

Vollzähligkeitsschätzungen von Krebsregisterdaten in Deutschland

Zusammenfassung

Das 1995 in Deutschland in Kraft getretene Bundeskrebsregistergesetz verpflichtete alle Bundesländer, epidemiologische Krebsregister einzurichten. Dieses inzwischen schon wieder ausgelaufene Gesetz hat die Krebsregistrierung in den Bundesländern stark vorangetrieben und zur Etablierung zahlreicher epidemiologischer Register geführt. Insbesondere für die sich im Aufbau befindlichen Krebsregister ist es wichtig, Vollzähligkeitsschätzungen durchzuführen, da ein Register nur mit einem ausreichend hohen Erfassungsgrad aussagekräftige Daten zu Krebsinzidenzen liefern kann. Neben direkten Methoden zur Vollzähligkeitsschätzung besteht die Möglichkeit, den erreichten Erfassungsgrad eines Registers indirekt zu schätzen. Hierzu werden Inzidenzen für das betreffende Einzugsgebiet auf der Basis der Daten eines anderen Krebsregisters geschätzt und mit den dort tatsächlich erhobenen Daten verglichen. Dabei wird die im Referenzregister geltende Beziehung zwischen Inzidenz und Mortalität auf die zu untersuchende Region übertragen unter Berücksichtigung der dortigen, aus der Todesursachenstatistik bekannten Mortalität. Die Datenbasis beruht im Wesentlichen auf saarländischen Krebsregisterdaten, die für einen langen Zeitraum vorliegen und, belegt durch zahlreiche Studien, hinsichtlich der Vollzähligkeit und Qualität internationalen Standards genügen. Die „Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland“ hat sich auf dieses Verfahren zur Vollzähligkeitsschätzung geeinigt. Die Dachdokumentation Krebs als Mitglied dieser Arbeitsgemeinschaft führt die entsprechenden Analysen

für die epidemiologischen Krebsregister Deutschlands durch. Die wichtigsten diesbezüglichen Ergebnisse werden im Folgenden präsentiert.

Schlüsselwörter

Epidemiologische Krebsregister · Vollzähligkeit · Log-lineare Modelle

Die Schätzung der Vollzähligkeit von Krebsregisterdaten ist gegenwärtig gerade in Deutschland eine überaus wichtige Aufgabe. Die Häufigkeit von Krebskrankheiten kann nur aus Daten bevölkerungsbezogener Krebsregister exakt ermittelt werden. Mithilfe der Register können z. B. regionale Besonderheiten oder sinkende bzw. steigende Trends bei den Erkrankungszahlen erkannt werden, oder sie können als Basis für die Evaluation spezieller Programme zur Vorsorge oder Früherkennung dienen. Wichtigste Voraussetzung für eine solche Nutzung ist die Vollzähligkeit der Registerdaten, d. h. die Erfassung aller Krebserkrankungen in der beobachteten Bevölkerung. Abgesehen vom bundesweiten Kinderkrebsregister in Mainz, kann in Deutschland bis dato allein das saarländische Krebsregister über einen langen Zeitraum für alle Krebskrankheiten als vollzählig angesehen werden. Mit dem 1995 in Kraft getretenen Bundeskrebsregistergesetz [1], das alle Bundesländer verpflichtete, bis 1999 epidemiologische Krebsregister

einzurichten, hat die Krebsregistrierung in Deutschland eine neue Qualität erreicht. Inzwischen haben sich in allen Bundesländern bevölkerungsbezogene Krebsregister etabliert, und bereits länger bestehende Krebsregister sind strukturellen Veränderungen unterworfen worden. Diese erheben inzwischen – zumindest für einige ausgewählte Krebskrankheiten und für einen begrenzten Zeitraum – den Anspruch auf Vollzähligkeit ihrer Daten. Diesen Anspruch gilt es zu prüfen, insbesondere um die Daten möglichst vieler Register schnell zu nutzen und somit auch eine breitere Datenbasis für die Ermittlung der Krebshäufigkeit in Deutschland zu erhalten.

Datenlage und Methodik

Zur Schätzung des Erfassungsgrades von Krebsregisterdaten gibt es kein eindeutiges Standardinstrument. Häufig werden Capture-Recapture-Verfahren eingesetzt [2], die genaue Angaben zu den Meldequellen erfordern. Neben solchen direkten Methoden zur Vollzähligkeitsschätzung, die allein auf der Datenbasis des Registers selbst beruhen, besteht die Möglichkeit, den erreichten Erfassungsgrad indirekt zu schätzen. Hierzu schätzt man Inziden-

© Springer-Verlag 2003

Dr. J. Haberland
Dachdokumentation Krebs, Abt. Epidemiologie
und Gesundheitsberichterstattung,
Robert Koch-Institut, Seestraße 10,
13353 Berlin
E-Mail: Haberlandj@rki.de

J. Haberland · D. Schön · J. Bertz · B. Görsch

Estimating the completeness of German cancer registries

Abstract

In 1995, a federal cancer registry law became effective in Germany obliging all federal states to set up epidemiological cancer registries. As a result, numerous cancer registries have been established. However, it is essential to know the degree to which all relevant cases are recorded within a registry – particularly for those registries established recently – as only a high degree of completeness can prevent variations in the efficacy of registry procedures from distorting comparative studies. In this study, the completeness of the cancer registries was evaluated indirectly. Incidences for geographical areas covered by a cancer registry were estimated using incidence/mortality ratios of another registry known to have complete data sets. Log-linear models, successfully applied in a separate study, were fitted to incidence/mortality ratios using both the age and sex-specific data of the reference registry. These estimates were then compared with the actual data collected by the newly established registry, allowing the level of completeness to be assessed. The German Epidemiological Cancer Registry Study Group has agreed to this method and the „Dachdokumentation Krebs“ – as a member of this group – performs the analyses. The first results of the study are presented in this contribution.

Keywords

Cancer registration · Completeness · Log-linear models

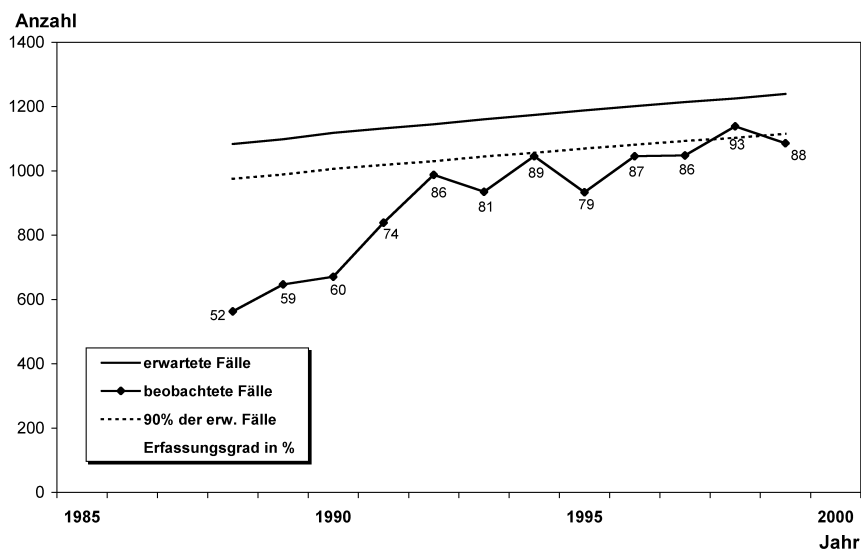


Abb. 1 ▲ Entwicklung der erfassten und geschätzten Anzahl der an Brustkrebs erkrankten Frauen in Hamburg (ohne DCO-Fälle). (DCO Death-Certificate-Only, d. h. Krebsfälle, die dem Register erst aufgrund einer Todesmeldung bekannt werden)

zen für das betreffende Einzugsgebiet unter Zuhilfenahme von Daten anderer Register und vergleicht diese mit den tatsächlich erhobenen Daten. Indirekte Verfahren – wie das hier vorgestellte – benötigen als Voraussetzung ein seit langem gut arbeitendes Referenzregister, dessen Vollzähligkeit durch andere Studien ausreichend belegt ist. Diese Bedingung ist für Deutschland mit den Daten des saarländischen Krebsregisters erfüllt. Hier erfolgt die Registrierung von Krebserkrankungen seit Beginn der 70er-Jahre mit einer hohen Kontinuität und Stabilität. Zahlreiche Studien belegen, dass die dortigen Krebsregisterdaten hinsichtlich Vollzähligkeit und Qualität internationalen Standards genügen (z. B. [2]).

Unterstellt man, dass sich die Krebserkrankungsrisiken in einer zu untersuchenden Region und einer Referenzregion derart unterscheiden, dass sich die Differenzen an den jeweiligen Mortalitätsunterschieden messen lassen, so kann man die Quotienten von Inzidenz und Mortalität verwenden, um Krebserkrankungshäufigkeiten zu schätzen [3]. Unter der Annahme, dass die Todesursachenstatistik in allen Regionen gleichartig erhoben wird und sich Diagnostik und Therapie nicht wesentlich unterscheiden, lassen sich damit Inzidenzen anderer Regionen unter Verwendung der dortigen Mortalität sinnvoll schätzen. Zur Schätzung der Vollzähligkeit wurde ein log-linearer

Modellansatz mit polynomialen Trends für die Quotienten aus Inzidenz und Mortalität angewendet, der sich schon in einer französischen Studie [4] bewährt hat und bereits zur Schätzung bundesweiter Inzidenzen in Deutschland erfolgreich eingesetzt wurde [5]. Durch die explizite Modellierung zeitlicher Inzidenzentwicklungen lassen sich Veränderungen in den Erfassungsgraden eines zu untersuchenden Krebsregisters gut beobachten. Als Grundlage benötigt dieser Ansatz jedoch Daten, die über einen möglichst langen Zeitraum erhoben wurden.

Die Inzidenzschätzung erfolgt jeweils nach Geschlecht und für eine ausgewählte Krebslokalisation. Sie beruht auf mehreren Schritten. Der erste Schritt besteht in einer Mortalitätsanalyse. Da die altersspezifischen Mortalitätsraten des Referenzregisters – insbesondere für seltenere Krebslokalisationen – relativ starke Schwankungen über die Zeit aufweisen, werden diese Daten zunächst unter Einbeziehung der bundesweiten Mortalität mittels eines log-linearen Modells geglättet:

$$\ln E(M_{irt}) = \ln(P_{irt}) + \alpha_{ir} + \lambda_{ir}t + \eta_{ir}t^2$$

Dabei bezeichnet M die Anzahl der Gestorbenen in der Altersklasse i , der Region r zum Zeitpunkt t und P die dazugehörige durchschnittliche Bevölkerung. Das Modell unterstellt sowohl für das Referenzregister als auch für die übrigen

Bundesländer individuelle, altersspezifische polynomiale Trends. Im Ansatz von Colonna et al. [4] wurde vereinfachend angenommen, dass die Trends – abgesehen von Niveaudifferenzen – in beiden Regionen gleichartig verlaufen:

$$\ln E(M_{irt}) = \ln(P_{irt}) + \beta_i + \gamma_i t + \delta_i t^2 + \varepsilon_r$$

Eigene Untersuchungen zeigen, dass diese Vereinfachung überaus sinnvoll ist, insbesondere bei selteneren Krebskrankheiten, bei denen Trendverläufe teilweise schwer zu beurteilen sind.

Das aus der Mortalitätsanalyse letztlich ausgewählte Modell führt zu geglätteten Mortalitätsdaten (\hat{M}_{it}), die ihrerseits zur Modellierung der eigentlich interessierenden Inzidenz (I_{it}) im Referenzregister in einem zweiten Schritt verwendet werden:

$$\ln E(I_{it}) = \ln(\hat{M}_{it}) + \beta_i + \gamma_i t + \delta_i t^2$$

Mittels der auf der Basis der Daten der Referenzregion geschätzten Modellgleichung kann man nun auch Inzidenzschätzungen für andere Regionen ermitteln, indem man die dortige Mortalität in das Modell einbindet. Dabei hat es sich als sinnvoll erwiesen, auch die Mortalität der Untersuchungsregion vorab zu glätten:

$$\ln E(M_{it}) = \ln(P_{it}) + \beta_i + \gamma_i t + \delta_i t^2$$

Zur Beurteilung des Erfassungsgrades werden alle Erst- und Folgeerkrankungen in die Inzidenzen einbezogen. DCO-Fälle, also diejenigen Krebsfälle, die dem Register erst aufgrund der Todesmeldungen bekannt werden und zu denen auch nach Rückfragen keine zusätzlichen Informationen eingehen, bleiben unberücksichtigt. Neu eingerichtete Register weisen einen erhöhten DCO-Anteil auf, da in diesen Registern Neuerkrankungen erst ab dem Zeitpunkt der Einrichtung registriert werden können. Vor Beginn der Registrierung aufgetretene Erkrankungsfälle (insbesondere bei Krebskrankheiten mit längerer Überlebensdauer) können demnach erst durch die Todesmeldung bekannt werden. Die in diesen Registern gemessene Inzidenz, die sich wie international üblich aus den zu Lebzeiten erfassten Fällen (reine Erkrankungsmeldungen) und den DCO-Fällen zusammensetzt, würde deshalb überschätzt werden. In dieser Untersuchung wird der Erfassungsgrad der Krebsregisterdaten daher nur anhand der reinen Erkrankungsmeldungen beurteilt. Als vollzählig werden Registerdaten akzeptiert, wenn sie den international geforderten Schwellenwert von 90% der erwarteten Fälle übersteigen.

Um zu vermeiden, dass die zu untersuchenden Register dauerhaft an einem einzigen Referenzregister (dem

saarländischen Krebsregister) gemessen werden, wurde mit der Bildung eines Krebsregisterdatenpools begonnen, in den alle Registerdaten eingehen, die nach dieser Methode als ausreichend vollzählig angesehen werden. Zuletzt wird jedes Register gegen den Datenpool geprüft, sodass regionale Besonderheiten bezüglich der Krebsinzidenz und -mortalität in den Analysen an Bedeutung abnehmen. Beispielsweise haben Änderungen in der Klassifikation der Harnblasenkarzinome [6], die im Saarland auch rückwirkend konsequent umgesetzt wurden, sicherlich dazu beigetragen, dass die dortigen Erkrankungs-raten seit Anfang der 90er-Jahre stark rückläufig sind [7] im Unterschied zu den Erkrankungs-raten anderer Regionen, bei denen diese Entwicklung erst später einsetzt.

Die „Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland“ hat sich auf das hier beschriebene Verfahren zur Vollzähligkeitsschätzung geeinigt. Als Mitglied dieser Arbeitsgemeinschaft hat die Dachdokumentation Krebs die entsprechenden Untersuchungen für die häufigsten Krebslokalisationen vorgenommen. Die Analysen wurden mittels des Statistik-Programmpakets GLIM durchgeführt [8]. Mit diesem nicht mehr ganz zeitgemäßen und damit unkomfortabel zu bedienenden Programm lassen sich

Tabelle 1

Maximal erreichter Erfassungsgrad (in %) in den Jahren 1997 bis 1999 bezüglich ausgewählter Krebslokalisationen in den untersuchten Krebsregister-Einzugsgebieten

Krebslokalisation Region/Geschlecht	Prostata Männer	Brustdrüse Frauen	Lunge		Dickdarm, Mastdarm		Krebs gesamt	
			Männer	Frauen	Männer	Frauen	Männer	Frauen
Berlin	43	80	72	72	47	43	65	63
Brandenburg	60	93	81	85	60	55	74	74
Bremen	36	89	76	88	64	53	64	77
Hamburg	83	93	>95	>95	73	67	95	83
Meckl.-Vorp.	72	>95	90	83	72	69	85	85
Reg. Bez. Münster (Nordrhein-Westf.)	55	82	>95	>95	67	69	82	84
Ortenaukreis (Baden-Würt.)	81	>95	94	83	>95	>95	>95	>95
Rheinland-Pfalz	59	86	41	38	69	66	63	67
Sachsen	64	>95	92	>95	78	72	86	88
Sachsen-Anhalt	34	68	39	38	36	35	46	50
Schleswig-Holstein	63	88	84	77	78	77	79	79
Thüringen	49	89	72	68	48	42	57	64
Weser-Ems (Niedersachsen)	48	79	44	46	65	56	57	58

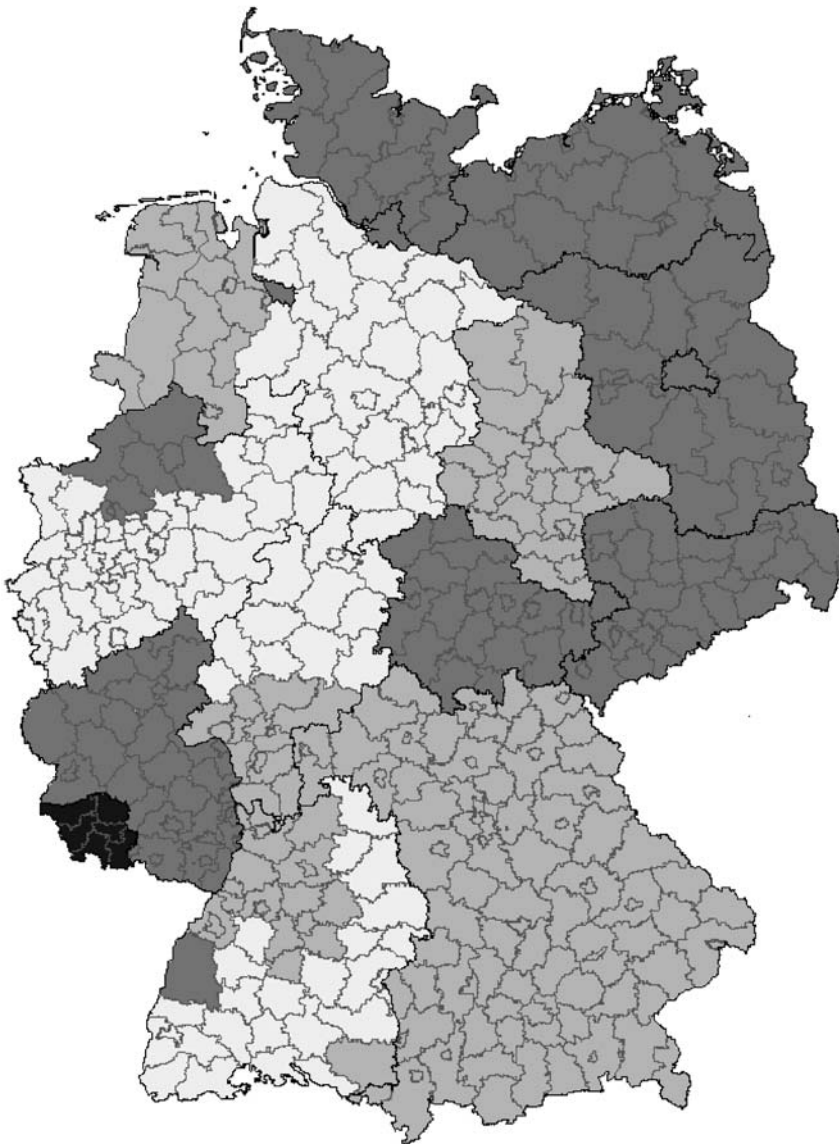
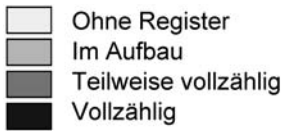


Abb. 2 ▲ Stand der Krebsregistrierung in Deutschland auf der Basis einer Datenanalyse bis 1999

die teilweise komplexen Designmatrizen der Modelle – im Gegensatz zu neueren Statistikprogrammen – relativ einfach erzeugen.

Ergebnisse

Abbildung 1 zeigt beispielhaft für die häufigste Krebslokalisation bei Frauen, d. h. für das Mammakarzinom, die Entwicklung der beobachteten und der mittels log-linearer Modelle geschätzten Anzahl der zu erwartenden Erkrankten ohne DCO-Fälle in Hamburg. Hier trat

1985 das Hamburgische Krebsregistergesetz in Kraft, weshalb dieser Zeitpunkt als Neubeginn des wohl ältesten deutschen Krebsregisters gewertet wird. Auch wenn das saarländische Register in den letzten 20 Jahren einen absoluten und auch altersstandardisierten Anstieg in der Brustkrebsinzidenz aufzeigt, kann man angesichts dieser strukturellen Veränderungen des hamburgischen Krebsregisters sicherlich davon ausgehen, dass der dort beobachtete starke Anstieg in den absoluten Erkrankungszahlen maßgeblich auf seine verbesserte Erfas-

sungstätigkeit zurückzuführen ist. Eine Schätzung der Erkranktenzahlen mittels der hier vorgestellten Methoden zeigt tatsächlich nur noch einen leichten Anstieg in den absoluten Erkrankungszahlen. Die Differenzen zwischen der erwarteten und der beobachteten Anzahl an Erkrankten, die indirekt den erreichten Erfassungsgrad im betreffenden Jahr schätzen, sind als Prozentangaben in der Grafik aufgeführt. Demnach ist das Register 1998 bezüglich Brustkrebs als vollzählig zu akzeptieren, da es in diesem Jahr den international geforderten Schwellenwert von 90% überschreitet.

Der im jeweils letzten Beobachtungsjahr häufig festgestellte leichte Rückgang im Erfassungsgrad wird durch Nacherfassungen im Laufe der weiteren Registertätigkeit meist kompensiert. Daher weist Tabelle 1 nicht allein den im Jahr 1999 geschätzten Erfassungsgrad auf, sondern stellt den von den untersuchten Krebsregistern über die letzten 3 Beobachtungsjahre (1997–1999) maximal erreichten, geschätzten Erfassungsgrad für die häufigsten Krebslokalisationen dar. Auffällig sind die starken Schwankungen in den Schätzungen zwischen den einzelnen Lokalisationen, die auch in anderen Studien zu dieser Thematik zu beobachten sind [9]. Vollzähligkeitsaussagen bezüglich eines Krebsregisters sollten sich daher nur auf einzelne Lokalisationen beziehen. Die lokalisationspezifischen Schwankungen erklären die teilweise starken Unterschiede zwischen Männern und Frauen im Hinblick auf Krebs gesamt. In Hamburg mit besonders deutlichen Unterschieden zwischen den Geschlechtern wird nach dieser Schätzung gerade der Lungenkrebs, die bei Männern neben Prostatakrebs häufigste Krebserkrankung, sehr gut erfasst.

Abgesehen von dem nicht das gesamte Krebsregistereinzugsgebiet betreffenden Ortenaukreis in Baden-Württemberg mit einer geringen durchschnittlichen Bevölkerungsgröße von rund 400.000 Einwohnern, weist nur Hamburg im Hinblick auf Krebs gesamt bei Männern einen ausreichend hohen Erfassungsgrad auf. Bezüglich Lungenkrebs sind für beide Geschlechter neben Hamburg weitere Register als vollzählig anzusehen. Gleiches gilt für die Erfassung des Brustkrebses bei Frauen. Hingegen zeigt beim Dickdarm-, Mastdarm- und Prostatakrebs noch kein Re-

gister – mit Ausnahme des Ortenaukreises – einen ausreichend hohen Erfassungsgrad.

Neben den in Tabelle 1 aufgeführten häufigsten Tumoren wurden zahlreiche weitere wichtige Krebserkrankungen auf die Vollzähligkeit ihrer Registrierung untersucht. Abbildung 2 zeigt eine zusammenfassende grafische Darstellung dieser Analysen. In der Untersuchung wurden die Kreise Deutschlands in Krebsregisterinzugsgebiete und in Regionen, die nicht von Krebsregistern abgedeckt werden, unterteilt. Bei den Krebsregistereinzugsgebieten wurde weiter nach Regionen differenziert, in denen sich die Registrierung noch im Aufbau befindet bzw. für keine der untersuchten Krebslokalisationen vollzählig ist, und nach Regionen, die für einige bzw. für alle Krebserkrankungen einen ausreichend hohen Erfassungsgrad aufweisen. Den letztgenannten Zustand hat in Deutschland bisher nur das saarländische Krebsregister erreicht.

Abschließende Bemerkungen

Der hier verwendete Modellansatz hat sich schon in einer französischen Studie bewährt und führt auch in der vorliegenden Untersuchung zu plausiblen Ergebnissen. Eine Überprüfung der Zuverlässigkeit dieser Methode – wie bei Colonna et al. [4] mittels Kreuzvalidierung geschehen – ist mit deutschen Krebsregisterdaten gegenwärtig nicht durchführbar, da zu diesem Zweck die Existenz mehrerer, seit langem gut arbeitender, vergleichbarer epidemiologischer Krebsregister erforderlich ist. Die Grundannahme vergleichbarer Überlebensraten in den einzelnen Regionen ist bei dem relativ homogenen Gesundheitssystem Deutschlands sicherlich vertretbar.

Durch die Berücksichtigung der Mortalität in den Modellen können die Ergebnisse jedoch durch bestimmte Ereignisse, wie beispielsweise einem Klassifikationswechsel von der ICD'9 auf die ICD'10, beeinflusst werden. In der Vergangenheit führten Klassifikationswechsel vereinzelt zu Strukturbrüchen in der Todesursachenstatistik.

Das Verfahren geht von der idealisierten Vorstellung eines hundertprozentigen Erfassungsgrades im Referenzregister aus. Wie die Anwendungen der direkten Capture-Recapture-Verfahren auf saarländische Krebsregisterdaten zeigen, besitzen diese Daten zwar einen ausreichend hohen Erfassungsgrad von über 90%, eine hundertprozentige Vollzähligkeit wird jedoch nicht erreicht [2].

Die Abschätzung der Vollzähligkeit beruht allein auf den reinen Erkrankungsmeldungen an die Krebsregister. DCO-Fälle bleiben hier bewusst unberücksichtigt. Daher sind die aus dieser Untersuchung resultierenden erwarteten Erkrankungszahlen nicht direkt vergleichbar mit den von den Registern regelmäßig publizierten Inzidenzen, die gemäß internationalen Gepflogenheiten die DCO-Fälle enthalten. Ein nachträgliches Hinzufügen der DCO-Fälle zu den erwarteten Erkrankungszahlen ist nicht sinnvoll, da gerade junge Register – wie schon erwähnt – oft überhöhte DCO-Anteile aufweisen.

Bei der Verwendung der Inzidenzen aus im Aufbau befindlichen Registern, die jedoch schon einen Erfassungsgrad von 90% und mehr erreicht haben, ist trotz des hohen Erfassungsgrades kritische Zurückhaltung geboten. Aufgrund des teilweise überhöhten DCO-Anteils, der sich aus den schon vor Einrichtung des Registers aufgetretenen Erkrankungsfällen ergibt, wird die Inzidenz für diese Register oft überschätzt, da sie sich aus den reinen Erkrankungsmeldungen und dem DCO-Anteil zusammensetzt.

Die nach dieser Methode als vollzählig akzeptierten Krebsregisterdaten können mit den saarländischen Daten aggregiert werden und als Datenbasis zur Schätzung deutschlandweiter Inzidenzen dienen. Dadurch werden regionale Besonderheiten bezüglich des Auftretens und des Registrierens von Krebserkrankungen vermindert, sodass die bundesweiten Krebsinzidenzen verlässlicher geschätzt werden können.

Literatur

1. Gesetz über Krebsregister (Krebsregistergesetz – KRG) Bundesgesetzblatt Teil I, 79:3351 vom 11.11.94
2. Brenner H, Stegmaier C, Ziegler H (1995) Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. *J Epidemiol Community Health* 49:426–430
3. Parkin DM, Pisani P, Ferlay J (1993) Estimates of the worldwide incidence of eighteen major cancers in 1985. *Int J Cancer* 54:594–606
4. Colonna M, Grosclaude P, Faivre J et al. (1999) Cancer registry data based estimation of regional cancer incidence: application to breast and colorectal cancer in French administrative regions. *J Epidemiol Community Health* 53:558–564
5. Haberland J, Bertz J, Görsch B, Schön D (2001) Krebsinzidenzschätzungen für Deutschland mittels log-linearer Modelle. *Gesundheitswesen* 63:556–560
6. Grundmann E, Hermanek P, Wagner G (1997) *Tumorhistologieschlüssel*, 2. Aufl. Springer, Berlin Heidelberg New York Tokyo, S 40f
7. Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland (2002) *Krebs in Deutschland*, 3. erweiterte, aktualisierte Ausgabe. Eigenverlag, Saarbrücken, S 76ff
8. Francis B, Green M, Payne C (1994) *The GLIM System, Release 4 Manual*. Clarendon Press, Oxford
9. Parkin DM, Wabinga H, Namboose S (2001) Completeness in an African cancer registry. *Cancer Causes Control* 12:147–152