

## 2 Methodische Aspekte

### 2.1 Schätzung des Erfassungsgrades in den epidemiologischen Krebsregistern (Vollzähligkeitsschätzung)

Der Nutzen bevölkerungsbezogener Daten zum Krebsgeschehen hängt wesentlich von der Vollzähligkeit der Erfassung aller neu auftretenden Krebserkrankungen ab. Daher schätzt das Zentrum für Krebsregisterdaten (ZfKD) jährlich den Erfassungsgrad der epidemiologischen Krebsregister in Deutschland, seit 2010 für alle Bundesländer. Die Schätzung erfolgt mit Hilfe eines international gebräuchlichen Vollzähligkeitsindikators, dem Verhältnis von Mortalität (Sterblichkeit) zu Inzidenz (Erkrankungshäufigkeit). Dieses Verhältnis (M/I-Index) kann unter der Voraussetzung, dass sich Diagnostik und Therapie und damit auch die Überlebenseinsichten von Krebspatientinnen und -patienten innerhalb Deutschlands nicht wesentlich unterscheiden, für die jeweilige Krebsdiagnose als regional weitgehend konstant angenommen werden. Mit Hilfe des M/I-Index in einer als vollzählig angenommenen Referenzregion und unter Verwendung der regionalen Mortalität wird die Inzidenz in der jeweiligen Untersuchungsregion geschätzt und mit den dort tatsächlich erhobenen Daten verglichen. Nur über Todesbescheinigungen identifizierte (DCO-) Fälle werden hierbei nicht berücksichtigt. Auch die Vollzähligkeit der Register der Referenzregion wird anhand des Vergleichs mit den Erwartungswerten geschätzt.

Für die Referenzregion wurden vor einigen Jahren folgende Einschlusskriterien definiert:

- ▶ Flächendeckende Krebsregistrierung seit mindestens zehn Jahren
- ▶ Vollzähligkeit für »Krebs gesamt« in den letzten zehn Jahren im Mittel über 90 % (nach der früheren Schätzmethode des RKI) und über 80 % für alle Einzeljahre
- ▶ Anteil von DCO-Fällen (nur über Todesbescheinigung registrierte Erkrankungsfälle) für »Krebs gesamt« in den letzten zehn Jahren oder ab dem sechsten Jahr nach Beginn der Erfassung im Mittel unter 15 %

Die Referenzregion wurde in den letzten Jahren aufgrund der Datenlage mehrfach leicht angepasst, aktuell setzt sie sich aus den Krebsregistern aus Bayern, Bremen, Hamburg, Niedersachsen, Nordrhein-Westfalen und Schleswig-Holstein zusammen.

Nach dem oben beschriebenen Prinzip werden Erwartungswerte für je sechs Altersgruppen (jeweils für Frauen und Männer getrennt) und 18 Diagnosegruppen berechnet.

Bei zu geringer Mortalität in der Untersuchungsregion (durchschnittlich weniger als fünf Sterbefälle pro Jahr) wurde für die entsprechende Altersgruppe anstelle des Quotienten aus Inzidenz und Mortalität die modellierte Inzidenz in der Referenzregion verwendet, um die erwartete Zahl der Neuerkrankungen zu berechnen. Der geschätzte Erfassungsgrad für jede Diagnosegruppe ergibt sich aus dem Quotienten der über alle Altersgruppen aufsummierten beobachteten und erwarteten Fallzahlen. Die Vollzähligkeit für »Krebs gesamt« wird wiederum über die Summation der beobachteten und erwarteten Werte für alle Diagnosegruppen geschätzt.

Limitationen des beschriebenen Verfahrens bestehen vor allem dann, wenn die Mortalität einer Krebsart insgesamt bzw. im Verhältnis zur Inzidenz gering ist (Hodenkrebs, malignes Melanom, Schilddrüsenkrebs), oder wenn das reale Verhältnis von Mortalität zu Inzidenz sich zwischen Regionen unterscheidet. Dies kann z. B. der Fall sein, wenn Früherkennungsuntersuchungen in den Bundesländern unterschiedlich stark wahrgenommen oder, wie beim Mammographie-Screening, zu unterschiedlichen Zeitpunkten eingeführt werden. Auch eine regional unterschiedliche Verteilung von Tumorstadien oder verschiedenen Subtypen einer Krebsdiagnose (etwa beim Schilddrüsenkrebs) kann zu Verzerrungen führen.

Nach der aktuellen Vollzähligkeitsschätzung für das Diagnosejahr 2020 erreichen unter 16 Bundesländern 13 Krebsregister bei beiden Geschlechtern für »Krebs gesamt« einen Erfassungsgrad von über 95 %. Ein weiteres Register liegt zwischen 90 % und 95 %. In den beiden verbleibenden Registern wurde ein Erfassungsgrad zwischen 80 % und 90 % erreicht.

### 2.2 Schätzung der bundesweiten Neuerkrankungszahlen (Inzidenzschätzung)

Die dargestellten bundesweiten Neuerkrankungszahlen (Inzidenzen) ergeben sich zum Teil aus den gezählten Erkrankungen und zum Teil aus den Ergebnissen eines gemischten Poisson-Regressionsmodells. Hierzu wurden zunächst Register und Diagnosejahre identifiziert, in denen folgende Qualitätskriterien bezüglich »Krebs gesamt« erfüllt wurden: Erfassung seit mindestens 10 Jahren, geschätzte Vollzähligkeit über die letzten fünf Jahre  $\geq 90\%$ . Zudem darf der jährliche DCO-Anteil in diesen fünf Jahren 15 % nicht

überschreiten. In Jahren, in denen Register die Qualitätskriterien erfüllten, wurden die dort gezählten Neuerkrankungen als Referenzdaten für die Regression verwendet und unverändert in die bundesweite Inzidenzschätzung übernommen. In Jahren, in denen Register die Qualitätskriterien nicht erfüllen, wurden die Inzidenzen mit Hilfe des Regressionsmodells geschätzt.

Bei der Regression wurde die Inzidenz durch die krebsspezifische Mortalität, die Bevölkerungsgröße und das Diagnosejahr modelliert. Zusätzlich wurden Unterschiede zwischen den Inzidenzraten in den Registern durch registerspezifische Parameter (Achsendurchschnitte als zufällige Effekte) im Modell berücksichtigt. Die Regression wurde, stratifiziert nach Geschlecht, Diagnose und Altersgruppe, an die Referenzdaten angepasst. Die so ermittelten Inzidenzen wurden, wie die Referenzdaten zum Schätzen der Regressionsparameter genutzt, damit Besonderheiten auch von diesem Register abgebildet werden konnten.

Die bundesweite jährliche Inzidenz ergibt sich aus der Summation der gezählten Fälle in den Registern, die die Qualitätskriterien in dem Diagnosejahr erfüllt haben, und der geschätzten Fälle in den anderen Registern. Der Anteil der gezählten Inzidenzen an der bundesweiten Inzidenz wird sich mit zunehmender Vollständigkeit der Register erhöhen und ermöglicht einen fließenden Übergang vom »Schätzen« zum »Zählen« bei entsprechender Steigerung der Datenqualität.

Die Schätzung der Häufigkeit der nicht-melanotischen Hautkrebsformen (ICD-10 C44) wurde grundsätzlich mit der gleichen Methode vorgenommen. Allerdings liegen für diese Diagnosen aufgrund der geringen Mortalität keine Schätzungen für die Vollständigkeit der Erfassung vor. Die Referenzregion bestand aus neun Registern (Schleswig-Holstein, Hamburg, Niedersachsen, Bremen, Nordrhein-Westfalen, Hessen, Rheinland-Pfalz, Saarland und Sachsen), deren Daten zumindest für einen gewissen Zeitraum als ausreichend vollzählig eingeschätzt wurden. Aufgrund der Datenlage wurden die Angaben über die Fallzahlen auf den Zeitraum 2006 bis 2020 beschränkt. Die Schätzung der Inzidenz des nicht-melanotischen Hautkrebs ist insgesamt mit einer größeren Unsicherheit verbunden als die Ergebnisse für andere Krebsdiagnosen. Bei der Darstellung der Häufigkeit von Krebserkrankungen insgesamt (Kapitel 3.1) sind die nicht-melanotischen Hautkrebsformen, wie international üblich, nicht berücksichtigt.

## 2.3 Zu Indikatoren und Darstellungsweisen

Im Folgenden werden die in den Ergebniskapiteln verwendeten Maßzahlen und grafischen Darstellungen erläutert.

### Altersspezifische und rohe Raten

Die altersspezifische Rate wird berechnet, indem die Zahl aufgetretener Krebskrankheiten bzw. Todesfälle durch Krebs in einer bestimmten Altersgruppe durch die entsprechende Anzahl von Frauen oder Männern dieses Alters in der Bevölkerung dividiert wird. Die grafische Darstellung dieser Raten zeigt, getrennt nach Geschlecht, den Zusammenhang zwischen Lebensalter und Erkrankungshäufigkeit. Die altersspezifischen Inzidenzraten werden als jährliche Neuerkrankungen pro 100.000 Einwohner der jeweiligen Altersgruppe und pro Jahr angegeben. Bezogen auf die Gesamtbevölkerung werden nach dem gleichen Prinzip die »rohen« Erkrankungs- und Sterberaten berechnet. Sie zeigen die Krankheitslast in einer Bevölkerung an. Sind aber, im Gegensatz zu altersstandardisierten Raten, stark abhängig von der Altersstruktur einer Bevölkerung und eignen sich daher weniger gut für die Beurteilung zeitlicher Trends und regionaler bzw. internationaler Unterschiede.

### Altersstandardisierte Raten

Wie die Darstellung der altersspezifischen Inzidenz von Frauen und Männern in dieser Broschüre zeigt, nimmt die Krebserkrankungsrate in der Regel mit steigendem Lebensalter erheblich zu. Will man daher Inzidenz oder Mortalität in verschiedenen Ländern und Regionen oder in derselben Bevölkerung zu unterschiedlichen Zeitpunkten miteinander vergleichen, müssen zunächst mit Hilfe der Altersstandardisierung Unterschiede im Altersaufbau der zu vergleichenden Bevölkerungen herausgerechnet werden. Hierfür erfolgt zunächst eine Gewichtung der beobachteten altersspezifischen Raten mit den (fiktiven) Bevölkerungsanteilen einer gewählten »Standardbevölkerung«. Anschließend werden die gewichteten Raten über alle Altersgruppen aufsummiert. Die so berechnete altersstandardisierte Rate gibt an, wie hoch die Zahl der Neuerkrankungen oder Sterbefälle pro 100.000 Personen in der beobachteten Bevölkerung wäre, wenn diese die gleiche Altersstruktur wie die gewählte Standardbevölkerung hätte. In der vorliegenden Broschüre wurde die »alte Europa-standardbevölkerung« verwendet.

### Erkrankungs- und Sterberisiken

Altersspezifische Inzidenz- und Mortalitätsraten lassen sich auch als Maß für das alters- und geschlechtsspezifische Risiko interpretieren, innerhalb eines

Jahres an Krebs bzw. an einem bestimmten bösartigen Tumor zu erkranken oder zu versterben. Um diese Form der Risikokommunikation anschaulicher zu gestalten, wurde in Abhängigkeit von Alter und Geschlecht das Risiko berechnet, innerhalb der nächsten zehn Jahre bzw. jemals an einem bestimmten Tumor erstmals zu erkranken bzw. an diesem zu versterben. Die Angaben erfolgen neben der üblichen Prozentdarstellung in natürlichen Bezugsgrößen, d. h. als eine oder einer pro N Personen des gleichen Alters und Geschlechts. Dabei wurden so genannte »konkurrierende Risiken« mit einbezogen, d. h. es wurde berücksichtigt, dass z. B. ein 75-jähriger Mann mit einer gewissen Wahrscheinlichkeit innerhalb der nächsten zehn Jahre aus einem anderen Grund als einer Krebserkrankung versterben kann. Ähnlich wurde auch das »Lebenszeitrisiko«, also das Risiko, innerhalb des gesamten Lebens an einem Tumor zu erkranken, berechnet. Bei den Berechnungen gehen allerdings nur die jeweils aktuellen Raten (Inzidenz- und Sterberaten sowie allgemeine Lebenserwartung) ein. Eine Prognose über die zukünftige Entwicklung dieser Werte wird damit nicht getroffen. Außerdem sind diese Ergebnisse als durchschnittliche Werte für die deutsche Bevölkerung zu sehen, individuelle Risiken können aufgrund des Vorhandenseins oder Fehlens bestimmter Risikofaktoren erheblich abweichen. Für die Berechnungen wurde das vom National Cancer Institute in den USA entwickelte Programm »DevCan«<sup>1</sup> benutzt.

### Internationaler Vergleich

Um die Höhe der geschätzten Krebsinzidenz und der Krebsmortalität in Deutschland international einordnen zu können, wurden aktuelle altersstandardisierte Neuerkrankungs- und Sterberaten der Nachbarländer Deutschlands sowie aus England, Finnland, Schweden und den USA herangezogen. Ein Quellenachweis findet sich im Anhang (Kapitel 5.7), dort sind auch eventuell abweichende Zeiträume vermerkt. Die entsprechenden Angaben wurden ohne Prüfung auf Plausibilität bzw. Vollzähligkeit übernommen, so dass in einigen Fällen eine Unterschätzung insbesondere der Inzidenzraten möglich erscheint. Bei einigen Ländern weicht die Gruppierung der Diagnosen nach ICD-10 etwas von der für Deutschland verwendeten ab, was die Vergleichbarkeit gegebenenfalls einschränkt (siehe entsprechende Fußnoten).

### Mittleres Erkrankungs- und Sterbealter

Das mittlere (mediane) Erkrankungs- und Sterbealter wurde nach einer Näherungsformel aus den nur für 5-Jahres-Altersgruppen vorliegenden Ergebnissen der Inzidenzschätzung bzw. aus der amtlichen Todesursachenstatistik des Statistischen Bundesamtes berechnet.

### Mortalität (Sterblichkeit)

Die Krebsmortalität beruht auf der Zahl der jährlichen Krebstodesfälle nach der amtlichen Todesursachenstatistik. Die Todesfälle werden dazu alters- und geschlechtsspezifisch der zu Grunde liegenden Todesursache zugeordnet. Die Sterberate ergibt sich durch den Bezug der jährlichen Sterbefälle auf die Größe der Bevölkerung. Die Raten werden auf 100.000 Personen bezogen. In der vorliegenden Broschüre werden die absolute Zahl an Sterbefällen sowie rohe als auch altersstandardisierte Sterberaten (alte Europastandardbevölkerung) von 1999 bis 2021 berichtet. Datenquelle ist die offizielle Todesursachenstatistik des Statistischen Bundesamtes ([www.gbe-bund.de](http://www.gbe-bund.de)).

### Regionaler Vergleich

Die geschätzten altersstandardisierten Inzidenzraten (alte Europastandardbevölkerung) aus den Bundesländern für den Zeitraum 2019 bis 2020 werden im Vergleich zur entsprechenden Schätzung für Deutschland wiedergegeben, für die Referenzregionen entsprechen sie der erfassten Inzidenz (siehe Kapitel 2.2). Für den gleichen Zeitraum wird die altersstandardisierte Mortalität nach Lokalisation und Geschlecht für alle Bundesländer im Vergleich zur bundesweiten Mortalität dargestellt.

### Überlebensraten

Die Ergebnisse der Überlebenszeitanalysen in dieser Broschüre beschreiben die durchschnittlichen Überlebensaussichten von bei Diagnosestellung über 15 Jahre alten Personen nach einer bestimmten Krebsdiagnose. Hierfür wurden absolute und relative Überlebensraten von bis zu 10 Jahren nach Diagnosestellung berechnet. Absolute Überlebensraten stellen den Anteil der Patientinnen und Patienten dar, die zu einem bestimmten Zeitpunkt nach ihrer Diagnose noch leben. Ein absolutes 5-Jahres-Überleben von 80 % bedeutet zum Beispiel, dass 80 von 100 an einer bestimmten Krebsart erkrankte Personen die ersten fünf Jahre nach ihrer Diagnose überlebt haben.

Relative Überlebensraten bilden dagegen die krebsbedingte Sterblichkeit ab, indem der Quotient aus dem absoluten Überleben der Krebspatientinnen und -patienten und dem erwarteten Überleben in der allgemeinen Bevölkerung gleichen Alters und Geschlechts berechnet wird. Ein relatives 5-Jahres-Überleben von 100 % bedeutet demnach, dass innerhalb von fünf Jahren nach einer Krebsdiagnose genauso viele betroffene Personen verstorben sind, wie auch ohne diese Diagnose zu erwarten gewesen wäre. Das erwartete Überleben wurde mit der sogenannten Ederer-II-Methode anhand der bundesdeutschen Periodensterbetafeln des Statistischen Bundesamtes berechnet.

Auf der Basis vorher festgelegter Kriterien zur Datenqualität wurden als Grundlage für die aktuellen Überlebenszeitberechnungen die Daten aus Schleswig-Holstein, Hamburg, Niedersachsen, Bremen, dem Saarland, Sachsen sowie aus dem Regierungsbezirk Münster (Nordrhein-Westfalen) eingeschlossen.

Dargestellt wurden außerdem relative 5-Jahres-Überlebensraten nach Tumorstadium (und Geschlecht). Für diese Auswertungen wurden nur Fälle einbezogen, für die sich zwischen der 7. und 8. Auflage der TNM-Klassifikation keine wesentlichen Änderungen ergeben hatten. Für die übrigen Diagnosen wurden andere Stratifizierungsmerkmale gewählt.

Um möglichst aktuelle Überlebensaussichten zu schätzen, wurde die sogenannte Periodenmethode verwendet. Diese berücksichtigt das Überleben von Personen, die während einer bestimmten Zeitperiode (hier: 2019 bis 2020) gelebt haben.

Die angegebene Spanne des 5- und 10-Jahres-Überlebens stellt das jeweilige niedrigste bzw. höchste Überleben in den oben genannten Regionen dar, wobei hierfür nur Regionen mit einem Standardfehler des geschätzten Überlebens von unter 7,0 berücksichtigt wurden. Wenn dieses Kriterium von weniger als vier Regionen erreicht werden konnte, wurde auf die Darstellung der Spannbreite verzichtet. Die dargestellte Spannbreite spiegelt nach bisheriger Einschätzung wahrscheinlich nur zu einem sehr geringen Anteil Unterschiede in der Versorgungsqualität wider: Unterschiede in der Qualität des Mortalitätsabgleichs bzw. beim DCO-Anteil können ebenso eine Rolle spielen wie zufallsbedingte Schwankungen, vor allem in den kleineren Bundesländern. Auch methodische Unterschiede zwischen den Registern, vor allem die nicht überall durchgeführte Nachrecherche von DCO-Fällen (»Follow-back«), können die Ergebnisse beeinflussen. Die Berechnung des Überlebens 10 Jahre nach Diagnose basiert auf deutlich kleineren Fallzahlen als die Berechnung des Überlebens 5 Jahre danach. Aus diesem Grund weist das registerspezifische 10-Jahres-Überleben eine größere statistische Unsicherheit auf als das 5-Jahres-Überleben. Deshalb kann es vorkommen, dass die Werte in der Spanne des relativen 10-Jahres-Überlebens geringfügig über den entsprechenden Werten beim 5-Jahres-Überleben liegen.

Insgesamt ist für die angegebenen Schätzwerte der Überlebensraten in Deutschland zumindest bei den Krebserkrankungen mit ungünstiger Prognose eher von einer leichten Überschätzung auszugehen, was allerdings wahrscheinlich auch für die meisten international veröffentlichten Ergebnisse zutrifft.

### Verteilung der Tumorstadien

Die Ausbreitung eines soliden bösartigen Tumors bei Diagnosestellung in den Jahren 2019 bis 2020 nach Geschlecht wurde anhand der TNM-Klassifikation (8. Auflage) ausgewertet. Dargestellt wurden die

UICC-Stadien (I bis IV), die neben der Größe bzw. Ausbreitung des Primärtumors (T) auch den Lymphknotenstatus (N) und eine eventuell vorliegende Fernmetastasierung (M) berücksichtigen. Fehlende Angaben zu M wurden als Mo (keine Metastasen) gewertet, während fehlende Angaben zu N in den meisten Fällen zu einem fehlenden UICC-Stadium führen. Der Anteil fehlender Werte schließt auch solche Fälle mit ein, für die aufgrund der Histologie kein TNM bzw. UICC-Stadium vorgesehen ist, dies betrifft u. a. Sarkome. Für die Verteilung der Tumorstadien gingen die Daten aus allen Registern ein.

### Prävalenz (bis 25 Jahre nach Diagnose)

Die Prävalenz bezeichnet die Anzahl der Personen, die zu einem gegebenen Zeitpunkt (hier 31.12.2020) leben und vorher an Krebs erkrankt sind. Dabei berücksichtigt beispielsweise die 5-Jahres-Prävalenz nur Personen, bei denen innerhalb der letzten 5 Jahre Krebs diagnostiziert wurde. Die Hochrechnung der Prävalenz erfolgte nach der Methode von Pisani<sup>2</sup> aus der Inzidenzschätzung für Deutschland (siehe Kapitel 2.2) und aus Überlebensraten, die nach der Kaplan-Meier-Methode hergeleitet wurden. Für die Herleitung der 25-Jahres-Prävalenz wurde die Inzidenzschätzung bis auf das Jahr 1995 ausgedehnt. Bei der Berechnung der Überlebensraten wurden Daten von Personen ausgeschlossen, die mit hoher Wahrscheinlichkeit verstorben sind, deren Tod bzw. Sterbedatum dem Krebsregister jedoch nicht übermittelt wurde (z. B. nach Umzug in ein anderes Bundesland)<sup>3</sup>. Weiter wurden Überlebensraten für den Anteil der Inzidenzmeldungen korrigiert, die nur durch Todesbescheinigungen (DCO-Fälle) dokumentiert sind. Aufgrund der Korrekturen konnten mit Ausnahme von Berlin und Sachsen-Anhalt alle Register für die Überlebenszeitschätzungen verwendet werden.

### Weitere Auswertungen

Für einige Lokalisationen finden sich in diesem Bericht oder auf der Website des Zentrums für Krebsregisterdaten ([www.krebsdaten.de](http://www.krebsdaten.de)) zusätzliche Auswertungen, beispielsweise zur Histologie oder der genaueren Tumorlokalisation. Diese Auswertungen beruhen in dieser Ausgabe auf den Daten aller bevölkerungsbezogenen Krebsregister.

### Referenzen

- 1 DevCan: *Probability of Developing or Dying of Cancer Software, Version 6.9.0 Surveillance Research Program, Statistical Methodology and Applications*, National Cancer Institute, 2012. <http://surveillance.cancer.gov/devcan/>
- 2 Pisani, P., Bray, F., & Parkin, D. M. (2002). *Estimates of the world-wide prevalence of cancer for 25 sites in the adult population*. *Int J Cancer*, 97(1), 72 – 81. doi: 10.1002/ijc.1571
- 3 Dahm S, Barnes B, Kraywinkel K (2023) *Detection of missed deaths in cancer registry data to reduce bias in long-term survival estimation*. *Front Oncol Mar 9*; 13. doi: 10.3389/fonc.2023.1088657