platform Into molecular testing, treatment and outcome of (non-)Small cell lung carcinoma Patients. [last accessed 2023 Dec 14]. Available from: <https://www.pei.de/SharedDocs/awb/nis-0301-0400/0356.html>

Corresponding author:

Prof. Dr. Annika Waldmann
 Institut für Sozialmedizin und Epidemiologie, Universität
 zu Lübeck, Ratzeburger Allee 160 (Haus 409), 23562
 Lübeck, Germany, Phone: +49 451 50051240
 Annika.Waldmann@uksh.de

Please cite as

Waldmann A, Labohm L, Baltus H, Eisfeld C, Jansen L, Löhden I, Nennecke A, Oesterling F, Pritzkeleit R, Katalinic A. Zusammenführung von umfangreichen Daten und Vollständigkeit der Informationen zu Diagnose, Behandlung und zum Erkrankungsverlauf aus mehreren populationsbezogenen Krebsregistern in Deutschland – erste Erfahrungen am Beispiel von Lungenkrebs. *GMS Med Inform Biom Epidemiol.* 2025;21:Doc02. DOI: 10.3205/mibe000274, URN: urn:nbn:de:0183-mibe0002746

This article is freely available from

<https://doi.org/10.3205/mibe000274>

Published: 2025-05-30

Copyright

©2025 Waldmann et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.